

SKIN DISEASE DETECTION USING K-NEAREST NEIGHBORS (KNN) ALGORITHM

Faculty1¹, Student1², Student2³, Student4⁴

¹Asst.Professor, ^{2, 3, 4}U.G.Student

Department of Computer Science and Engineering
Vel Tech Multi Tech Dr.Rangarajan Dr.Sakunthala Engineering College
Faculty1@veltechmultitech.org

Abstract- *Researches on identification of Skin diseases from the digital images are increasing due to multidimensional challenges of the domain. Skin diseases such as Melanoma and Carcinoma are often quite hard to detect at an early stage and it is even harder to classify them separately. Recently, it is well known that, the most dangerous form of skin cancer among the other types of skin cancer is melanoma because it is much more likely to spread to other parts of the body if not diagnosed and treated early. In order to classify these skin diseases, “Support Vector Machine (SVM)” and “Convolutional neural network” (CNN) a Machine Learning Algorithm can be used. In this paper, we propose a k-nearest neighbors (KNN) algorithm method to identify whether a given sample is affected with skin disease or not. This proposed model will predict most common skin disease like Chronic Dermatitis, Lichen Planus, Pityriasis Rosea, Psoriais and Seborrheic Dermatitis etc.,*

Keywords- k-nearest neighbors, Support vector machine Convolutional neural network, Feature extraction, Scikit-Learn, flask, scipy tool, Image capture, Image classifier.

1. INTRODUCTION

Skin diseases are one of those set of diseases whose number has been largely increasing day by day. Only in India, about 200 million people suffer from one or the other forms of skin diseases. People often neglect skin diseases and do not take necessary treatment. This is especially seen in rural and economically backward areas due to many factors such as lack of awareness, poverty and lack of resources etc. this is even higher when it comes to the case of Melanoma skin cancer. It is reportedly found that about 132,000 melanoma skin cancers occur globally each year. When the people tend to approach a physician, it is quite difficult for the physician in order to exactly detect the type of skin diseases the patient is getting affected with. Especially when it comes to the diseases like

Melanoma, Chronic Dermatitis, Lichen Planus, Pityriasis Rosea, Psoriasis and Seborrheic Dermatitis etc. However, k-nearest neighbors (KNN) algorithm is a better technique to classify than Neural Networks and SVM because they have a strong founding theory. KNN reach the global optimum due to quadratic programming, they have no issue for choosing a proper number of parameters.

2. LITERATURE SURVEY

Tanzina Afroz Rimi, Nishat Sultana and Md. Ferdouse Ahmed Foysal “Skin Diseases Detection Using Convolutional Neural Network (CNN)” Daffodil International University, Dhaka, Bangladesh, 2020

In above article author’s proposed, Skin is the most powerful protection of important organs in the human body. It acts as a shield to protect our internal body to get damaged. But this important part of the human body can be affected by so serious infections caused by some fungus or viruses or even dust too. Around the world, millions of people suffer from various skin diseases. so they came up to predicting process with CNN classifier.

LING-FANG LI, XU WANG, NEAL N. XIONG, YONG-XING DU AND BAO-SHAN LI “Deep Learning in Skin Disease Image Recognition: A Review” Inner Mongolia

University of Science and Technology, Baotou 014010, China, 2020

In above study, we review 45 research efforts on the identification of skin disease by using deep learning technology since 2016. Authors analyze these studies from the aspects of disease type, data set, data processing technology, data augmentation technology, model for skin disease image recognition, deep learning framework, evaluation indicators, and model performance. For Prediction used K-NN classifier.

Kumar, N. V., Kumar, P. V., Pramodh, K., & Karuna, Y. (2019). “Classification of Skin diseases using Image processing and SVM”. 2019.

In this study, recently, it is well known that, the most dangerous form of skin cancer among the other types of skin cancer is melanoma because it is much more likely to spread to other parts of the body if not diagnosed and treated early. In order to classify these skin diseases, “Support Vector Machine (SVM)” a Machine Learning Algorithm can be used.

