Computer Assisted Methods in Engineering and Science
30(2): 247–256, 2023, doi: 10.24423/cames.479

This article belongs to the *Special Issue on Scientific Computing and Learning Analytics for Smart Healthcare Systems* edited by Dr. C. Chakraborty, Dr. S. Barbosa and Dr. L. Garg

Classification and Detection of Skin Disease Based on Machine Learning and Image Processing Evolutionary Models

Dibyahash BORDOLOI[1)*], Vijay SINGH[1)],
Karthikeyan KALIYAPERUMAL[2)], Mahyudin RITONGA[3)],
Malik JAWARNEH[4)], Thanwamas KASSANUK[5)],
Jose QUIÑONEZ-CHOQUECOTA[6)]

[1)] *Department of Computer Science and Engineering, Graphic Era Deemed to be University, India*

[2)] *Department of Information Technology, School of Informatics and Electrical Engineering, Ambo University, Ethiopia*

[3)] *Universitas Muhammadiyah Sumatera Barat, Indonesia*

[4)] *Faculty of Computing Sciences, Gulf College, Oman*

[5)] *Faculty of Food and Agricultural Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand*

[6)] *Universidad Nacional del Altiplano, Peru*

[*] *Corresponding Author e-mail: dibyahash.cse@geu.ac.in*

Skin disorders, a prevalent cause of illnesses, may be identified by studying their physical structure and history of the condition. Currently, skin diseases are diagnosed using invasive procedures such as clinical examination and histology. The examinations are quite effective and beneficial. This paper describes an evolutionary model for skin disease classification and detection based on machine learning and image processing. This model integrates image preprocessing, image augmentation, segmentation, and machine learning algorithms. The experimental investigation makes use of a dermatology data set. The model employs the machine learning methods: the support vector machine (SVM), the k-nearest neighbors (KNN), and random forest algorithms for image categorization and detection. This suggested methodology is beneficial for the accurate identification of skin disease using image analysis. The SVM algorithm achieved an accuracy of 98.8%. The KNN algorithm achieved a sensitivity of 91%. The specificity of KNN was 99%.

Keywords: skin disorders, machine learning, classification, image enhancement, image segmentation, disease detection.

1. Introduction

Skin diseases are a common source of illnesses that may be recognized by examining the condition's physical structure and history. Skin diseases are now diagnosed using invasive methods such as clinical examination and histology. When medicine and technology are used together, they can produce good results in the future. Artificial intelligence (AI) could be used to process a lot of data and find timely, relevant and knowledge-based information so that people can make good decisions about their health. The skin protects from injuries, harmful pathogens (such as bacteria), ultraviolet (UV) rays and other factors. It is the first line of defense. It has to bear the brunt of being vulnerable to a lot of diseases because of where it is and how big it is. It has been a long time since skin problem detection methods and medications to cure them have changed. In addition, a cure has yet to be found for many skin illnesses and conditions [1, 2].

When it comes to skin diseases, dermatologists have been having trouble because they do not have the right information, which is what is needed. Skin diseases or illnesses linked to them have had an effect on overall health at both individual and global levels. When researchers looked at the Global Burden of Skin Diseases project, they found that there are fifteen types of skin diseases that are the most common. They also looked at how these diseases affect people around the world in terms of their health. Eczema, psoriasis, acne, pruritis and impetigo are some of the common types of skin infections that people develop. Molluscum warts, urticaria, viral warts, scabies, fungal skin disorders, abscesses, bacterial skin disorders and cellulitis are also common. Other skin and subcutaneous diseases take up the rest of the diseases [3, 4].

Many professionals in the medical business study symptomatic imaging, such as X-rays, magnetic resonance images (MRIs), computer tomography (CT) scans, positron emission tomography (PET) scans, and other forms of images, as well as electrocardiogram (ECG) and electromyography (EMG) signal data. A better grasp of the disease's classification would help clinicians provide the best care possible to those suffering from it. Machine learning designs can address highly crucial difficulties by spontaneously detecting the input data characteristics, and deep learning models can adapt to changes in the problem that they seek to solve. Machine learning models will use the concluded data to find and analyze properties in the data patterns that have not been presented. This will save a lot of time because even low-computing models will be able to perform it [5].

