International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 12 | December - 2022 Impact Factor: 7.185 ISSN: 2582-3930

SKIN TYPE CLASSIFIER

Aditya Chaurasia[1], Nihal Jha[2], Sangeeta Parshionikar[3]

[1]*Student, Electronics and Computer Science, Fr. Conceicao Rodrigues College of Engineering*
[2] *Student, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering*
[3] *Professor, Computer Engineering, Fr. Conceicao Rodrigues College of Engineering*

---***---

Abstract - Variations in skin diseases largely fall into the subset of skin types. Classification of Skin Types plays an essential part in the classification of skin diseases. The paper proposes an end to-end machine learning model used to classify skin type as dry, oily, or normal using Python Libraries. Skin type classification is done using Gray-Level Co-Occurrence Matrix (GLCM) for texture identification and Light Gradient Boosting Machine (LGBM). Red Green Blue (RGB) images are first resized and then converted into Gray-Scale images and then features like energy, correlation, dissimilarity, homogeneity, contrast, and entropy are extracted from a given image using GLCM. Such features give us a basic understanding of the type and texture of the image. The GLCM of the image is calculated by changing the distance and angle of the pixels within the image in a regular interval. These features after extraction are fed to the LGBM algorithm. The trained model is then deployed using Flask.

***Key Words*:** features extraction, GLCM, LGBM, skin type

1.INTRODUCTION

With cosmetic companies spending billions on marketing their products which purportedly brightens the skin, it has made people inquisitive to find their skin type. The sudden surge in the curiosity to know the skin type is not at all surprising. It is preferable to diagnose skin disorders early on in order to prevent them from spreading, and it is preferable to classify them according to the skin type they belong to. As it is the outermost layer of the human body it is prone to adverse states and conditions. Long-term skin disease can cause irritants, allergies, immune system issues, and a variety of other problems [5]. Several more authors use various machine learning techniques to classify skin diseases [4] and types. Classification of skin types will give a better understanding of their skin and the respective precaution they need to follow. Nowadays humans are more conscious about their skin types and are eager to know their skin type. This proposed model can also help dermatologists in predicting the skin type of a person by using machine learning. A computer program can autonomously catch the type of skin from an image that is digital, captured from a digital camera system with less stress on the region that is of interest, the patient or the victim can recognize the type of skin from the comfort of their home [4]. Normal, oily and dry are the three types in which the type of skin can be classified. Rough, itchy, flaky, or scaly are the characteristics of dry skin. The spots appear in different parts of the face of different people. Oily skin, on the other hand, can clog pores and contribute to more acne breakouts. Python supports powerful libraries like cv2, tensorflow, numpy, and keras. Figure 1 depicts the proposed method's flow chart. There are five sections to this study. Section 2 mentions the dataset. Section 3 explains the feature analysis; type of images, and types of features extracted from the grayscale image. Section 4 presents Light Gradient Boosting Machine (LGBM) and its application. Section 5 presents the final results and the deployed model and finally, Section 6 presents the conclusion.