In above most of the skin disease took-up the neural network and support vector machine algorithm. But, CNN makes predictions by looking at an image and then checking to see if certain components are present in that image or

not. If they are, then it classifies that image accordingly. If the certain components are not present, then it will ignore the image and doesn't classify it. Similarly at SVM, we have less accuracy prediction problems.

3. THE PROPOSED SYSTEM

In our proposal, we use the concept of Python - Flask and we are implementing with the help of Scikit-Learn & Pandas (**SciPy Toolkit**) in python. Prediction accuracy is high, robust working when training datasets contain errors. For separation of (disease affected skin and healthy skin), we use the concept of on K-NN classifier algorithm.

K-nearest neighbors (KNN) Algorithm: The k-nearest neighbors (KNN) algorithm is a simple, supervised (set of label training data) machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.

Scikit-learn: provide a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions,

encouraging academic and commercial use. Command to install **pip install scikit-learn**.

Pandas: pandas is a python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis/manipulation tool available in any language. It is already well on its way toward this goal. Command to install pandas in desk **pip install pandas**.

Python-Flask: Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common. Command to install pandas in desk **pip install flask**.

4. WORKING OF PROPOSED SYSTEM

K-nearest neighbors (KNN) algorithm uses ‘feature similarity’ to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps.

First step: For implementing any algorithm, we need dataset. During the first step of KNN, we must load the training as well as test data. Second step: Next, we need to choose the value of K i.e. the nearest data points. K can be any integer. For each point in the test data do the following, Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.

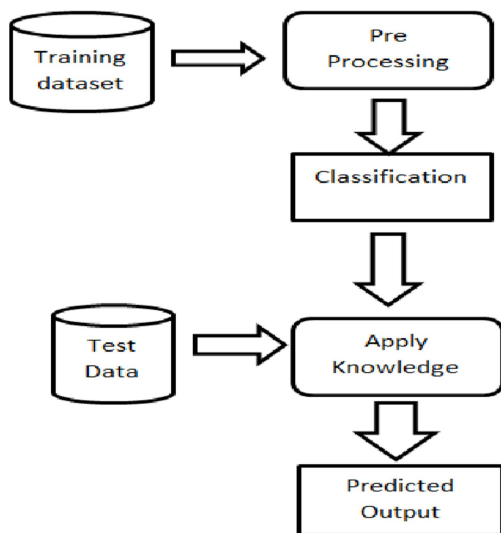


Fig 1. Block diagram of the proposed system

So KNN is an exception to general workflow for building/testing supervised machine learning

models. In particular, *the model* created via KNN is just the available labeled data, placed in some metric space. In other words, for KNN, there is no training step because there is no model to build. Template matching & interpolation is all that is going on in KNN.

Neither is there a validation step. Validation measures model accuracy against the training data as a function of iteration count (training progress). Over fitting is evidenced by the upward movement of this empirical curve and indicates the point at which training should cease. In other words, because no model is built, there is nothing to validate.

But you can still test--i.e., assess the quality of the predictions using data in which the targets (labels or scores) are concealed from the model.

But even testing is a little different for KNN versus other supervised machine learning techniques. In particular, for KNN, the quality of predictions is of course dependent upon amount of data, or more precisely the density (number of points per unit volume)--i.e., if you are going to predict unknown values by averaging the 2-3 points closest to it, then it helps if you have points close to the one you wish to predict. Therefore, *keep the size of the test set small*, or better yet use k-fold cross-validation or leave-one-out cross-validation, both of which give you more thorough model testing but not at the cost of reducing the size of your KNN neighbor population.

4.1 Advantages:

- It is very simple algorithm to understand and interpret.
- It is very useful for nonlinear data because there is no assumption about data in this algorithm.
- It is a versatile algorithm as we can use it for classification as well as regression.
- It has relatively high accuracy but there are much better supervised learning models than KNN.

