iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VIII **Month of publication:** August 2022

DOI: https://doi.org/10.22214/ijraset.2022.46138

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VIII August 2022- Available at www.ijraset.com

Prediction of Skin Diseases Using Machine Learning

Mr. B. Suman[1], N. Harika[2], B. Sruthi[3], B. Bhagyasree[4]
[1]*Associate Professor,* [2, 3, 4]*Undergraduate Student, Department of Computer Science Engineering,Sridevi Women's Engineering College, Hyderabad, Telangana*

Abstract: ***Dermatological diseases are found to induce a serious impact on the health of millions of people as everyone is affected by almost all types of skin disorders every year. Since the human analysis of such diseases takes some time and effort, and current methods are only used to analyze singular types of skin diseases, there is a need for a more high-level computer-aided expertise in the analysis and diagnosis of multi-type skin diseases. This paper proposes an approach to use computer-aided techniques in Machine learning such as Ensemble Algorithm and Data Mining Algorithms to predict skin diseases real-time and thus provides more accuracy than other techniques.***

I. INTRODUCTION

A. *Purpose*

Skin diseases are the most common among them especially prone to spread and can prove to be fatal leading to skin cancer if not treated in its earlier stages. The occurrence of skin cancer is now increasing in numbers than the incidence of other new types of cancer of the lung, breast combined. The most common human malignancy is primarily diagnosed visually, beginning with an initial clinical screening and followed potentially by dermo copy analysis, a biopsy and histopathological examination. Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions

B. *Scope*

Since the human analysis of skin diseases takes some time and effort, and current methods are only used to analyze singular types of skin diseases, there is a needfor a more high-level computer-aided expertise in the analysis and diagnosis of multi- type skin diseases. By using the appropriate methods, the dataset is studied and then byapplying various techniques and algorithms the skin disease can be predicted. Comparison among algorithms helps to achieve the best one which provides high accuracy

C. *Model Diagram/Overview*

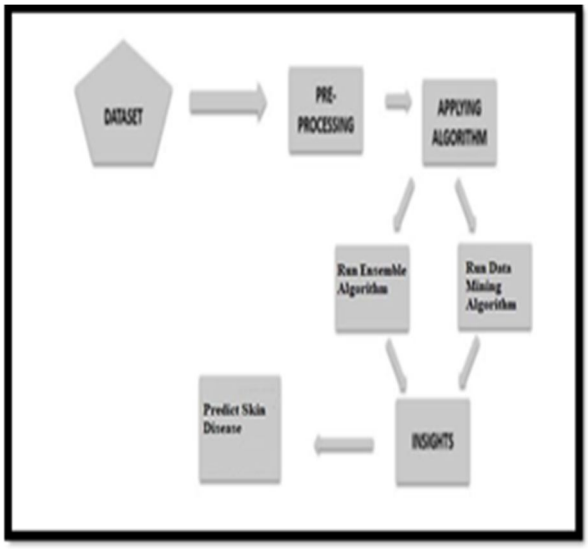


Fig. Model Diagram

©IJRASET: All Rights are Reserved

II. SYSTEM ANALYSIS

A. *Existing System*

In existing system, human analysis of such diseases takes some time and effort, and current methods are only used to analyze singular types of skin diseases.

DISADVANTAGES OF EXISTING SYSTEM

1) It takes some time and effort.
2) It is cost effective.

B. *Problem Statement*

Medical study demonstrates that different skin disease observation techniques are being used. However, there is still a great need to classify skin diseases at an early point. Machine learning algorithms have the potential to have an impact on early detection of skin diseases.

C. *Proposed System*

This paper proposes an approach to use computer-aided techniques in Machine learning such as Ensemble Algorithm and Data Mining Algorithms to predict skin diseases real-time and thus provides more accuracy than other techniques.

ADVANTAGES OF PROPOSED SYSTEM

1) It provides more accuracy than other technique

III. SYSTEM REQUIREMENT SPECIFICATION

A. *Functional Requirements*

1) Data Collection
2) Data Pre-processing
3) Training And Testing
4) Modelling & Predicting

B. *Non-Functional Requirements*

Non- functional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, thesite should load in 3 seconds when the number ofsimultaneous users are > 10000.

