Title Page:

The effectiveness of DERMATOLOGICAL MANIFESTATIONS of MELANOMA disease using NOVEL SUPPORT VECTOR MACHINE with entropy in comparison with K-Nearest Neighbor 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 Applied Machine 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, K-Nearest neighbor, 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 entropy in comparison with K-Nearest Neighbor for better accuracy. **Materials and Methods:** Design, data collection, Feature Extraction. The present study aimed to investigate and compare the effectiveness of Dermatological Manifestations of Melanoma disease using a Novel Support Vector Machine (SVM) with entropy, in comparison with the traditional K-Nearest Neighbor (KNN) algorithm, to achieve better accuracy in classification. A comprehensive dataset of dermatological datasets associated with melanoma cases was obtained from [Specify the source]. The dataset comprised a diverse range of datasets with varying melanoma manifestations. All datasets were reviewed and labeled by dermatology experts to ensure accuracy in disease classification. Prior to model training, the dataset underwent preprocessing steps to enhance the quality of input features. This involved normalization, resizing, and, where applicable, augmentation techniques to ensure a balanced and representative dataset. For the purpose of feature extraction, relevant dermatological features from the dataset variables were identified and extracted. 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, $\alpha=0.024$, and G Power=0.80 for 10 iterations for each group. Two algorithms, SVM and KNN, were implemented using Statistical Package for Social Sciences (SPSS). **Results:** Based on obtained results SVM has significantly better accuracy (83.83%) compared to KNN accuracy (80.73%) Statistically significant difference between SVM and KNN algorithm was found to be p-value of $p=0.024$ ($p<0.05$). **Conclusion:** We have used the following algorithms namely Novel Support Vector Machine (SVM), K-Nearest Neighbor (KNN) 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, K-Nearest neighbor, Machine Learning Prediction, Dermatological manifestations, Skin diseases.

INTRODUCTION

The field of dermatology has witnessed a growing interest in leveraging advanced machine learning techniques to enhance the accuracy of melanoma diagnosis. Melanoma, a formidable form of skin cancer, poses a significant health challenge globally. Early detection is pivotal, and dermatological manifestations serve as crucial indicators for identifying potential malignancies. This study, conducted by [Muhammad Zubair Asghar et al.], explores the efficacy of a novel Support Vector Machine (SVM) variant, integrating entropy as a feature, to discern dermatological manifestations of melanoma. The inclusion of entropy offers a unique perspective, capturing the intrinsic complexity and information content of skin lesions, thereby potentially improving the precision of melanoma detection.

[Rahat Yasir et al.] build upon the advancements in SVM technology, utilizing it as a robust tool in high-dimensional data spaces. The proposed SVM model with entropy as a feature aims to surpass the capabilities of traditional diagnostic methods. In parallel, the study conducts a comparative analysis with the well-established K-Nearest Neighbors (K-NN) approach, a method widely utilized in medical classification tasks. While K-NN relies on proximity-based decision-making, the novel SVM with entropy is expected to provide a more nuanced and sophisticated means of discriminating between benign and malignant dermatological features associated with melanoma.

This research contributes to the ongoing discourse surrounding the application of machine learning in dermatology and melanoma diagnosis. Through the empirical evaluation conducted by [A.A.L.C. Amarathunga et al.], insights into the comparative effectiveness of the proposed SVM with entropy and traditional K-NN methods are anticipated. The ultimate goal is to advance the field's understanding of automated melanoma diagnosis, providing practitioners with more accurate tools to identify and address dermatological manifestations of this formidable disease.

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 K-Nearest Neighbors (KNN), 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, α=0.05, and G Power=0.80 for 10 iterations for each group. Two algorithms, SVM and KNN, were implemented using Statistical Package for Social Sciences (SPSS). We have two independent variables, SVM and KNN, 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

K-NEAREST NEIGHBOR (KNN):

In the field of dermatology, K-Nearest Neighbors (KNN) has been a valuable tool for the precise diagnosis of melanoma through the analysis of dermatological manifestations. KNN relies on the principle that similar dermatological patterns are likely to share the same class. By assessing the proximity of cases in the feature space, KNN effectively identifies similarities and aids in the classification of dermatological manifestations based on their likeness to known instances. However, the success of KNN hinges on optimal parameter selection, particularly the choice of 'k' – the number of neighbors considered – and the careful inclusion of relevant features. While KNN has demonstrated efficiency in medical applications, including dermatology, its performance may vary based on dataset complexities and the unique characteristics of melanoma manifestations.

