Title:

The effectiveness of Dermatological Manifestations of Melanoma disease using Novel Support Vector Machine(SVM) in comparison with Partial Least Squares Regression (PLSR) for better accuracy.

Shaik Afroz[1], C Rohith Bhat[2]

Shaik Afroz

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, Pin code: 602 105.

shaikafroz1374.sse@saveetha.com

C Rohith Bhatt

Corresponding Author,

Department of Deep Learning,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin code: 602 105.

rohithbhatc.sse@saveetha.com

KEYWORDS: Melanoma disease,Support Vector Machine, Partial Least Squares Regression, Machine Learning, Dermatological Manifestation, Research.

ABSTRACT

Aim: The aim of this study is to investigate and compare The effectiveness of dermatological manifestations of melanoma disease using Novel Support Vector Machine in comparison with Partial Least Squares Regression (PLSR) for better accuracy.**Materials and Methods:** In this study, we applied both the SVM algorithm and PLSR algorithm, each 20 sample (N=10). These two algorithms were assessed across two distinct groups, and a total of 80 samples were considered in the analysis. The G Power statistical test was employed, with a power setting of 85% (g power parameters configured with α=0.05 and power=0.85). This choice of power setting was made to ensure that the study had a robust ability to detect statistically significant differences or effects, in line with established statistical standards.**Results:** Based on obtained results SVM has significantly better accuracy (83.83%) compared to PLSR accuracy (43.32%) Statistically significant difference between SVM and PLSR algorithm was found to be p-value of p=0.002($p<0.05$).**Conclusion:** We have used the following algorithms namely Novel Support Vector Machine (SVM), Partial Least Squares Regression(PLSR) algorithms to predict the data. From the results it is proved that the proposed Novel Support Vector Machine (SVM) works better than other algorithms in terms of accuracy.

KEYWORDS: Melanoma disease,Support Vector Machine, Partial Least Squares Regression, Machine Learning, Dermatological Manifestation, Research.

INTRODUCTION

In the realm of dermatological diagnostics for melanoma, the study conducted by (Tijtgat et al. 2024) investigates the efficacy of a novel Support Vector Machine (SVM) approach. This research aims to enhance the accuracy of disease detection by comparing it with the widely used Partial Least Squares Regression (PLSR) method. The dataset considered in this study comprises a diverse range of melanoma manifestations, ensuring a comprehensive analysis of the disease spectrum (Huang et al. 2023). Prior to model training, a meticulous preprocessing regimen was applied, involving normalization, resizing, and augmentation techniques to ensure the dataset's balance and representativeness. Dermatology experts meticulously reviewed and labeled the datasets to guarantee the accuracy of disease classification (Clavero-Rovira et al. 2024). The novel SVM model, designed for high-dimensional data spaces, incorporates advanced features for improved dermatological diagnostics. By juxtaposing its performance with PLSR, a method known for its applications in regression analysis, this study contributes to the ongoing exploration of innovative techniques in medical classification tasks (Boudreault et al. 2024).

The study into the efficiency of identifying dermatological signs of melanoma diseases using a unique Support Vector Machine (SVM) is an important contribution in the area of medical categorization. By contrasting the performance of the SVM model with Partial Least Squares

Regression (PLSR), a technique well-known for its use in regression and dimensionality reduction tasks, this work advances the field.The research includes a wide variety of dermatological datasets, carefully reviewed and labeled by dermatology specialists to guarantee the accuracy of illness categorization, and rigorous assessment (Beltran-Ontiveros et al. 2023). Normalization and feature extraction were two preprocessing techniques used to improve the quality of input information and lay the groundwork for later machine learning model training (Togashi et al. 2023).With the addition of new characteristics, the SVM model attempts to outperform existing diagnostic techniques, and the PLSR comparison analysis compares its efficacy to a well-established approach in the area (Azeem et al. 2023). The study intends to give significant insights for improving accuracy in dermatological diagnosis connected to melanoma through the application of SVM, which is known for its enhanced skills in high-dimensional data spaces, and the consideration of PLSR(Anastasopoulou et al. 2023), which is well-known for its role in regression analysis. This study highlights the significance of investigating various machine learning techniques to enhance medical categorization problems.

