FRAMEWORK FOR THE DEVELOPMENT OF A BACKGROUND PIGMENT RESPONSIVE SKIN DISEASE IDENTIFICATION AND CLASSIFICATION SYSTEM

A. A. ADIGUN[1], O. A. ODEJOBI[2], R. G. KOMOLAFE[3], M. A. IBRAHIM[4]
[1, 2, 3,4]Osun State University, Osogbo, Nigeria
[1]adepeju.adigun@uniosun.edu.ng, [2] oodejobi@yahoo.com, [3]rotimikomolafe@yahoo.com,
[4]ibrahima@uniosun.edu.ng

Keywords: Feature Selection, Classification, Skin Pigment, Light Skin, Dark Skin.

Abstract: *Skin disease is one of the common illnesses that affect human beings. It affects all cultures, occurs at all ages, and affects between 30 percent and 70 percent of persons. Ineffective classification of skin disease can be caused by bias in the dataset; also, the image background could affect the skin pigment and also affect in carrying out an identification of skin disease. Current progress in pattern recognition has led to success in the development of computerized skin image analysis. In particular, skin disease classification models have gained the feat higher than qualified dermatologists. However, no attempt has been made to assess the steadiness in performance of pattern recognition models across populations with varying skin pigment. In this research, an approach to estimate pigment in benchmark skin disease datasets, and examine if model performance is dependent on this measure. We will also propose the use of two skin disease datasets, one for the Light skin dataset, and it is a collection of dermatoscopic images, and another for the Dark skin dataset which will be obtained locally within Osun State. Feature selection algorithms proposed to be used are Information Gain and Chi-Square, after which we proposed to apply Decision Tree and Random Forest for classification, and then the performance will be evaluated. Our outcome will serve as an evaluation model that will help in improving the accuracy of skin disease detection.*

1. INTRODUCTION

As pattern recognition technology is appropriate more often functional to hold decisions that shape the lives of many, understanding how to precisely evaluate hidden dataset characteristics and demographic illustration is of fast-increasing importance to prevent the likely unhelpful cost of data bias [30] hence referred to as dataset bias. Data bias is a significant subject because it is one of the causes of the pattern recognition-based systems placing certain groups of people at a logical disadvantage [31]. Identification and improvement of the unnecessary preconception all through the machine learning channel is essential to construct learning systems that are trusted in their eventual domains of deployment [32] Skin diseases persist to bear considerable negative impacts on human health worldwide. Skin diseases such as skin cancer account for about 7 percent of new cancer cases worldwide [33] with an outlay to the American healthcare system that passed 8 billion in 2011. In skin cancer cases, there is proof of various result disparities for ethnicity: although people of dark skin are in the region of 20 to 30 times less likely to develop melanoma than light skin persons, for definite melanoma sub-types they have lower or higher [34] survival rate. Some studies have found that for people of dark skin, the identification of skin cancer might happen at a more difficult stage, leading to lesser rates of survival and low outcomes [35]. However, augmented screening with presently existing diagnostic resources also has risks and can result in major harms, like avoidable surgeries, deformity, disability, and overdiagnosis [36].

Pattern recognition technologies have been studied in the perspective of dermatology image examination for years, and numerous appraise articles have been on paper [37], but have failed when applied to dark skin e.g. racist soap dispenser in Facebook office [56] and racist criminal facial recognition machine in America [57]. The achievement of pattern recognition models has directed to studies where it is applied to dermatological use cases [38] Models using Random forest and Decision trees have been used with problems such as skin cancer diagnosis and were found to do better than trained dermatologists in controlled settings and datasets [39] However, as nearly all of the widely available datasets of skin images come from light-skinned populations, due to the severe disparities in disease incidence, there are concerns about how to best collect data, train, and evaluate models for dark-skinned populations [40]. Also, because of the considerable risks of damage from over-diagnosis with better screening in low-risk dark skin populations, there is a need to better distinguish between critical and stable presentations of disease [30].

In this study, a background-sensitive technique to skin pigment pattern recognition that is responsive to dark skin diseases pattern manifestation will be investigated. Several methods and techniques for the identification and classification of patterns associated with human skin diseases have been reported in computing literature. However, the efficacy of these methods and techniques is yet to be confirmed for patterns manifesting in the diseases of different grades of skin pigment. This study emerges in the context of recent failures recorded in the application of modern pattern recognition technologies to a different grade of skin pigment subjects. This study will develop a background sensitive technique to skin pigment pattern recognition that is responsive to patterns manifesting in dark skin diseases.

In computer science, information stored in the database with the incoming data can be matched with pattern recognition technology. Sometimes the question comes that, "What is the difference between pattern recognition and machine learning? "It is a type of machine learning; that makes it an essential part of the whole process of machine learning. It helps the algorithms to ascertain regularities in the enormous numbers of data and facilitates how to classify it into different categories [48].

