2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC) | 978-1-6654-9056-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/iSSSC56467.2022.10051620

SKIN DISEASE CLASSIFICATION USING MACHINE LEARNING AND DATA MINING ALGORITHMS

Dr V Vasudha Rani,
Associate Professor,
Dept. of CSE, GMR Institute of Technology, Rajam, AP
vasudharani.v@gmrit.edu.in

Dr. G.Vasavi
Associate Professor
BV Raju Institute of Technology
vasavi.g@bvrit.ac.in

Balajee Maram,
School of Engineering and Technology, Chitkara University, Baddi, Himachal Pradesh, India.
maram.balajee@chitkarauniversity.edu.in

ABSTRACT: **Skin is an extraordinary human structure. As a result of inherited traits and environmental variables, skin conditions are the most prevalent worldwide. People frequently neglect the effects of skin diseases in their initial stages. It commonly experienced both well-known and rare diseases. Identifying skin diseases and their kinds in the medical field is a very difficult process. It can be very challenging to identify the precise type of disease because of the intricacy of human skin complexion as well as the visual proximity effect of the conditions. As a result, it's critical to identify and categorize skin diseases as soon as they are discovered. The most ambiguous and challenging field in science is therefore the detection of human skin diseases. For segmentation and diagnosis, ML techniques are frequently employed in the biomedical industry. These techniques decide using features extracted from photos as their input. To obtain high classification accuracy, it is crucial to select appropriate feature extraction techniques along with appropriate Machine Learning (ML) approaches. The classification of skin diseases is discussed in this analysis using ensemble data mining approaches and ML algorithms. In this method, four distinct ML techniques are used to categorize the various kinds of diseases while ensemble approaches are used to increase the classification reliability of skin diseases.**

KEYWORDS: Skin disease, classification, Machine Learning, Ensemble data Mining techniques

I. INTRODUCTION

Skin can increase its skin barrier function by using fluids to stop lipid degradation in the epidermis. The epidermis, which covers a surface area of 21 to 23 square feet, is the most important biological organ in the human body. It aids in thermoregulation, permits sensibility, heat, and cold emotions, and shields various important organs from external injury, pathogens, and the atmosphere. However, a variety of external and hereditary factors can affect the skin. Skin conditions are the most prevalent among humans. Infectious skin disorders, bacterial skin disorders, and contact dermatitis abnormalities are the three types of skin conditions that have the biggest effects on human skin. A skin condition may cause changes to the skin's texture or colour [1].

The skin plays a crucial role in controlling body temperature and guards from viruses, bacteria, allergies, and fungi. However, many people experience skin issues that have a variety of underlying reasons. The most prevalent skin problems are psoriasis, ringworm, alopecia, and eczema [2]. Skin conditions affect practically all age categories of people. Because of altering environments and habits, skin disease rates have increased.

Skin conditions are chronic, infectious, and occasionally can result in skin cancer. People with darker skin tones are approximately 20 to 30 times more likely to develop melanoma than those who have lighter skin tones, it has been discovered that people with darker skin tones either have a higher or lower mortality risk for specific types of melanoma, depending on their skin tone. In order to administer the appropriate treatment, it is essential to identify a skin lesion correctly [3]. Skin problems are frequently fairly difficult to identify at an initial stages, and it is much more difficult to categorize them independently. Melanoma is currently known as the most serious type of skin cancer among all others because, if not detected and treated promptly, it has a far higher propensity to spread to other regions of the body [4].

It is simple to cure fungal and allergy diseases if they are correctly diagnosed and identified in their initial stages. However, in the case of viral illnesses, prompt detection is crucial. Skin illnesses can develop as a undetected germs, microbes modifying the texture of the skin, or microbes producing pigment. Chronic skin conditions sometimes develop into cancerous tissues. Skin conditions need to be treated right away to reduce their growth and spread [5].

Skin problems can be brought on by a variety of variables, including biological ones like insect bites,

978-1-6654-9056-6/22/$31.00 @2022 IEEE

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on November 20,2024 at 16:26:59 UTC from IEEE Xplore. Restrictions apply.

allergic conditions, and even viral infections, as well as physical ones like light, temperature, and friction. Skin illnesses can also develop as a result of environmental and hereditary factors [6]. Differences in skin tone, the presence of artifacts like hair, air bubbles, uneven illumination, and the lesion's physical position can all make lesion imaging more challenging. Additionally, most lesions differ in terms of colour, texture, form, size, and placement within an image frame.