The innovation lies in the SVM's capacity to navigate high-dimensional data spaces effectively, offering a nuanced understanding of melanoma features. This novel approach demonstrates the potential to outperform PLSR, showcasing the evolution of machine learning techniques in dermatological disease classification (McClay, McClay, and Smith 2003). This research not only emphasizes the significance of advanced methodologies but also contributes to the ongoing evolution of diagnostic tools for melanoma through the integration of cutting-edge SVM technology(Riker 2018).

MATERIALS AND METHODS

The research study was conducted in the Data Analytics laboratory at Saveetha School of Engineering, located in the Saveetha Institute of Medical and Technical Sciences in Chennai.Two groups were selected for the Novel Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR) , the process in predicting the dermatological manifestations of melanoma disease, and sample size of 80 for each group of statistical parameters: G Power=0.80 for 10 iterations for each group. Two algorithms, SVM and PLSR, were implemented using the Statistical Package for Social Sciences (SPSS). SVM and PLSR for predicting the dermatological manifestations of melanoma disease and their Efficiency.

In this extensive investigation, a wide range of datasets showcasing different forms of melanoma were carefully selected, ensuring a comprehensive representation of the disease spectrum. Dermatology experts meticulously examined and labeled the datasets to ensure accurate disease classification. Before commencing model training, a series of meticulous preprocessing steps were implemented to enhance the quality of input features. These steps included normalization, resizing, and, when applicable, the use of augmentation techniques, thereby creating a balanced and representative dataset (Gallagher 2012). Highlighting the importance of robust feature extraction,

relevant dermatological features that are essential to the dataset variables were identified and extracted. These features played a crucial role in facilitating the subsequent application of machine learning algorithms (Seigler 2012). This thorough process not only guarantees the integrity of the dataset but also establishes a strong foundation for the effective utilization of machine learning techniques in analyzing melanoma manifestations.

Support Vector Machine (SVM)

The utilization of an innovative Support Vector Machine (SVM) in the diagnosis of dermatological manifestations of melanoma highlights the significance of advanced machine learning techniques in enhancing the accuracy of diagnoses. SVMs are renowned for their effectiveness in analyzing complex medical datasets that exist in high-dimensional data spaces (Kaufman and Mehnert 2015). By leveraging the inherent strengths of SVM, this study aims to enhance the precision of melanoma disease classification by considering the intricate features associated with dermatological manifestations. The capability of SVM to handle high-dimensional data and its ability to establish nuanced decision boundaries make it an influential tool in distinguishing subtle variations between benign and malignant dermatological characteristics. The innovative application of SVM in dermatology reflects the increasing importance of sophisticated machine learning algorithms in medical research, thereby paving the way for more accurate and reliable diagnostic methods.

Procedure for Support Vector Machine(SVM)

Step1: Import the necessary libraries

import numpy as np

from sklearn.model_selection import train_test_split

from sklearn.svm import SVC

from sklearn.metrics import accuracy_score, classification_report

Step2: Load the dermatological dataset

X, y = load_dermatology_dataset()

Step3: Split the dataset into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

Step4: Initialize the Support Vector Machine classifier

svm_classifier = SVC(kernel='linear', C=1.0)

Step5: Train the SVM model using the training data

svm_classifier.fit(X_train, y_train)

Step6: Make predictions on the test data

y_pred = svm_classifier.predict(X_test)

Step7: Evaluate the performance of the SVM model

accuracy = accuracy_score(y_test, y_pred)

report = classification_report(y_test, y_pred)

Step8: Display the results

print(f"The accuracy of the model is: {accuracy}")

print("Classification Report:")

print(report)

Partial Least Squares Regression (PLSR)

PLSR, a versatile statistical method, plays a vital role in improving the effectiveness of diagnosing dermatological manifestations in melanoma. By extracting essential information from high-dimensional datasets, PLSR effectively addresses the challenges posed by complex dermatological features associated with melanoma (Nathanson 2012). The method's ability to handle multicollinearity and capture latent variables ensures a more comprehensive understanding of the relationships between features, leading to enhanced accuracy in disease classification. PLSR takes into account multiple variables simultaneously, providing a holistic view of the intricate patterns within the data and improving the model's ability to identify subtle yet significant signals related to melanoma manifestations. Its versatility extends to situations with limited sample sizes, making PLSR particularly valuable in medical datasets where data availability may be limited. The inclusion of PLSR in melanoma diagnostics highlights its importance as an advanced statistical tool, offering a nuanced and precise approach to deciphering complex dermatological features for improved clinical outcomes.