4.2 KNN (K-nearest neighbors) Classification

K Nearest Neighbor (KNN) is a very simple, easy to understand, versatile and one of the topmost machine learning algorithms. KNN used in the variety of applications such as finance, healthcare, political science, handwriting detection, image recognition and video recognition. In Credit ratings, financial institutes will predict the credit rating of customers. In loan disbursement, banking institutes will predict whether the loan is safe or risky. In political science, classifying potential voters in two classes will vote or won't vote. KNN algorithm used for both classification and regression problems. KNN algorithm based on feature similarity approach.

KNN is a non-parametric and lazy learning algorithm. Non-parametric means there is no assumption for underlying data distribution. In other words, the model structure determined from the dataset. This will be very helpful in



Fig 4.2.1 Collection of diseased images

Practice where most of the real world datasets do not follow mathematical theoretical assumptions. Lazy algorithm means it does not need any training data points for model generation. All training data used in the testing phase. This makes training faster and testing phase slower and costlier. Costly testing phase means time and memory. In the worst case, KNN needs more time to scan all data points and scanning all data points will require more memory for storing training data.

K is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2. When $K=1$, then the algorithm is known as the nearest neighbor algorithm. This is the simplest case. Suppose P_1 is the point, for which label needs to predict. First, you find the one closest point to P_1 and then the label of the nearest point assigned to P_1 .

5. EXPERIMENTAL RESULT

5.1 LOGIN VALIDATION TO APP

Allowing users to log in to your app is one of the most common features you'll add to your web application. This article will cover how to add authentication to your Flask app with the Flask-Login package, we build some sign-up and login pages that allow users to log in and access protected pages that users who aren't logged in can't see. We'll grab information from the user model and display it on our protected pages when the user logs in to simulate what a profile would look like.

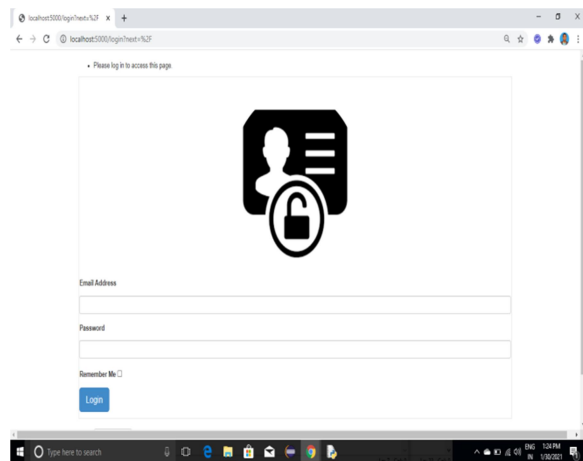


Fig 5.1 Login Validation

Our app will use the Flask app factory pattern with blueprints. We'll have one blueprint that handles everything auth related, and we'll have another for our regular routes, which include the index and the protected profile page.

5.2 CAPTURE IMAGE TO UPLOAD

A view function is the code you write to respond to requests to your application. Flask uses patterns to match the incoming request URL to the view that should handle it. The view returns data that Flask turns into an outgoing response.

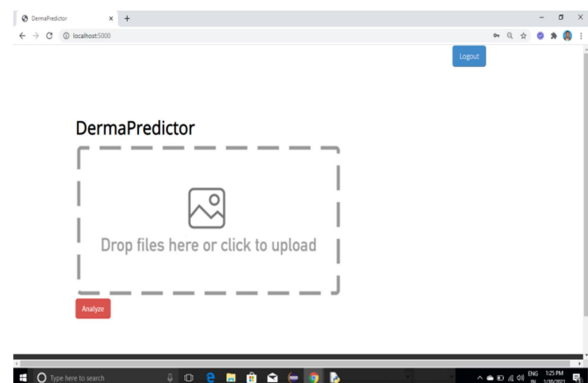


Fig 5.2 Capture image to upload

A while ago many developers had the idea to read the incoming file in small chunks and store the upload progress in the database to be able to poll the progress with JavaScript from the client. If the image is already capture than we can drag and drop into input derma predictor.