The process of digitally capturing an object's visual qualities is known as image capture [6]. An image is a name given to this type of item. There are many different ways to use and exhibit it, and it can be printed and displayed in various sizes or locations. Images are enhanced, segmented, features extracted, and matched as part of the "process". The practice of improving photographs to better display or analyze the results is known as "image enhancement". Noise reduction, sharpening and/or lightening of the image, cropping the image, and histogram equalization are all examples of image improvement [7]. The image segmentation approach helps to reduce processing time and energy consumption by choosing the region of interest in an image that benefits from image enhancement methods. For this objective, it is appropriate to use image segmentation procedures such as thresholding, color-based segmentation, transform, and texture approaches. In order to make a decision, you must first select a region of interest and then extract the appropriate features from that area. Different methods can be used to extract different sorts of color and/or texture information from an image. To handle challenges such as object observation and identification, content-dependent image retrieval and quality categorization, characteristics extraction and matching approaches can be coupled. The key color or pattern information is converted into feature values using feature extraction techniques, which serve as a link between the image data and the decision-making model. After that, machine learning algorithms use the feature values as inputs.

This article describes a skin disease categorization and prediction model based on machine learning and image processing. Image addition, image enhancement, image subdivision, categorization, and prediction are necessary to classify and detect skin disorders properly. Section 2 provides a literature review, Sec. 3 presents methodology, Sec. 4 encompasses result analysis, and Sec. 5 contains the conclusion.

2. Literature survey

In [8], the authors collected 813 images from patients with five different skin conditions, which would aid in the diagnosis of these conditions. Red, green, and blue (RGB) color means were extracted from each image using the median filter, image sharpening, and binary masking techniques. An artificial neural network (ANN) classifier was used to examine these color values, and 90% classification accuracy was obtained.

Herpes, paederus dermatitis, and psoriasis were studied using 10 reference samples and 20 test samples [9]. When the target area was narrowed down using a combination of the median filter and a marker-controlled watershed technique with clustering, it was possible to extract the gray level cooccurrence matrix (GLCM) texture features. The four GLCM characters were examined using the SVM categorization and achieved 85%, 90%, and 95% accuracy for the studied herpes, paederus dermatitis, and psoriasis cases, respectively.

Images of eczema, impetigo, and psoriasis were used in [10] to classify healthy and diseased skin regions. The Dermnet skin disease atlas database was used to gather these images. Contrast enhancement, median filtering, and maximum entropy thresholding methods were used to process images in the beginning and pick regions of interest. The feedforward approach of ANN was used to extract and analyze GLCM texture features. This technique has an overall accuracy rate of 80%, with a sensitivity of 71.4% and a specificity of 87.5%.

A study [11] investigated the performance of a convolutional neural network (CNN) tool for categorizing skin diseases vs characterizing skin lesions. Six web databases, AtlasDerm, Danderm, Derma, DermQuest and Dermanet, were used to gather 75 665 photos. For disease and lesion classification, these images were utilized to instruct a multi-class CNN for disease and another for lesion classification. Using fine-tuning learning, 27.6% of the top-1 and 57.9% of the top-5 perfection values were achieved, with an average accuracy rate of 0.42.

Kumar and Singh used the KNN classification algorithm to construct a tool for detecting different skin diseases. Their study evaluated bloody skin, skin burns, skin cancer, skin allergies, and normal skin. Histogram equalization was used to improve the images, and then the hue, saturation, and intensity value (HSV) color histogram and the speeded-up robust features (SURF) blob detection algorithms were used to extract important features. Finally, the KNN classifier was used to examine these feature values and showed good diagnostic performance [12].

Skin illnesses may be recognized and classified using a variety of current methods, many of which are automated. Epidermal detection of skin illnesses does not need radiological imaging technologies to be used in most diagnostic procedures. Through image processing procedures, such as improvement, equalization, enhancement, edge detection, and segmentation [13–15], the existing condition can be identified.