1) Usability requirement
2) Serviceability requirement
3) Manageability requirement
4) Recoverability requirement
5) Security requirement
6) Data Integrity requirement
7) Capacity requirement

C. *Hardware Requirements*

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user.

Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

1) Operating system: Windows 11, Linux
2) Processor: minimum intel I3
3) Ram: minimum 4 GB
4) Hard disk: minimum 250 GB

D. *Software Requirements*

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraintsand user documentaton.

The appropriation ofrequirements and implementation constraints gives the general overview of the project in regards to what the areas of strength and deficit are and how to tackle them.

1) Python Idel 3.7 version
2) Anaconda 3.7
3) Jupiter
4) Google colab

IV. SYSTEM DESIGN

A. System Architecture

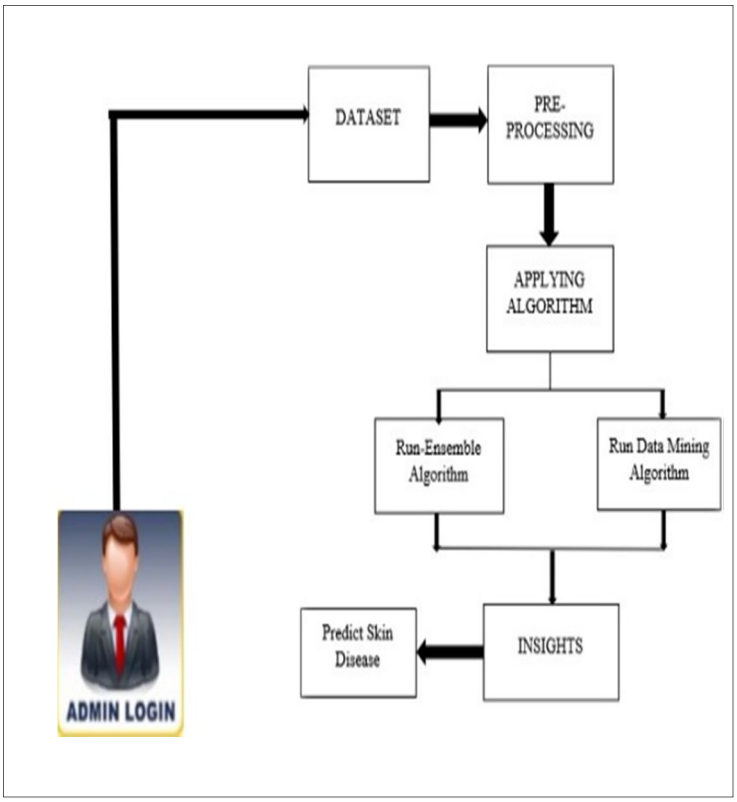


Fig System Architecture

B. System Components (Modules)

In this paper the author going to perform the following functions

1) Upload dataset: using this module we will upload dataset
2) Train & Test Split: Using this module we will split dataset into 80% trainset and 20% test set . We have used 80% trainset to train Ensemble and data mining algorithms and then apply 20% test data on trained model to calculate accuracy, precision, recall and FSCORE.
3) Run Ensemble Algorithm: Using this module we have train Bagging, Adaboost and Gradient boosting classification algorithms.
4) Run Data Mining Algorithms: using this module we have trained various data mining algorithms such as SVM, Random Forest and Decision Tree Algorithm.
5) Accuracy Comparison Graph: Using this module we are displaying accuracy, precision, recall and FMEASURE graph between all algorithms

C. Data Flow Diagram

The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.