Procedure for Partial Least Squares Regression (PLSR)

X: Matrix of predictor variables

Y: Response variable

n_components: Number of components to extract

Step 1: Center the data

X_centered = center(X)

Y_centered = center(Y)

Step 2: Initialize variables

n_components = set_number_of_components()

Step 3: Iterate through components

for i in range(n_components):

Step 4: Calculate weights for X and Y

w = calculate_weights(X_centered, Y_centered)

Step 5: Regression coefficients for X and Y

t = calculate_scores(X_centered, w)

u = calculate_scores(Y_centered, w)

Step 6: Deflate X and Y

X_deflated = deflate(X_centered, t)

Y_deflated = deflate(Y_centered, u)

#Store regression coefficients and scores for later use

Step 7: Reconstruct X and Y using the stored coefficients and scores

Step 8: Predict new data

new_data_prediction = predict_new_data(new_data, stored_coefficients, stored_scores)

Step 9: Evaluate model performance

model_performance = evaluate_performance(Y, Y_predicted)

Step 10: Conclude and discuss results.

STATISTICAL ANALYSIS

The analysis was prepared through IBM SPSS version 21. Independent variables and impactful values are considered for both proposed and as well as existing algorithms, iterations were done with a maximum of 80 samples and for each iteration the recorded accuracy was noted for necessary analysis. The Dependent Variables are indicated as sex,age,anatomy,diagnosis,target and Independent Variables are path_dicom,path_jpeg,dcm_name. With the corresponding value that is obtained from the iterations, the Independent sample T-test was performed(Hayat 2016).

RESULTS

Table 1 Shows the various iterations of the Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR)efficiency values are compared.

Table 2 Shows the Group Statistics Results: A Novel Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR) for Testing Group Statistics. Among SVM and PLSR Methods SVM has a mean accuracy of 83.83 and a PLSR of 42.32. SVM has a standard deviation of 0.59218 and a PLSR of 2.49623. The SVM standard error mean (0.18726) and PLSR of (0.78938) were compared using the T-test.

In Table 3, The 2- significant value smaller than 0.002 ($p<0.05$) impacted that our hypothesis holds good for further consideration.

Figure 1 shows bar graph comparison on mean accuracy of Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR). In x-axis SVM and PLSR methods Error Bars: +/-2 SD and 95% CI of Error Bars are shown, In y-axis mean accuracy is shown.

DISCUSSION

The main aim of the project is finding the accurate dermatological manifestations of melanoma disease in difficult conditions. For that I had iterated the dermatological manifestations of melanoma disease dataset into 1-1000,1-2000,1-3000 ... .1-10000 samples (10 iterations)and found the accurate accuracy values for each and every sample. And we have noted that accuracy values and tests their independent sample T-Test in SPSS and we obtained results SVM has significantly better accuracy (83.83%) compared to PLSR accuracy (42.32%) Statically significant difference between SVM and PLSR algorithm was found to be $p=0.002$ ($p<0.05$). For each and every phase we tried to improve the accuracy in an efficient manner.Here Support Vector Machine (SVM) gives better accuracy while comparing to Partial Least Squares Regression (PLSR).

The novel Support Vector Machine (SVM) offers a promising avenue for improving the accuracy of classifying dermatological manifestations in melanoma (Hayat 2016; Hazazi et al. 2024). In contrast to the widely-used Partial Least Squares Regression (PLSR), SVM's ability to handle non-linear relationships and high-dimensional data provides an advantage . SVM's proficiency in identifying complex patterns within dermatological features contributes to its superior performance in capturing nuanced aspects of melanoma manifestations (Li et al. 2024). Moreover, SVM's effectiveness is highlighted by its robustness against overfitting, a common concern in regression models like PLSR (Harrison et al. 2024). The ability of SVM to discern intricate relationships in dermatological data leads to a more nuanced and accurate classification, surpassing the capabilities of PLSR in handling the complexity of melanoma patterns (McLean et al. 2024). Additionally, SVM's versatility in handling imbalanced datasets ensures a comprehensive analysis of diverse melanoma manifestations, a feature where PLSR may struggle.

