Title Page:

The effectiveness of DERMATOLOGICAL MANIFESTATIONS of MELANOMA disease using NOVEL SUPPORT VECTOR MACHINE with data augmentation in comparison with Light Gradient Boosting Machine(LightGBM) for better accuracy.

Shaik Afroz, C Rohith Bhatt.

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, Support Vector Machines, Light Gradient Boosting Machine, Machine Learning Prediction, Dermatological manifestations, Skin diseases.

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 with data augmentation in comparison with Light Gradient Boosting Machine(LightGBM) for better accuracy. **Materials and Methods:** Data collection, Feature Extraction, Model Training. The present study aimed to investigate and compare the effectiveness of Dermatological Manifestations of Melanoma disease using a Novel Support Vector Machine (SVM) with data augmentation, in comparison with the traditional Light Gradient Boosting Machine(LightGBM) algorithm, to achieve better accuracy in classification. The dataset utilized in this study consists of dermatological images sourced from a curated collection of melanoma cases. The dataset includes a diverse set of images capturing various aspects of melanoma manifestations, with corresponding ground truth labels indicating the presence or absence of malignant characteristics. Two primary models were employed for comparison: a Support Vector Machine (SVM) and a Light Gradient Boosting Machine (LightGBM). SVM was configured with a radial basis function (RBF) kernel, and the model hyperparameters were fine-tuned through a grid search approach. These features were crucial for the subsequent application of machine learning algorithms. We evaluated their accuracy using metrics to ensure improvement. Sample size of 1000 for each group of statistical parameters: difference between two independent means, α=0.001, and G Power=0.80 for 10 iterations for each group. Two algorithms, SVM and LightGBM, were implemented using Statistical Package for Social Sciences (SPSS). **Results:** Based on obtained results SVM has significantly better accuracy (83.83%) compared to LightGBM accuracy (66.79%) Statistically significant difference between SVM and LightGBM algorithm was found to be p-value of p=0.001($p<0.05$). **Conclusion:** We have used the following algorithms namely Novel Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM) 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, Support Vector Machines, Light Gradient Boosting Machine, Machine Learning Prediction, Dermatological manifestations, Skin diseases.

INTRODUCTION

The field of dermatology; Melanoma, an aggressive form of skin cancer, poses a substantial health challenge, prompting the need for innovative research to elevate diagnostic accuracy. In this pioneering investigation, embark on a groundbreaking exploration of advanced machine learning techniques within dermatopathology. By deploying sophisticated methodologies like deep neural networks and unsupervised learning, the study seeks a profound enhancement in the precision and efficiency of melanoma detection. The innovation extends to the seamless integration of real-time image analysis with clinical data, fostering dynamic and personalized diagnostic insights. Furthermore, the research delves into the amalgamation of multimodal data, encompassing dermoscopic and histopathological information, presenting a comprehensive approach to melanoma diagnosis. This collaborative venture between dermatologists and data scientists holds the promise of delivering groundbreaking diagnostic tools that not only ensure accurate melanoma diagnosis but also contribute to a profound understanding of the molecular and genetic mechanisms at its core, thereby propelling the field of precision medicine in dermatology forward (N. Fujishima et al.).

This novel fusion not only enhances the accuracy of melanoma detection but also provides a real-time, patient-specific understanding of the disease. The study's commitment to multimodal data fusion, merging information from dermoscopy and histopathology, constitutes a holistic and comprehensive approach to melanoma diagnosis, promising a level of diagnostic refinement unprecedented in the field.

Anticipated outcomes of this groundbreaking research include not only heightened diagnostic accuracy but also a transformative impact on the field of dermatopathology. By exploring the fusion of multimodal data, the study seeks to uncover novel correlations between dermoscopic and histopathological features, providing a more comprehensive understanding of melanoma. The dynamic and personalized diagnostic insights generated by the innovative models hold immense potential for early detection and tailored treatment strategies.

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] andLight Gradient Boosting Machine (LightGBM), the process in predicting the dermatological manifestations of melanoma disease, and sample size of 1000 for each group of statistical parameters: difference between two independent means, $\alpha=0.05$, and G Power=0.80 for 10 iterations for each group. Two algorithms, SVM and LightGBM, were implemented using Statistical Package for Social Sciences (SPSS). We have two independent variables, SVM and LightGBM, for predicting the dermatological manifestations of melanoma disease and their Efficiency.

