

DEVELOPMENT OF A SKIN LESION CLASSIFICATION SYSTEM EMPLOYING A K-NEAREST NEIGHBOR ALGORITHM

M Nancy¹

¹Dept. of CS, Fatima College, India.

ABSTRACT

In the realm of medical health, accurate disease diagnosis is paramount, with dermatology posing particular challenges. Dermatologists often rely on extensive testing, patient history reviews, and data analysis for precise diagnoses. Consequently, there is a need for a swift and reliable method to ensure accurate diagnoses. While various machine learning approaches have been explored for dermatological diagnosis, many lack high accuracy. This study introduces a MATLAB-based system designed to swiftly identify and classify skin lesions as normal or benign. Employing the K-nearest neighbor (KNN) method for classification, the system achieves a remarkable accuracy of 98% in distinguishing between normal and pathological skin lesions.

1. INTRODUCTION

Dermatology, the study of skin diseases, relies on visual diagnosis, utilizing imaging like ultrasound, dermatoscopy, and reflectance confocal microscopy. Challenges in dermatology include managing vast personal and imaging data. Over the past two decades, artificial intelligence and machine learning have revolutionized medical diagnostics and treatment, extending to various domains such as energy markets. Dermatoscopy, a key method for detecting skin cancers, benefits from automated analysis using technologies like convolutional neural networks (CNNs), which excel in classifying skin pathologies.

In skin disease classification, innovative techniques like linear discriminant analysis, support vector machine (SVM), naive Bayes, K-nearest neighbor (KNN), and deep learning algorithms are widely employed. The crucial feature selection step ensures effective analysis. The healthcare system as a whole embraces technologies like deep neural networks and machine learning for pattern recognition, enhancing diagnostic accuracy. Notable examples include neural networks like multilayer perceptron (MLP)-artificial neural network (ANN) combined with optimization algorithms like world cup and grey wolf, demonstrating high efficiency in melanoma detection.

Computer-based methods and machine learning algorithms prove beneficial in healthcare due to data mining, logical analysis, and assisted feature selections. These technologies offer comparable or superior results to human capabilities with efficient data management. Optimization processes in medical imaging involve uncertainty quantification (UQ), crucial for decision-making. Ensemble machine learning techniques and Bayesian approximations leverage UQ in computer vision and medical image analysis. Novel models, such as binary residual feature fusion with Monte Carlo dropout, enhance medical image classification. Additionally, uncertainty-aware models contribute to successful cancer detection in breast histology images.

In this study, the proposed system prioritizes accurate and efficient skin lesion diagnosis, employing KNN for classifying normal and malignant skin lesions. KNN stands out for its time efficiency, interpretability, and versatility, making it a highly accurate algorithm for classification and regression tasks.

2. SYSTEM OVERVIEW

The system goes through four separate stages. Our method entails modifying test photos using a variety of morphological procedures to reduce noise, enhance features inside infected regions, and increase image contrast for better feature detection. Adaptive thresholding is used to segment data. We use a variety of statistical techniques to extract features, including mean (Fast Fourier Transform), standard deviation, histogram-based metrics, edge-based pixel count for area and hole detection, and edge-based logarithmic pixel count. These characteristics are critical in identifying various illnesses. We use a KNN classifier to predict the malignancy of skin lesions based on statistical textural characteristics. This entails classifying skin photos in the test set by comparing them to a training set that consists of photographs.

3. METHODS

The skin classification process comprised several steps:

Image Acquisition: Dermoscopic images were used as input, captured using a dermatoscope instrument, which facilitates the diagnosis of skin diseases. A database of skin diseases can be established for convenient online access.

Preprocessing: The skin images underwent preprocessing to address issues such as pathological noise and varied texture backgrounds. This involved two main steps: hair removal and image enhancement. The objective was to

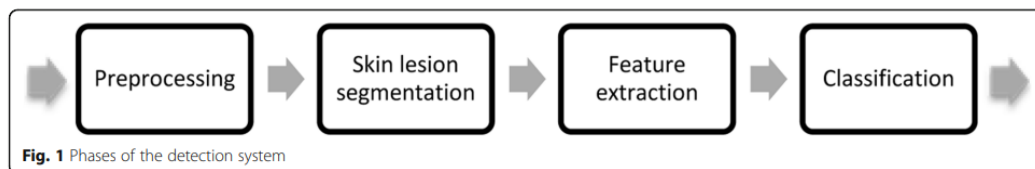
enhance image quality, making skin lesion segmentation easier. The morphological 'closing' operation was employed to filter out image shape and structure.

