Isabela State University Linker:
Journal of Engineering, Computing, and Technology
Volume 1, Issue 1
ISSN (Print): 3082-3676 ISSN (Online): 3082-3684
DOI: https://doi.org/10.65141/ject.v1i1.n8

Skin Classification Via RGB Values and Machine Learning Algorithms

Vince R. Gallardo[1], Edwin R. Arboleda[2], James Patrick M. Dones[3], John Mark M. Panganiban[4]
Department of Computer and Electronics Engineering, Cavite State University, Philippines [1,2,3,4]
edwin.r.arboleda@cvsu.edu.ph

RESEARCH ARTICLE INFORMATION	ABSTRACT
Received: May 26, 2023 **Reviewed:** July 20, 2023 **Accepted:** June 03, 2024 **Published:** June 29, 2024 	This research proposed a skin and non-skin classification method using R, G, and B color space and machine learning algorithms. The methodology created a dataset of skin and non-skin samples and used MATLAB's Classification Learner App with 25 different algorithms. The Weight KNN training model was identified as the most accurate and fastest, achieving 100% classification accuracy with 0.20881 seconds of time accuracy. This approach has potential applications in medical image analysis, computer vision, and facial recognition. The study suggests the Weight KNN training model is the most effective and accurate for skin and non-skin classification using R, G, and B color space and machine learning algorithms.

Keywords: *classification learner app, MATLAB, RGB color space, skin and non-skin, Weight KNN*

Introduction

The challenge of tolerating varying lighting conditions within an image sequence for skin-color tracking systems has been extensively studied in the literature (Rahmat et al., 2016; Subban & Mishra, 2013). To address this issue, color spaces that are luminance invariant have been proposed to provide robustness (Dwina et al., 2018). One such color space is the RGB color space, which represents the three types of photoreceptors in the human eye, corresponding to the red, green, and blue spectrum. The RGB color space is a three-dimensional coordinate system that uses the Cartesian coordinate system to describe the three primary colors (Rahmat et al., 2016). Color space-based models have been widely used for skin-like region detection and can provide a quick and efficient method for identifying skin regions before moving on to more complex tasks such as face and body identification and tracking (Almohair et al., 2014; Dahal et al., 2016). The main objective of this paper was to propose a skin and nonskin classification method with high accuracy and processing time using the R, G, and B color space constructed using the red, green, and blue primary colors. The differentiation of skin and non-skin areas was achieved using this color space (Joseph & Panicker, 2017; Lameski et al., 2019; Liihuhqw et al., 2019; Liu et al., 2019; Ninh et al., 2019). To classify the skin and non-skin samples, the researchers used the Weight KNN algorithm, which was trained on a dataset consisting of thousands of random non-skin samples and different skin textures. The Weight KNN algorithm achieved high accuracy and processing time, making it an effective and efficient method for skin and non-skin classification.

Methods

Skin and Non-Skin Samples T

This study used 4000 samples from the UCI Machine Learning Repositories, with an equal number of 2000 samples for skin and non-skin categories. The samples were distinguished from each other based on their RGB colors, which enabled an analysis of the differences in RGB values between skin and non-skin samples.

Figure 1. *Skin and Non-Skin Samples*

Extraction of Samples

The RGB values were extracted from the skin and non-skin samples to evaluate whether they could be distinguished using the MATLAB Classification Learner App. This study utilized different approaches such as model-based, threshold-based, and region-based approaches (Abuse, 2014) to classify the data. Specifically, the region-based approach was utilized to segment the data.

Classification

In order to differentiate between skin and non-skin samples, the researchers employed a comprehensive set of 25 learning algorithms from the powerful MATLAB Classification Learner App (CLA). These algorithms were put to the test with the RGB color attributes data extracted from the skin and non-skin datasets, which were then used to train the CLA in five (5) folds. Specifically, 1600 data points were used to train the skin and non-skin datasets, while 400 were reserved for testing and validation. To ensure the best classifier for these samples, we carefully weighed the accuracy and training time of each algorithm, ultimately identifying the optimal solution for this important task.

Ethical Considerations

It is important to consider ethical protocols when conducting research involving human or animal subjects. In this study, skin and non-skin samples were collected from a public dataset and a machine learning algorithm was used to classify them. No human or animal subjects were involved in the study.

Results and Discussion

The 4000 samples and 5-fold cross-validation were used to input the data into the MATLAB Classification Learner App. This means that for training, 1600 skin and non-skin data samples were used, and for testing and validation, 400 data samples were used. The result of the classification obtained from the CLA is presented in Table 1.

The table shows the classification results for the 25 machine-learning algorithms. When two or more algorithms have the same highest accuracy, the training time of those models is considered, and the model with the fastest training time is chosen as the best. The shortest training time would be considered since the data shows that six of them had the same accuracy of 100%. The Weight KNN model achieved the highest accuracy with the shortest training time in this analysis, with 100% accuracy and 0.20881 training time. RUS Boosted Trees had the lowest accuracy, with 50% accuracy and 0.31 training time. The MATLAB CLA feature Principal Component Analysis (PCA) was disabled in the classification.