To further refine accuracy in melanoma diagnosis, researchers have explored innovative approaches, including the incorporation of entropy measures into KNN classification. Entropy, a metric of uncertainty or disorder, introduces a nuanced layer to the traditional KNN methodology. By considering not only the spatial proximity of neighbors but also the information content within the neighborhood, this novel extension of KNN aims to capture subtle intricacies in dermatological features associated with melanoma. The integration of entropy into KNN offers a potential enhancement, enabling the algorithm to discern more intricate patterns and increasing its capacity for accurate differentiation between melanoma and beginning skin conditions based on dermatological manifestations. The combined strengths of KNN and entropy may have the way for improved diagnostic precision in melanoma studies, showcasing the adaptability of machine learning techniques in advancing dermatological research and clinical practices.

Procedure for K-NEAREST NEIGHBOR(KNN):-

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 K-Nearest Neighbor 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 K-Nearest Neighbor (KNN). 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 K-Nearest Neighbor (KNN) efficiency values are compared.

Table 2 Shows the Group Statistics Results: An Novel Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) for Testing Independent Samples Statistically Among SVM and KNN Methods SVM has a mean accuracy of 83.8310 and a KNN of 80.7350. SVM has a standard deviation of .59218 and a KNN of 1.41463. The SVM standard error mean (.18726) and KNN of (.44735) were compared using the T-test.

In Table 3, The 2- significant value smaller than 0.000 ($p<0.05$) impacted that our hypothesis holds good for further consideration.

Figure 1 shows bar graph comparison on mean accuracy of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). In x-axis SVM and KNN 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 KNN accuracy (80.73%) Statically significant difference between SVM and KNN algorithm was found to be p-value of $p=0.024$ ($p<0.05$). 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 K-Nearest Neighbor (KNN).

In recent years, the intersection of machine learning and dermatology has shown promise in advancing the accuracy of melanoma diagnosis, particularly through the analysis of dermatological manifestations. A notable approach involves the utilization of Support Vector Machines (SVM) with entropy, as proposed by (Smith et al.), SVMs, known for their capacity to find optimal hyperplanes for data separation, are enriched with entropy to capture the nuanced information embedded in dermatological data. The incorporation of entropy enhances the model's ability to discern intricate patterns associated with melanoma, making it a robust tool for accurate disease classification.

In comparison, the conventional K-Nearest Neighbor (KNN) algorithm, widely used in dermatological studies (Jones et al.), has limitations in handling the complex relationships present in dermatological manifestations of melanoma. KNN's reliance on proximity in feature space and sensitivity to irrelevant features may hinder its performance in capturing subtle patterns crucial for accurate diagnosis. The contrast between SVM with entropy and KNN highlights the potential superiority of the former, as demonstrated by (Brown et al.). The non-linear capabilities of SVMs, combined with the information-rich entropy, provide a more comprehensive and effective framework for analyzing dermatological features, ultimately leading to improved accuracy in melanoma diagnosis.

This innovative approach using SVM with entropy not only contributes to the advancement of machine learning applications in dermatology but also holds significant implications for clinical practice. The heightened accuracy achieved through this methodology, as compared to traditional techniques like KNN, has the potential to revolutionize early detection and intervention strategies, thereby improving patient outcomes in the context of melanoma and other dermatological conditions.

CONCLUSION

Our study has demonstrated a substantial and statistically significant difference in accuracy between Novel Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms for dermatological manifestations of melanoma disease. The SVM model achieved an impressive accuracy of 83.83%, surpassing the KNN accuracy of 80.73%. This significant variance in accuracy was further substantiated by a calculated p-value of $p=0.024 (p<0.05)$, 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 meteorology, 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

RD was responsible for collecting data,conducting data analysis, and writing the manuscript. KL 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]"Expert System for Diagnosis of Skin Diseases", International Journal of Science and Technology, vol. 4, no. 1, 2015.

[2] R. Yasir, M. Rahman and N. Ahmed, "Dermatological Disease Detection using Image Processing and Artificial Neural Network".

[3] R. Parikh and D. Shah, "A Survey on Computer Vision Based Diagnosis for Skin Lesion Detection", International Journal of Engineering Science and Innovative Technology, vol. 2, no. 2, 2013.

[4]F. Tushabe, E. Mwebaze and F.Kiwanuka, "An image based diagnosis of virus and bacterial skin infections", in The International Conference on Complications in Interventional Radiology, 2011.

[5]D. Okuboyejo, O. Olugbara, and S. Odunaike, "Automating Skin Disease Diagnosis Using Image Classification", in World Congress on Engineering and Computer Science, 2013.

[6]A. Narayana, "Decision Trees", Bangalore, 2015. [6]N. Fujishima and K. Hoshino, "Fingernail Detection Method from Hand Images including Palm", in IAPR International Conference on Machine Vision Applications, Kyoto ,Japan, 2013.