How Pattern Recognition Works

Pattern recognition is a procedure that considers the accessible data and tries to detect if there is any regularity in it. There are two main parts:

i. Explorative fraction - the algorithms looking for general patterns
ii. Descriptive fraction - the algorithms classify the found patterns

Types of Pattern Recognition Theory

i. Template matching theory
   The inward bound sensory information is evaluated directly to templates stored in the long-term memory. These templates are stored in the course of our past knowledge and learning.
   E.g. A A A are all known as the letter A
ii. Prototype matching theory
   Prototype means a notion of normal characteristics of a particular matter. For example, a notion of a small animal with feathers, beak, two wings and can fly is a prototype notion of a sparrow, hen, eagle, etc. Prototype matching unlike template matching does not stress a perfect match between the inward bound stimuli and the stored notion in the brain.
iii. Feature analysis theory
   In this theory, the visual system breaks down the inward bound stimuli into their features and processing the information. Some features sometimes are more important for recognition than others. Feature analysis proceeds through four stages.
   a. Detection
   b. Pattern dissection
   c. Feature comparison in memory
   d. Recognition

Pattern recognition allows us to read, understand, recognize things, and also appreciate music. These theories are applied to different activities and areas where pattern recognition is examined [49]

The research aims to design, implement and evaluate skin disease identification and classification system. The specific objectives are to:

i. Specify the features for characterizing dark skin disease pattern
ii. Design a skin disease identification system
iii. Implement the objective (ii)
iv. evaluate the objective (iii)

The scope of this research work is based on skin disease identification and classification system. The limitations of the research are;

i. Two skin disease datasets (light and Dark) will be used.
ii. Eczema and Pimple are skin diseases in both datasets.
iii. That the system will not consider albino skin disease.
iv. That the disease site can be on any part of the body.
v. That the detection can be done with one image.

Related Works

i. Three types of skin diseases were identified from skin images observed, the images were preprocessed and segmented by using histogram analysis, then features were extracted which are then used to classify the type of skin disease in the image. With the help of the image under process and the techniques of data mining, it provided the user treatments based on the results obtained in a shorter period than the existed methods. The process was conducted on the different skin patterns and analyzed to obtain the results that can be used to identify the type of skin disease the user is suffering from. The data helped in the early detection of skin diseases and their cure. The results obtained are classified according to the given prototype, and diagnosis accuracy assessment is performed to provide users with efficient and fast results. [6]

ii. A study of predicting skin diseases using the Naïve Bayesian classifier and concluded that it was a good classifier in predicting skin diseases. The orange data mining tool was used and the dataset was obtained from the UCI repository. The study only investigated NB. [7]

iii. Deep learning algorithms to help diagnose four common cutaneous diseases based on dermoscopy images. To facilitate decision-making and improve the accuracy of their algorithm, the summary of classification/diagnosis scenarios based on domain expert knowledge and semantically represented them in a hierarchical structure. The algorithm achieved an accuracy of 87.25 ± 2.24% in their test dataset with 1067 images. The semantic summarization of diagnosis scenarios helped to further improve the algorithm to facilitate future computer-aided decision support. [22]

iv. A method that uses techniques related to computer vision to distinguish different kinds of dermatological skin abnormalities. Various types of Deep learning algorithms (Inception_v3, MobileNet, Resnet, xception) were employed for feature extraction and learning algorithm for training and testing purposes. Their architecture considerably increases efficiency by up to 88 percentages. Furthermore, by using ensemble features mapping, combining the models trained using Inception V3, MobileNet, Resnet, Xception a voting-based model was an ensemble and thereby increased the efficiency. [17]

v. A new method to classify skin disease, which applied five different data mining techniques and then developed an ensemble approach that consists of all the five different data mining techniques as a single unit. Informative Dermatology data were used to analyze different data mining techniques before classifying the skin disease then; an ensemble machine learning method is applied. The ensemble method, which is based on machine learning, was tested on Dermatology datasets and classifies the type of skin disease into six different classes C1: psoriasis, C2: seborrheic dermatitis, C3: lichen planus, C4: pityriasis rosea, C5: chronic dermatitis, C6: pityriasis Rubra. Python was used to find the prediction on the skin diseases dataset; the accuracy and sensitivity of the five different data mining techniques were calculated. The results showed that the dermatological prediction accuracy of the test data set is 98.64%.[16]

vi. An efficient skin disease identification approach using an enhanced deep neural network model was proposed. The database images are segmented using an enhanced level set approach-based segmentation. Feature extraction is carried out for all the images to retrieve the feature vector using GLCM. Finally, a dragonfly optimization-based deep neural network is utilized for the classification of skin diseases. The system is implemented in the working platform of MATLAB, evaluation metrics such as accuracy, sensitivity, and specificity were used to show the system efficiency. The segmented images produce 98% accuracy while classifying the skin database images as normal and abnormal for various metrics. [20]