Table 1. Classification Results

Classifier Type	Accuracy (%)	Training Time (Seconds)
Fine Tree	99.9	0.86466
Medium Tree	9.99	0.16457
Coarse Tree	99.8	0.1686
Linear Discriminant	99.8	0.24729
Quadratic Discriminant	99.9	0.18657
Logistic Regression	99	0.52308
Gaussian Naïve Bayes	74.2	0.17011
Kernel Naïve Bayes	97.4	1.3067
Linear Support Vector Machine (SVM)	98.9	5.0428
Quadratic SVM	99.8	1.2625
Cubic SVM	100	0.4611
Fine Gaussian SVM	100	0.44086
Medium Gaussian SVM	100	0.38789
Coarse Gaussian SVM	98.7	0.53611
Fine KNN	100	0.26183
Medium KNN	99.9	0.21435
Coarse KNN	98.7	0.32179
Cosine KNN	99.8	0.34901
Cubic KNN	99.9	0.24998
Weight KNN	100	0.20881

The scatterplot of the Weight KNN model is displayed in Figure 2. The skin and non-skin data samples are distinguishable by their respective colors, with blue representing skin samples, and orange representing non-skin samples. Interestingly, among the six classifiers that yielded a 100% accuracy, namely the Cubic SVM, Fine Gaussian SVM, Fine KNN, Subspace KNN, and Weight KNN, the Weight KNN boasted the shortest training time of just 0.20881 seconds.

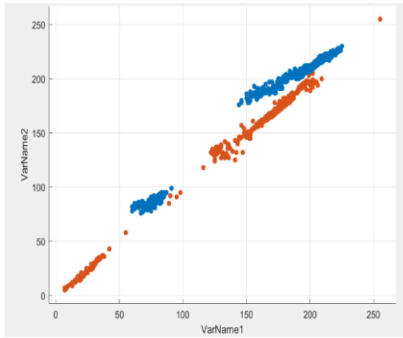

Figure 2. *Scatter Plot of Data Samples Using Weight KNN*

Figure 3 shows that the skin and non-skin samples both achieved 100 percent accuracy in the confusion matrix.

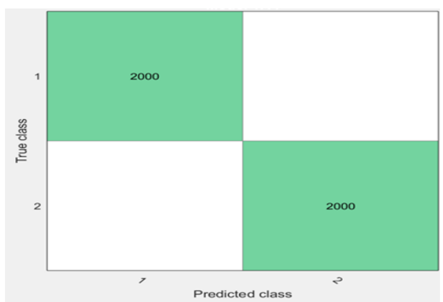

Figure 3. *Confusion Matrix of Skin and Non-Skin Using Weight KNN*

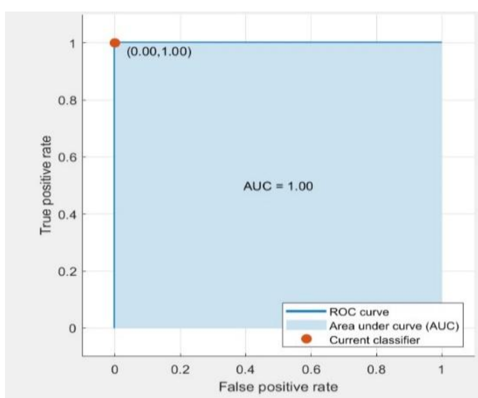

Figure 4. *ROC Curve Using Weight KNN*

It can be seen in Figure 4 that for the Weight KNN model, the skin and non-skin samples were all correctly classified. This means that this model Weight KNN model can identify the samples of skin and non-skin samples without any error.

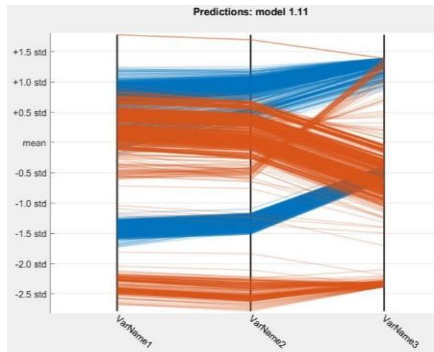

Figure 5. *Parallel Coordinates Plot Using Weight KNN*

Figure 5 shows the parallel coordinates plot of Weight KNN. The two classes were separated by color (see Figure 1). It can be seen that the skin samples are above and some of them are between the non-skin samples. Although the samples are overlapping, the Weight KNN was able to separate and differentiate the two classes. The MATLAB CLA feature Principal Component Analysis (PCA) was disabled in the classification.

Conclusion and Future Works

The main goal of this study was to determine the optimal machine-learning technique for skin and non-skin classification. The Weight KNN model was found to be the most suitable approach, with 100% accuracy and a training time of only 0.20881 second. This outperformed the Fuzzy Decision Tree (FDT) technique, which achieved an average accuracy of 94.10%.