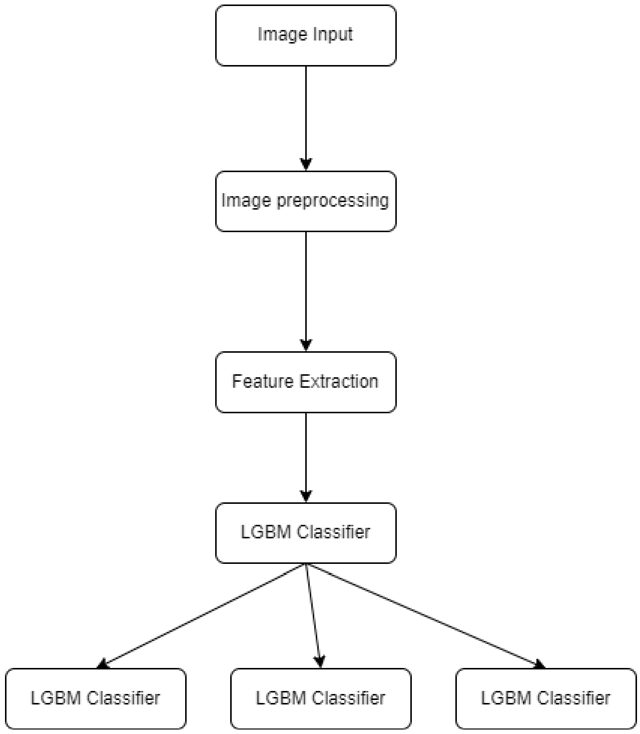

Fig -1: Block Diagram for the proposed method

2. DATASET

The dataset comprises images of dry, oily and normal skin. We gathered the datasets from different online websites to ensure that we are able to get different types of dry, oily and normal skin images. The total number of images used for the sake of training the model is 508. The number of images kept for testing the accuracy is 119. Such a large number of images in the data helps in achieving high accuracy. The train to test split ratio is approximately 80% and 20% respectively.

3. FEATURE ANALYSIS

Color and texture information are the most prominent features used to visually describe and identify skin types and skin diseases. Color information is critical for distinguishing one skin type from another. Color histograms, colour correlograms, colour descriptors, and GLCM can all be used to extract these colour features [2]. The texture information conveys the skin's complex visual patterns as well as spatially organized entities like brightness, colour, shape, and size.

3.1 Grey Level Co-Occurrence Matrix

The co-occurrence matrix illustration was proposed by R.M. Haralick, it researches the texture and its spatial dependence on the grey level [3]. Co-occurrence matrix or gray-level co-occurrence matrix is defined as a matrix of an image, as the distribution of co-occurring values of pixels at a given offset. A texture analysis and research method, also applied in a broad spectrum of applications, which include medical image analysis. Fig. 2 GLCM is formed from grey co-matrix of a 4 by 5 image where θ = 0 and D=1 [6].

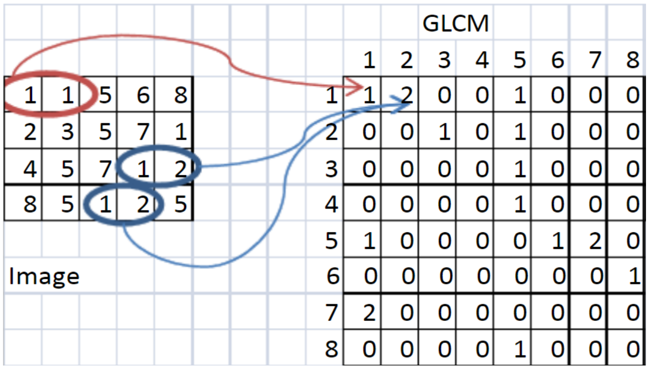

Fig -2: GLCM calculation [13]

3.2 Feature Extraction

It is a critical stage in texture analysis since it extracts textural properties from an image for future processing, such as recognition or segmentation [1]. The six textural features retrieved from the grey-level co-occurrence matrix are correlation, contrast, homogeneity, entropy, dissimilarity, and energy. The amount of grey levels in the image, and pd(i,j) is the (i, j)th element in GLCM, expressing the probability of pixel pairs being present at a given distance and angle is Ng [6].

3.2.1 *Energy*

- It is a metric for picture homogeneity that may be derived using the normalized grey level co-occurrence matrix.
- It is an appropriate metric for detecting texture image disturbance.
- In the GLCM, it calculates the summation of the elements that are squared.
- Uniformity or the angular second moment are other names for it.

$$Energy\ (d,\theta) = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} [p_d^{\theta}](i,j)^2 \qquad (1)$$

3.2.2 *Entropy*

- It is a metric of randomness. The entropy of an image is a measure of its complexity. Entropy is higher in complex texture values.
- The absence of data or a statement in a signal conveyed, as well as the image data, is measured by entropy.

