

YUNI- THE PCOS DETECTOR

Submitted in partial fulfillment of the requirements
for the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

Dikshita Basu (38110129)
Deepika R (38110122)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with Grade "A" by NAAC**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI - 600 119**

MAY – 2022



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600119

www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **DIKSHITA BASU (38110129)** and **DEEPIKA R (38110122)** who carried out the project entitled "**MR . MURARI DEVAKANNAN**" under my supervision from October 2021 to May 2022.

Internal Guide

MR.MURARI DEVAKANNAN.

Head of the Department

Dr. S. Vigneshwari M.E., Ph.D., and Dr. L. Lakshmanan M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **DIKSHITA BASU (Reg No:37110129)** and **DEEPIKA R (Reg No: 38110122)** hereby declare that the Project Report entitled "**YUNI- THE PCOS DETECTOR**" done by us under the guidance of **MR. MURARI DEVANKANNAN.** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in 2018-2022.

DATE:

PLACE:

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E.,Ph.D., Dean**, School of Computing **Dr.S.Vigneshwari M.E., Ph.D.** and **Dr.L.Lakshmanan M.E., Ph.D.** , Heads of the Department of Computer Science and Engineering for providing us necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and a deep sense of gratitude to my Project Guide **MR. MURARI DEVAKANNAN.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Polycystic Ovary Syndrome or PCOS is an endocrine disorder that occurs in women of reproductive age. The condition once detected cannot be cured but treatment can help relieve its effects.

The exact cause of PCOS is still unknown but there are certain factors that portray the risk of getting PCOS. The factors that result in this syndrome are obesity, insulin resistance, blood pressure, depression, inflammation. The symptoms include: hirsutism, Oligo-ovulation, acne, heavy bleeding, skin darkening. Using the causes and symptoms, a model is prepared in order to accept them as features and outputs the presence or absence of this condition. The machine learning models used for supervised classification are Logistic Regression. The reason behind building multiple models is to find out the best one for the given dataset, in the known scope of knowledge

- web- flask server
- access to all public
- refer the nearest gynecologist

TABLE OF CONTENTS

Chapter No.	TITLE	Page No.
	ABSTRACT	v
	LIST OF FIGURES	viii
	LIST OF ABBREVIATIONS	ix
1	INTRODUCTION	1
	1.1. OVERVIEW	2
	. MACHINE LEARNING	3
	MACHINE LEARNING STRATEGIES	3
	SUPERVISED LEARNING	3
	UNSUPERVISED LEARNING	4
	DEEP LEARNING	
2	LITERATURE SURVEY	6
	RELATED WORK	6
3	METHODOLOGY	8
	EXISTING SYSTEM	8
	PROPOSED SYSTEM	8
	OBJECTIVE	8
	SOFTWARE AND HARDWARE REQUIREMENTS	9
	SOFTWARE REQUIREMENTS	9
	HARDWARE REQUIREMENTS	9
	LIBRARIES	9
	PROGRAMMING LANGUAGES	10
	PYTHON	10
	HTML	
	CSS	
	3.5.4. DOMAIN	10
	SYSTEM ARCHITECTURE	11
	ALGORITHMS USED	11

	MULTI LAYER PERCEPTRON	11
	LOGISTIC REGRESSION	12
	MODULES	13
DATASET COLLECTION	13	
TRAIN AND TEST THE MODELS	13	
	DEPLOY THE MODELS	13
4.	RESULTS AND DISCUSSION	14
	WORKING	14
5.	CONCLUSION	15
	CONCLUSION	15
	REFERENCES	16
	APPENDICES	17
	A. SOURCE CODE	17
	B. SCREENSHOTS	20
	C. PLAGIARISM REPORT	24
	D. JOURNAL PAPER	25

LIST OF ABBREVIATIONS

ABBREVIATIONS	EXPANSION
ML	Machine Learning
DL	Deep Learning
MLP	Multi-Layer Perceptron
LG	Logistic Regression

CHAPTER 1

INTRODUCTION

PCOS (Polycystic ovary syndrome) is one of the most common hormonal disorders which most women suffer once in a lifetime during their reproductive age. This paper considers the issue when it affects more than 70% of the population of females, it is not that known among the people. Many times females ignore the symptoms and only recognize it when it gets very serious when one has to take drastic measures against it. PCOS is a hormonal disorder where women develop some follicles in their ovaries called cysts and these follicles cause ovaries to fail in releasing eggs. Women suffering from PCOS can release excessive androgen (male hormone) and can suffer from extended or infrequent menstrual periods .

This paper focuses on the female population to make them aware about this disorder so that they can take the symptoms into consideration as not everyone goes to screenings as this disorder can be diagnosed through it, considering the symptoms can be seen as a cause to go for screenings. Up until 2019, all that has been done regarding PCOS are always based on ultrasound images and image recognition techniques to diagnose the disorder using particle swarm optimization, automated detection using follicle recognition, segmentation using polycystic ovary using ultrasound images, data mining techniques or speed gradient method.

Out of all the methods one of the problems is they are considering the symptoms but focusing on the ultrasound images or recognizing follicles through any other screening method and then giving the result that determines whether a person is suffering from PCOS or not. The researcher is predicting the percentage of chances that PCOS can occur in a female only through the symptoms using back propagation algorithm in neural network and genetic algorithm as they are known to provide a good result. These two methods can also help us recognize the disorder in the early reproductive age of a female (9-15 years) to warn them about their future.

Polycystic Ovary Syndrome is a disease that occurs in women during their child-bearing years. The reproductive organs of women called ovaries that produce progesterone and estrogen-hormones that regulate the menstrual cycle, are affected.