However, this study was limited to using the RGB color space and may not be generalizable to other color spaces. Further research can investigate other color spaces and additional features for classification. Additionally, the dataset used in this study was restricted to skin and non-skin samples and may not be representative of other image classification tasks. Thus, future studies using larger and more diverse datasets are recommended.

Moreover, several recommendations should be considered for future works. Firstly, alternative machine learning techniques such as SVM or neural networks can be explored to determine their effectiveness in achieving higher level of accuracy or faster training times compared to the Weight KNN model used in this study. Secondly, other color spaces and image features such as texture or shape can be investigated to improve the accuracy of the classification models. Thirdly, multi-class classification can be explored to further categorize skin into different groups based on skin tone or texture. Fourthly, larger and more diverse datasets can be used to test the generalizability of the Weight KNN model and other machine-learning techniques for skin and non-skin classification. Finally, comparing the performance of different models that use various combinations of color spaces and image features can provide insights about the most effective features and combinations for this task.

Overall, this study presents a reliable and efficient approach to skin and non-skin classification using machine learning algorithms. The Weight KNN model has potential applications in various fields, including medical image analysis, facial recognition technology, and computer vision.

References

[1] Abuse, C. (2014). *Color-based skin segmentation: An evaluation of the state of the art* (Frerk Saxen & Ayoub Al-Hamadi). In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 4467–4471). IEEE.

[2] Al-Mohair, H. K., Mohamad-Saleh, J., & Suandi, S. A. (2014). Color space selection for human skin detection using color-texture features and neural networks. In *Proceedings of the International Conference on Computer and Information Sciences (ICCOINS 2014) – A Conference of World Engineering, Science and Technology Congress, ESTCON 2014* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCOINS.2014.6868362

[3] Dahal, B., Alsadoon, A., Prasad, P. W. C., & Elchouemi, A. (2016). Incorporating skin color for improved face detection and tracking system. In *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation* (pp. 173-176). IEEE. https://doi.org/10.1109/SSIAI.2016.7459203

[4] Dwina, N., Arnia, F., & Munadi, K. (2018). Skin segmentation based on improved thresholding method. In *1st International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON 2018)* (pp. 95–99). IEEE. https://doi.org/10.1109/ECTI-NCON.2018.8378289

[5] Joseph, S., & Panicker, J. R. (2016, August 12–13). Skin lesion analysis system for melanoma detection with an effective hair segmentation method. In *Proceedings of the 2016 International Conference on Information Science (ICIS 2016)* (Vol. 1, pp. 91–96). IEEE. https://doi.org/10.1109/INFOSCI.2016.7845307

[6] Lameski, J., Jovanov, A., Zdravevski, E., Lameski, P., & Gievska, S. (2019). Skin lesion segmentation with deep learning. In *Proceedings of EUROCON 2019 – 18th International Conference on Smart Technologies* (pp. 1–5). IEEE. https://doi.org/10.1109/EUROCON.2019.8861636

[7] Liihuhqw, D. R. Q., Frp, T. T., Vr, U., Wkh, W., Fdq, V. V., Wkh, G., Prvw, I., Frpsrqhqw, F., Eh, F. D. Q., Dqg, U., Vsdfh, W. K. H., & Space, A. R. G. B. C. (2019). Research and improvement of skin color segmentation method based on different color space. In *Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (AICA 2019)* (pp. 112–116).

[8] Liu, L., Mou, L., Zhu, X. X., & Mandal, M. (2019). Skin lesion segmentation based on improved U-Net. In *Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE 2019)* (pp. 1–4). IEEE. https://doi.org/10.1109/CCECE.2019.8861848

[9] Ninh, Q. C., Tran, T. T., Tran, T. T., Anh Xuan Tran, T., & Pham, V. T. (2019). Skin lesion segmentation based on modification of SegNet neural networks. In *Proceedings of the 2019 6th NAFOSTED Conference on Information and Computer Science (NICS 2019)* (pp. 575–578). IEEE. https://doi.org/10.1109/NICS48868.2019.9023862

[10] Rahmat, R. F., Chairunnisa, T., Gunawan, D., & Sitompul, O. S. (2016). Skin color segmentation using multi-color space threshold. In *Proceedings of the 2016 3rd International Conference on Computer and Information Sciences (ICCOINS 2016)* (pp. 391–396). IEEE. https://doi.org/10.1109/ICCOINS.2016.7783247

[11] Subban, R., & Mishra, R. (2013). Face detection in color images based on explicitly-defined skin color model. In D. Zhou, X. Luo, & S. Hu (Eds.), *Communications in Computer and Information Science* (Vol. 361, pp. 570–582). Springer. https://doi.org/10.1007/978-3-642-36321-4_54

Acknowledgment

The author would like to express their sincere gratitude to the creator of the dataset utilized in this study. The dataset proved to be an invaluable resource, allowing for the successful completion of the research.