vii. An approach to estimate skin tone in standard skin disease datasets, and study if model performance is reliant on this measure. Explicitly, individual typology angle (ITA) was used to estimate skin tone in dermatology datasets. They used the distribution of ITA values to better realize skin color illustration in two standard datasets: a) the ISIC 2018 Challenge dataset, a compilation of dermatoscopic images of skin lesions for the discovery of skin cancer b) the SD-198 dataset that is a compilation of clinical images covering a wide selection of skin diseases. In estimating ITA, they developed segmentation models to separate non-diseased areas of skin. The result showed that the greater part of the data in the two datasets had ITA values between 34.50 and 480, which are related to lighter skin, and is regular with under-representation of darker-skinned populations in these datasets. No measurable connection between the performance of the machine learning model and ITA values, although more comprehensive data are required for additional validation. [29]

viii. The influence of the feature selection approach in the prediction of skin diseases using data mining techniques was discussed. Dataset was gotten from the UCI repository; Information Gain and Chi-square were the feature selection algorithms used to reduce features of the skin diseases dataset. The classification was done using Random Forest, C4.5 Decision Trees, and Functional Tree. WEKA data mining tool was used to generate the result. Experimental results of the developed predictive model on skin diseases have revealed that the feature selection algorithms did not necessarily improve the accuracy and sensitivity of these algorithms and in situations where they brought an improvement; it was just a little about 1 percent. [18]

ix. CNN structure for the skin image diagnosis of three common skin diseases (Melanoma, Nevus, Seborrheic Keratosis) and had constructed a dataset consisting mainly of skin disease images. The overall accuracy is 71% and the results demonstrated that CNNs can recognize and classify skin diseases. [52]

Research Gap

In the research of feature selection approach to predict skin diseases [18], but in the research, skin diseases data used were gotten for the light skin, this makes it not to predict accurately skin disease from the dark skin. This research will propose the use of datasets from both light and dark skin to complement the gap in the above-mentioned research.

2. METHODOLOGY

The expected variables to achieve this research work are;

i. Light skin Dataset Eczema (90), Pimple (95)
- **i.** Dermatoscopic images
- **ii.** gotten from a digital dermoscopy
- **iii.** low-level noise and consistent background illumination.

ii. Dark skin Dataset Eczema (60) Pimples (60)
- **i.** Clinical images
- **ii.** gotten from local hospitals

Hypothesis

H1 (Alternative Hypothesis)

That the combination of feature selection and classification algorithms will greatly enhance the performance level of skin disease classification.

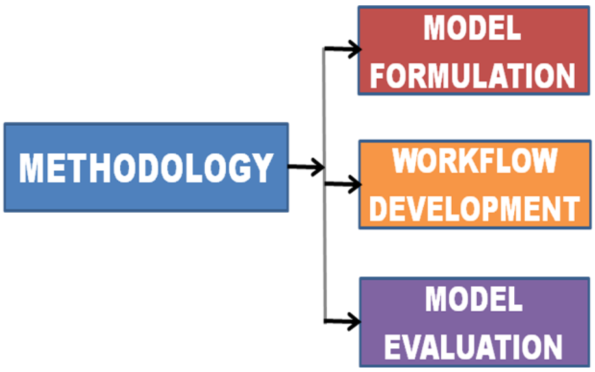

Figure 1: Research Structure

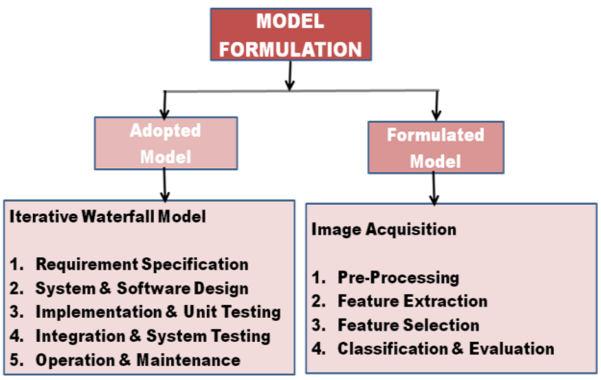

Figure 2: Model Formulation Structure

This research will have these approaches; first, enumeration of the skin diseases that are available in the datasets. With this, normalization of the image to enhance the local contrast will be carried out, and then segmentation of skin images to extract non-disease regions, to distinguish disease area from the background. Second, feature selection algorithms will be used to give the subset feature of the original datasets. Thereafter, the performance evaluation of the classification across Dark and Light skin categories with and without feature selection algorithms will be enumerated.

