

Disease Prediction Website

Final Year Project

Supervised by,
Dr. Manish Gaur
Dr. Jasvant Kumar

By,
1805210050 (Shivansh Verma)
1805210065 (Yashika Tyagi)
1900520109005 (Shubhi Agarwal)

Motivation

- Diabetes and skin diseases are among the most neglected/poorly diagnosed diseases prevalent in the majority of the population of the globe around.
- Hence we have tried to built a website ,using which people showing probable symptoms can at least be made aware about their present health status and By uploading the skin images, they can at least get the idea of the skin diseases which may have the most probability of occurrence.

Features of Website

- User can login or Register through our site.
- Then after authentication from the backend user gets directed to the home page.
- It has two tabs-
 - Check skin-disease where user can upload the skin image and get the predicted result accordingly.
 - Another tab is for Check Diabetes where user just check mark some listed symptoms and get the predicted results.

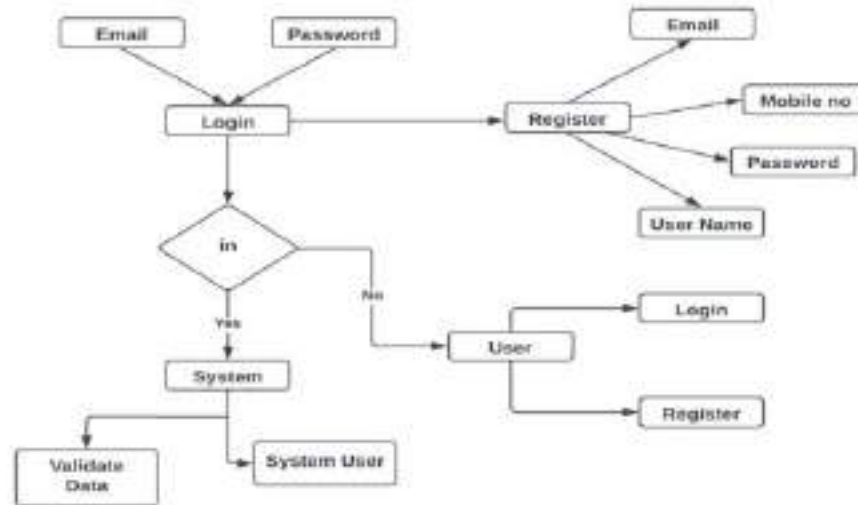
Symptoms taken in Diabetes

- 1.Polyupesia
- 2.Polyurea
- 3.Sudden weight loss
- 4.Visual blurring
- 5.Irritability
- 6.Partial Paresis
- 7.Weakness
- 8.Genital Thrush
- 9.Delayed hearing
- 10.Itching
- 11.Muscle Stiffness
- 12.Obesity
- 13.Gender
- 14.Aplopecia

Focused Skin Disease for Now

- 1.Acne
- 2.Vitiligo
- 3.Hairloss
- 4.Nail fungus
- 5.Poison Ivy and other contact Dermatitis
- 6.Melanome
- 7.Warts Molluscum
- 8.Clear Skin

Website ER Diagram

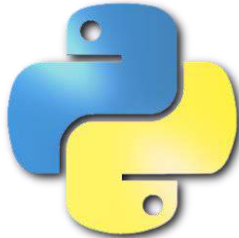


Technology Used



- ❑ **Flask**- Flask is a small and lightweight Python web framework that provides useful tools and features that make creating web applications in Python easier. It gives developers flexibility and is a more accessible framework for new developers since you can build a web application quickly using only a single Python file.

- ❑ **mySQL** - MySQL is a widely used relational database management system (RDBMS). MySQL is free and open-source.