The image processing technologies such as morphological image functions for skin observation [16] are also used to classify skin disorders. Thresholding generates a binary image, and this binary image determines a lot of the morphological methods such as region opening, region closing, disintegration, and expansion. Based on the image's texture, morphological procedures may not be adequate for assessing the damaged region's growth. Using the genetic algorithm (GA), a classification system for skin diseases is developed and implemented.

The findings of diagnosing skin illnesses using CNN are promising [17]. However, working with photos taken on a smartphone or digital camera is difficult since CNN models are neither scaling nor rotation invariant. Both neural network techniques need enormous amounts of training data to obtain the model's high performance, which in turn necessitates a substantial amount of computing effort [18]. It is harder to adapt the neural network models since they are more abstract. There is a remarkable rise in the number of qualified parameters in ANNs as the image resolution increases, which means training becomes more difficult. Deterioration and explosion of the gradient are common problems with the ANN model, and data observations using CNN do not indicate the object's size or magnitude.

Combining numerous prediction models improves the accuracy of skin disease classification using ensemble models [19, 20]. Because of overfitting issues, ensemble models do not operate well with unknown disparities between the considered sample and the population. Classification of skin disorders using a deep neural network model [21, 22] has shown impressive results. Nevertheless, the results of the experiments reveal that the model is unsuitable for images with many lesions. To achieve a respectable degree of accuracy, deep neural network models need a significant amount of training, which takes additional processing effort.

3. Methodology

Variables in camera phone photographs might alter segmentation results. To acquire the best results, the image must be enhanced and filtered. This involves three steps: scaling, noise reduction, and image enhancement. Digital photographs may include noise and other abnormalities. A basic thresholding job may become complicated due to inadequate capture. Therefore, a clean image is essential. Image noise is a random change in an image's lighting or color statistics. Additionally, other types of noise may be detected in images, such as Gaussian, shot and quantization noises. Filters such as median and Wiener filters may help remove these noises. Various morphological techniques may minimize noise. Gaussian filtering can smooth images whereas median filtering can change pixel brightness.

In the present study, the Gaussian filter is used to minimize noise. The intensity significance of each pixel is replaced by a weighted average of its neighboring pixels' intensities. This approach may smooth images while preserving their edges. The image is smoothed by Gaussian kernels using standard deviation [23].

A detailed description of the proposed methodology is shown in Fig. 1.

Images are improved after filtering to boost the human eye's ability to comprehend information. This enables more complex image processing. The produced image has a constant intensity distribution and histogram, which improves visual contrast. In images with tight contrast levels, using this approach regularly enhances the universal contrast. This strategy makes the histogram's intensities more uniform [24]. This helps low-contrast regions stand out. Histogram equalization effectively spreads out frequent intensity values.

This study uses region-dependent segmentation, which separates the ROI into portions based on texture or pattern. k-means clustering splits diverse interpretations into k clusters, with the local mean serving as a cluster pattern. The method uses the sum of the groups represented by the variable k to find

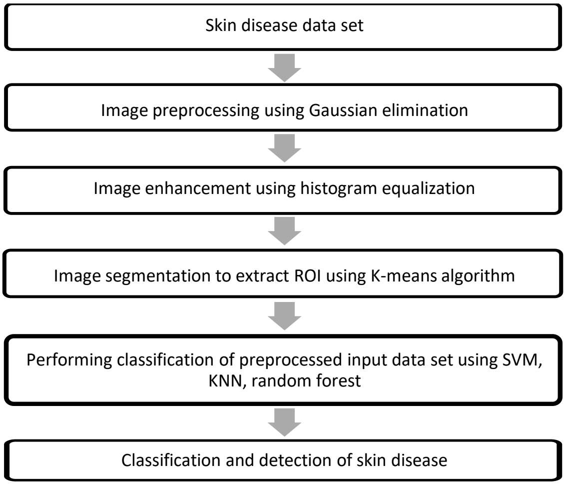

Fig. 1. Machine learning and image processing based skin disease classification and detection.

groups in data. It uses squared Euclidean distances to get the nearest data. The approach assigns each data point to a single k group based on the specified attributes. The data are categorized by feature similarity [25].