5.3 UPLOAD IMAGE TO PREDICT.

A common feature in web applications is to let users upload files to the server. The HTTP protocol documents the mechanism for a client to upload a file in RFC_1867, and our favorite web framework Flask fully supports it, but there

are many implementation details that fall outside of the formal specification that are unclear for many developers. Things such as where to store uploaded files, how to use them afterwards, or how to protect the server against malicious file uploads generate a lot of confusion and uncertainty.

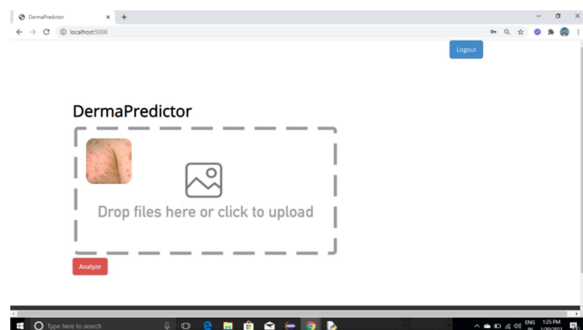


Fig 5.3 uploaded image to predict

From a high-level perspective, a client uploading a file is treated the same as any other form data submission. In other words, you have to define an HTML form with a file field in it.

5.4 PREDICTION OF SKIN DISEASE

The **KNN** algorithm uses 'feature similarity' to **predict** the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. The **output** is a class membership (predicts a class a discrete value). An object is classified by a majority vote of its neighbors, with the object being assigned to the

class most common among its k nearest neighbors.

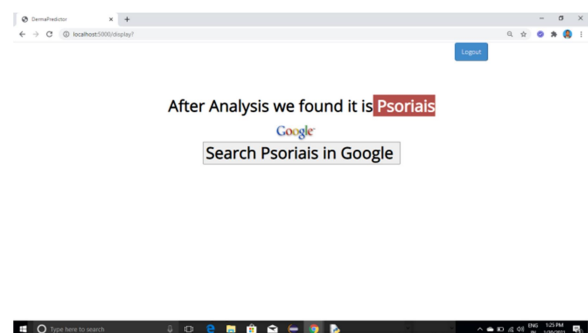


Fig 5.4 Prediction of skin disease

For determining skin disease, we need to analyze multiple medical parameters. We had collected few parameters as required. After analyzing input data from the users, it will determine the present and absence of skin disease.

CONCLUSION

We proposed an effective method with accurate value of wounded area based on region growth segmentation. KNN is very simple algorithm to understand and interpret. It is very useful for nonlinear data because there is no assumption about data in this algorithm.

It is a versatile algorithm as we can use it for classification as well as regression. It has relatively high accuracy but there are much better supervised learning models than KNN. Prediction accuracy is high and robust.

5. REFERENCES

[1] Tanzina Afroz Rimi, Nishat Sultana and Md. Ferdouse Ahmed Foysal “Skin Diseases Detection Using Convolutional Neural Network” Daffodil International University, Dhaka, Bangladesh, 2020

[2] LING-FANG LI, XU WANG, NEAL N. XIONG, YONG-XING DU AND BAO-SHAN LI “Deep Learning in Skin Disease Image Recognition: A Review” Inner Mongolia University of Science and Technology, Baotou 014010, China, 2020.

[3] Kumar, N. V., Kumar, P. V., Pramodh, K., & Karuna, Y. (2019). “Classification of Skin diseases using Image processing and SVM”. 2019

[4] Nibaran Da, Anabik Pal, Sanjoy Mazumder, Somenath Sarkar, Dwijendranath Gangopadhyay, Mita Nasipuri “An SVM based skin disease identification using Local Binary Patterns” 2013.