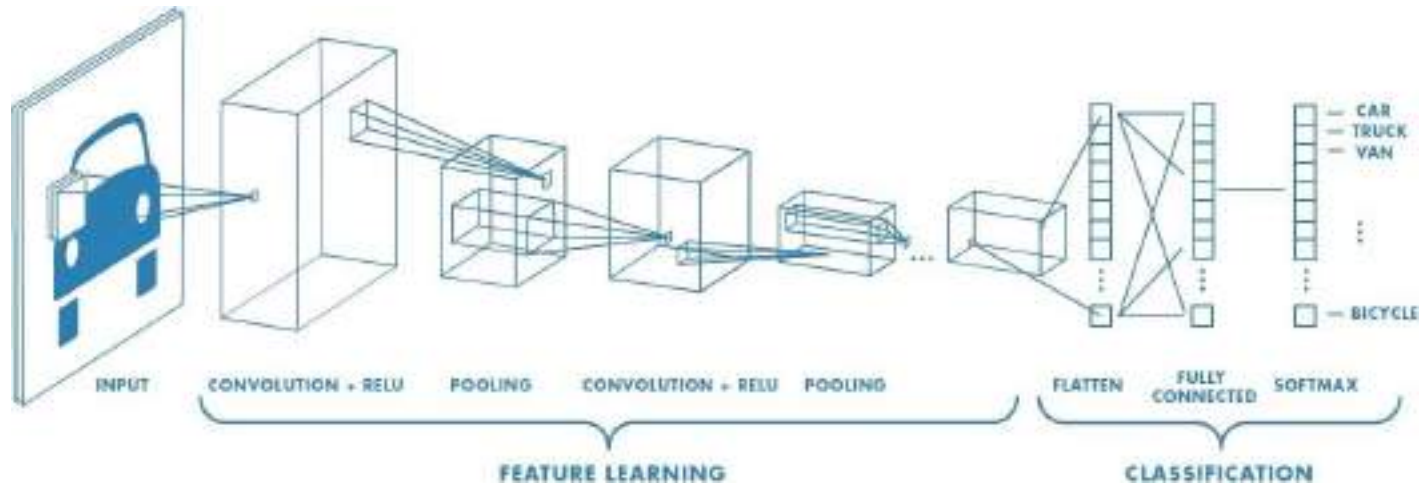


- ❑ **Python**- Python is a popular multi-purpose programming language widely used for its flexibility, as well as its extensive collection of libraries, which are valuable for analytics and complex calculations