Ovaries also produce little amounts of androgen also called as male hormones. The basic features of PCOS are:

- Cysts in ovaries.
- High levels of hormone: androgen.
- Irregular Periods
- Excessive body hair growth

Since the condition is a syndrome, it has a collection of symptoms that suggest its presence. These symptoms play a vital role in detection of this condition. Along with these symptoms, causes which may lead to the risks of the syndrome can also be considered. It is necessary to detect PCOS as early as possible since it holds the risk of infertility, diabetes, endometrial cancer and cardiovascular disease at a later stage of the condition. Here, a couple of machine learning models are built in order to determine the presence of PCOS. Since the dataset does classify if the condition is present or not, supervised machine learning algorithms are used called: K-Nearest Neighbor (K-NN) and Logistic Regression. The former is a distance based technique and the latter is probability based, this is why these two techniques, that are poles apart, are used and their accuracy is compared. Logistic Regression is more accurate with accuracy 92% while K-NN's is 90.74%. But we tried to get more accuracy than above two techniques.so we tried to use another algorithm which is Multi layer Perceptron from DL.which gave us the accuracy of 98%.

OVERVIEW

This report discusses the result of the work done in development of "YUNI-PCOS DETECTOR" using machine learning and python. It aims at the development of an application that assists females to check whether they are prone to pcos or not, and can get some food tips and gynecologist nearby .

There are also some remote villages which lack medical facilities. The dataset was processed in DL models Multi Layer Perceptron. While processing the data, symptoms are given as input and the disease was received as an output. This project helps to get the idea about the disease of an individual based on the symptoms he/she have, and get the treatment easily by contacting the concerned doctor.

1.2.MACHINE LEARNING

Machine learning could be a subfield of computer science (AI). The goal of machine learning typically is to know the structure of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process approaches.

In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular vary. thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

1.3 Machine Learning Strategies

In machine learning, tasks square measure typically classified into broad classes. These classes square measure supported however learning is received or however feedback on the educational is given to the system developed. Two of the foremost wide adopted machine learning strategies square measure supervised learning that trains algorithms supported example input and output information that's tagged by humans, and unattended learning that provides the algorithmic program with no tagged

information so as to permit it to search out structure at intervals its computer file.

Supervised Learning

In supervised learning, the pc is given example inputs that square measure labeled with their desired outputs. The aim of this technique is for the algorithmic program to be ready to “learn” by comparison its actual output with the “taught” outputs to search out errors, and modify the model consequently. Supervised learning thus uses patterns to predict label values on extra unlabelled information. For example, with supervised learning, an algorithm may be fed data with images of sharks labeled as fish and images of oceans labelled as water. By being trained on this data, the supervised learning algorithm should be able to later identify unlabelled shark images as fish and unlabelled ocean images as water.

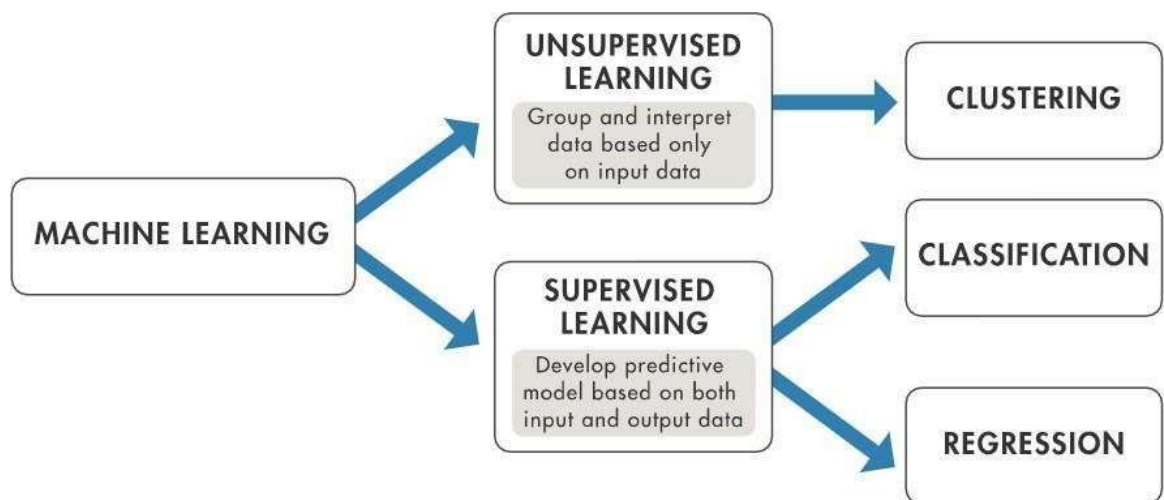
A common use case of supervised learning is to use historical information to predict statistically probable future events. It's going to use historical stock exchange info to anticipate approaching fluctuations, or be used to filter spam emails. In supervised learning, labeled photos of dogs are often used as input files to classify unlabeled photos of dogs.

Unattended Learning

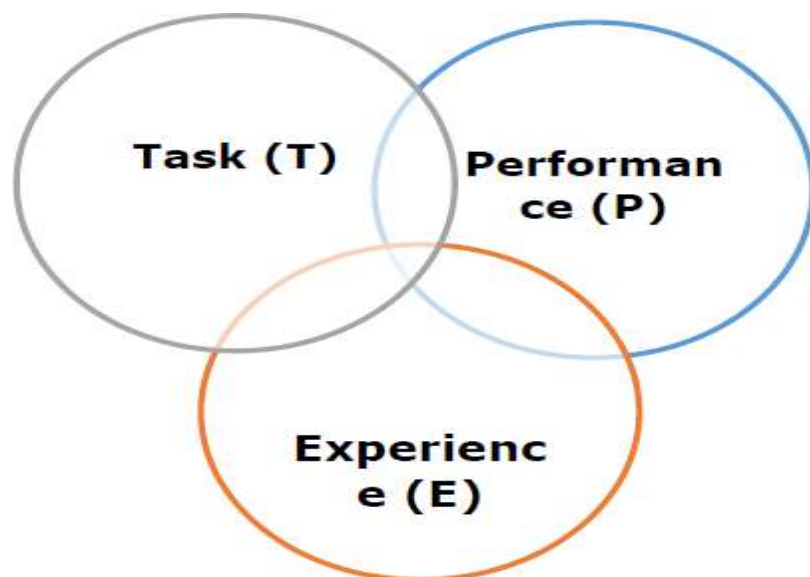
In unattended learning, information is unlabeled, that the learning rule is left to seek out commonalities among its input file. The goal of unattended learning is also as easy as discovering hidden patterns at intervals a dataset, however it should even have a goal of feature learning, that permits the procedure machine to mechanically discover the representations that square measure required to classify data.