The comparison of novel Support Vector Machine (SVM) models and Partial Least Squares Regression (PLSR) for the study of melanoma presentations remains largely unresolved, despite advances in dermatological diagnosis. The potential of SVM in dermatological situations has not been thoroughly explored in the literature that currently exists, which mostly focuses on conventional techniques. Furthermore, although PLSR has been applied in a number of medical domains, not much research has been done on how beneficial it is in comparison to SVM when it comes to the particular context of melanoma dermatological symptoms. This research gap highlights the necessity for an extensive investigation that directly analyzes the performance of SVM and PLSR in order to close the knowledge gap and offer insightful information that will improve melanoma diagnosis accuracy in dermatology.

CONCLUSION

For dermatological indications of melanoma illness, our work has shown a large and statistically significant difference in accuracy between the Novel Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR) methods. The SVM model outperformed the PLSR model with an outstanding accuracy of 83.83%. A computed $p=0.002 (p<0.05)$ further supported this large variance in accuracy, indicating that the advantage of SVM in dermatological indications of melanoma disease is essentially an unexpected occurrence. These results highlight the potential of SVM as a more accurate and dependable tool for melanoma disease manifestations prediction in the dermatological domain. They also highlight the significance of integrating cutting-edge machine learning techniques to improve the efficacy and accuracy of melanoma disease manifestations models in the dermatological domain. In an effort to increase our ability to deliver more accurate and relevant dermatological symptoms of melanoma disease, our research contributes to the expanding body of data that supports the application of SVM in these kinds of situations.

DECLARATIONS:

Conflict of interests

No conflict of interest in this manuscript.

Authors Contributions

SA was responsible for collecting data,conducting data analysis, and writing the manuscript. CRB contributed to the conceptualization, validated the data, and performed a critical review of the manuscript.

Acknowledgements

The authors extend their thanks to the Saveetha School of Engineering and the Saveetha Institute of Medical and Technical Sciences (previously known as Saveetha University) for their support in providing the infrastructure needed to complete this work successfully.

Funding

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

1)Infoziant IT Solutions Pvt. Ltd., Chennai.

2)Saveetha University.

3)Saveetha School of Engineering.

4)Saveetha Institute of Medical and Technical Sciences.