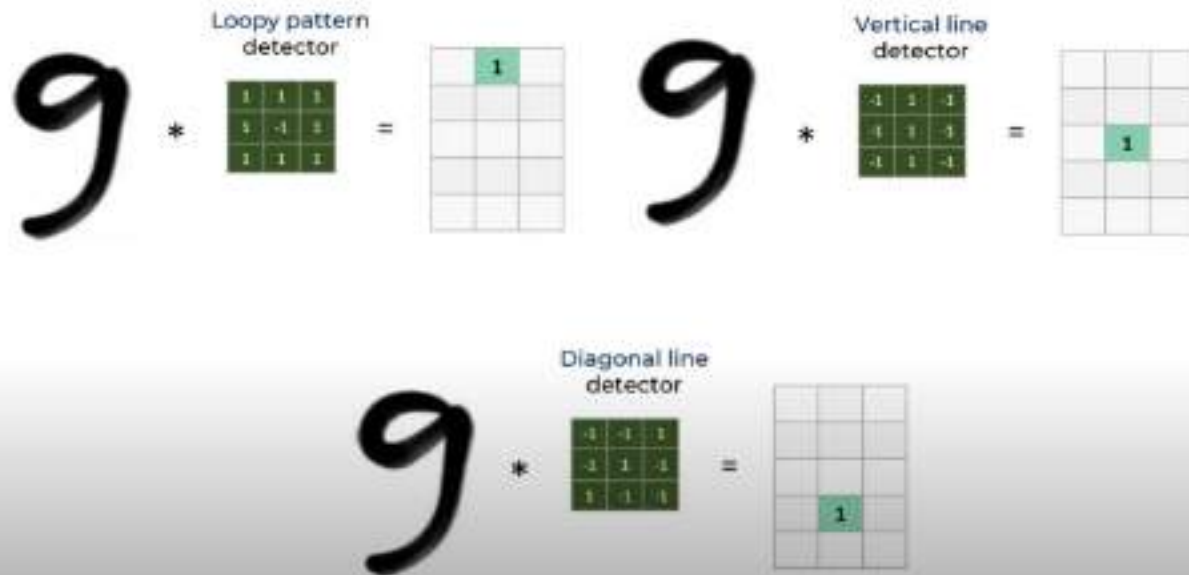
METHODOLOGY

❑ Skin diseases

CNN- In deep learning, a convolutional neural network (CNN/ConvNet) is a class of deep neural networks, most commonly applied to analyze visual imagery.

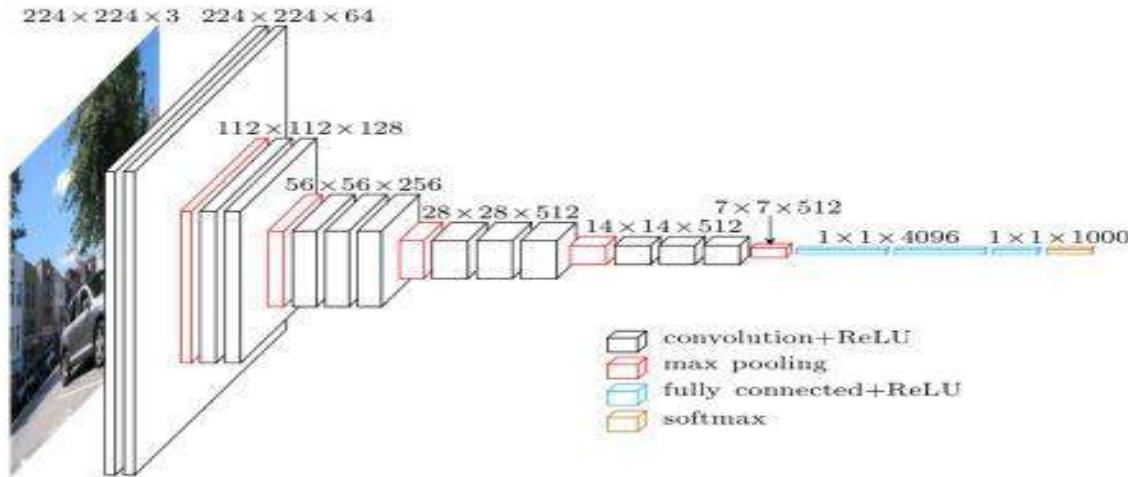


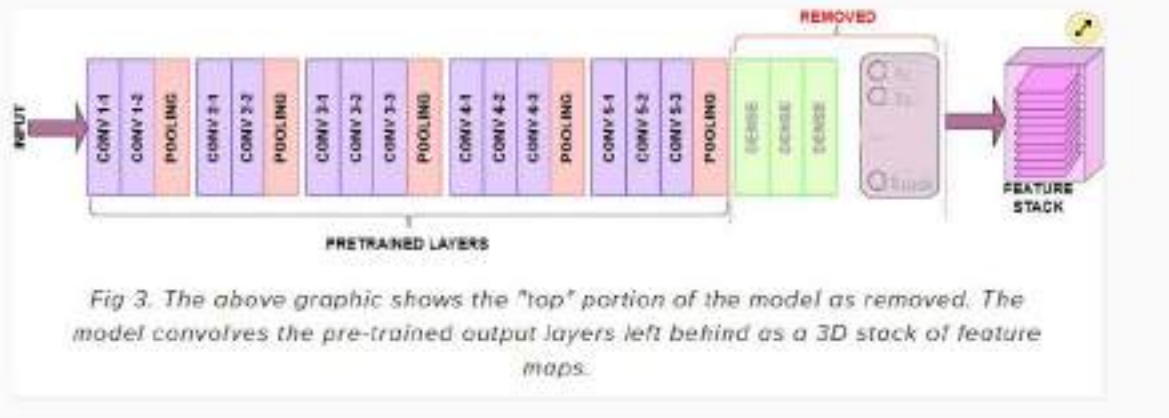
How does CNN work?