Unsupervised learning is usually used for transactional information. You will have an oversized dataset of consumers and their purchases, however as a person's you'll probably not be able to add up of what similar attributes will be drawn from client profiles and their styles of purchases.

With this information fed into Associate in Nursing unattended learning rule, it should be determined that ladies of a definite age vary UN agency obtain unscented soaps square measure probably to be pregnant, and so a promoting campaign associated with physiological condition and baby will be merchandised.



MACHINE LEARNING CLASSIFICATION

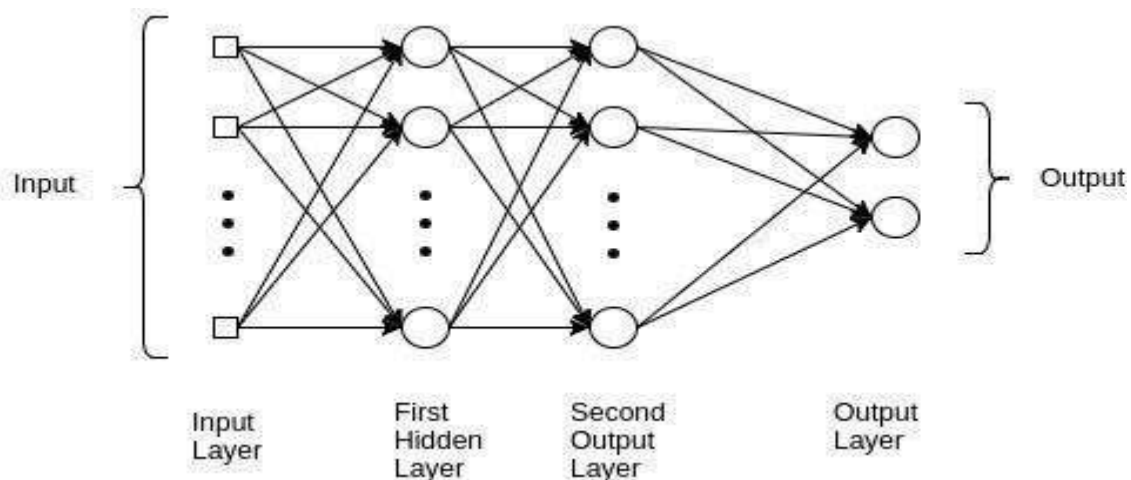


MACHINE LEARNING TASK

DEEP LEARNING

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans: learn by example. Deep learning is a key technology behind driverless cars, enabling them to recognize a stop sign, or to distinguish a pedestrian from a lamppost. It is the key to voice control in consumer devices like phones, tablets, TVs, and hands-free speakers. Deep learning is getting lots of attention lately and for good reason. It's achieving results that were not possible before.

In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. Deep learning models can achieve state-of-the-art accuracy, sometimes exceeding human-level performance. Models are trained by using a large set of labeled data and neural network architectures that contain many layers.



DEEP LEARNING

CHAPTER 2

LITERATURE SURVEY

2.1 RELATED WORK

Zhang,2018 ,Tried to predict and identify the risk of pcos through Hyperprolactinemia, Cushing's syndrome and non-classic congenital adrenal hyperplasia algorithm.

Norman,2007,paper determines the data analysis on the Risk of Obesity and Overweight in a Women study on the disorder and its three diagnostic criteria in depth giving us insights on not just pcos.

P. Mehrotra, 2011, used machine learning algorithms like Bayes and Logistic Regression .Performance of both the models was compared and the best classifier is used to predict the above diseases.

B.Cahyono Adiwijaya, M.s. Automated Detection of PCOS using follicle recognition(fr) has tried to recognize PCOS based on the FR through pre-processing to increase the quality of image.

SharvariS.Deshpande ,Asmita's implementations of Convolutional Neural Network on Pcos classifies patients into Pcos and non-Pcos class by using convolutional neural network algorithms.

Honglei Jing,Xiang Xu, Jinzhu Wang, provided a prospect that Pcos can be prevented at an early age .

Xiangyang Liu, Juan Bao, Yan, used various types of machine learning techniques. In that Rf algorithm is found superior.

Akinrotimi,Akinyemi Omololu, paper determined that patient suffering from obese are more likely to get hypertension and more.so by controlling obesity in patients makes them more responsive towards clomiphene.

Hector F. Escobar-Morreale, paper shows for cysts developed in the ovary and symptoms are, in which ultrasound images are one of the best methods which is used.

S.Sheela,M Sumathi, Used an Icv algorithm that can give us the right result in less computational time and iterations.

H Prasanna Kumar, S Srinivasan , Exploration On PCOS Explained various types of data mining techniques that can be useful in getting the right results.

CHAPTER 3

METHODOLOGY

EXISTING SYSTEM

From the above literature survey, We have inferred that all the systems existing predict either have less accuracy by implementing various algorithms on the particular datasets.

After implementing various algorithms, the most accurate one is selected and it is used for prediction of PCOS. Sometimes, we may get confused about what algorithm to use. Also, all the systems find only the particular Pcos and not the Pcos based on the symptoms.

PROPOSED SYSTEM

We are planning to pre-process the data more thoroughly to get more accurate results. Build a couple of machine learning model and improve the accuracy. We are planning to use Logistic Regression and try build a more accurate model. We also plan to do more extensive hyperparameter tuning of the ML algorithms for better performance. We plan to embed the model in a web app and make it more accessible to the general people.

OBJECTIVE OF PROJECT

Through this project we will ask basic information from users regarding their health conditions. The data will be run through the detector. The detector finds out if a female is prone to PCOS. If she requires medical care, she will be suggested home remedies and nearby gynecologist address.

The user can visit the doctors to get proper treatment places and get further information that the individual might be suffering from. This project helps to get the idea about an individual based on the symptoms she has, and get the treatment easily by contacting the concerned doctor.