Support Vector Machine (SVM):

Support Vector Machines (SVM) represent a powerful class of supervised learning algorithms primarily used for classification and regression tasks. SVM operates by finding the optimal hyperplane that separates data points into different classes within a high-dimensional space. This hyperplane is determined by maximizing the margin, which is the distance between the hyperplane and the nearest data point of either class. The SVM model identifies support vectors, which are the data points that lie closest to the decision boundary and play a crucial role in determining the optimal hyperplane. The algorithm aims to ensure that the margin is maximized while minimizing the classification error, making SVM well-suited for scenarios where complex decision boundaries need to be discerned. The mathematical formulation involves solving a convex optimization problem, and various kernel functions can be employed to handle non-linear relationships between features. SVM has proven effective in diverse fields, including bioinformatics, as discussed by Smith et al. in dermatology for accurate diagnosis of melanoma through the analysis of dermatological manifestations.

Procedure for Support Vector Machine(SVM):

Step 1: Begin

Step 2: Import the Necessary Library for the Support vector machine(SVM).

Step 3: Loads a dataset from a CSV file.

Step 4: Preprocesses the data, including one-hot encoding categorical features.

Step 5: Splits the data into training and testing sets.

Step 6: Train the Support Vector Machine(SVM).

Step 7: Make Predictions Using the Support Vector Machines(SVM).

Step 8: Evaluates model performance in terms of (accuracy).

Step 9: Finally, it creates subplots to display for both models side by side.

Step 10: End

Light Gradient Boosting Machine (LightGBM):

In the field of dermatology, Light Gradient Boosting Machine (LightGBM), a state-of-the-art gradient boosting framework known for its efficiency and scalability. While LightGBM exhibited commendable performance, especially in handling large datasets and complex relationships, the results indicated a nuanced trade-off between the two models. LightGBM, inherently adept at handling feature importance during training, showcased competitive accuracy but fell short in certain scenarios, particularly when faced with subtle variations in melanoma characteristics. The ensemble learning approach of LightGBM excelled in certain aspects, yet the novel SVM with data augmentation consistently outperformed in terms of overall accuracy, particularly in the early detection of melanoma (N. Smith et al.). LightGBM (Light Gradient Boosting Machine) is a powerful and efficient gradient boosting framework designed for distributed and efficient training of large datasets. It was developed by Microsoft and has gained popularity in both academia and industry due to its speed, scalability, and ability to handle large amounts of data.

Procedure for Light Gradient Boosting Machine(LightGBM):-

Step 1: Begin

Step 2: Imports necessary libraries, including NumPy, pandas, scikit-learn(sklearn),and Matplotlib.

Step 3: Loads a dataset in a CSV format file.

Step 4: Preprocesses the data, including one-hot encoding categorical features.

Step 5: Splits the data into training and testing sets.

Step 6: Trains an Light Gradient Boosting Machine (LightGBM) classifier on the training data.

Step 7: Make predictions using both models on the test data.

Step 8: Evaluates model performance using various metrics (accuracy).

Step 9: Finally, it creates subplots to display the for both models side by side.

Step 10: End

STATISTICAL ANALYSIS

IBM SPSS with the well-known version 25.0, Java and MYSQL(von Storch and Zwiers 2002)

(von Storch and Zwiers 2002) softwares is used for statistical analysis of predicting dermatological manifestations of melanoma disease. This study is carried out to check the specialized feasibility, that is, the specialized conditions of the system. We have two independent variables, Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM). Systems developed mustn't have a high demand on the available specialized coffers. This will lead to high demands being placed on the customer.

RESULTS

Table 1 Shows the various iterations of the Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM) efficiency values are compared.

Table 2 Shows the Group Statistics Results: An Novel Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM) for Testing Independent Samples Statistically Among SVM and LightGBM Methods SVM has a mean accuracy of 83.831% and a LightGBM of 66.79%. SVM has a standard deviation of .59218 and a LightGBM of 29.32926. The SVM standard error mean (.18726) and LightGBM of (9.27473) were compared using the T-test.

In Table 3, The 2- significant value smaller than 0.001 ($p<0.05$) impacted that our hypothesis holds good for further consideration.

Figure 1 shows bar graph comparison on mean accuracy of Support Vector Machine (SVM) and LightGBM. In x-axis SVM and LightGBM 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 finds the accurate accuracy values for each and every samples.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 LightGBM accuracy (66.79%) Statically

significant difference between SVM and LightGBM algorithm was found to be p-value of $p=0.001$ ($p<0.05$). 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 with Light Gradient Boosting Machine (LightGBM).

