2023 4th International Conference for Emerging Technology (INCET)
Belgaum, India. May 26-28, 2023

Skin Disease Classification using Machine Learning based Proposed Ensemble Model

Bisahu Ram Sahu
Department of IT & CS,
Dr. C.V. Raman University
Kota, Bilaspur, (C.G.) India
bisahu.sahu@gmail.com

Akhilesh Kumar Shrivas
Department of CSIT,
Guru Ghasidas Vishwavidyalaya
Bilaspur, (C.G.) India
akhilesh.mca29@gmail.com

Abhinav Shukla
Department of IT & CS,
Dr. C.V. Raman University
Kota, Bilaspur, (C.G.) India
amitshukla9611@gmail.com

2023 4th International Conference for Emerging Technology (INCET) | 979-8-3503-3575-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/INCET57972.2023.10170128

Abstract —**Skin disease is a major issue of global health problem affecting a large amount of persons. The advancement of dermatological diseases categorization has grown more accurate in recent years due to the rapid growth of technology and the use of various machine learning techniques. Therefore the creation of machine learning methods that can accurately differentiate between the classifications of skin diseases is one of the great importance. This research work focuses on the classification of different kinds of skin diseases using machine learning techniques. In this research, we introduce a novel approach that makes use of four distinct data mining techniques like support vector machine (SVM) , k-nearest neighbor (KNN) , random forest (RF) and , naïve bayes (NB) algorithm. This research work proposed an ensemble model that is combination of SVM, KNN, RF and NB using voting scheme. The proposed model classified the skin disease into five different classes that are Acne, Skin allergy, Nail fungus, Hair loss, and Normal skin. The proposed ensemble model used on skin disease classification that gives better performance over other classifier algorithms. The proposed ensemble model achieved highest 97.33% of accuracy as compared to others.**

Keywords—Naive Bayes, Support Vector Machines, Skin Disease, Acne, Ensemble method, Max Voting Scheme, Classification.

I. INTRODUCTION

Nowadays, classification of skin disease play very impoartant and cruvial role among people all over world. Various authors used different machine learning techniques to develop the robust model for skin disease classification. This research work have used machine learning based classification techniques for developing proposed ensemble model for classification of skin disease.The technology can match a dermatologist's results using sophisticated computer methods and a sizable dataset, raising the bar for research and medical excellence. Machine learning techniues are capable to develop a robust model to classify the various types of skin diseases.The diagnosis of skin disease is very complicated task due to variety of human skin. The other cause of the diagnosis of skin disease is lack of expertise. To solve the problem of skin disease diagnosis, machine learning technqiues play very important role. Now a days machine learning based model are playing improrant role to analysis of healthcare images and develop a robust model for classification of healthcare diseases like skin disease. Now a days, some effective tools and libraries are using by diiferent researchers like Scikit-Image, OpenCV and NumPy[19]. The main objective of this reserch work is to build a robust and effective model that classify the skin and non skin disease images uisng machine learning based proposed model. Authors[1] developed a novel algorithm GLCM for classification of skin disease. The proposed GLCM model given satisfactory aaccuracy with skin disease dataset. Author[2] suggested machine learning methods KNN and SVM classification algorithms and fuzzy clustering with 50 sample images of dematology dataset. They soggested KNN with fuzzy clustering method as compared to SVM with fuzzy clustering where soggested algorthmg given better accuracy as 91.2%. Author [3] developed mobile based dermatology detection application based on Image Processing that system worked even in remote areas. DL and image processing delivered real-time skin disease detection algorithm based on accuracy. They compared Support Vector Machine (SVM) and Convolutional Neural Networks (CNN). Author [4] focused to identify whether a given skin disease is infected by Melanoma or not. They trained the SVM algorithm with labeled dermatology data using a suitable kernel and achieved about 90% accuracy. In paper [5] authors combines Convolutional Neural Network with Support Vector Machine classifier to develop a Mobile Android Application (CNN-SVM -MAA), where skin disease dataset consists 3000 images which collected from Beni-Suef University Hospital, Cairo University Hospital and different web sites as well.Mobile android application provides help to doctor for prescribe patient with high accuracy. Author [6] used KNN classification algorithm and merged two FST like LBP and GLCM. First, dematology data set is collected from International cancer centre at Neyoor. This dataset contains 20 images including 10 images are benign and 10 are melanoma. Finally, develop the robust model uisng KNN classifier with segementaion and FST and achieved accuracy rate of 93%. Author [7] compared the performance of SVM and NB algorithms with skin disease dataset.They used real time photograph as input of infected region and achieved real-world testing accuracy as 77.23% and 83% with SVM and NB respectively. Author [8] used CNN and SVM algorithms for obtaining highest accurate results with infected area of skin images . CNN and SVM algorithm is performed on this dermatology test images and finally after tuning many more times of the model then author achieved 90% of accuracy with CNN. Author [9] used KNN data mining techniques to classify between normal skin and malignant skin lesions diseases. They used MATLAB graphical user interface platform for experimental work and they achieved the highest accuracy of 98% in KNN. Author [10] developed a system which recognize the disease based on input symptoms. They collected 10 skin disease images with the help of expert doctor based on symptoms of patient from the department of Skin G. S. M. C.