VGG16

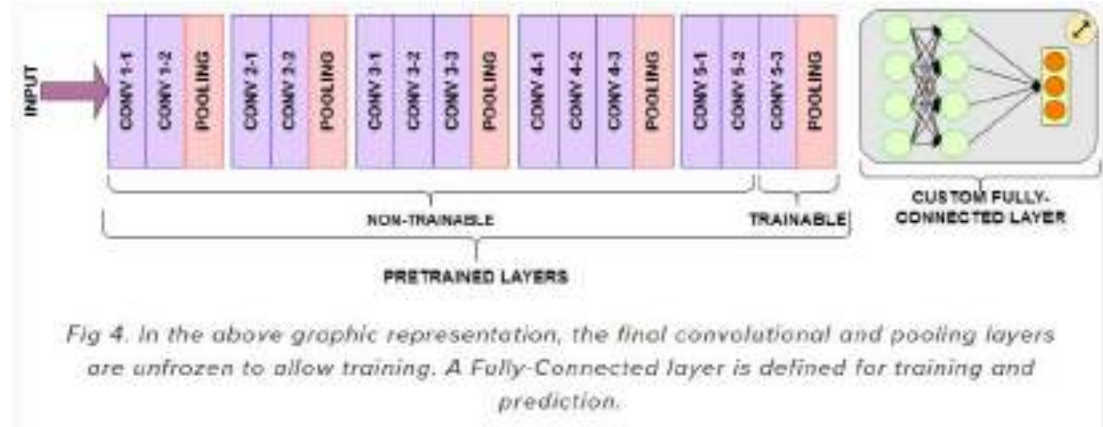
VGG16- VGG-16 is a convolutional neural network that is 16 layers deep. You can load a pre-trained version of the network trained on more than a million images from the ImageNet database.