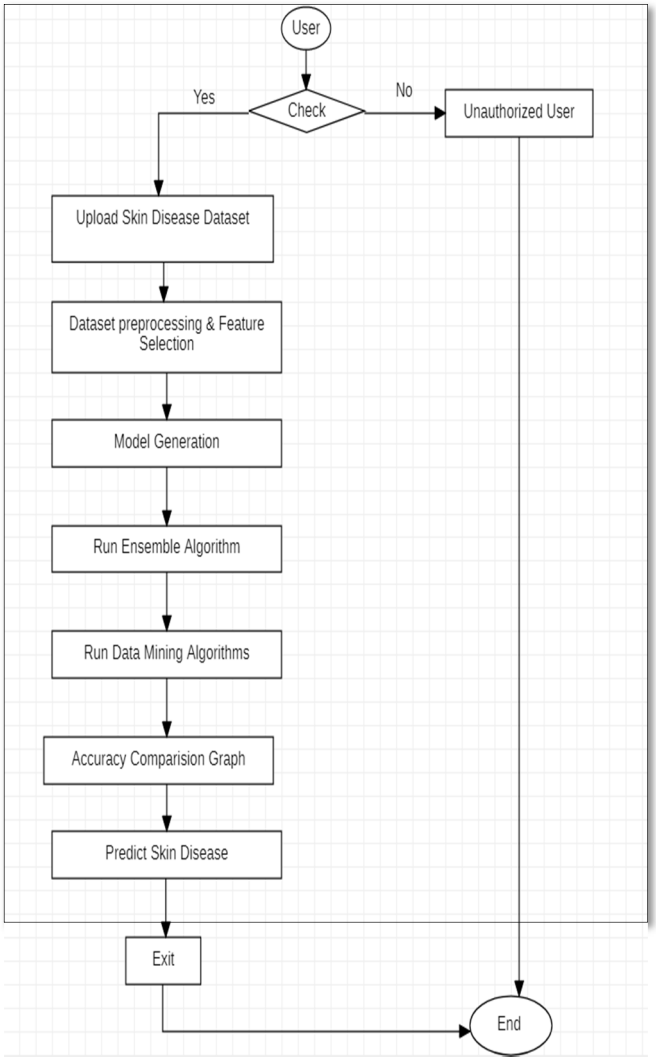
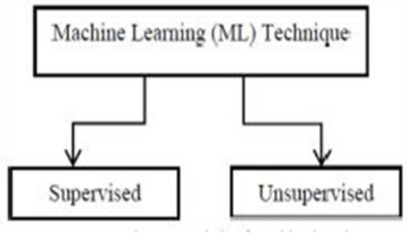


Fig Data Flow Diagram

V. ALGORITHMS

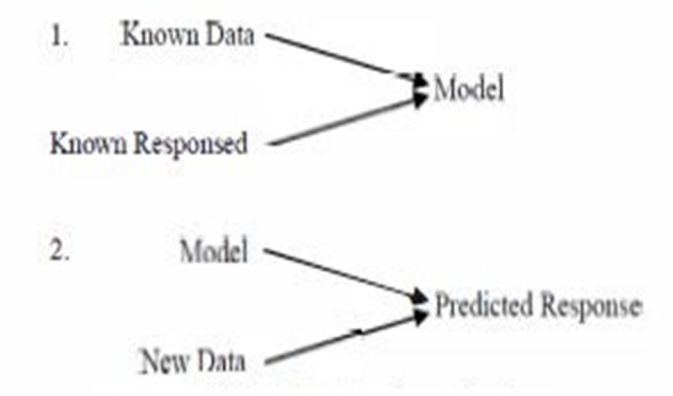
A. Algorithms Used In This Paper Include

Machine Learning (ML) Technique Machine learning (ML) technique is based on data set (Labeled Data Set). In this technique, a machine learning classifier is trained as input and then using the trained sample prediction, unknown classes are classified. There are two main areas in machine learning technique: the supervised and unsupervised learning technique.



SUPERVISED LEARNING TECHNIQUE Supervised learning technique is a machine learning technique. This technique is also called classification methods. It means that the supervised learning technique trains the model with some labeled data set and then it will produce prediction output in new data samples.

Unsupervised Technique Unsupervised technique is also called a cluster technique. In this method, there is no need of complete labeled data sets.