KEM Hospital, Parel Mumbai. Authors used ANN, KNN,SVM,DT,RF and these 5 machine learning algorithm were trained on the symptom's skin image dataset and they got 100%,98%,99%,97.7% and 100% accuracy with ANN,KNN,SVM,DT,RF respectively. Authors[11] suggested ensemble of SVM and ANN where proposed esenble model achieved better accuracy as compared to SVM and ANN. Authors[14] compared the performance of different classifiers like RF, SVM, LR, KNN and NB for skin disease classification where NB achieved highest accuracy. Another authors[15] proposed an ensemble model where our proposed ensemble model achived better accuracy ac compared to individual classifiers.

The main purpose of these investigations is differential analysis of Acne, Skin allergy, Nail fungus, Hair loss, and Normal skin disease. In this research work, an attempt is made to combine four independent data mining techniques like Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), and K-Nearest Neighbor using voting scheme ensemble technique. These four data mining techniques are combined to create an ensemble model that can forecast skin conditions. Using the same dataset of skin diseases, each of the four approaches is used separately. After that, an ensemble model is created using these machine learning techniques. The final prediction outcomes demonstrate that, the proposed outcome is compared with individuals data mining strategies, where our proposed ensemble model achieved better performance with given skin image dataset.

II. Methodology and Proposed Architecure

This research work focused on the classification of skin disease using different types of machine learning techniques. The machine learning techniques have ability to learn from past experience datasets. Figure 1 show that generic architecture of proposed model where RF, NB, SVM and KNN classifiers are used to develop a proposed ensemble model. This research work proposed an ensemble model that achieved better performance as compared to others.

A. Support Vector Machine (SVM)

SVM is a supervised machine learning technique that is used for classification and prediction problems [2]. Many researchers have used classification techniques in many domains for classification and regression problems. This algorithm plots every data point as a point in n-dimensional space, where n is the number of features and each feature's value is a specific coordinate value. We then perform classification by identifying the hyper-plane that effectively distinguishes the five classes. Support Vectors are nothing more than an individual observation's coordinates. The technology which best separates the five classes is support vector machine.

B. Random Forest (RF)

RF is a type of decision tree algorithm that is used for classification of data samples into different categories. RF is an ensemble classifiers that is combine with many decision tree. This classifier is robust and able to handle the noise as well as missing values. RF classifier will not over fit the model when there are more trees in the forest and for categorical values as well[10].

C. Naive Bayes Algorithm (NB)

NB is statistical technique that work based on Bayes' theorem. It is used for classification and prediction of data. Bayes theorem based on the following formulas: P(H/X)=[P(X/H). P(H)]/P(X) where X is the sample set, H is the hypothesis that X belongs to class C,P(X/H) is the posterior probability of H. NB is suitable approach for classification of categorical data [7].

D. K-Nearest Neighbor (KNN)

KNN is an unsupervised machine learning technique that is used for classification and prediction of data. KNN is classifying the data based on the nearest point of similar object. KNN is mostly used for classification purpose, but it can be used for regression problem also. The term majority of voting used in various research works[1].