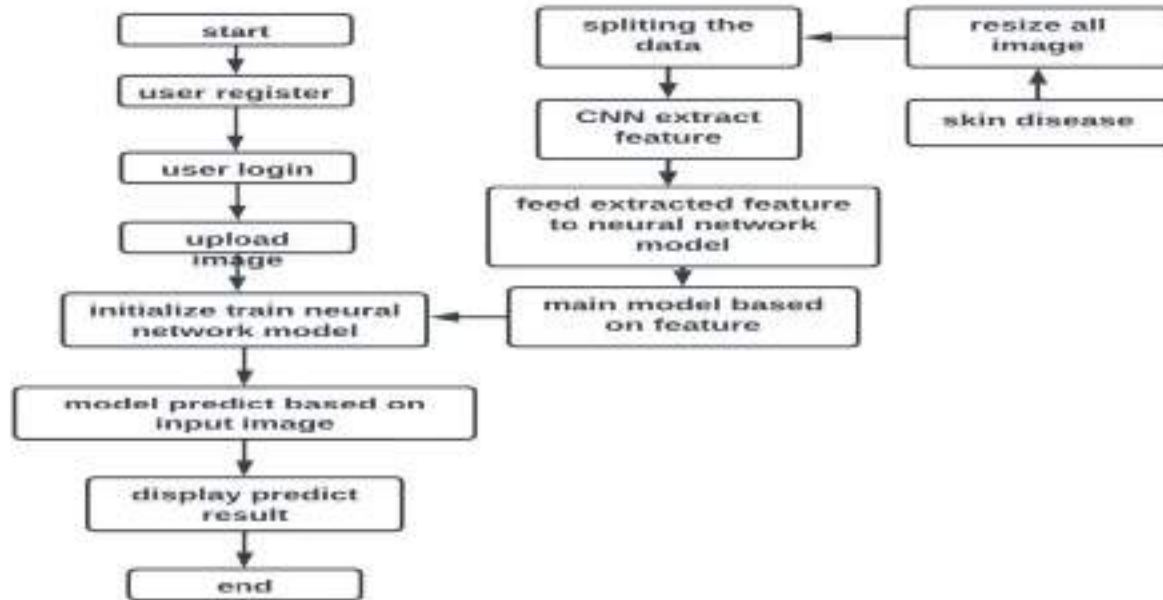


Removing top layers

Introducing custom layers

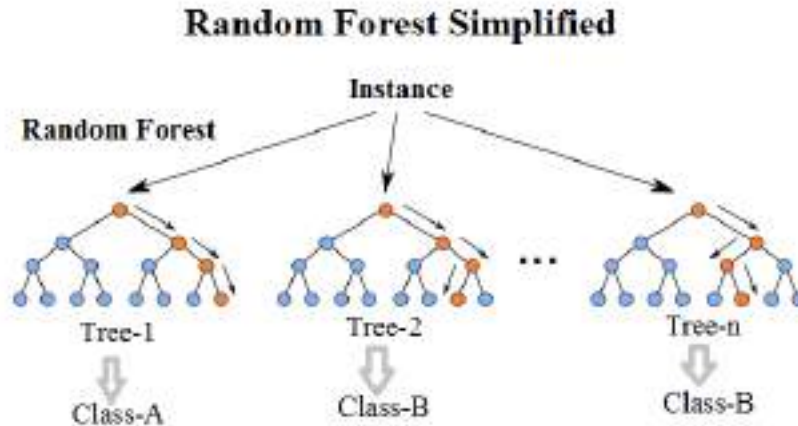


Skin Disease Flowchart



❑ Diabetes

Random Forest-Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning**, which is a process of *combining multiple classifiers to solve a complex problem and to improve the performance of the model*.



Dataset

[illegible]

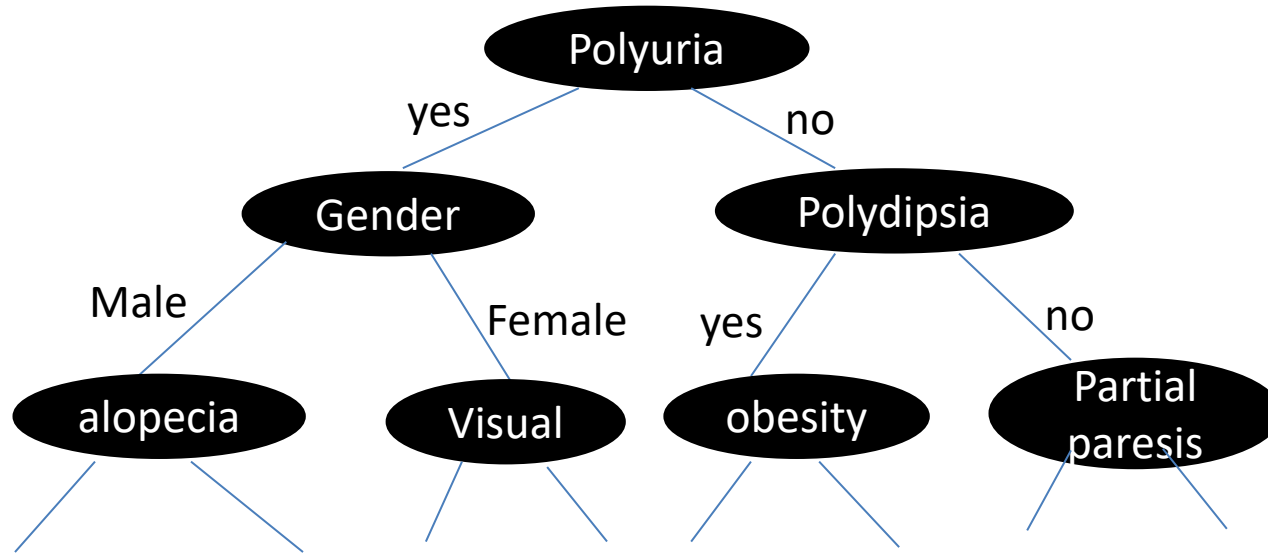
Calculate Information Gain for each Attribute