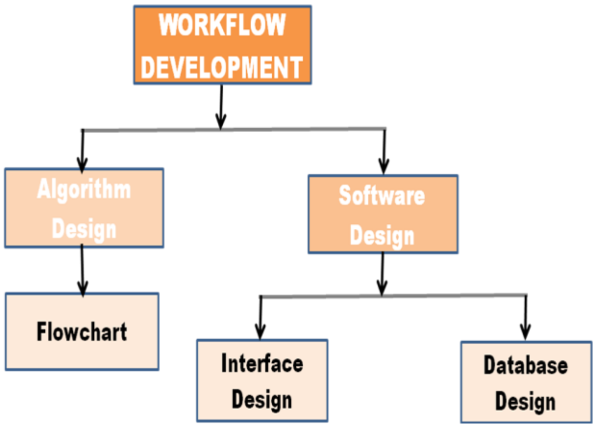

Figure 3: Workflow Development Structure

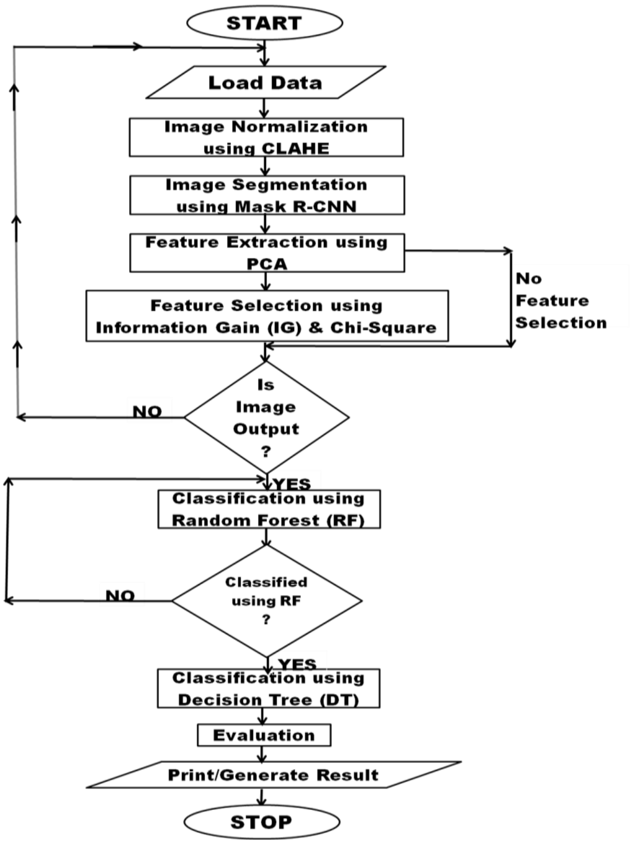

Figure 4: Proposed System Algorithm

Data Acquisition

Likewise, two datasets will also be adopted for this research work; one will be obtained from ISIC 2018 Challenge dataset (light skin). Light skin dataset is a collection of dermatoscopic images, which are gotten from a digital dermoscopy with low-level noise and consistent background illumination. The Light skin dataset consists of 185 dermatoscopic images that are publicly available in the ISIC archive [42] it has two skin diseases; Eczema (90), pimple (95).

Figure 5: Eczema disease (Light skin)

Figure 6: Pimple disease (Light skin)

The second dataset to be adopted will be obtained from the State hospital in Osogbo, Osun State (dark skin). The dark skin dataset contains 120 clinical images from four skin diseases; Eczema (60) pimple (60).

Figure 7: Eczema disease (Dark skin)

Figure 8: Pimple disease (Dark skin)

K-fold cross-validation technique will be adopted to estimate the accuracy of the system. Validate on the test set.

i. Pre-processing

i. Normalization
ii. Contrast Limited Adaptive Histogram Equalization (CLAHE) will be adopted to normalize the acquired images. The reason is to improve the local contrast and enhancing the definition of edges in each region of the acquired images. [47]
iii. Segmentation
iv. Mask R-CNN model will also be adopted for segmentation of skin disease from non-diseased skin. It is adopted Mask R-CNN because it was one of the best performing skin disease boundary segmentation models, and it has shown to be effective and efficient in doing semantic segmentation. [55]

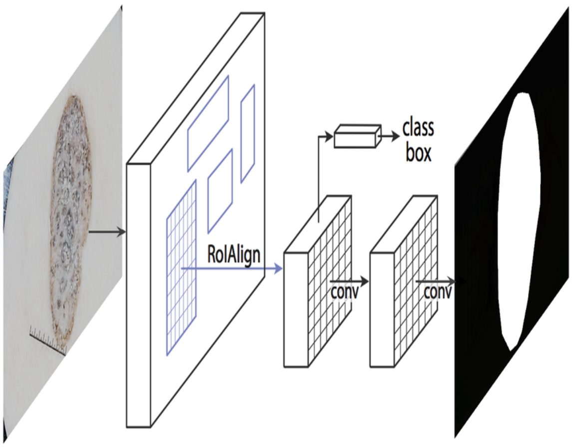

Figure 9: Mask R-CNN Segmentation Framework

ii. Feature Extraction

Feature extraction is a dimensionality reduction process, with an initial set of the raw data that is divided and reduced to more controllable groups.