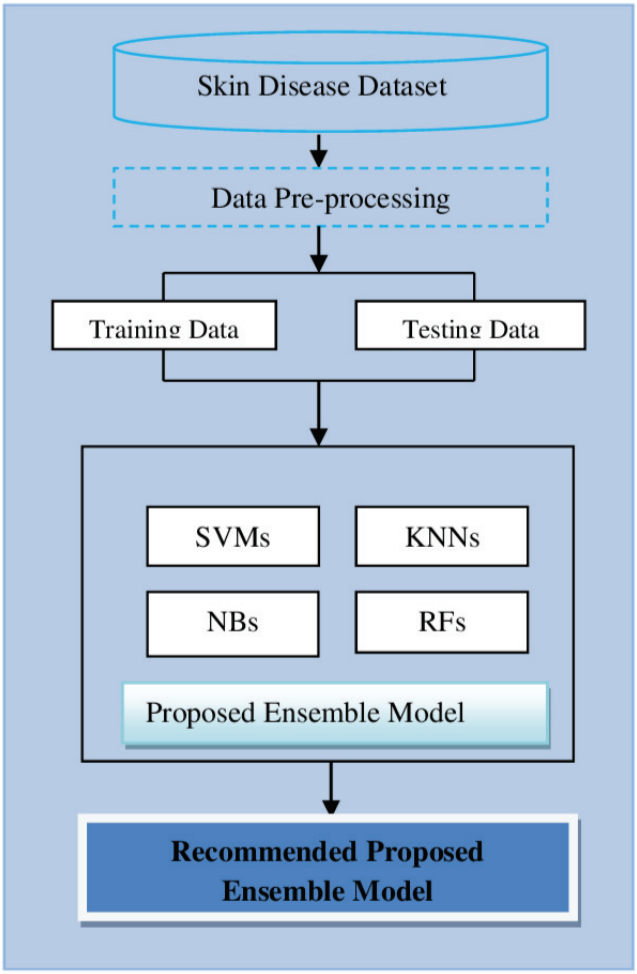

Fig. 1. Flow diagram of methodological approach

The above figure shows that flow of proposed ensemble model that is combination of SVM, KNN, NB and RF. In this flow architecture, the skin disease dataset is divided into training and testing for train and test the proposed ensemble model and individulas clasifiers.

III. Dataset

We have taken dermatology dataset from kaggle data science website .It contains 750 images with 5 different types of classes are Acne, Skin allergy, Nail fungus, Hair loss, and Normal skin. We have used 80% of the images for training, and 20% images for testing in which 600 images for training and 150 images for testing that play effective role to develop the robust model for classification of skin disease [20].

A. Data Preprocessing

Data preprocessing play very effective role to improve the performance and develop the robust model. The first step in data preprocessing uses a (i) data-driven strategy to select the relevant features from datasets (ii) The collected data from open source repository is not completely preprocessed data, it contains noise, missing and inconsistent values that decrease the performance of models.(iii) Data transformation also play very important role to transform of data sample into suitable format and size that will be useful for mining and achieve better results.

B. Ensembles Method

An ensemble method play very important role to combine the two or more classifier and achieve the better accuracy as compared to base classifiers. This research work used an ensemble approach for classification of different kinds of skin disease. The figure 2 shows that an ensemble approach where training samples are supplied to individuals classifiers C1, C2 …Cn and P1, P2 …P_n are performance of C1, C2.. Cn classifiers respectively. The voting scheme is used ensemble the above individual classifiers and gives the final prediction or performance of ensemble classifier.

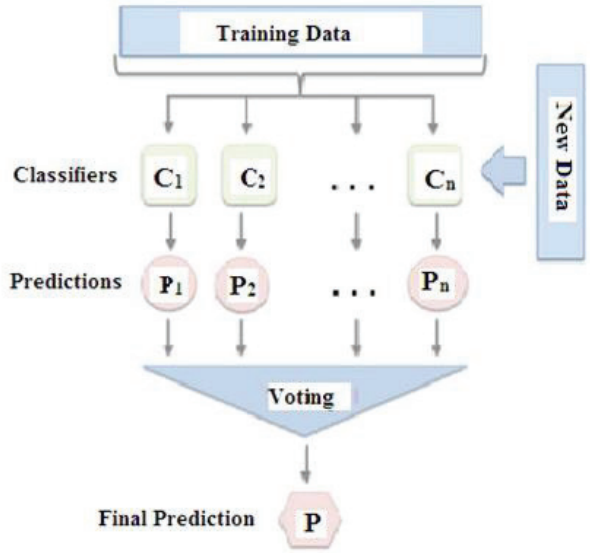

Fig. 2. Ensemble Approach

IV. Experimental Approch

In this experiment, we have used different classification techniques for skin disease classification with python software. These research works have used KNN, SVM, NB and RF as individual's classifiers for classification of skin disease. We have also compared the proposed ensemble model with others classifiers like KNN, SVM, NB and RF where proposed ensemble classifier that achieved better accuracy as compared to others. The proposed ensemble model is combination of KNN, SVM, NB and RF. The accuracy of proposed ensemble model is shown in Table I where proposed ensemble model achieved better performance as compared to others. Figure 3 shows that graphical representation with accuracy of existing and proposed ensemble model. Table II shows that classification report of proposed ensemble model. Confusion matrix play very important role to evaluate the performance of model. Figure 4 shows that the confusion matrix of proposed ensemble model. Table III shows that our proposed ensemble model gives better accuracy as compared to other previous research work done by many researchers. Our proposed ensemble model achieved 97.33% of accuracy which is highest among others.