$$Entropy = \sum_{i=1}^{Ng-1} \sum_{j=1}^{Ng-1} P_{ij} * log(P_{ij}) \qquad (2)$$

3.2.3 *Correlation*

- Correlation tells us how closely two pixels are related.
- Correlation is used to assess the linear dependency of grey levels of surrounding pixels.
- Digital Image Correlation is an optical method that measures changes in images in 2D and 3D using tracking and image registration techniques. Its range is -1 to 1. where -1 represents perfect negatively correlated data, 0 represents uncorrelated data, and 1 represents perfect positively correlated data [6].

$$Correlation = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{ij\ P_d^{\theta}(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \qquad (3)$$

3.2.4 *Homogeneity*

- It assesses consistency of GLCM's nonzero entries [7].
- It returns a value indicating the proximity of the GLCM element distribution to the GLCM diagonal. It has a value range of 0 to 1.

$$Homogeneity\ (d,\theta) = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{1}{1+(i-j)^2} P_d^{\theta}(i,j) \qquad (4)$$

3.2.5 *Dissimilarity*

- The dissimilarity feature computes the distance between two items in the ROI [8].
- It computes the grey level mean difference in the distribution of the image [8].
- A higher number indicates a bigger gap in intensity levels between neighbouring pixels [8].
- Equation of Dissimilarity

$$Dissimilarity = \sum_i \sum_j |i-j| P(i,j) \qquad (5)$$

3.2.6 *Contrast*

- The contrast of a picture is a measure of its spatial frequency, i.e., the intensity of each pixel as well as the neighbours across the image [8].
- It also determines how much variation there is in the image at a local level.
- The contrast value is a visual measure of variation between grey degrees in an image area [9].

$$Contrast\ (d,\theta) = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} |i-j|^2\, P_d^{\theta}(i,j) \quad (6)$$

4. Light Gradient Boosting Machine (LGBM)

For classification, the Light Gradient Boosting Machine, which is built on a Decision Tree algorithm. It grows the tree leaf by leaf rather than level by level [11]. LGBM finds the best split candidates using a histogram-based method. It produces a sorted result with the best attribute values. It effectively addresses the issue of over-fitting. To improve training, LGBM employs the Gradient-based One-Side Sampling (GOSS) sampling algorithm [10]. To deal with sparsity in datasets, LGBM employs the Exclusive Feature Bundling technique. Instead of the traditional level-wise algorithm, LightGBM employs a leaf-wise algorithm with depth limitation, which helps to improve accuracy and inhibits over-fitting [12].

5. RESULTS

The images are converted from RGB to grayscale. It calculates features of the images by changing the angle by 45 degrees of every pixel and also by changing the distance of every pixel for each image. The data frame containing these feature values is trained to the LGBM classifier. As shown in the Fig.3, the model is deployed on localhost using Flask. The confusion matrix is shown below.

Table -1: Confusion Matrix

	1(YES)	0(NO)
1(YES)	74	1
0(NO)	13	17

Precision: Denotes the percentage of correctly predicted positive cases [14].

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall: The fraction of positive instances that were correctly identified, also termed as the True Positive Rate [14].

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

The precision of the model is 0.98, and the recall is 0.85, where TP represents True Positives, TN denotes True Negatives, FP denotes False Positive, and FN denotes False Negative. The model is very efficient in classifying skin types into dry and oily. The accuracy of the model is approximately 87% it successfully classifies skin types into dry and oily. It is evident through the confusion matrix.