$$\begin{aligned}\text{Entropy}(\text{dataset}) &= -(7/18)\log(7/18) - (11/18)\log(11/18) = 0.29021 && S[+7, -11] \\ \text{Value}(\text{gender}) &= \text{Male, Female} \\ \text{Entropy}(\text{male}) &= -(4/15)\log(4/15) - (11/15)\log(11/15) = 0.251 && S(\text{male})[+4, -11] \\ \text{Entropy}(\text{female}) &= -(3/3)\log(3/3) - (0/3)\log(0/3) = 0 && S(\text{female})[+3, -0] \\ \text{Gain}(\text{gender}) &= \text{Entropy}(\text{dataset}) - \sum |S_v|/|S| * \text{Entropy}(S_v) = 0.081\end{aligned}$$

Similarly we can calculate Information gain of all attributes:

Gain(polyuria)=0.16	Gain(Delayed hearing)=0.01
Gain(polyudesia)=0.023	Gain(Itching)=0.036
Gain(sudden Weight loss)=0.09	Gain(Muscle Stiffness)=0.092
Gain(visual Blurring)= 0.12	Gain(Alopecia)=0.054
Gain(Irritability)=0.04	Gain(Obesity)=0.102
Gain(Partial Paresis)=0.202	
Gain(Weakness)=0.079	
Gain(Genital trush)=0.075	

Now the attribute with maximum gain is made root node. Here, Polyurea is made the root node.



Leave node will contain the classes to which the user belong.