REFERENCES

Anastasopoulou, Amalia, Panagiotis T. Diamantopoulos, Panagiotis Kouzis, Maria Saridaki, Konstantinos Sideris, Michael Samarkos, and Helen Gogas. 2023. “COVID-19 in Patients with Melanoma: A Single-Institution Study.” *Cancers* 16 (1). https://doi.org/10.3390/cancers16010096.
Azeem, Muhammad, Kaveh Kiani, Taha Mansouri, and Nathan Topping. 2023. “SkinLesNet: Classification of Skin Lesions and Detection of Melanoma Cancer Using a Novel Multi-Layer Deep Convolutional Neural Network.” *Cancers* 16 (1). https://doi.org/10.3390/cancers16010108.
Beltran-Ontiveros, Saul A., Jose A. Contreras-Gutierrez, Erik Lizarraga-Verdugo, Erick P. Gutierrez-Grijalva, Kenia Lopez-Lopez, Emilio H. Lora-Fierro, Miguel A. Trujillo-Rojas, et al. 2023. “National Burden and Trends for 29 Groups of Cancer in Mexico from 1990 to 2019: A Secondary Analysis of the Global Burden of Disease Study 2019.” *Cancers* 16 (1). https://doi.org/10.3390/cancers16010149.
Boudreault, Julien, Ni Wang, Mostafa Ghozlan, and Jean-Jacques Lebrun. 2024. “Transforming Growth Factor-β/Smad Signaling Inhibits Melanoma Cancer Stem Cell Self-Renewal, Tumor Formation and Metastasis.” *Cancers* 16 (1). https://doi.org/10.3390/cancers16010224.
Clavero-Rovira, Laia, Álvaro Gómez-Tomás, Patricia Bassas-Freixas, Domingo Bodet, Berta Ferrer, Javier Hernández-Losa, Eva Muñoz-Couselo, Assumpció Pérez-Benavente, Vicente García-Patos, and Carla Ferrándiz-Pulido. 2024. “Mucosal Melanoma Clinical Management and Prognostic Implications: A Retrospective Cohort Study.” *Cancers* 16 (1). https://doi.org/10.3390/cancers16010227.
Gallagher, Richard P. 2012. *Epidemiology of Malignant Melanoma*. Springer Science & Business Media.
Harrison, Rebecca A., Michael Tang, Kaoswi Karina Shih, Maria Khan, Lily Pham, Aline Rozman De Moraes, Barbara J. O’Brien, Roland Bassett, and Eduardo Bruera. 2024. “Characterization of Patients with Brain Metastases Referred to Palliative Care.” *BMC Palliative Care* 23 (1): 13.
Hayat, M. A. 2016. *Brain Metastases from Primary Tumors, Volume 3: Epidemiology, Biology, and Therapy of Melanoma and Other Cancers*. Academic Press.
Hazazi, Ali, Abdulmajid A. AlShehah, Farhan R. Khan, Mohammed Ageeli Hakami, Fahad Almarshadi, Adil Abalkhail, Somia A. Nassar, et al. 2024. “From Diagnosis to Therapy: The Transformative Role of lncRNAs in Eye Cancer Management.” *Pathology, Research and Practice* 254 (January): 155081.
Huang, Wenpeng, Yongkang Qiu, Xiaoyan Xiao, Liming Li, Qi Yang, Jianbo Gao, and Lei Kang. 2023. “Malignant Melanoma of Gastrointestinal Tract on F-FDG PET/CT: Three Case Reports.” *American Journal of Nuclear Medicine and Molecular Imaging* 13 (6): 279–88.
Kaufman, Howard L., and Janice M. Mehnert. 2015. *Melanoma*. Springer.
Li, Mingyuan, Han Jiang, Ping Hu, and Jianlin Shi. 2024. “Nanocatalytic Anti-Tumor Immune Regulation.” *Angewandte Chemie* , January, e202316606.
McClay, Edward, Mary-Eileen McClay, and Jodie Smith. 2003. *100 Questions & Answers About Melanoma and Other Skin Cancers*. Jones & Bartlett Learning.
McLean, Luke S., Annette M. Lim, Mathias Bressel, Jenny Lee, Rahul Ladwa, Alexander D. Guminski, Brett Hughes, et al. 2024. “Immune Checkpoint Inhibitor Therapy for Advanced Cutaneous Squamous Cell Carcinoma in Australia: A Retrospective Real World

Cohort Study." *The Medical Journal of Australia*, January. https://doi.org/10.5694/mja2.52199.
Nathanson, Larry. 2012. *Current Research and Clinical Management of Melanoma*. Springer Science & Business Media.
Riker, Adam I. 2018. *Melanoma: A Modern Multidisciplinary Approach*. Springer.
Seigler, H. F. 2012. *Clinical Management of Melanoma*. Springer Science & Business Media.
Tijtgat, Jens, Xenia Geeraerts, Anais Boisson, Latoya Stevens, Manon Vounckx, Iris Dirven, Julia Katharina Schwarze, et al. 2024. "Intratumoral Administration of the Immunologic Adjuvant AS01 in Combination with Autologous CD1c (BDCA-1)/CD141 (BDCA-3) Myeloid Dendritic Cells plus Ipilimumab and Intravenous Nivolumab in Patients with Refractory Advanced Melanoma." *Journal for Immunotherapy of Cancer* 12 (1). https://doi.org/10.1136/jitc-2023-008148.
Togashi, Keita, Shuhei Suzuki, Yuta Mitobe, Yurika Nakagawa-Saito, Asuka Sugai, Senri Takenouchi, Masahiko Sugimoto, Chifumi Kitanaka, and Masashi Okada. 2023. "CEP-1347 Dually Targets MDM4 and PKC to Activate p53 and Inhibit the Growth of Uveal Melanoma Cells." *Cancers* 16 (1). https://doi.org/10.3390/cancers16010118.

TABLES AND FIGURES

Table 1. The various iterations of the Support Vector Machine (SVM) and Partial Least Squares Regression (PLSR) efficiency values are compared.