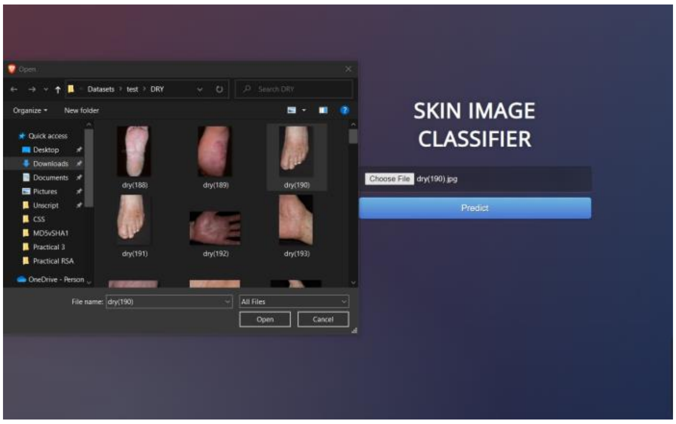

Fig -3: UI of the application

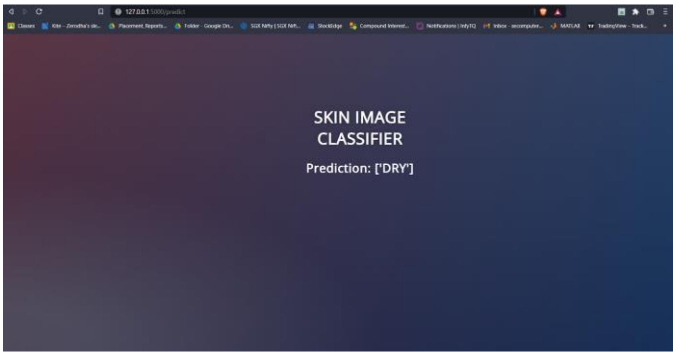

Fig -4: Output of the application

6. CONCLUSION

Skin Type Classification is an important step in the detection of skin diseases and disorders, creating a model for this was essential. Skin type classification was successfully executed using GLCM and LGBM. An accuracy of 87% achieved. Further extraction of features in a more efficient way and refining the LGBM model can increase the accuracy even more. Thus, enabling growth and further research.

REFERENCES

1. S. W. . -C. Lam, ”Texture feature extraction using gray level gradient based co-occurrence matrices,” 1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929), 1996.
2. T. Goswami, V. K. Dabhi and H. B. Prajapati, ”Skin Disease Classification from Image - A Survey,” 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020.
3. G. Preethi and V. Sornagopal, ”MRI image classification using GLCM texture features,” 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), 2014.
4. H. Janoria, J. Minj and P. Patre, ”Classification of Skin Disease from Skin images using Transfer Learning Technique,” 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020.
5. O. K. Pal, ”Skin Disease Classification: A Comparative Analysis of KNearest Neighbors (KNN) and Random Forest Algorithm,” 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), 2021.
6. S. Singh, D. Srivastava and S. Agarwal,” GLCM and its application in pattern recognition,” 2017 5th International Symposium on Computational and Business Intelligence (ISCBI), 2017.
7. A. Gebejes and R. Huertas, “Texture characterization based on greylevel co-occurrence matrix,” Proc.Conf. Informatics and Manage. Sci., vol. 2, March 2013.
8. P. K. Mall, P. K. Singh and D. Yadav, ”GLCM Based Feature Extraction and Medical X-RAY Image Classification using Machine Learning Techniques,” 2019 IEEE Conference on Information and Communication Technology, 2019.
9. E. Resita, R. Sigit, T. Harsono and R. Rahmawati, ”Color RGB and Structure GLCM Method to Feature Extraction System in Endoscope Image for The Diagnosis Support of Otitis Media Disease,” 2020 International Electronics Symposium (IES), 2020.
10. F. Alzamzami, M. Hoda and A. E. Saddik, ”Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation,” in IEEE Access, vol. 8, pp. 101840-101858, 2020.
11. G. Shobana and K. Umamaheswari, ”Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling,” 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021.
12. D. Zhang and Y. Gong, ”The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure,” in IEEE Access, vol. 8, pp. 220990-221003, 2020.
13. http://in.mathworks.com/help/images/ref/graycoprops.html
14. ”http://www2.cs.uregina.ca/∼dbd/cs831/notes/confusionmatrix/confusionmatrix.html”