TABLE I. ACCURACY OF PROPOSED AND EXISITING METHODS

Algorithms	Accuracy
KNN	73.66%
SVM	96.57%
NB	47.00%
RF	96.68%
Proposed Ensemble Model	**97.33%**

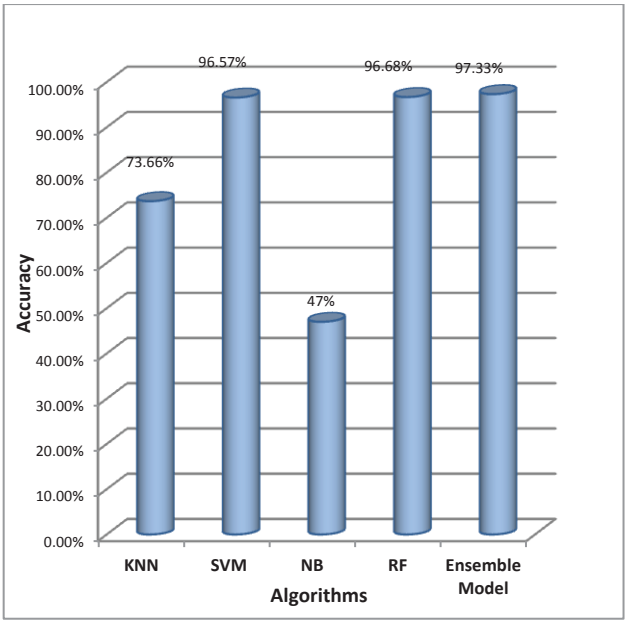

Fig. 3. Accuracy of existing and proposed ensemble model

TABLE II. CLASSIFICATION REPORT OF PROPOSED ENSEMBLE MODEL

CLASSIFICATION REPORT				
	precision	recall	f1-score	support
Acne	0.96	1.00	0.98	101
Hairloss	0.98	1.00	0.99	105
Nail Fungus	0.94	0.98	0.96	153
Normal	1.00	0.9	0.95	111
Skin Allergy	1.00	0.98	0.99	11
accuracy			0.97	600
macro avg	0.99	0.99	0.99	600
weighted avg	0.97	0.97	0.97	600
accuracy_score	**97.33%**			

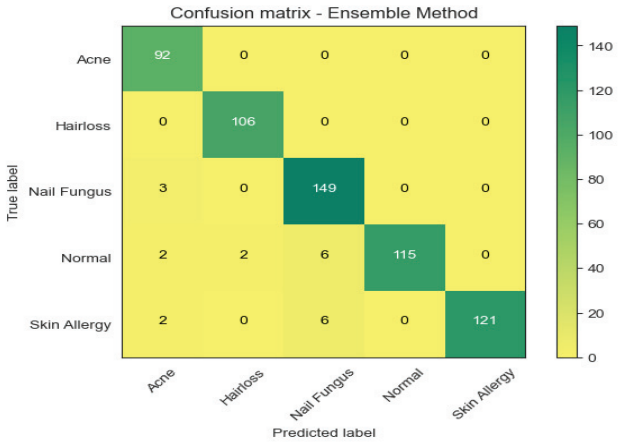

Fig. 4. Confusion matrix of proposed ensemble model

TABLE III. COMPARATIVE ANALYSIS OF PROPOSED ENSEMBLE MODEL WITH PREVIOUS WORK

Reference	Year	Proposed Methods by different authors	Classification accuracy (Percentage)
[12]	2017	ANN	92.5
[13]	2018	GLCM+SVM	90
[16]	2019	GLCM+SVM	89.65
[18]	2022	CNN+SVM	91.04
[19]	2022	SVM	90.7
Our Proposed Ensemble Model			**97.33%**

V. CONCLUSION

This research work uses skin disease dataset consists of five classes like Acne, Skin allergy, Nail fungus, Hair loss, and Normal skin and used four different classification techniques like RF, SVM, NB and K-NN. We have also compared the performance of proposed ensemble model with individual's classifiers. The proposed ensemble is combination of SVM, NB, RF and SVM where proposed ensemble model achieved highest 97.33% of accuracy as compared to other individual classifiers and previous proposed model done by different authors. Finally, we recommended that our proposed ensemble model achieved satisfactory performance for classification of skin disease as compared to others.