SOFTWARE AND HARDWARE REQUIREMENTS

Software Requirements:

- ✓ PyCharm

Hardware Requirements:

- ✓ Processor: Intel Core i5
- ✓ RAM: 8GB
- ✓ OS: Windows

Libraries:

- ✓ **Pickle**- Pickling” is the process whereby a Python object hierarchy is converted into a byte stream, and “unpickling” is the inverse operation, whereby a byte stream (from a binary file or bytes-like object) is converted back into an object hierarchy.
- ✓ **Sklearn**- Scikit Learn also known as sklearn is a open source library for python programming used for implementing machine learning algorithms. It features various classification, clustering, regression machine learning algorithms. In this it is used for importing machine learning models, get accuracy, get confusion matrix.
- ✓ **Pandas**- Library of python which can be used easily. It gives speed results and also easily understandable. It is a library which can be used without any cost. We have used it for data analysis and to read the dataset.
- ✓ **Matplotlib**- Library of python used for visualising the data using graphs, scatterplots and so on. Here, we have used it for data visualisation.

- ✓ **Numpy-** Library of python used for arrays computation. It has so many functions. We have used this module to change a 2-dimensional array into a contiguous flattened array by using the `ravel` function.
- ✓ **MLP Classifier** -stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network. Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLPClassifier relies on an underlying Neural Network to perform the task of classification.

PROGRAMMING LANGUAGES

PYTHON

Python is the best programming language fitted to Machine Learning. In step with studies and surveys, Python is the fifth most significant language yet because it is the preferred language for machine learning and information science. It's owing to the subsequent strengths that Python has –

- ✓ **Easy to be told and perceive-** The syntax of Python is simpler; then it's comparatively straightforward, even for beginners conjointly, to be told and perceive the language.
- ✓ **Multi-purpose language** – Python could be a multi-purpose programming language as a result of it supports structured programming, object-oriented programming yet practical programming.
- ✓ **Support of the open supply community** – As being an open supply programming language, Python is supported by awfully giant developer community. Because of this, the bugs square measure simply mounted by the Python community. This characteristic makes Python terribly strong and adaptative.

HTML:

HTML (**HyperText Markup Language**) is the most basic building block of the Web. It defines the meaning and structure of web content. HTML uses "markup" to annotate text, images, and other content for display in a Web browser.

CSS:

Cascading Style Sheets is a style sheet language used for describing the presentation of a document written in a markup language such as HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

DOMAIN

Machine learning/ Deep Learning could be a subfield of computer science (AI). The goal of machine learning typically is to know the structure of information and match that data into models which will be understood and used by folks. Although machine learning could be a field inside technology, it differs from ancient process approaches. In ancient computing, algorithms are sets of expressly programmed directions employed by computers to calculate or downside solve. Machine learning algorithms instead give computers to coach on knowledge inputs and use applied math analysis so as to output values that fall inside a particular variable. Thanks to this, machine learning facilitates computers in building models from sample knowledge so as to modify decision-making processes supported knowledge inputs.

SYSTEM ARCHITECTURE

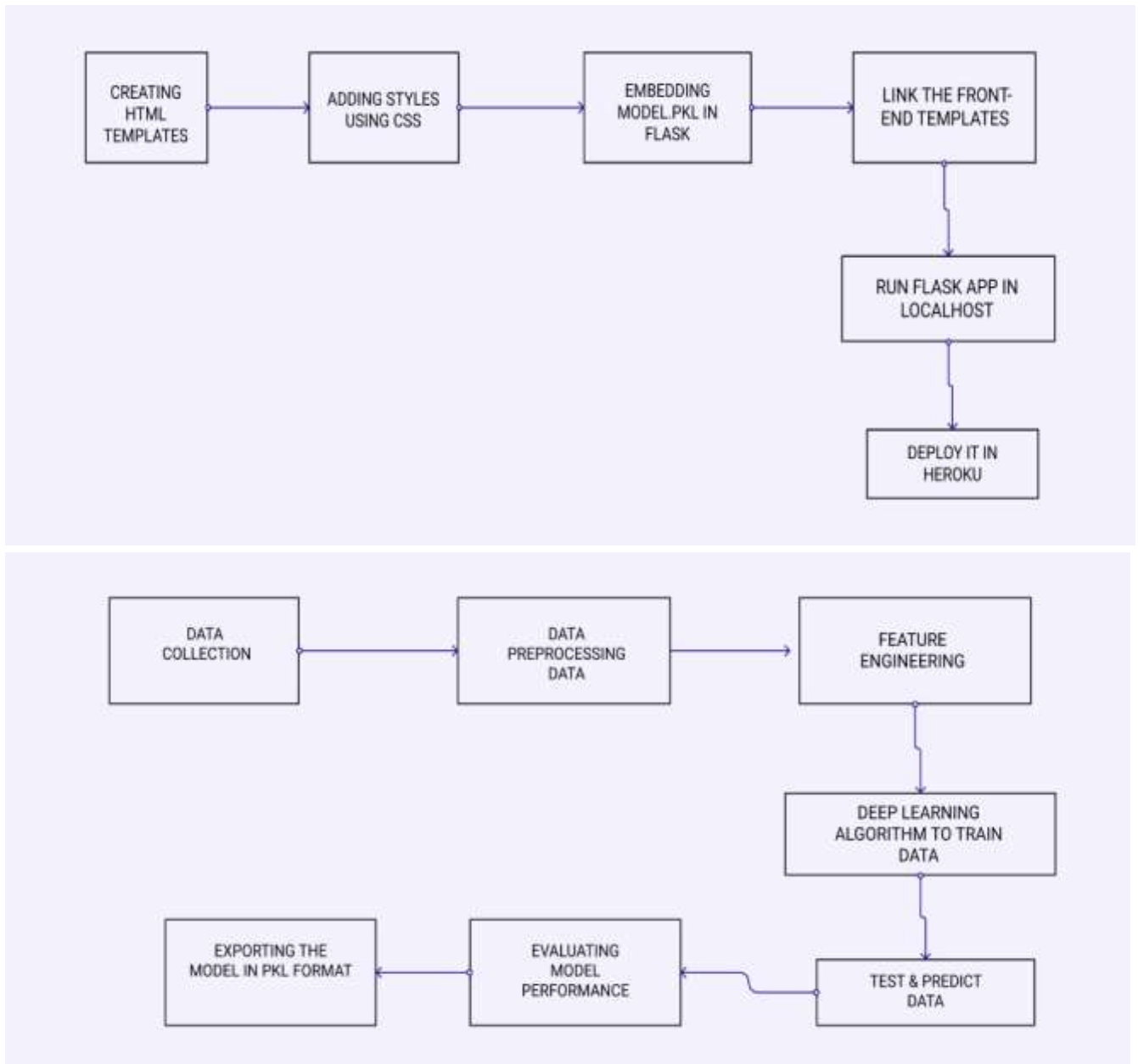


Fig 3.1 System Architecture

ALGORITHMS USED

MULTI-LAYER PERCEPTRON

Our model consists of three Multilayer Perceptron layers in a Dense layer. The first and second are identical, followed by a Rectified Linear Unit (ReLU) and Dropout activation function.

A multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network. MLP is a deep learning method.

MLP is widely used for solving problems that require supervised learning as well as research into computational neuroscience and parallel distributed processing. Applications include speech recognition, image recognition and machine translation.

LOGISTIC REGRESSION

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability. Logistic regression predicts the output of a categorical dependent variable.

Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used.

Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which

predicts two maximum values (0 or 1).The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

MODULES

Project contains three parts:

DATASET COLLECTION.

TRAIN AND TEST THE MODEL.

DEPLOY THE MODELS.

- ❑ **Dataset Collection-** We had collected dataset from kaggle notebooks. The dataset contains the symptoms and the corresponding disease. It contains 4920 rows and 133 columns.
- ❑ **Train and Test the model-** We had used the Multi-Layer Perceptron as a model to train the dataset. After training, we tested the model and found its accuracy.
- ❑ **Deploy the models-** Deployed Multi-Layer Perceptron by creating interface to get the name, symptoms of an individual. By this, we will get the disease and accuracy of the model as the output. We have also created a Food Tips bar using Html which helps an individual to get What food can be intake to have a proper diet and to avoid the risk of Pcos .

Following are the steps to do this project (use Pycharm):

1. Collect the dataset.
2. Import the necessary libraries.
3. Visualize the dataset.
4. Train the dataset using MLP and LG.

5. Test the model and find the accuracies of both.
6. Deploy the model- a) Multi-Layer Perceptron
7. Predict the Pcos based on the symptoms given by the user.

CHAPTER 4

RESULTS AND DISCUSSION

Doctors and medical professionals are always required in case of an emergency. In the current situation of COVID-19, where essential resources are unavailable and people are also not willing to go outside in fear of spreading virus, our prediction system will be very helpful for finding the If they have pcas or not based on the symptoms in the early stage and get the correct diagnosis.This also helps in reduction of the cost and give the correct and fast result.

WORKING

In this way Deep learning when implemented in healthcare can help in satisfying the individual and also take care of them easily. Multi-Layer Perceptron, which is the most easy model, helps to get the idea about whether an individual has Pcas with symptoms showing she has, and get the treatment easily by contacting the concerned doctor. We have also given Food tips to lead a healthy life.

Algorithm	Accuracy
Multi-Layer Perceptron	98%

Table 4.1 Accuracies

CHAPTER 5

5.1. CONCLUSION

Polycystic Ovary Syndrome (PCOS) is a medical condition which causes hormonal disorder in women in their childbearing years.

The hormonal imbalance leads to a delayed or even absent menstrual cycle. Women with PCOS majorly suffer from excessive weight gain, facial hair growth, acne, hair loss, skin darkening and irregular periods leading to infertility in rare cases.

Our proposal seems to perform better than the respective others. Hence, Multi-layer Perceptron is used for our final hypothesis which will predict the results using test data with accuracy of 98%.

The project presented the technique of Detecting Whether they have Pcos or not based on the symptoms of an individual patient. Almost all the ML models gave good accuracy values but the most accurate one is selected in DL and the filled form given by it is considered as the symptom of an individual. Once the Pcos is Detected, we could easily guide them to a nearby hospital to get a full checkup . If they don't have it, they can get food tips to manage a healthy life .

Doctors and medical professionals are always required in case of an emergency. In the current situation of COVID-19, where sufficient facilities and resources are unavailable, our Detection system can prove to be helpful and can be used in the diagnosis Whether they have Pcos or not.

REFERENCES

1. Mehrotra, P., Chatterjee, J., Chakraborty, C., Ghoshdastidar, B., & Ghoshdastidar, S. (2011). Automated screening of Polycystic Ovary Syndrome using machine learning.
2. Revised 2003 consensus on diagnostic criteria and long term health risks related to polycystic ovary syndrome (PCOS) . Human Reproduction, Volume 19, Issue 1, January 2004, Pages 41-47.
3. Supervised Learning-Classification Using K-Nearest Neighbors (KNN). (2019). Python® Machine Learning, 205–220. doi:10.1002/9781119557500.ch9
4. Bichler, Martin and Kiss, Christine, "A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management" (2004). AMCIS 2004 Proceedings. 230.
5. Murat KORKMAZ,Selami GÜNEY,Şule Yüksel YİĞİTERTHE, "Importance of Logistic Regression Implementation In The Turkish LiveStock Sector And Logistic Regression Implementations.
6. B.Cahyono,Adiwijaya,M.S. Mubarak,U.N. Wisesty "An Implementation of Convolutional Neural Network on PCOS.
7. Sharvari S.Deshpande,Asmita Wakankar, "Automated Detection of PolyCystic Ovarian Syndrome Using Follicle Recognition",IEEE International Conference on Advanced Communication Control and Computing Technologies.
8. Neural Network",<https://www.investopedia.com/terms/n/neuralnetwork.asp>,Oct 2019
9. Jing Li, Ji-hang Cheng, Jing Yuan Shi, Fei Huang, "Brief Introduction of Back Propagation(BP) Neural Network Algorithm and Its Improvement" Springer-Verlag Berlin Heidelberg ,2012,553-554 .
10. Back propagation algorithm", [www.towardsdatascience.com/](http://www.towardsdatascience.com/understanding-backpropagation-algorithm-7bb3aa2f95fd) understanding – backpropagation -algorithm-7bb3aa2f95fd.
11. Xiangyang Liu, Juan Bao, Yan Jiang, Zhunping Ke, Chengbo Wang "College Graduates Employment Prediction Based on Principal Component Analysis and the Combined Adaptive Boosting and the Back Propagation Neural Network Algorithm" 11th International Conference on Natural Computation(ICNC), 2015 .