Each data point is expressed by a k-dimensional point in an SVM model (where k denotes the number of features). The value of a point is obtained from the value of each section. This is followed by the selection of a suitable hyperplane that differentiates the classes. Vapnik introduced SVM, which has stimulated the curiosity of academics all over the world. The inserted data gathered by an SVM classifier is frequently separated into two groups. Training data are used to develop a model that may be used to categorize test data. A difficulty known as multi-class classification may develop from time to time. To solve it, we will need to develop a lot of binary classifiers [26].

Random forest, an ensemble learning approach that is often used in classification tasks, is a good example. The algorithm builds numerous decision trees throughout the training process, and the mode of the outputs from each one is the output of the forest, which is then used to categorize the data. The conclusion is that random forest lower the variance of the overall model and control overfitting by creating decision trees from random samples of the training data generated by the algorithm [27].

Many researches have used the KNN classifier, which has been quite successful. Pattern recognition categorizes items using a range of algorithms developed specifically for this purpose. When employing KNN, objects are classified by comparing them to similar items in the surrounding area. KNN is all about event-based learning, for which it is most known. Using a locally estimated function to delay calculations until categorization allows faster processing. When there is no previous knowledge about the data's passage, the KNN classification strategy is the simplest classification approach. KNN is a pattern recognition algorithm that is commonly utilized. Many analysts have observed that KNN computing produces great results on a wide range of data sets, which they have uncovered [28].

4. Result analysis

For experimental analysis, the dermatology data set [29] is used. This data set comprises a total of 34 attributes. Thirty-three attributes are linear valued and one attribute is nominal valued. A total of 336 instances are available in this particular data set. It contains records related to six skin diseases: psoriasis, seborrheic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris.

Three parameters: accuracy, sensitivity and specificity are used in this study to compare the performance of different algorithms:

$$\begin{aligned}\text{Accuracy} &= (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}),\\ \text{Sensitivity} &= \text{TP}/(\text{TP} + \text{FN}),\\ \text{Specificity} &= \text{TN}/(\text{TN} + \text{FP}),\end{aligned}$$

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

The sensitivity and specificity parameters are also utilized to calculate the performance of machine learning algorithms. Sensitivity and specificity of SVM, random forest and logistic regression for skin disease detection are shown below in Fig. 2, which presents an overall comparison of classifiers. The SVM algorithm

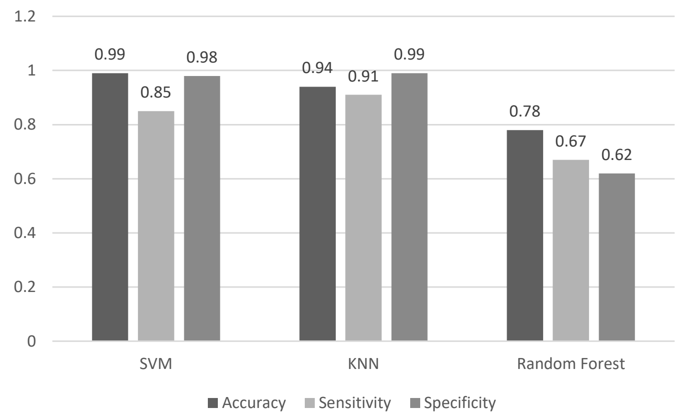

Fig. 2. Result comparison of classifiers for skin disease detection.

achieved an accuracy of 98.8%, the KNN algorithm achieved a sensitivity of 91%, and the specificity of KNN was 99%.

The performance comparison between different classifiers is shown in Fig. 2. Three parameters of accuracy, specificity and sensitivity are used for the comparative study. SVM performs better as far as accuracy is concerned. SVM and KNN perform equally well as far as specificity is concerned. KNN performs better as far as sensitivity is concerned.