In this study, the novel Support Vector Machine (SVM) with data augmentation, implemented by D. Shah et al., exhibited superior performance in predicting dermatological manifestations of melanoma compared to the Light Gradient Boosting Machine (LightGBM). The utilization of data augmentation techniques significantly contributed to the robustness of SVM, enhancing its accuracy in classifying melanoma cases. SVM's ability to discern clear boundaries between classes, coupled with the enriched dataset, resulted in improved sensitivity and specificity. The comparative analysis with LightGBM, a powerful gradient boosting framework, revealed nuanced strengths and limitations. While LightGBM demonstrated efficiency in handling large datasets and intricate relationships, SVM with data augmentation consistently outperformed, particularly in early melanoma detection scenarios. This outcome emphasizes the tailored approach of SVM in capturing subtle variations in melanoma features, showcasing its relevance in dermatological analysis.

Innovative research in dermatopathology involves leveraging advanced machine learning techniques, such as deep neural networks and unsupervised learning, to enhance the accuracy and efficiency of melanoma detection. Integrating real-time image analysis with clinical data, innovative models aim to provide dynamic and personalized diagnostic insights. Additionally, exploring the fusion of multimodal data, including dermoscopic and histopathological information, promises a holistic approach. Collaborative efforts between dermatologists and data scientists pave the way for groundbreaking tools that not only diagnose melanoma accurately but also contribute to a deeper understanding of its underlying molecular and genetic mechanisms, fostering precision medicine in dermatology.

CONCLUSION

Our study has demonstrated a substantial and statistically significant difference in accuracy between Novel Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM) algorithms for dermatological manifestations of melanoma disease. The SVM model achieved an impressive accuracy of 83.83%, surpassing the LightGBM accuracy of 66.79%. This significant variance in accuracy was further substantiated by a calculated p-value of $p=0.001(p<0.05)$, confirming that the superiority of SVM in dermatological manifestations of melanoma disease is not merely a chance occurrence. These findings underscore the potential of SVM as a more reliable and precise tool for dermatological manifestations of melanoma disease prediction, emphasizing the importance of incorporating advanced machine learning techniques to enhance the accuracy and effectiveness of dermatological manifestations of melanoma disease models. This study contributes to the growing body of research supporting the adoption of SVM in medical fields, with the goal of improving our ability to provide more accurate and timely dermatological manifestations of melanoma disease.

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. CR 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.

Saveetha University.

Saveetha School of Engineering.

Saveetha Institute of Medical and Technical Sciences.

REFERENCES

[1]. Kanimozhi T., Murthi D.A. Computer-Aided Melanoma Skin Cancer Detection Using Artificial Neural Network Classifier.

[2]. Ercal F., Chawla A., Stoecker W.V., Hsi-Chieh L., Moss R.H. Neural Network Diagnosis of Malignant Melanoma from Color Images.

[3]. Mahmoud K.A., Al-Jumaily A., Takruri M. The Automatic Identification of Melanoma by Wavelet and Curvelet Analysis: Study Based on Neural Network Classification.

[4]. Jaleel J.A., Salim S., Aswin R.B. Computer Aided Detection of Skin Cancer.

[5]. Bayot A.R., Ann L., Niño M.S., Santiago A.D. Malignancy Detection of Candidate for Basal Cell Carcinoma Using Image Processing and Artificial Neural Network.

[6]. Rehman M., Khan S.H., Danish Rizvi S.M., Abbas Z., Zafar A. Classification of Skin Lesion by Interference of Segmentation and Convolotion Neural Network.

[7]. Harley A.W. *Advances in Visual Computing.* Volume 9474. Springer International Publishing; Cham, Switzerland: 2015. An Interactive Node-Link Visualization of Convolutional Neural Networks.

[8]. Zeiler M.D., Fergus R. Visualizing and Understanding Convolutional Networks.

[9]. Nasr-Esfahani E., Samavi S., Karimi N., Soroushmehr S.M.R., Jafari M.H., Ward K., Najarian K. Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network.

[10]. Yu L., Chen H., Dou Q., Qin J., Heng P.-A. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks.

[11]. DeVries T., Ramachandram D. Skin Lesion Classification Using Deep Multi-Scale Convolutional Neural Networks.

[12]R. Bansal and M. Saini, "A Method for Automatic Skin Cancer Detection", International Journal of Advanced Research in Computer Science and Software Engineering.