Their features can be processed easily, and still able to describe the real data set with their accuracy and originality. Principal Component Analysis (PCA) will be adopted for the extraction of features. The reason is that it reduces the number of features by constructing a new; smaller number of variables that capture a significant portion of the information found in the original features.

iii. Feature selection

In machine learning, feature selection is the process of selecting a subset of relevant features for use in model building. The main reasons for using feature selection are:

i. Training faster the machine learning algorithm,
ii. Complexity reduction of a model and easier to interpret,
iii. Accuracy improvement of a model.

Information Gain and Chi-square will be adopted as feature selection algorithms.

a. Chi-square

The Chi-Square test is used to check how well the observed values for a given distribution fit with the distribution when the variables are independent. The reason for using chi-square is that each entity cannot fit in more than one category. [53] A chi-square test a null hypothesis about the affiliation between two variables. Chi-Square Test for Feature Selection

i. Define Hypothesis.
ii. Build a Contingency table.
iii. Find the expected values.
iv. Calculate the Chi-square statistic.
v. Accept or Reject the Null Hypothesis.

b. Information Gain

Information gain can also be used for feature selection, by evaluating the gain of each variable in the context of the target variable. Information gain is the reduction in entropy or transformation of a dataset and is mostly used in decision trees training. [54]

iv. Classification

a. Decision Tree

A decision tree or classification trees predict responses to data. In predicting a response, the decisions in the tree from the root node down to a leaf node will be followed. The response is the leaf node; the Classification trees give normal responses, like 'true' or 'false'. Each step in a prediction involves checking the value of one predictor (variable) [51] For example; here is a simple classification tree:

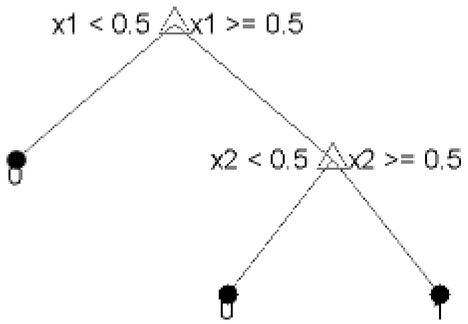

Figure 10: Decision Tree

The tree predicts classifications based on two predictors, x1 and x2, to predict, start at the top node, it is represented by a triangle (Δ). The first decision is if x1 is smaller than 0.5, then, follow the left branch, and the tree classifies the data as type 0. If x1 is greater than 0.5, then it will follow the right branch to the lower-right triangle node. In figure 10, if x2 is smaller than 0.5, then it will follow the left branch that the tree classifies the data as type 0. If not, it will follow the right branch that the tree classifies the data as type 1.

b. Random Forest

The random forest creates many decision trees. The decision trees are created depending on the random selection of data and also the selection of variables randomly. The dependent variable class is determined by the class based on many decision trees. The random forest algorithm being most of the decision trees predicts the correct classes for most of the given data. The voting for each observation can be done, and the class of the observation can be determined based on the results of the voting. This voting and classification are considered to be much nearer to the exact classification. [50]

Evaluation

The system will be simulated on MATLAB, and the goal is to investigate how classification with or without feature selection performs across skin diseases. The accuracy and sensitivity of the classification will be enumerated.

i. Accuracy:- The percentage of instances that are correctly classified is called accuracy while the percentage of instances that are not correctly classified is obtained by subtracting the correctly classified instances from 100.
Accuracy = TP + TN/ TP + TN + FP + FN
ii. Sensitivity:- This is the proportion of people who have the disease and were rightly classified as having the disease. It can be called recall or true positive rate.
Sensitivity = TP/ TP + FN