Segmentation: Medical image segmentation aims to distinguish tumors from surrounding context. The skin lesion was segmented using thresholding, wherein pixels with intensity values exceeding the threshold were assigned a foreground value, and the remaining pixels were designated background values. Adaptive thresholding, adjusting the threshold dynamically, was utilized to refine the segmentation.

Feature Extraction: Effective segmentation necessitates feature extraction, as lesions exhibit diverse sizes and properties. A feature vector converted the skin mole into intensity values, with the number of extracted properties determining its dimensions. Various statistical properties, such as mean value, were computed to generate these features.

Classification: Skin lesions were identified and classified based on their characteristics. While various classifiers have been explored, KNN classifiers demonstrated superior performance, hence their selection for this system. KNN offers advantages such as shorter execution time and higher accuracy compared to other methods like hidden Markov models, kernel processes, and SVM classifiers.

This methodology aims to accurately classify skin lesions, leveraging advanced preprocessing, segmentation, feature extraction, and classification techniques.



Proposed system:

The proposed system utilizes Nearest Neighbor classification, a straightforward technique in image space. In this method, the test image is assigned a label based on the nearest point in the learning set within the image space. Typically, Euclidean distance measurement is employed to calculate distances between data points in the image, with each pixel assigned a distance. After feature extraction, the features are directly fed into classifiers or machine learning tools for distribution into two classes, constituting two phases: training and testing.

During the training phase, patterns of benign and malignant images, along with their respective features and class labels, are used to train the classifiers. Approximately 40 melanoma (type = 1) and 40 normal (type = 0) skin images are employed for this purpose, with the feature attributes extracted previously being crucial for training. These data points are plotted in the feature space. In the testing phase, an unknown test pattern is classified based on the information obtained during the training process and plotted in the feature space. The feature space represents an abstract space where each sample image is depicted as a point in an n-dimensional space, with the number of features determining its size. The training database comprises a total of 40 images. The optimal K value for the KNN classifier is determined using the accuracy plot method.