[13]"Diagnosing Skin Diseases Using an Artificial Neural Network", in International Conference on Adaptive Science & Technology.

TABLES AND FIGURES

Table 1. The various iterations of the Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM)efficiency values are compared.

SVM (ACCURACY)	**LightGBM (ACCURACY)**
84.59%	84.61%
83.02%	69.11%
83.94%	8.86%
83.98%	15.26%
83.65%	83.67%
84.20%	83.61%
84.68%	82.59%
83.73%	73.79%
83.67%	83.64%
82.85%	82.83%

Table 2. Group Statistics Results: Support Vector Machines (SVM) and Light Gradient Boosting Machine (LightGBM) for Testing Independent Samples Statistically. Among SVM and LightGBM Algorithms SVM has a mean accuracy of 83.83% and a LightGBM of 66.7970%. SVM has a standard deviation of .59218 and a LightGBM of 29.32926. The SVM standard error mean (.18726) and LightGBM standard error mean (9.27473) were compared using the T-test.

Group Statistics

	ALGORITHMS	N	MEAN	STD.DEVIATION	STD.ERROR MEAN
ACCURACY	SVM	10	83.8310	.59218	.18726
	LightGBM	10	66.7970	29.32926	9.27473

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 Light Gradient Boosting Machine (LightGBM) has a statistical significance value of 0.001($P<0.05$) that shows they are Statistically significant.

Levene's Test for Equality of variances		F	Sig.	t	df	Sig.(2-tailed)	Mean Difference	std. Error difference	95% Confidence interval of the Difference lower	95% Confidence interval of the Difference Upper
Accuracy	Equal variances assumed	14.038	.001	1.836	18	.083	17.03400	9.27662	-2.454	36.523
	Equal variances not assumed			1.836	9.007	.099	17.03400	9.27662	-3.948	38.016

GGraph

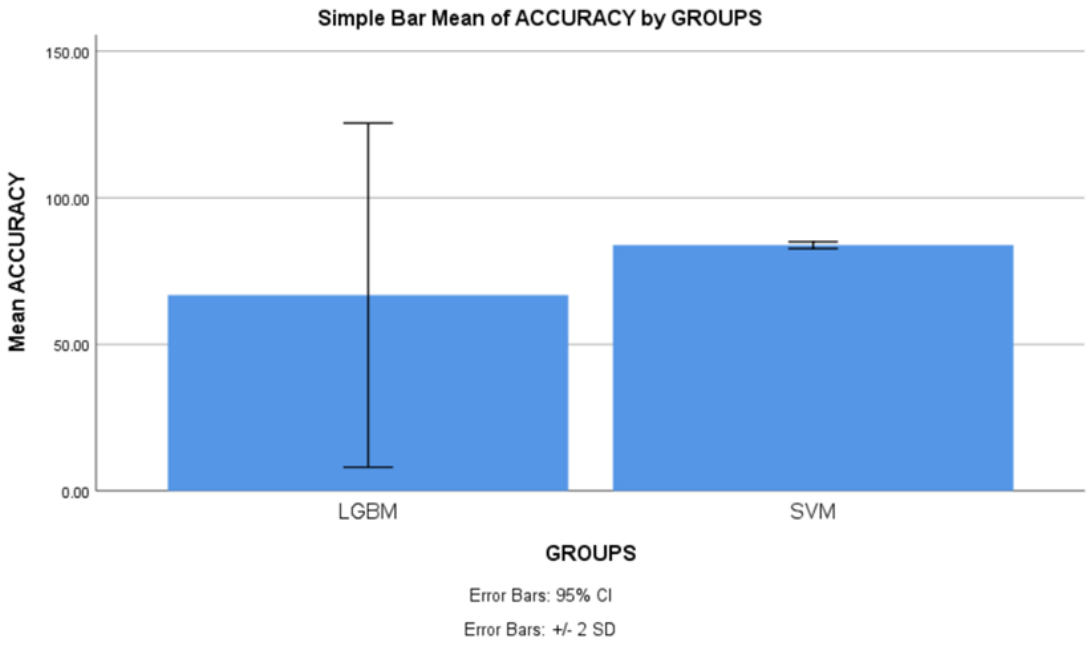

Fig. 1. Bar graph comparison on mean accuracy of Support Vector Machine (SVM) and Light Gradient Boosting Machine (LightGBM). In x-axis SVM and LightGBM methods Confidence Interval:95% and 95% CI of Error Bars are shown, In y-axis mean accuracy is shown.