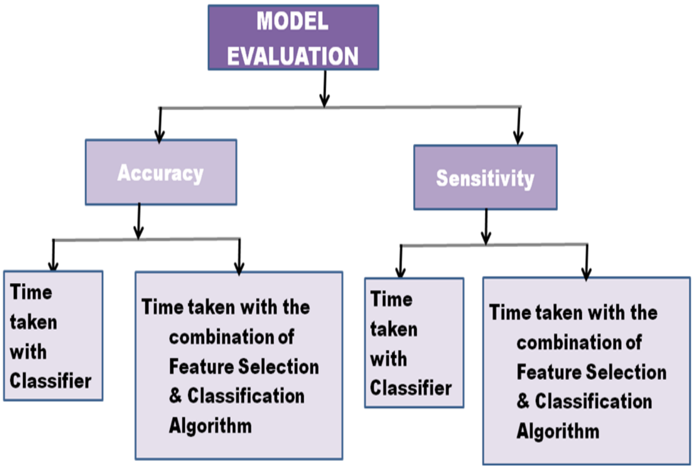

Figure 11. Model Evaluation Structure

3. RESULTS AND DISCUSSION

Taking into consideration different grades of skin pigment and the influence of feature selection on decision tree classifiers in identifying skin disease, the expected result should be able to identify and classify different skin diseases not minding the skin pigment. However, two skin disease datasets will be used, one for the light skin and the second for the dark skin, to avoid data bias. Both datasets will be pre-processed (Normalization and Segmentation) and feature extraction will be done after which feature selection will be carried out. Classification of the data into their categories will follow; lastly, evaluation of the system will be done with the Accuracy and Sensitivity metrics. The degree of accuracy and sensitivity when used only classifiers will be analyzed. Likewise, the degree of accuracy and sensitivity when feature selections were applied before classification will be enumerated.

Therefore, the expected result should be more accurate than the previous research especially, in taking both Dark and Light skin diseases into consideration.

4. CONCLUSIONS

This research work is expected to improve the accuracy of skin disease detection by removing data bias. Also, it will show the effect of feature selection in skin disease classification. Moreover, it will serve as a basis for skin disease classification for further research.

5. REFERENCES

[1]. Ankush Rai and Jagadeesh Kannan R. - "Central Processing Unit-Graphics Processing Unit Computing Scheme for Multi-object Tracking in Surveillance", Asian Journal of Pharmaceutical and Clinical Research. DOI: 10.22159/ajpcr.2017.v10s1.19651, 2017.

[2]. Adeyemo, O O, and T O Adeyeye. "Comparative Study of ID3 / C4 . 5 Decision Tree and Multilayer Perceptron Algorithms for the Prediction of Typhoid Fever." *African Journal of Computing & ICT* 8(1): 103–12. 2015.

[3].Aneeshkumar, A S, and C Jothi Venkateswaran. "Relevance Study of Data Mining for the Identification of Negatively Influenced Factors in Sick Groups." *Procedia - Procedia Computer Science* 47: 101–8. 2015. http://dx.doi.org/10.1016/j.procs.2015.03.188.

[4].Bindushree, D C. "Prediction of Cardiovascular Risk Analysis and Performance Evaluation Using Various Data Mining Techniques: A Review." *International Journal of Engineering Research* 5013(5): 796–800, 2016.

[5].Chandna, Deepali. "Diagnosis of Heart Disease Using Data Mining Algorithm." *International of Computer Science and Information Technologies* 5(2): 1678–80, 2014.

[6].Ginghi, Er, Ansari Nausheen, Ansari Zoya, and Shaikh Ruhin. "An Innovative Approach for Skin Disease Detection Using Image Processing and Data Mining." *International Journal of Innovative Research in Computer and Communication Engineering* 5(4): 8135–41, 2017.

[7].Karthik, Sistla, Bhimana Rahul, Mukesh Vibhudi, and Tarun Tej. "Prediction of Dermatological Condition Using Naive Bayesian Classification." *International Journal of Pharmacy & Technology* 9(1): 28988–94, 2017.

[8].Kaur, Parminder, and Aditya Khamparia. "Classification of Liver Based Diseases using Random Tree." *International Journal of Advances in Engineering & Technology* 8(3): 306–13, 2015.

[9].Lebbe, Ahamed, Sayeth Saabith, Elankovan Sundararajan, and Azuraliza Abu Bakar. "Comparative Study on Different Classification

Techniques for Breast Cancer Dataset." *International Journal of Computer Science and Mobile Computing* 3(10): 185–91, 2014.

[10]. Parikh, Krupal S, Trupti P Shah, Rahulkrishna Kota, and Rita Vora. “Diagnosing Common Skin Diseases Using Soft Computing Techniques.” *International Journal of Bio-science and Bio-Technology* 7(6): 275–86, 2015.

[11]. Patil, R V, S S Sannakki, and V S Rajpurohit. “A Survey on Classification of Liver Diseases Using Image Processing and Data Mining Techniques.” *International Journal of Computer Sciences and Engineering* 5(3): 29–34, 2017.

[12]. Sumithra, R., Suhil, M., & Guru, D. S. Segmentation and classification of skin lesions for disease diagnosis. *Procedia Computer Science*, *45*, 76-85, 2015.

[13]. Sudha, J., Aramudhan, M., & Kannan, S. Development of a mathematical model for skin disease prediction using response surface methodology. *Biomedical Research*. 2017.