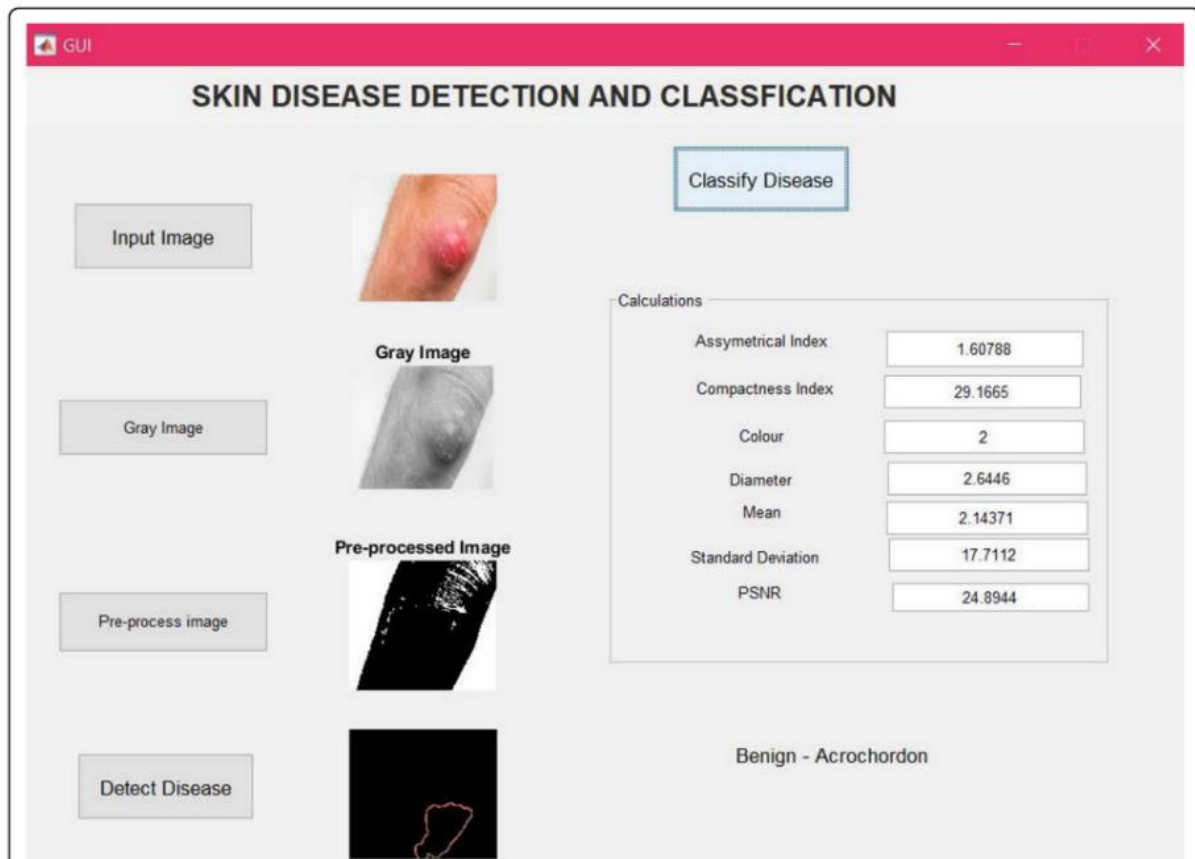
4. RESULTS AND DISCUSSION

The system employs a graphical user interface (GUI), which simplifies interaction and offers features like "point and click support" to the user. The MATLAB GUI is chosen for its efficiency in acquiring and uploading images, extracting features, and performing segmentation. The GUI contains various buttons for different purposes, facilitating image uploading, preprocessing, disease detection, and classification. Users can directly upload images into the application, and the results of each step can be displayed on the GUI screen itself. The workflow involves capturing the affected skin region using the input image button, followed by preprocessing to enhance image clarity. Subsequently, the grey image button is activated for contrast enhancement and conversion to grayscale. The preprocessing image button initiates contrast enhancement using histogram equalization and grayscale conversion, distinguishing each pixel from its neighbor. Disease detection involves feature extraction to highlight relevant details in the image, simplifying the subsequent detection or classification step. Upon successful execution of all preceding processes, the "classification of the disease" button becomes enabled (as depicted in Fig. 5). When pressed, this button generates output results comprising various calculated parameters such as the asymmetrical index, mean value, compactness index, color, diameter, standard deviation value, and peak signal-to-noise ratio. These calculations play a crucial role in determining the type of skin disease present in the input image.

Through the implementation of the proposed system, we achieved an impressive accuracy of 98%. Comparative analysis with previous studies demonstrates that our system exhibits remarkably high accuracy, comparable to the

best-performing methods. Notably, the proposed system offers several advantages. Firstly, it is easy to implement and operates swiftly, without requiring a training period, owing to its reliance on the KNN algorithm. Additionally, the system seamlessly accommodates new data without compromising accuracy. However, it is worth noting some drawbacks, such as its limited effectiveness with large datasets and susceptibility to noise present in the dataset.

To further enhance the system's performance, future iterations could explore the utilization of ensemble learning methods or evolutionary algorithms, promising even higher accuracy and faster results.



5. CONCLUSION

Automated systems leveraging machine learning play a pivotal role in aiding physicians to accurately classify skin lesions. The proposed system comprises four primary phases: image preprocessing, segmentation, feature extraction, and classification. Implemented in MATLAB, the system is underpinned by the KNN algorithm and is accessible via a user-friendly GUI, facilitating a step-by-step classification process with visualizations of the statistical features utilized for classification. Through testing, the system demonstrated an impressive accuracy of 98%. Looking ahead, enhancements to the system could involve integrating ensemble learning techniques or evolutionary algorithms to ensure further improvements in accuracy and efficiency.

6. REFERENCES

- [1] Abdullah AA, Hatem MQ (2018) Audio transmission through Li-Fi technology. *Int J Civil Eng Technol* 9(7):853–859
- [2] Ferreira IG, Weber MB, Bonamigo RR (2021) History of dermatology: the study of skin diseases over the centuries. *An Bras Dermatol* 96(3):332–345.
- [3] Gao W, Darvishan A, Toghani M, Mohammadi M, Abedinia O, Ghadimi N (2019) Different states of multi-block based forecast engine for price and load prediction. *Int J Electr Power Energy Syst* 104:423–435.
- [4] Ghadimi N, Akbarimajd A, Shayeghi H, Abedinia O (2018) Two stage forecast engine with feature selection technique and improved metaheuristic algorithm for electricity load forecasting. *Energy* 161:130–142.
- [5] Abdar M, Fahami MA, Chakrabarti S, Khosravi A, Plawiak P, Acharya UR et al (2021) BARF: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification. *Inform Sci* 577:353–378.
- [6] Senousy Z, Abdelsamea MM, Gaber M, Abdar M, Acharya RU, Khosravi A et al (2021) MCUa: Multi-level context and uncertainty aware dynamic deep ensemble for breast cancer histology image classification. *IEEE Trans Biomed Eng.*