Several skin illnesses have symptoms that can take a long time since they may develop for years prior to getting noticed, to treat. As a result, computer-based disease diagnosis comes into play since it can produce a quicker and more accurate outcome than human assessment employing laboratory techniques. Many fields, including face identification, fingerprint recognition, tumor detection, and segmentation, apply ML-based investigations. Machine Learning techniques have shown promise in the classification and prediction of a variety of diseases. In the realm of medicine, ML techniques are frequently applied. To discover the highest degree of accuracy in forecasting the disease, a variety of disease diagnosis classification techniques are developed.

After looking at the many characteristics of the disease, numerous ML algorithms are built for forecasting the various types of sickness at an early stage. These algorithms are generally applicable to a variety of conditions including diabetes, cancer, thyroid illness, erythemato-squamous disorders, breast cancer, and many others [7]. Ensemble Data Mining Methods, sometimes referred to as Committee Techniques or Model Combiners, are ML techniques that make use of numerous models to increase prediction accuracy beyond what any one algorithm might do on its own.

In this analysis, Ensemble data mining approaches and Machine Learning techniques for skin disease classification is presented. Literature Survey is presented in Section II, Classification of Skin Disease is explained in Section III, Result Analysis is presented in Section IV and Conclusion in Section V.

II. LITARATURE SURVEY

V.R. Balaji et. al., [8] demonstrates the identification and categorization of skin diseases employing the dynamic graph cut technique and the Naive Bayes (NB) classifier. In this investigation, they plan to divide skin lesions using a unique dynamic graph cut technique, followed by the categorization of skin diseases employing a NB classifier.

Nawal Soliman ALKolifi ALEnezi et. al., [9] offers A Skin Disease Detection Method Employing ML and Image Processing. They provided a technique for identifying skin issues based on image processing. This technique uses a digital photograph to identify the kind of disease through image analysis. This method is straightforward, quick, and does not call for expensive tools beyond a lens and computer.

Jessica Velasco et. al., offers A MobileNet CNN-Based Smartphone Skin Disease Classification. To further increase the effectiveness of the MobileNet, several sampling techniques and input data pre-processing were investigated. An accuracy of 84.28 percent was attained using the under sampling approach and the standard pre-processing of the input data. While achieving 93.6 percent accuracy with an unbalanced dataset and the default pre-processing of the incoming data.

Anurag Kumar Verma et. al., [10] used ensemble approaches to use ML algorithms to segment different types of skin diseases, and they then used a feature selection method to compare the findings. The subset generated by the feature selection approach is contrasted with the ensemble technique.

V. Pugazhenthi et. al., [11] Detection And Classification Of Skin Diseases is explained. Employing image processing as well as categorization techniques, the major goal of this analysis is to increase the diagnostic systems accuracy. A picture taken using a camera is used as an input for the suggested system. Contrast Enhancement as well as Grayscale Conversion would be used as part of the pre-processing of this image to make it segmentation-suitable. If there is a disease, this method uses the Decision tree approach to identify it as Melanoma, Leprosy, or Eczema.

Saad Albawi et. al., [12] Deep Neural Network-based robust skin disease detection and classification is shown. This analysis suggested a novel classification system for three different skin diseases, including melanoma, Nevus, and Atypical. The source's skin picture is pre-processed using an adaptive filtering technique to remove pointless noisy parts. For precise segmentation of lesions, this segmentation chooses the following zone to expand adaptively.

There are three openly accessible image recognition architectures—InceptionV3, InceptionResnetV2, and Mobile Net with alternations for the applicability to skin diseases and correctly diagnoses the skin illness dependent on the highest voting from the three networks.

Kumari A et. al., [13, 14] provides a method for diagnosing skin diseases using colour and texture features in images. In this investigation, a new recognition approach was used to detect three types

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on November 20,2024 at 16:26:59 UTC from IEEE Xplore. Restrictions apply.

of skin illnesses, including herpes, dermatitis, and psoriasis. Dr.V.Vasudha Rani et. al. found that by increasing the training data set size through augmentation of images, improves the performance of deep neural network models.