[14]. Pathan, S., Prabhu, K. G., & Siddalingaswamy, P. C. Techniques and algorithms for computer-aided diagnosis of pigmented skin lesions—a review. *Biomedical Signal Processing and Control*, *39*, 237-262, 2018.

[15]. Lim, H. W., Collins, S. A., Resneck Jr, J. S., Bolognia, J. L., Hodge, J. A., Rohrer, T. A., ... & Nerenz, D. R. The burden of skin disease in the United States. *Journal of the American Academy of Dermatology*, *76*(5), 958-972, 2017.

[16]. Verma, A. K., Pal, S., & Kumar, S. Classification of skin disease using ensemble data mining techniques. *Asian Pacific Journal of Cancer Prevention: APJCP*, *20*(6), 2019.

[17]. Patnaik, S. K., Sidhu, M. S., Gehlot, Y., Sharma, B., & Muthu, P. Automated Skin Disease Identification using Deep Learning Algorithm. *Biomedical and Pharmacology Journal*, *11*(3), 1429-1437, 2018.

[18]. Abikoye, O. C., Komolafe, R. G., & Aro, T. O. Performance Evaluation of Feature Selection Algorithms on Skin Disease Prediction, 2019.

[19]. Bhadula, S., Sharma, S., Juyal, P., & Kulshrestha, C. Machine Learning Algorithms based Skin Disease Detection.

[20]. Melbin, K., & Raj, Y. J. V. An Enhanced Model for Skin Disease Detection using Dragonfly Optimization-based Deep Neural Network. In *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC)* (pp. 346-351). IEEE, 2019.

[21]. Rathod, J., Wazhmode, V., Sodha, A., & Bhavathankar, P. Diagnosis of skin diseases using Convolutional Neural Networks. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1048-1051). IEEE, 2018.

[22]. Zhang, X., Wang, S., Liu, J., & Tao, C. Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. *BMC medical informatics and decision making*, *18*(2), 59, 2018.

[23]. Liao, H., & Luo, J. A deep multi-task learning approach to skin lesion classification. *arXiv preprint arXiv:1812.03527*, 2018.

[24]. Kuhn, C., Swartzwelder, S., & Wilson, W. *Buzzed: The straight facts about the most used and abused drugs from alcohol to ecstasy*. WW Norton & Company, 2019.

[25]. Marques JS, Barata C, Mendonca T. On the role of texture and color in the classification of dermoscopy images. In: Annual international conference of the IEEE engineering in medicine and biology society (EMBC). 2012.

[26]. Khan MA, Tallha A, Muhammad S, Aamir S, Khursheed A, Musaed A, Syed IH, Abdualziz A. An implementation of normal distribution-based segmentation and entropy-controlled features selection for skin lesion detection and classification. BMC Cancer 18(1):638, 2018.

[27]. Naeem S, Riaz F, Hassan A, Miguel Tavares C, Nisar R. Description of visual content in dermoscopy images using joint histogram of multiresolution local binary patterns and local contrast. In: Proceedings of 16th international conference on intelligent data engineering and automated learning (IDEAL 2015), Poland

[28]. Khan MA, Sharif M, Akram T, Bukhari SA, Nayak RS. Developed newton-raphson based deep features selection framework for skin lesion recognition. Pattern Recognit Lett 129:293–303, 2020.

[29]. Kinyanjui, N. M., Odonga, T., Cintas, C., Codella, N. C., Panda, R., Sattigeri, P., & Varshney, K. R. Estimating skin tone and effects on classification performance in dermatology datasets. *arXiv preprint arXiv:1910.13268*, 2019.

[30]. Veronica Rotemberg, Allan Halpern, Stephen W. Dusza, and Noel C. F. Codella. The role of public challenges and data sets towards algorithm development, trust, and use in clinical practice. Seminars in Cutaneous Medicine and Surgery, 38(1): E38–E42, March 2019

[31]. Solon Barocas and Andrew D. Selbst. Big data’s disparate impact. California Law Review,104(3):671–732, June 2016

[32]. Rachel K. E. Bellamy, Kuntal Dey, Michael Hind, Samuel C. Hoffman, Stephanie Houde, Kalapriya Kannan, Pranay Lohia, Sameep Mehta, Aleksandra Mojsilovi´c, Seema Nagar, Karthikeyan Natesan Ramamurthy, John Richards, Diptikalyan Saha, Prasanna Sattigeri, Moninder Singh, Kush R. Varshney, and Yunfeng Zhang.

Think your artificial intelligence software is fair? Think again. IEEE Software, 36(4):76–80, July 2019

[33]. Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6):394–424, November/December 2018.