12. D.Whitley,T.Stakeweather and C.Bogart "Genetic Algorithm and Neural Networks:optimizing connections and connectivity"Elsevier Science Publishers B.V(North Holland),1990:351-356.
- 13.Honglei Jing,Xiang Xu,Jinzhu Wang "Research on Genetic Neural Network Algorithm and its Application" International Conference on Virtual Reality and Intelligent Systems,2018:223-226 .
- 14.Akinrotimi,AkinyemiOmololu and Oladele,Rufus Olalere "Predictive Analysis in Health Data Using Back Propagation Neural .
15. S.Sheela,M.Sumathi "Study and Theoretical Investigation on PCOS",IEEE International Conference on Computational Intelligence and Computing Research,2014
16. Robert L.Rosenfield "Clinical Review:Identifying Children at Risk for Polycystic Ovary Syndrome " The Journal of Clinical Endocrinology & Metabolism,2007,787-78.

APPENDICES

A) SOURCE CODE

MODEL.PY

```
y = np.array(y)
X= np.array(X)
y = y.reshape(-1,1)
sns.heatmap(df.corr(), annot=True, cmap='Blues',linewidth=0.2)
fig = plt.gcf()
fig.set_size_inches(10,8)
plt.title("correlation matrix")
plt.show()

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33,
random_state=20)
from sklearn.neural_network import MLPClassifier
mlp = MLPClassifier(hidden_layer_sizes=(5,8,5), activation='relu', solver='adam',
max_iter=1000)
mlp.fit(X_train,y_train.ravel())

predict_train = mlp.predict(X_train)
predict_test = mlp.predict(X_test)
pickle.dump(mlp, open('model.pkl','wb'))
model = pickle.load(open('model.pkl','rb'))
from sklearn.metrics import classification_report,confusion_matrix
acc = accuracy_score(y_test, predict_test)
print(confusion_matrix(y_test,predict_test))
print("MLP Classifier Accuarcy : ", accuracy_score(y_test, predict_test))
print(classification_report(y_test,predict_test))
# plot_confusion_matrix of knn
array = [[605,44],
         [8,476]]
```

```

df_confusion_matrix = pd.DataFrame(array, index = [i for i in "01"],
                                     columns = [i for i in "01"])
plt.figure(figsize = (2,2))
sns.heatmap(df_confusion_matrix, annot=True)
plt.title('Confusion matrix of MLP ', y=1.1)
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
plt.show()

```

```

from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 40) # n_neighbors means k
knn.fit(X_train, y_train)
prediction = knn.predict(X_test)
acc2 = knn.score(X_test, y_test)
print("{} KNN Score: {:.2f}%".format(2, knn.score(X_test, y_test)*100))
print(classification_report(y_test,prediction ))
print("Confusion Matrix :\n" ,confusion_matrix(prediction,y_test))

```

```

# plot_confusion_matrix of knn
array = [[576,86],
         [73,398]]
df_confusion_matrix = pd.DataFrame(array, index = [i for i in "01"],
                                     columns = [i for i in "01"])
plt.figure(figsize = (2,2))
sns.heatmap(df_confusion_matrix, annot=True)
plt.title('Confusion matrix of KNN', y=1.1)
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
plt.show()

```

```

from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 20)
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)
acc1 = accuracy_score(y_test, y_pred)
print(confusion_matrix(y_test,y_pred ))
print(classification_report(y_test,y_pred ))
print ("3. Logistic regression Accuracy : ", accuracy_score(y_test, y_pred))
# plot_confusion_matrix of knn
array = [[611,38],
         [10,474]]
df_confusion_matrix = pd.DataFrame(array, index = [i for i in "01"],
                                   columns = [i for i in "01"])
plt.figure(figsize = (2,2))
sns.heatmap(df_confusion_matrix, annot=True)
plt.title('Confusion matrix of Logistic regression', y=1.1)
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
plt.show()

```

```

from lightgbm import LGBMClassifier
model = LGBMClassifier(max_bin= 2, learning_rate=2.5, num_leaves=2)
model.fit(X_train, y_train)
preds=model.predict(X_test)

```

```

from sklearn.metrics import accuracy_score
print(confusion_matrix(y_test,preds ))
print(classification_report(y_test,preds))
acc3 = accuracy_score(y_test, preds)
print ("4. LGBM Accuracy : ", accuracy_score(y_test, preds))
# plot_confusion_matrix of knn

```

```

array = [[618,31],
         [16,468]]
df_confusion_matrix = pd.DataFrame(array, index = [i for i in "01"],
                                   columns = [i for i in "01"])
plt.figure(figsize = (2,2))
sns.heatmap(df_confusion_matrix, annot=True)
plt.title('Confusion matrix oF LGBM', y=1.1)
plt.ylabel('Actual label')
plt.xlabel('Predicted label')
plt.show()

```

```

labels = ["mlp", "LR ", "KNN", "LGBM" ]
usages = [acc, acc1, acc2, acc3]

```

```

y_positions = range(len(labels))
plt.bar(y_positions, usages)
plt.xticks(y_positions, labels)
plt.ylabel("Accuracy")
plt.title("model selection")
plt.show()

```

INDEX.HTML

```

<!DOCTYPE html>

<html lang="en">
<head>
  <title>PCOS predictor</title>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <link href="https://fonts.googleapis.com/css?family=Montserrat"
rel="stylesheet">