III. CLASSIFICATION OF SKIN DISEASE

This section presents the block diagram of proposed model in Fig. 1. The input skin image is taken from Data for Skin Disease. The UCI (University of California Irvine) ML repository provided the database for this investigation. In a summary, this dataset was created to investigate skin illnesses and categorize them. There are 34 variables in this dataset; 33 of them are linear, and one is a nominal variable.

One of the fundamental approaches to these disorders diagnosis is biopsy. When completing the many classes of detection of these diseases, dermatologists encounter another challenge: a disease may also have characteristics of some other class of diseases during the initial stage. Patients were initially evaluated for 12 clinical aspects, and then employing skin disease samples, 22 histological qualities were evaluated.

They have to solve a few issues with colour contrast and image resolution that crop up during data loading in order to get a good reliability of skin disease categorization. They have a component in the application that takes responsibility for this to solve the issue. Before uploading the images to server for further processing, the image resizing software in Python does all of the work. Therefore, the primary goal of this stage is to eliminate backdrop noises from the skin disease image. The format and edges of the skin infection image are then improved through post-processing.

Color feature: The Scikit-image package's colour De-convolution principle is employed to segment cropped images in order to avoid non-lesion areas. Immuno-histochemical staining separation has been accomplished utilizing the Hematoxylin, Eosin, and DAB (Diamino Benzidine) segmentation approach, which is also the method utilized in this analysis. In this investigation, a variety of segmentation techniques were examined, including multiple thresholding methods, color-based segmentation, background subtraction, and skin colour modeling. However, the HED segmentation algorithm produced the best outcomes. In ML-based classification tools, colour characteristics are one of the key parts of disease diagnosis according to clinical standards;

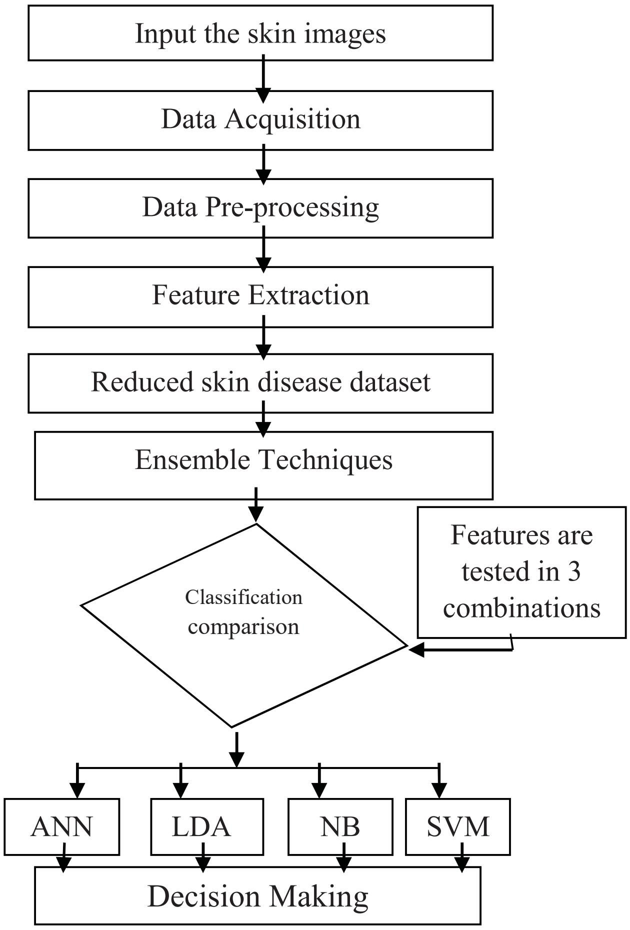

Fig. 1: THE BLOCKDIAGRAM OF PRESENTED SKIN DISEASE CLASSIFICATION MODEL

The colour features are significant. Red, Green, and Blue (RGB) colour space is employed in this investigation due to its widespread use and role as the standard colour space for digital photographs. Separating the RGB hues, the average result was recorded for further investigation.

Texture feature: The pre-defined function in the Scikit-image package is used to extract GLCM features. It is additionally referred to as a grey level spatial dependence matrix. It analyses the texture using statistical techniques while taking the spatial relationship among pixels into account. Employing the GLCM method, the succeeding properties are generated.