REFERENCES

[1] A. Nosseir and M. A. Shawky, "Automatic Classifier for Skin Disease Using k-NN and SVM," in Proceedings of the 8th International Conference on Software and Information Engineering, Cairo Egypt, pp. 259–262,2019.

[2] D. M. Jagdish, S. P. G. Guamangate, M. A. G. López, J. A. De, L. Cruz-Vargas, and M. E. R. Camacho, "Advance Study Of Skin Diseases Detection Using Image Processing Methods,"vol. 9,pp. 997-1007,2022.

[3] M. J. Hajgude, A. Bhavsar, H. Achara, and N. Khubchandani, "Skin Disease Detection Using Image Processing with Data Mining and Deep Learning," International Research journal of Engineering and Technology(IRJET),vol. 06,pp.4363-4366 ,2019.

[4] N. V. Kumar, P. V. Kumar, K. Pramodh, and Y. Karuna, "Classification of Skin diseases using Image processing and SVM,"International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India,pp. 1–5,2019.

[5] A. A. Elngar, R. Kumar, A. Hayat, and P. Churi,"Intelligent System for Skin Disease Prediction using Machine Learning," J. Phys. Conf. Ser., vol.1,2021.

[6] A.G.Shiji,K.J.Malar,"Malanoma detection in dermatology using knn," International Research Journal of Modernization in Engineering Technology and Science(IRJMETS),vol.03,pp.3482-3486,2021.

[7] V.S.R.K.Sarma,V.Chandrahaas,"skin disease detection using image processing and machine learning (svm) ,"Neuro Quantology, vol.20,pp.7071 - 7077,2022.

[8] L.A,A.K.Jaiswal,"Detection and Classification of Skin Diseases,"International Research Journal of Engineering and Technology (IRJET),vol.07,pp.3658- 3663,2020.

[9] M. Q. Hatem, "Skin lesion classification system using a K-nearest neighbor algorithm,"Vis. Comput. Ind. Biomed. Art, vol. 5,2022.

[10] S. Kolkur, D. D. R. Kalbande, and D. V. Kharkar, "Machine Learning Approaches to Multi-Class Human Skin Disease Detection," International Journal of Computational Intelligence Research(IJCIR),vol.14,pp.29-39,2018.

[11] D.K.Sharma, H.S.Hota. "Data mining techniques for prediction of different categories of dermatology diseases", Academy of Inform Mgmt Sci J.,vol.14, pp.103-15,2013.

[12] M. Koklu and I. A. Ozkan, "Skin Lesion Classification using Machine Learning Algorithms," International Journal of Intelligent System and Applications in Engineering(IJISAE), vol. 4, pp. 285– 289,2017.

[13] L. Wei, Q. Gan, and T. Ji, "Skin Disease Recognition Method. Based on Image Color and Texture Features," Comput. Math. Methods Med.,vol.2018,pp. 1–10, 2018.

[14] M. Sudha, B. Poorva, and PG Scholar, School of Information Technology & Engineering, VIT Vellore, India., "Predictive Tool for Dermatology Disease Diagnosis using Machine Learning Techniques," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8,pp. 355–360,2019.

[15] A. K. Verma, S. Pal, and S. Kumar, "Classification of Skin Disease using Ensemble Data Mining Techniques," Asian Pac. J.Cancer Prev., vol. 20, pp. 1887–1894,2019.

[16] Azeem Iqbal, "Skin Disease Classification using a Multiclass SVM Classifier," 2019.

[17] H. Nisar, Y. K. Ch'ng, and Y. K. Ho, "Automatic Segmentation And Classification Of Eczema Skin Lesions Using Supervised Learning,"IEEE Conference on Open Systems (ICOS), Kota Kinabalu, Malaysia, pp. 25–30,2020.

[18] M.Kalaiyarivu,N.J.Nalini,"Skin Disease Classification Using Combined Machine Learning And Deep Learning Models," Journal of Pharmaceutical ,vol.13,pp.3248-3256,2022.

[19] S.A.Aldera,M.T.B.Othman,"A Model for Classification and Diagnosis of Skin Disease using Machine Learning and Image Processing Techniques," International Journal of Advanced Computer Science and Applications (IJACSA),Vol. 13, pp-252- 259,2022.

[20] Web source. https://www.kaggle.com/ Accessed on Dec., 2022.