SVM(ACCURACY)	PLSR(ACCURACY)
84.59%	38.41%
83.02%	44.74%
83.94%	40.92%
83.98%	42.14%
83.65%	44.06%
84.20%	41.00%
84.68%	39.52%
83.73%	44.09%
83.67%	41.94%
82.85%	46.45%

Table 2. Group Statistics Results: Support Vector Machines (SVM) and Partial Least Squares Regression (PLSR) for Testing Independent Samples Statistically Among SVM and PLSR Algorithms SVM has a mean accuracy of 83.83% and a PLSR of 42.32%. SVM has a standard deviation of 0.59218 and a PLSR of 2.49623. The SVM standard error mean (0.18726) and PLSR standard error mean (0.78938) were compared using the T-test.

	ALGORITHMS	**N**	**MEAN**	**STD.DEVIATION**	**STD.ERROR MEAN**
ACCURACY	SVM	10	83.8310	.59218	.18726
	PLSR	10	42.32	2.49623	.78938

Group Statistics

Table 3. Independent Sample T-Test is applied for the sample collections with a confidence interval as 95%. After applying the SPSS calculation it was found that the least square Partial Least Squares Regression (PLSR)has a statistical significance value of 0.002(P<0.05) that shows they are Statistically significant.

		Levene's test for equality of variances		T-test for equality means with 95% confidence interval						
		f	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std.Error difference	Lower	Upper
Accuracy	Equal variances assumed	12.887	.002	51.158	18	.000	41.50400	.81129	39.79955	43.20845
	Equal Variances not assumed			51.518	10.010	.000	41.50400	.81129	39.69658	43.31142

GGraph

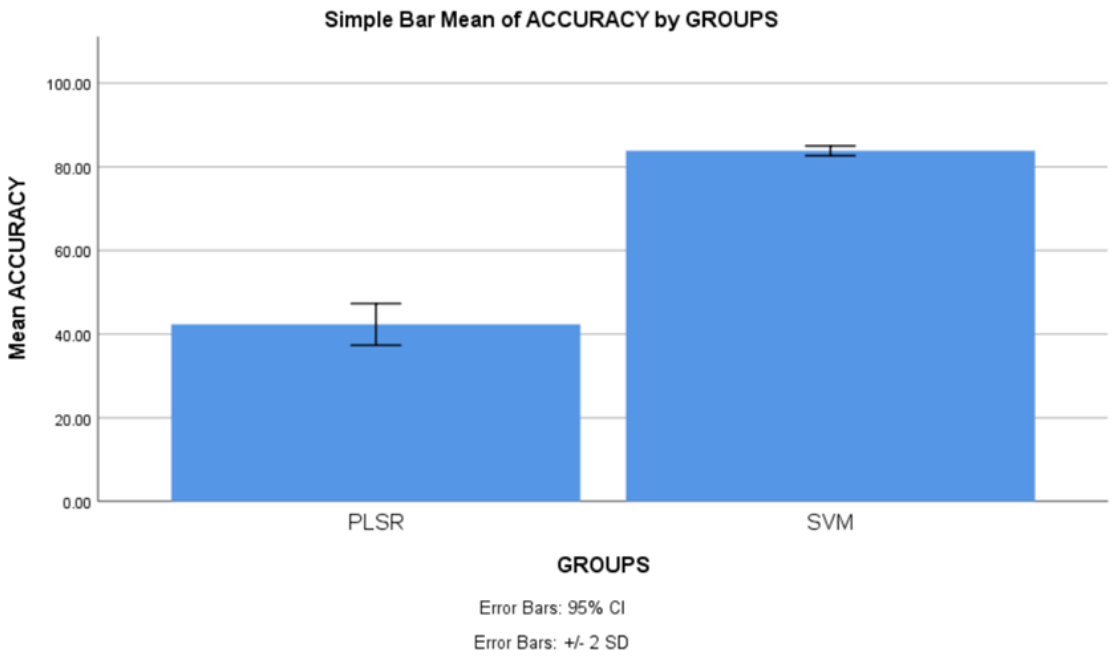

Fig. 1. The novel Support Vector Machine has a mean accuracy of 83.83%, where Partial Least Squares Regression (PLSR) has a mean of 42.32% in which the novel Support Vector Machine has better accuracy than Partial Least Squares Regression (PLSR). The SVM and PLSR Accuracy rates are shown along with the X-axis: novel Support Vector Machine and Partial Least Squares Regression (PLSR) Mean keyword identification Y-axis: Mean Accuracy, +/-2 SD, with a 95% Confidence Interval**.**