The gray-scale co-occurrence matrix's local variations are measured via contrast. Correlation calculates the likelihood that the chosen pixel pairs will appear. The value through which GLCM probabilities are enhanced and grows linearly far from the diagonal is called dissimilarity. In the GLCM, energy equals the total of the squared components. The degree of homogeneity indicates how closely distributed GLCM pieces are to the

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on November 20,2024 at 16:26:59 UTC from IEEE Xplore. Restrictions apply.

diagonal. The Angular Second Moment (ASM) determines the rotational acceleration measurements.

Four distinct classifiers get extracted features that have been tagged with the appropriate diseases for training as well as testing. Now, they simplify the dataset down to just 15 features and 366 cases, and then assess its categorization of skin diseases accuracy once more.

The accuracy of dataset for skin diseases is determined using the ensemble approach, which enhances algorithm efficiency. Using bagging classifiers and gradient boosting classifiers, they would mix four distinct ML techniques. Bagging is used to deal with bias-variance trade-offs and reduces the variance of a prediction model.

Bagging, often referred to as Bootstrap aggregating, is a ML model aggregation approach created to increase the consistency and accuracy of the algorithms and reduce variance in order to prevent over fitting when used in classification and regression approaches. It creates M duplicates of the source dataset D using N training samples $\{D_m\}_{m-1}^{M}$. With replacement sampling, every Dm is produced from D. The number of instances in every dataset Dm is identical to that in dataset D. These datasets are comparatively distinct from one another. It uses D1;......DM to train the models h1,......, and hM. Uses an averaged design that is as follows, as the final framework:

$$h = \frac{1}{M}\sum_{m-1}^{M} h_m \qquad (1)$$

An illustration of a generalized boosting technique is gradient boosting. Gradient boosting algorithm can be used for predicting not only continuous target variable (as a Regressor) but also categorical target variable (as a Classifier). For classification and regression issues, the gradient boosting ML technique provides a classification structure in a format of an ensemble of weaker classification methods, often decision trees. In view of the training data ((X1, Y1), (XN, YN). Set a constant as f_0's initial value. Perform the following for t = 1 to M: i) Calculate $g_t(x)$ negative gradient, ii) A novel base function, h(x, Ɵt), should be fitted iii) Identify the ideal gradient descent step length.

$$\rho_t : \rho_t = argmin\rho \sum_{i=1}^{N} \Psi\,[\mathrm{Y_i, f_{t-1}}\,(\mathrm{x_i}) + \rho h(\mathrm{x_i}, \theta_t)] \qquad (2)$$

Upgrade the function estimate$\mathrm{f_t} \leftarrow \mathrm{f_{t-1}} + \rho_t\, h(x, \theta_t)$. The following ML classifiers are chosen for the classification of different skin diseases.

A statistical structure called an Artificial Neural Network (ANN) is premised on the idea of a biological NN. Three layers comprise a neural network method: input, hidden, and output layers. Input layer is the initial layer. The capacity for problem resolution grows as hidden layers are added. AAN was put into execution using 100 neurons (the size of the hidden layer), an alpha value of 0.1 and 200 iterations. Although linear Discriminant analysis employs dimensionality reduction to identify a unique feature space to reflect the information in, it is a linear classification method.

The Bayes theorem is the source of Naive Bayes. It makes the assumption that a feature's existence in a class has nothing to do with any other feature. This investigation made use of Gaussian Naive Bayes. One of the popular classification methods, Support Vector Machine, uses a decision plane to divide a data set into multiple classes. With only one iteration, the linear kernel, the shrinking feature, and no randomization, they implemented it.

Three comparisons are made between features: the colour alone, the texture alone, and lastly the mixture of both. Four classifiers test all of the features. This approach effectively classifies different types of skin diseases namely melanoma, Plaque Psoriasis, Lichen Planus, Pityriasis Rosea. Four different ML classifiers performance is compared to obtain better ML classifier for accurate and effective skin disease classification.

IV. RESULT ANALYSIS

In this section, presented ensemble information mining methods and ML algorithms for skin disease classification result analysis are discussed. This presented skin disease classification approach is implemented using python.