B. Ensemble Algorithm

Empirically, ensembles tend to yield better results when there is a significant diversity among the models. Many ensemble methods, therefore, seek to promote diversity among the models they combine Although perhaps non-intuitive, more random algorithms (like random decision trees) can be used to produce a stronger ensemble than very deliberate algorithms (like entropy-reducing decision trees). Using a variety of strong learning algorithms, however, has been shown to be more effective than using techniques that attempt to dumb-down the models in order to promote diversity. It is possible to increase diversity in the training stage of the model using correlation for regression tasks or using information measures such as cross entropy for classification tasks.

C. Data Mining Algorithm

In today's world where data plays a major role, it's important to gather insights from it. Data mining techniques pave the way for programmers to find out these insights. Python is the most popular programming language that offers the flexibility and power for programmers and data scientists to perform data analysis and apply machine learning algorithms. In recent years, Python has become more popular for data mining due to the rise in the number of data analysis libraries. This article will showcase how different data mining techniques work using Python. We'll pick the most commonly used Python libraries for data analysis such as Matplotlib, NumPy for our examples. Classification (a type of supervised learning) helps to identify to which set of categories an observation belongs based on the training data set that contains the observations. The most common Python library used for classification is Scikit-Learn. Let's take an example dataset to identify fruits. The "size", "color" and "shape" will be the features of the fruit, and the different class labels will be "apple", "orange", "watermelon". For this article, we will use the decision tree and KNN (k-nearest neighbours) classifier classification methods.

VI. CONCLUSION

The proposed system is able to detect the skin disease with promising results combining computer vision and machine learning techniques. It can be used to help people from all over the world and can be used in doing some productive work. The tools used are free to use and are available for the user, hence, the system can be deployed free of cost. The application developed is light-weight and can be used in machines with low system specifications. It has also a simple user interface for the convenience of the user. The machine learning algorithms were successfully implemented**.**

VII. FUTURE ENHANCEMENT

Research and execution of limited medical information are accessible. If more real-time data are available in the future, the detection of skin disease can be explored with recent advances in AI and the benefits of diagnosis assisted with AI.

REFERENCES

[1] Damilola A. Okuboyejo, Oludayo O. Olugbara, and Solomon A. Odunaike, "Automating Skin Disease Diagnosis Using Image Classification," Proceedings of the World Congress on Engineering and Computer Science 2013 Vol II, WCECS 2013, 23-25 October, 2013, San Francisco, USA.
[2] R. Yasir, M. A. Rahman and N. Ahmed, "Dermatological disease detection using image processing and artificial neural network," 8th International Conference on Electrical and Computer Engineering, Dhaka, 2014, pp. 687- 690, doi: 10.1109/ICECE.2014.7026918.
[3] Ambad, Pravin S., and A. S. Shirat, "A Image Analysis System to Detect Skin Diseases," IOSR Journal of VLSI and Signal Processing (IOSR-JVSP), Volume 6, Issue 5, Ver. I (Sep. - Oct. 2016), PP 17-25.
[4] R S Gound, Priyanka S Gadre, Jyoti B Gaikwad and Priyanka K Wagh, "Skin Disease Diagnosis System using Image Processing and Data Mining,"

International Journal of Computer Applications 179(16):38-40, January 2018.