```

```
<script
src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
<script
src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
<style>
body {
  font: 20px Montserrat,serif;
  line-height: 1.8;
  color: #bc1abc ;
}
p {font-size: 16px;}
.margin {margin-bottom: 25px;}
.bg-1 {
  background-color: #ed93da; /* Green */
  color: #ffffff;
}
.bg-2 {
  background-color: #474e5d; /* Dark Blue */
  color: #ffffff;
}
.bg-3 {
  background-color: #ffffff; /* White */
  color: #555555;
}
.bg-4 {
  background-color: #ffffff; /* Black Gray */
  color: #fff;
}
.container-fluid {
  padding-top: 70px;
  padding-bottom: 70px;
}
```

```

.navbar {
  padding-top: 15px;
  padding-bottom: 15px;
  border: 0;
  border-radius: 0;
  margin-bottom: 0;
  font-size: 12px;
  letter-spacing: 5px;
}
.navbar-nav li a:hover {
  color: #bc1abc !important;
}
#container-v {
  overflow-y: hidden;
  width:100% !important;
}
.child_con {
  height: 500px;
  float: left;
  margin: 1em;
}
</style>
</head>
<body>

<!-- Navbar -->
<nav class="navbar navbar-default">
  <div class="container">
    <div class="navbar-header">
      <button type="button" class="navbar-toggle" data-toggle="collapse"
data-target="#myNavbar">
        <span class="icon-bar"></span>

```

```

        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
    </button>
    <a class="navbar-brand" href="#">Home</a>
</div>
<div class="collapse navbar-collapse" id="myNavbar">
    <ul class="nav navbar-nav navbar-right">
<!--#####
#####-->
        <!--CHANGED '#' to '/'-->
        <li><a href="/about">ABOUT PCOS</a></li>
<!--#####
#####-->
        <li><a href="#details">FIND ME</a></li>
        <li><a href="/practice">Prediction</a></li>
    </ul>
</div>
</div>
</nav>

<!-- First Container -->
<div id="about" class="container-fluid bg-1 text-center">
    <h3 class="margin">PCOS PREDICTOR</h3>
    
    <h3>PCOS is not a disease, it's a lifelong condition whose repercussions can be
mitigated through certain healthy changes in lifestyle, early detection and seeking
help from a professional. Aim of this project is generating awareness about PCOS.
</h3>
</br>
</div>

```

```

<!-- Third Container (Grid) -->
<div id="details" class="container-fluid bg-3 text-center">
  <h3 class="margin">CSEO62<br>DIKSHITA & DEEPIKA</h3><br>
  <div class="row">
    <div class="container-fluid bg-3 text-center" style="padding:0% 5% 0% 4%">
      <p style="text-align:center">COMPUTER SCIENCE AND ENGINEERING</p>
      <p style="text-align:center">GUIDE: Mr.MURARI DEVAKANNAN </p>
      <a href="XXX@gmail.com">Email Me</a>
    </div>
  </div>
</div>
</div>
</body>
</html>

```

ABOUT.HTML

```

<!DOCTYPE html>
<html lang="en">
<head>
  <title>PCOS PREDICTOR</title>
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <link rel="stylesheet"
    href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
  <link href="https://fonts.googleapis.com/css?family=Montserrat" rel="stylesheet">
  <script
    src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
  <script
    src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></scri
    pt>
  <style>
  body {

```

```
font: 20px Montserrat, sans-serif;
line-height: 1.8;
color: #bc1abc ;
}
p {font-size: 16px;}
.margin {margin-bottom: 25px;}
.bg-1 {
  background-color: #ed93da; /* Green */
  color: #ffffff;
}
.bg-2 {
  background-color: #474e5d; /* Dark Blue */
  color: #ffffff;
}
.bg-3 {
  background-color: #ffffff; /* White */
  color: #555555;
}
.bg-4 {
  background-color: #ffffff; /* Black Gray */
  color: #fff;
}
.container-fluid {
  padding-top: 70px;
  padding-bottom: 70px;
  text-align: left;
}
.navbar {
  padding-top: 15px;
  padding-bottom: 15px;
  border: 0;
  border-radius: 0;
```

```

margin-bottom: 0;
font-size: 12px;
letter-spacing: 5px;
}
.navbar-nav li a:hover {
color: #bc1abc !important;
}
#container-v {
overflow-y: hidden;
width:100% !important;
}
.child_con {
height: 500px;
float: left;
margin: 1em;
}
</style>
</head>
<body>

```

```

<!-- Navbar -->

```

```

<nav class="navbar navbar-default">
  <div class="container">
    <div class="navbar-header">
      <button type="button" class="navbar-toggle" data-toggle="collapse"
data-target="#myNavbar">
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
      </button>
      <a class="navbar-brand" href="/">Home</a>
    
```

```
</div>
<div class="collapse navbar-collapse" id="myNavbar">
  <ul class="nav navbar-nav navbar-right">
    <li><a href="/about">ABOUT PCOS</a></li>
    <li><a href="/practice">PREDICTION</a></li>
  </ul>
</div>
</div>
</nav>
```

```
<!-- First Container -->
```

```
<div id="about" class="container-fluid bg-1 text-center">
  <h1 class="margin">Polycystic ovary syndrome (PCOS)</h1>
  
  <h3 style="color:purple">
    <b> WHAT IS PCOS?</b>
    <p><h3 style="color:black"> PCOS is a common health problem caused by an
    imbalance of reproductive hormones. Women with PCOS produce
    higher-than-normal amounts of male hormones.
    The hormonal imbalance creates problems in the ovaries. The ovaries make the
    egg that is released each month as part of a healthy menstrual cycle.
    With PCOS, the egg may not develop as it should or it may not be released
    during ovulation as it should be, causing their body to skip menstrual periods
    and making harder for them to get pregnant. The ovaries may develop
    numerous small collections of fluid (follicles) and fail to regularly release eggs.
  </h3>
  </p>
  </br>
  <h3 style="color:purple"><b> SYMPTOMS PCOS</b></h3>

  <p><h3 style="color:black">
```

Some of the symptoms of PCOS include:</br>

</br>

Irregular menstrual cycle: One may miss periods or have fewer periods or periods may come every 21 days or more often. Some may even stop menstruating</br>

</br>

Too much hair:Excess hair on the face, chin, or parts of the body where men usually have hair. This is called "hirsutism." Hirsutism affects up to 70% of women with PCOS.