The Ensemble data mining techniques namely bagging and gradient boosting are utilized to enhance the performance of skin disease classification model. These ensemble approaches are applied to four different ML classifiers for effectively classifying the distinct kinds of skin infections. The four ML classifiers LDA (Linear Discriminant Analysis), ANN, SVM and NB classifiers performance is compared in terms of accuracy, specificity, sensitivity and precision for obtaining better classifier among them. The performance of these ML classifiers is evaluated like below:

TP: If a condition is appropriately categorized as positive and it genuinely is positive

TN: If a condition is deemed negative and is indeed negative.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on November 20,2024 at 16:26:59 UTC from IEEE Xplore. Restrictions apply.

FP: If a situation is labeled as positive when it is truly negative.

FN: If a instance is classified as negative but actually positive.

Accuracy: It is described as the proportion of accurate occurrences to all instances, and it is given as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

Sensitivity: It is also known as True Positive Rate (TPR) and is stated as and is understood to be the proportion of true positive instances to actual positive instances.

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

Specificity: It is characterized as the ration of true negative instances to the actual negative instances (i.e., FP + TN) and is expressed as

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

The performance metrics of presented ML classifiers are represented in Table 1.

TABLE 1: PERFORMANCE METRICS OF DIFFERENT ML

ML classifiers	Accuracy	sensitivity	specificity
NB	89.9	90.2	89.5
ANN	92.4	93.5	91.8
LDA	94.7	95.6	92.4
SVM	95.6	95.8	93.5

Among these four algorithms, SVM has better performance followed by LDA, ANN and NB. The SVM and LDA classifiers effectively classified the different types of skin diseases. The performance comparison between these models is shown in Fig. 2.

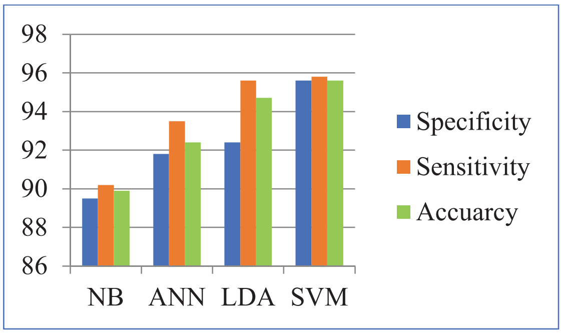

Fig. 2: PERFORMANCE COMPARISON BETWEEN DIFFERENT ML ALGORITHMS

The performance of presented ensemble data mining techniques and ML algorithms is compared with traditional ML model.

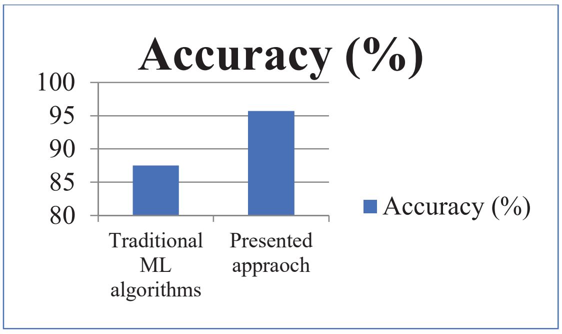

Fig. 3: PERFORMANCE COMPARISON OF TRADITIONAL ML MODEL AND PRESENTED MODEL

Hence, presented ensemble data mining techniques and ML algorithms for skin disease classification has high classification accuracy and it classify very effectively and accurately than traditional ML algorithms.

V. CONCLUSION

In this work, ensemble data mining techniques and ML algorithms for skin disease classification are presented. UCI Machine Learning repository dataset and skin disease data set are used in this approach. Bagging and gradient boosting ensemble methods are employed to enhance the performance of applied ML classifiers. The four different types of ML algorithms namely Linear Discriminant Analysis, Naïve Bayes, ANN and SVM are utilized and their performance is compared for obtaining better algorithms for skin disease classification. Distinct kinds of skin infections include melanoma, Plaque Psoriasis, Lichen Planus, Pityriasis Rosea are classified effectively through this model. The SVM algorithm has accuracy 95.6 which is better than LDA, ANN and NB. Compared to traditional ML algorithms, presented model with ensemble techniques and ML algorithms has high accuracy; therefore, presented model has effectively classified different types of skin diseases.