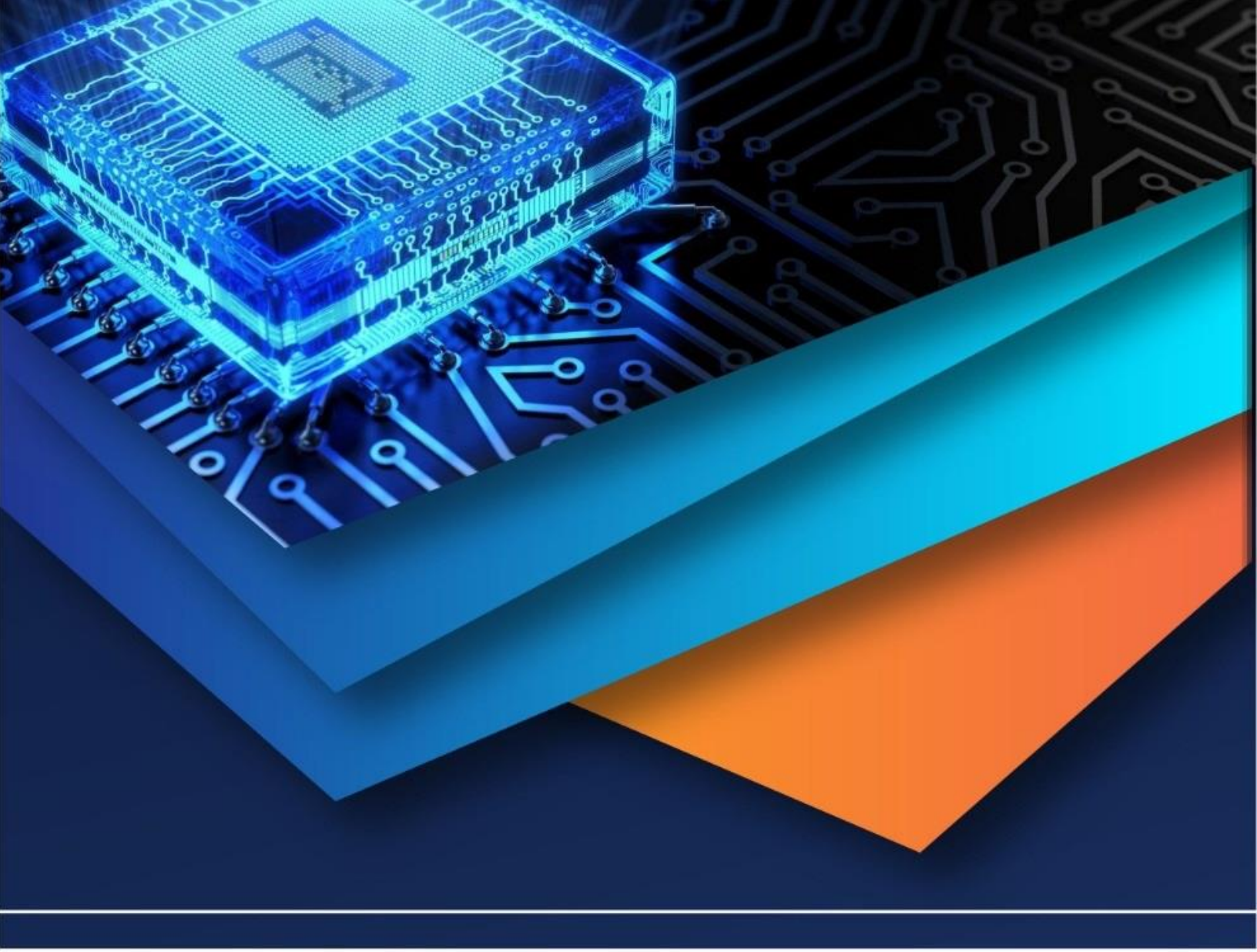
[5] V. B. Kumar, S. S. Kumar and V. Saboo, " Dermatological disease detection using image processing and machine learning," 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR), Lodz, 2016, pp. 1-6, doi: 10.1109/ICAIPR.2016.7585217.

[6] M. Shamsul Arifin, M. Golam Kibria, A. Firoze, M. Ashraful Amini and Hong Yan, "Dermatological disease diagnosis using color-skin images," 2012 International Conference on Machine Learning and Cybernetics, Xian, 2012, pp.1675-1680,doi: 10.1109/ICMLC.2012.6359626.

TEXTBOOKS:

- Programming Python, Mark Lutz
- Head First Python, Paul Barry
- Core Python Programming,R. Nageswara Rao
- Learning with Python, Allen BDowney WEBSITES:

1) https://www.w3schools.com/python/
2) .https://www.tutorialspoint.com/python/index.htm/
3) https://www.javatpoint.com/python-tutorial
4) https://www.geeksforgeeks.org/
5) https://www.python.org



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)