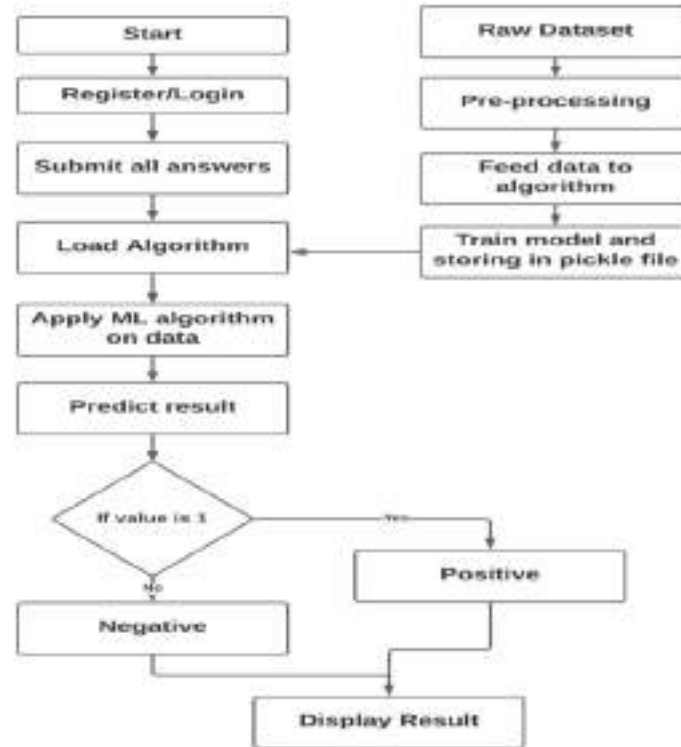
Diabetes

No
Diabetes

Diabetes

No
Diabetes

Diabetes Prediction Flowchart



Outcome of Skin disease prediction



Choose file

CHOOSE FILE

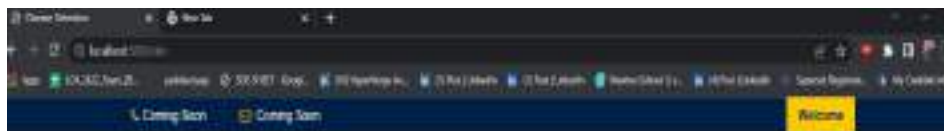
Submit



Choose file

CHOOSE FILE

Upload



Diabetes Detection

Home Check Diabetes Check Skin Diseases Logout


Choose file

CHOOSE FILE

Upload

Save_Skin_Detection_Image

Outcome of diabetes prediction

 **Diabetes Detection**

[Home](#) [Check Diabetes](#) [Check Skin Disease](#) [Logout](#)

Lets Predict:

What is your Age?

Do you have Polyuria?

Yes ☐

No ☒

Recently do you observe sudden weight loss?

Yes ☐

No ☒

Do you have Polyphagia?

Yes ☐

No ☒

Do you have Visual blurring?

Yes ☐

No ☒

Do you have Irritability?

Yes ☐

No ☒

Do you have Partial paresis?

Yes ☐

No ☒

Do you have Alopecia?

Yes ☐

No ☒

What is your Gender?

Male ☒

Female ☐

Do you have Polydipsia?

Yes ☐

No ☒

Do you feel any Weakness?

Yes ☐

No ☒

Do you have Genital thrush?

Yes ☐

No ☒

Do you have itching?

Yes ☐

No ☒

Do you have Delayed healing?

Yes ☐

No ☒

Do you have Muscle stiffness?

Yes ☐

No ☒

Do you have Obesity?

Yes ☐

No ☒

You don't have Diabetes. Please consult with Doctor for verification.

Conclusion

- The purpose of the project is to create early awareness about diseases and timely treatment of problems which might later prove to be fatal.
- Machine learning-based approach has been proposed for the forecast of diabetes. Machine learning classifiers have been utilized to characterize diabetes, with Random Forest (RF) and Naive bayes(NB).In our analysis, Random forest algorithm gives the accuracy up to 94% while Naive Bayes gave the accuracy up to 86%.
- Medical professionals and advanced instruments are needed to diagnose these skin diseases due to the lack of a feasible solution in the images of skin diseases. The proposed framework incorporates in-depth learning strategies such as CNN architecture and a predefined model VGG16.The accuracy is about 86%.

Limitations:

- Limited Dataset for prediction.
- Accuracy of the model decreased when number of diseases increased in skin disease prediction.
- Images in dataset may be noisy which makes our model to predict wrong answers.

Future Scope

- In future we aim to enhance the applicability of our website by including more number of diseases for example related to cancer, TB, and a lot of worlds neglected tropical diseases including meningitis, diphtheria etc.
- If accurate dataset is present then it is really easy to predict above mentioned diseases using neural networks and machine learning.
- An app could also be developed to increase the target audience.

References

1. <https://biomedpharmajournal.org/vol11no3/automated-skin-disease-identification-using-deep-learning-algorithm/>
2. G.H. Dayan, et al. Recent resurgence of mumps in the United States. New England Journal of Medicine, vol. 358, no. 15, pages 1580–1589, 2008. <https://doi.org/10.1056/NEJMoa0706589>
3. L. K. Tolentino, R. M. Aragon, W. R. Tibayan, A. Alvisor, P. G. Palisoc, and G. Terte. Diagnosis of Surrounding Fingernails Through Artificial Neural Network, Journal of Telecommunication, Electronic and Computer Engineering (JTEC), vol. 10, no. 1-4, pages 181-188, 2018.
4. <https://www.analyticsvidhya.com/blog/2022/01/diabetes-prediction-using-machine-learning/>

THANK YOU!