REFERENCES

[1] Samir Kumar Bandyopadhyay, Payal Bose, Amiya Bhaumik, Sandeep Poddar, "Machine Learning and Deep Learning Integration for Skin Diseases Prediction", International Journal of Engineering Trends and Technology Volume 70 Issue 2, 11-18, February, 2022, ISSN: 2231- 538, doi:10.14445/22315381/IJETT-V70I2P202

[2] Syeda Fatima Aijaz, Saad Jawaid Khan, Fahad Azim, Choudhary Sobhan Shakeel, and Umer Hassan, "Deep Learning Application for Effective Classification of Different Types of Psoriasis", Hindawi, Journal of Healthcare Engineering Volume 2022, Article ID 7541583, 12 pages, doi.org/10.1155/2022/7541583

[3] Viswanatha Reddy Allugunti, "Machine learning model for skin disease classification using convolution neural network",

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on November 20,2024 at 16:26:59 UTC from IEEE Xplore. Restrictions apply.

International Journal of Computing, Programming and Database Management 2022; 3(1): 141-147

[4] Ahmed A. Elngar, Rishabh Kumar, Amber Hayat, Prathamesh Churi, “Intelligent System for Skin Disease Prediction using Machine Learning”, Journal of Physics: Conference Series 1998 (2021) 012037, IOP Publishing doi:10.1088/1742-6596/1998/1/012037

[5] Parvathaneni Naga Srinivasu, Jalluri Gnana SivaSai, Muhammad Fazal Ijaz, Akash Kumar Bhoi, Wonjoon Kim and James Jin Kang, “Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM”, Sensors 2021, 21, 2852, doi.org/10.3390/s21082852

[6] Leelavathy S, Jaichandran R, Shobana R, Vasudevan, Sreejith S Prasad and Nihad, “ Skin Disease Detection Using Computer Vision And Machine Learning Technique”, European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 4, 2020

[7] Kritika Sujay Rao, Pooja Suresh Yelkar, Omkar Narayan Pise, Dr. Swapna Borde, “Skin Disease Detection using Machine Learning”, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, NTASU - 2020 Conference Proceedings, Volume 9, Issue 3

[8] V.R. Balaji, S.T. Suganthi, R. Rajadevi, V. Krishna Kumar, B. Saravana Balaji, Sanjeevi Pandiyan, “Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier”, ”, 2020 Elsevier Ltd, doi:10.1016/j.measurement.2020.107922

[9] Nawal Soliman ALKolifi ALEnezi, “ A Method of Skin Disease Detection Using Image Processing And Machine Learning”, 16 th International Learning & Technology Conference 2019, doi: 10.1016/j.procs.2019.12.090

[10] Anurag Kumar Verma, Saurabh Pal, Surjeet Kumar, “Comparison of skin disease prediction by feature selection using ensemble data mining techniques”, Informatics in Medicine Unlocked 16 (2019) 100202, doi.org/10.1016/j.imu.2019.100202

[11] V. Pugazhenthi, Sagar K. Naik, Amruta D. Joshi, Shreya S. Manerkar, Vinita U. Nagvekar, Kalpita P. Naik, Chinmay G. Palekar, K Sagar, “ Skin Disease Detection And Classification”, 2019, International Journal of Advanced Engineering Research and Science (IJAERS), 2019, 6 (5), pp.396-400. 10.22161/ijaers.6.5.53. hal-02141241

[12] Saad Albawi, Yasir Amer Abbas, Yasser Almadany, “Robust Skin diseases detection and classification using deep neural networks”, International Journal of Engineering &Technology, 7(4), 2018, 6473-6480, doi:10.14419/ijet.v.7i4.24178

[13] Kumari, A., Behera, R.K., Sahoo, B., Misra, S. (2022). Role of Serverless Computing in Healthcare Systems: Case Studies. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (eds) Computational Science and Its Applications – ICCSA 2022 Workshops. ICCSA 2022. Lecture Notes in Computer Science, vol 13380. Springer, Cham.

[14] Dr.V.Vasudha Rani, Dr. G.Vasavi, Dr. K.R.N. Kiran Kumar, A Detailed Review on Image Augmentation and Segmentation of Brain MRI Images Using Deep Learning, PAIDEUMA JOURNAL, Vol XIV Issue 8 2021, ISSN NO : 0090-5674.

Authorized licensed use limited to: VIT University- Chennai Campus. Downloaded on November 20,2024 at 16:26:59 UTC from IEEE Xplore. Restrictions apply.