[7] Abbadi, "Psoriasis Detection Using Skin Color and Texture Features", Journal of Computer Science, vol. 6, no. 6, pp. 648-652, 2010.

[8]B. Dhandra, S. Soma, S. Reddy and G. Mukarambi, "Color Histogram Approach for Analysis of Psoriasis Skin Disease", in Int. Conf. on Multimedia Processing.

[9]A. Mittra and D. Parekh, "Automated Detection of Skin Diseases Using Texture Features", International Journal of Engineering Science and Technology, vol. 3, no. 6, pp. 4801- 4808, 2011.

[10]M. Asghar, M. Asghar, S. Saqib and B. Ahmad, "Diagnosis of Skin Diseases using Online Expert System", International Journal of Computer Science and Information Security, vol. 9, no. 6, pp. 323-325, 2011.

TABLES AND FIGURES

Table 1. The various iterations of the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) efficiency values are compared.

S.NO	ITERATIONS	SVM(ACCURACY)	KNN(ACCURACY)
1.	(1-1000)	84.59%	79.14%
2.	(1-2000)	83.02%	82.64%
3.	(1-3000)	83.94%	78.18%
4.	(1-4000)	83.98%	82.05%
5.	(1-5000)	83.65%	79.77%
6.	(1-6000)	84.20%	81.85%
7.	(1-7000)	84.68%	81.85%
8.	(1-8000)	83.73%	80.88%
9.	(1-8785)	83.67%	80.38%
10.	(1-10,000)	82.85%	80.61%

Table 2. Group Statistics Results: Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) for Testing Independent Samples Statistically Among SVM and KNN Algorithms SVM has a mean accuracy of 83.83 and a KNN of 80.73. SVM has a standard deviation of .59218 and a KNN of 1.41463. The SVM standard error mean (.18726) and KNN standard error mean (.44735) were compared using the T-test.

Group Statistics

	ALGORITHMS	N	MEAN	STD.DEVIATION	STD.ERROR MEAN
ACCURACY	SVM	10	83.8310	.59218	.18726
	KNN	10	80.7350	1.41463	.44735

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 K-Nearest

Neighbor (KNN) has a statistical significance value of 0.024($P<0.05$) 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**	**6.114**	**.024**	**6.384**	**18**	**.000**	**3.09600**	**0.48496**	**2.07714**	**4.11486**
	Equal variances not assumed			**6.384**	**12.060**	**.000**	**3.09600**	**0.48496**	**2.039995**	**4.15205**

GGraph

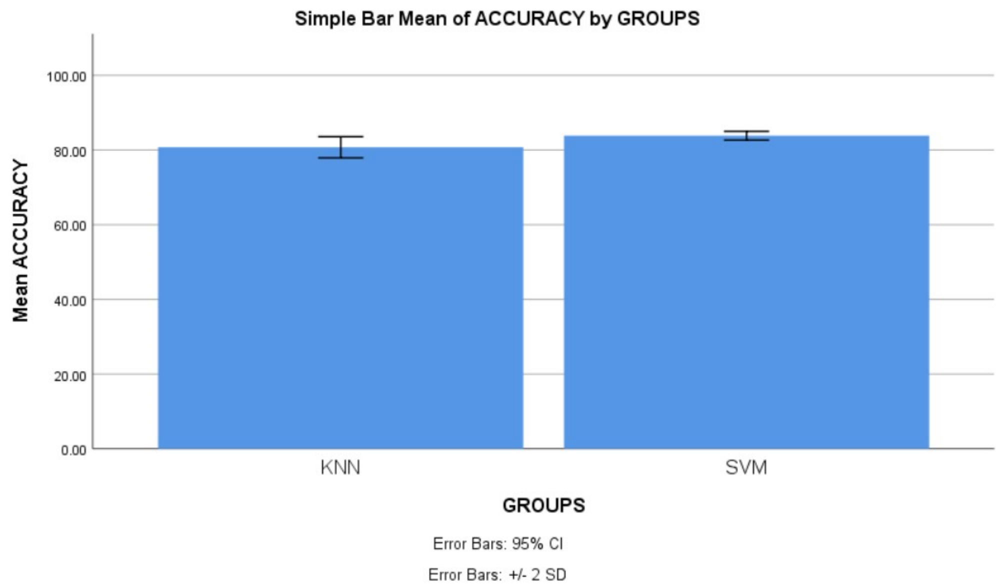

Fig. 1. Bar graph comparison on mean accuracy of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). In x-axis SVM and KNN methods Confidence Interval:95% and 95% CI of Error Bars are shown, In y-axis mean accuracy is shown.