5. Conclusion

Skin disorders are a prevalent source of illness, which can be detected by inspecting their physical appearance and history of the ailment. Skin disease is currently diagnosed using invasive approaches such as clinical examination and histological evaluation. These are highly effective and practical. However, these techniques necessitate subject matter expertise, require more labor and have a lower degree of trustworthiness of their results. This article presented an evolutionary model for skin disease categorization and detection, which depends on machine learning and image processing and is focused on the classification and detection of skin disorders. This model incorporated image preprocessing, image augmentation, segmentation, and machine learning algorithms, among others, for skin disease categorization and detection. For classification and detection, a dermatological data set was used, and the machine learning processes of SVM, KNN, and random forest, among others, were employed. The proposed strategy, which employed image analysis and machine learning techniques, is useful for correctly identifying skin illnesses. In the near future, a real-time skin disease detection system can be implemented by using internet of things (IoT) to collect skin disease images and then use machine learning analytics to detect skin disease in real time.

References

1. M.A.M. Almeida, I.A.X. Santos, Classification models for skin tumor detection using texture analysis in medical images, *Journal of Imaging*, **6**(6): 51, 2020, doi: 10.3390/jimaging 6060051.

2. A. Raghuvanshi, U.K. Singh, C. Joshi, A review of various security and privacy innovations for IoT applications in healthcare, [in:] Advanced Healthcare Systems, R. Tanwar, S. Balamurugan, R.K. Saini, V. Bharti, P. Chithaluru [Eds.], Chapter 4, pp. 43–58, 2022, doi: 10.1002/9781119769293.ch4.

3. V. Durga Prasad Jasti *et al.*, Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis, *Security and Communication Networks*, **2022**: Article ID 1918379, 2022, doi: 10.1155/2022/1918379.

4. W. Sae-Lim, W. Wettayaprasit, P. Aiyarak, Convolutional neural networks using MobileNet for skin lesion classification, [in:] *Proceedings of the 16th International Joint Conference on Computer Science and Software Engineering*, Chonburi, Thailand, 10–12 July, pp. 242–247, 2019, doi: 10.1109/JCSSE.2019.8864155.
5. D. Castillo, V. Lakshminarayanan, M.J. Rodríguez-Álvarez, MR images, brain lesions, and deep learning, *Applied Sciences*, **11**(4): 1675, 2021, doi: 10.3390/app11041675.
6. A. Raghuvanshi *et al.*, Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming, *Journal of Food Quality*, **2022**: Article ID 3955514, 2022, doi: 10.1155/2022/3955514.
7. V. Hemamalini *et al.*, Food quality inspection and grading using efficient image segmentation and machine learning-based system, *Journal of Food Quality*, **2022**: Article ID 5262294, 2022, doi: 10.1155/2022/5262294.
8. L. Bajaj, K. Gupta, Y. Hasija, Image processing in biomedical science, [in:] *Advances in Soft Computing and Machine Learning in Image Processing*, A.E. Hassanien, D.A. Oliva [Eds.], pp. 185–211, 2017, doi: 10.1007/978-3-319-63754-9_9.
9. L.S. Wei, Q. Gan, T. Ji, Skin disease recognition method based on image color and texture features, *Computational and Mathematical Methods in Medicine*, **2018**: 8145713, 2018, doi: 10.1155/2018/8145713.
10. M.N. Islam, J. Gallardo-Alvarado, M. Abu, N.A. Salman, S.P. Rengan, S. Said, Skin disease recognition using texture analysis, [in:] *IEEE 8th Control and System Graduate Research Colloquium (ICSGRC)*, pp. 144–148, 2017, doi: 10.1109/ICSGRC.2017.8070584.
11. H. Liao, Y. Li, J. Luo, Skin disease classification versus skin lesion characterization: Achieving robust diagnosis using multi-label deep neural networks, [in:] *23rd International Conference on Pattern Recognition (ICPR)*, pp. 355–360, 2016, doi: 10.1109/ICPR.2016.7899659.
12. S. Kumar, A. Singh, Image processing for recognition of skin diseases, *International Journal of Computer Applications*, **149**(3): 37–40, 2016, doi: 10.5120/ijca2016911373.
13. Deepalakshmi P., Prudhvi Krishna T., SiriChandana S., Lavanya K., P. Srinivasu, Plant leaf disease detection using CNN algorithm, *International Journal of Information System Modeling and Design*, **12**(1): 1, 21 pages, 2021, doi: 10.4018/ijismd.2021010101.
14. P. Naga Srinivasu, T. Srinivasa Rao, A.M. Dicu, C.A. Mnerie, I. Olariu, A comparative review of optimisation techniques in segmentation of brain MR images, *Journal of Intelligent & Fuzzy Systems*, **38**(5): 6031–6043, 2020, doi: 10.3233/jifs-179688.
15. A. Kumar, J. Kim, D. Lyndon, M. Fulham, D. Feng, An ensemble of fine-tuned convolutional neural networks for medical image classification, *IEEE Journal of Biomedical and Health Informatics*, **21**(1): 31–40, 2017, doi: 10.1109/JBHI.2016.2635663.
16. S. Kolkur, D.R. Kalbande, Survey of texture based feature extraction for skin disease detection, [in:] *2016 International Conference on ICT in Business Industry & Government*, pp. 1–6, 2016, doi: 10.1109/ICTBIG.2016.7892649.
17. B. Harangi, Skin lesion classification with ensembles of deep convolutional neural networks, *Journal of Biomedical Informatics*, **86**(1): 25–32, 2018, doi: 10.1016/j.jbi.2018.08.006.
18. J. Naranjo-Torres, M. Mora, R. Hernández-García, R.J. Barrientos, C. Fredes, A. Valenzuela, A review of convolutional neural network applied to fruit image processing, *Applied Sciences*, **10**(10): 3443, 2020, doi: 10.3390/app10103443.