[34]. Krishnaraj Mahendraraj, Komal Sidhu, Christine S. M. Lau, Georgia J. McRoy, Ronald S. Chamberlain, and Franz O. Smith. Malignant melanoma in African–Americans: A population-based clinical outcomes study involving 1106 African–American patients from the surveillance, epidemiology, and result (SEER) database (1988—2011). Medicine, 96(15):e6258, April 2017.

[35]. M. Gohara. Skin cancer: An African perspective. British Journal of Dermatology, 173(Suppl. 2):17–21, July 2015

[36]. Michael A. Marchetti, Esther Chung, and Allan C. Halpern. Screening for acral lentiginous melanoma in dark-skinned individuals. JAMA Dermatology, 151(10):1055–1056, October 2015

[37]. Mani Abedini, Qiang Chen, Noel Codella, Rahil Garnavi, Xingzhi Sun, M. Emre Celebi, Teresa Mendonca, and Jorge S. Marques. Accurate and scalable system for automatic detection of malignant melanoma. In M. Emre Celebi, Teresa Mendonca, and Jorge S. Marques, editors, Dermoscopy Image Analysis. CRC Press, 2015.

[38]. M. Emre Celebi, Noel Codella, and Alan Halpern. Dermoscopy image analysis: Overview and future directions. IEEE Journal of Biomedical and Health Informatics, 23(2):474–478, March 2019.

[39]. H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. Ben Hadj Hassen, L. Thomas, A. Enk, and L. Uhlmann. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, 29(8):1836–1842, August 2018.

[40]. Adewole S. Adamson and Avery Smith. Machine learning and health care disparities in dermatology. JAMA Dermatology, 154(11):1247–1248, November 2018

[41]. Kinyanjui, N. M., Odonga, T., Cintas, C., Codella, N. C., Panda, R., Sattigeri, P., & Varshney, K. R. Estimating skin tone and effects on classification performance in dermatology datasets. *arXiv preprint arXiv:1910.13268*, 2019.

[42]. Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, and Allan Halpern. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). arXiv:1902.03368, March 2019.

[43]. Xiaoxiao Sun, Jufeng Yang, Ming Sun, and Kai Wang. A benchmark for automatic visual classification of clinical skin disease images. In Proceedings of the European Conference on Computer Vision, pages 206–222, October 2016.

[44]. International Skin Imaging Collaboration. ISIC 2018: Skin lesion analysis towards melanoma detection, 2018. Available: https://challenge2018.isic-archive.com/.

[45]. Marcus Wilkes, Caradee Y. Wright, Johan L. du Plessis, and Anthony Reeder. Fitzpatrick skin type, individual typology angle, and melanin index in an African population. JAMA Dermatology, 151(8):902–903, August 2015.

[46]. Giuseppe R. Casale, Anna Maria Siani, Henri Diémoz, Giovanni Agnesod, Alfio V. Parisi, and Alfredo Colosimo. Extreme UV index and solar exposures at Plateau Rosà (3500 m a.s.l.) in Valle d'Aosta Region, Italy. Science of the Total Environment, 512–513:622–630, April 2015.

[47]. Li, C., Tang, S., Yan, J., & Zhou, T. Low-light image enhancement based on quasi-symmetric correction functions by fusion. *Symmetry*, *12*(9), 1561, 2020.

[48]. Eysenck, Michael W., and Mark T. Keane. Cognitive Psychology: A Student's Handbook. London: Psychology, Taylor & Francis Group, 2015. Print

[49]. Hillman, Keith. "Pattern Recognition and Your Brain." Psychology24.org. N.p., 21 Mar. 2016. Web. 01 Mar. 2017

[50]. Ali, Jehad, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. "Random Forests and Decision Trees." *International Journal of Computer Science Issues* 9(5): 272–78, 2012.

[51]. Karim, Hamid, and Khani Zand. "A Comparative Survey on Data Mining Techniques for Breast Cancer Diagnosis and Prediction." *Indian Journal of Fundamental and Applied Life Sciences* 5(1): 4330–39, 2015.

[52]. S. Malliga, G. Sherly Infanta, S. Sindoora, and S. Yogarasi. Skin Disease Detection and Classification using Deep Learning Algorithms, 2020.

[53]. Chen, Z., Zhao, P., Li, C., Li, F., Xiang, D., Chen, Y. Z., ... & Song, J. User Manual for iLearnPlus.

[54]. Mitchell, T. M. Machine learning. 1997.
[55]. Huang, C., Yu, A., Wang, Y., & He, H. Skin lesion segmentation based on mask r-CNN. 2020.
[56]. Futureis," This 'Racist soap dispenser' at Facebook office does not work for black people", web,https://www.youtube.com/watch?v=YJjv_OeiHmo
[57]. Alex Najibi, Racial Discrimination in Face Recognition Technology, web, https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/