19. A.K. Verma, S. Pal, S. Kumar, Classification of skin disease using ensemble data mining techniques, *Asian Pacific Journal of Cancer Prevention*, **20**(6): 1887–1894, 2019, doi: 10.31557/apjcp.2019.20.6.1887.

20. X. Zhang, S. Wang, J. Liu, C. Tao, Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge, *BMC Medical Informatics and Decision Making*, **18**(Suppl 2): Article number 59, 2018, doi: 10.1186/s12911-018-0631-9.

21. S.S. Yadav, S.M. Jadhav, Deep convolutional neural network based medical image classification for disease diagnosis, *Journal of Big Data*, **6**(1): Article number 113, 2019, doi: 10.1186/s40537-019-0276-2.

22. A.K. Gupta, C. Chakraborty, B. Gupta, Secure transmission of EEG data using watermarking algorithm for the detection of epileptical seizures, *Traitement du Signal*, **38**(2): 473–479, 2021, doi: 10.18280/ts.380227.

23. C. Wu, P. Lu, F. Xu, J. Duan, X. Hua, M. Shabaz, The prediction models of anaphylactic disease, *Informatics in Medicine Unlocked*, **24**: 100535, 2021, doi: 10.1016/j.imu.2021.100535.

24. A. Gupta, P. Bharat, Novel approaches in network fault management, *International Journal of Next-Generation Computing*, **8**(2): 115–126, 2017, doi: 10.47164/ijngc.v8i2.126.

25. N. Reza, I.S. Na, S.W. Baek, K.-H. Lee, Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images, *Biosystems Engineering*, **177**: 109–121, 2019, doi: 10.1016/j.biosystemseng.2018.09.014.

26. A. Kapoor, A. Gupta, R. Gupta, S. Tanwar, G. Sharma, I.E. Davidson, Ransomware detection, avoidance, and mitigation scheme: A review and future directions, *Sustainability*, **14**(1): 8, 2021, doi: 10.3390/su14010008.

27. M. Shabaz, U. Garg, Predicting future diseases based on existing health status using link prediction, *World Journal of Engineering*, **19**(1): 29–32, 2021, doi: 10.1108/wje-10-2020-0533.

28. H.Q. Yu, S. Reiff-Marganiec, Targeted ensemble machine classification approach for supporting IoT enabled skin disease detection, *IEEE Access*, **9**: 50244–50252, 2021, doi: 10.1109/ACCESS.2021.3069024.

29. Dermatology Data Set, [in:] D. Dua, C. Graff, UCI Machine Learning Repository, School of Information and Computer Science, University of California, Irvine, CA, 2019, https://archive.ics.uci.edu/ml/datasets/Dermatology.

Received January 20, 2022; revised version May 8, 2022; accepted May 15, 2022.