</br>

Acne: Pimples/acne on the face, chest, and upper back</br>

</br>

Thinning hair Excess hair loss on the scalp or male-pattern baldness</br>

</br>

Weight gain: Gaining excess weight difficulty losing weight</br>

</br>

Darkening of skin: Particularly along neck creases, in the groin, and underneath breasts</br>

</br>

Skin tags: Small excess flaps of skin in the armpits or neck area</br>

</br>

</h3>

</p>

<h3 style="color:purple"> TREATMENT</h3>

<h3 style="color:black">

Maintain a healthy weight</br>

</br>

Limit carbohydrates</br>

```
</br>
Be active</br>
</br>
Maintain healthy diet with high-fiber vegetables, lean protein foods and
anti-inflammatory foods and spices
</b>
</h3>
</h3>
</br>
</div>
```

PRACTICE.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
<title>PCOS predection</title>
<meta charset="utf-8">
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="stylesheet"
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
<link href="https://fonts.googleapis.com/css?family=Montserrat" rel="stylesheet">
<script
src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
<script
src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></scri
pt>
<style>
body {
font: 20px Montserrat, sans-serif;
line-height: 1.8;
color: #bc1abc ;
}
```

```
p {font-size: 16px;}
.margin {margin-bottom: 25px;}
.bg-1 {
  background-color: #ed93da; /* Green */
  color: #ffffff;
}
.bg-2 {
  background-color: #474e5d; /* Dark Blue */
  color: #ffffff;
}
.bg-3 {
  background-color: #ffffff; /* White */
  color: #555555;
}
.bg-4 {
  background-color: #2f2f2f; /* Black Gray */
  color: #fff;
}
.container-fluid {
  padding-top: 70px;
  padding-bottom: 70px;
}
.navbar {
  padding-top: 15px;
  padding-bottom: 15px;
  border: 0;
  border-radius: 0;
  margin-bottom: 0;
  font-size: 12px;
  letter-spacing: 5px;
}
.navbar-nav li a:hover {
```

```
    color: #bc1abc !important;
}
#container-v {
    overflow-y: hidden;
    width:100% !important;
}
.child_con {
    height: 500px;
    float: left;
    margin: 1em;
}
input[type=text] {

padding: 12px 20px;
margin: 8px 0;
box-sizing: border-box;
border: none;
background-color: white;
color: black;
}
.block {
display: block;
width: 10%;
border: none;
background-color: #04AA6D;
color: white;
padding: 14px 28px;
font-size: 16px;
cursor: pointer;
text-align: center;
}
```

```
label
{
width: 530px;
display: inline-block;
text-align: left;
}
```

```
</style>
</head>
<body>
```

```
<!-- Navbar -->
<nav class="navbar navbar-default">
  <div class="container">
    <div class="navbar-header">
      <button type="button" class="navbar-toggle" data-toggle="collapse"
data-target="#myNavbar">
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
        <span class="icon-bar"></span>
      </button>
      <a class="navbar-brand" href="/">Home</a>
    </div>
    <div class="collapse navbar-collapse" id="myNavbar">
      <ul class="nav navbar-nav navbar-right">
        <li><a href="/about">ABOUT PCOS</a></li>
        <li><a href="/practice">PREDICTION</a></li>
      </ul>
    </div>
  </div>
```

</nav>

<!-- First Container -->

<div class="login" style = "height: 850px;overflow-y: scroll;" >

<div id="practice" class="container-fluid bg-1 text-center">

<h1 class="margin">Polycystic ovary syndrome (PCOS)</h1>

<!--#####
#####-->

<h1>PCOS Predictor</h1>

<!-- Main Input For Receiving Query to our ML -->

<form action="{{ url_for('predict')}}"method="post">

<!--#####
#####-->

<!--REGEX FOR ERROR HANDLING-->

<label for="fname">How old are you?</label>

<input type="text" id="fname" name="Age (yrs)" placeholder="Age (yrs)"
pattern="^(0?[1-9]|[1-9][0-9]|[1][1-9][1-9]200)\$" title="Please enter a valid number
between 0 and 200." required="required" />

<label for="name">What is your BMI?</label>

<input type="text" id="name" name="BMI" placeholder="BMI"
pattern="^(0?[1-9]|[1-9][0-9]|[1][1-9][1-9]200)\$" title="Please enter a valid number

between 0 and 200." required="required" />

<!--#####
#####-->

<label for="cycle">Your menstrual cycle can say a lot about your health. Is your Menstrual cycle regular ?</label>

<input type="text" id="cycle" name="Cycle (R/I)" placeholder="Cycle (regular/irregular)" required="required" />

<label for="gain">We know this could be depressing but have you recently gained weight?</label>

<input type="text" id="gain" name="Weight Gain (Y/N)" placeholder="Weight Gain (yes/no)" required="required" />

<label for="hair">Do you get a lot of facial/chest hair?</label>

<input type="text" id="hair" name="Hair Growth (Y/N)" placeholder="Hair Growth (yes/no)" required="required" />

<label for="skin">Have you notice your skin darkened in the recent months?</label>

<input type="text" id="skin" name="Skin Darkening (Y/N)" placeholder="Skin Darkening (yes/no)" required="required" />

<label for="hair">Have you notice any Hair Loss or hair damage in recent days?</label>

<input type="text" id="hair" name="Hair Loss (Y/N)" placeholder="Hair Loss (yes/no)" required="required" />

<label for="pi">We know that pimples always pop up at the wrong times but is that frequent with you?</label>

<input type="text" id="pi" name="Pimples (Y/N)" placeholder="Pimples

(yes/no)" required="required" />

<label for="fa">Burger king could be awfully delicious at times but do you
consume fast-food frequently?</label>

<input type="text" id="fa" name="Fast food (Y/N)" placeholder="Fast food
(yes/no)" required="required" />

<label for="re">Do you do Regular Exercise?</label>

<input type="text" id="re" name="Regular Exercise (Y/N)" placeholder="Regular
Exercise (yes/no)" required="required" />

<button type="submit" class="block">Predict</button>

</form>

<p> </p>

{{ prediction_text }}

</div>

<div class="col-lg-6 pt-4 pt-lg-0">

<p>

Visit the Nearby Doctor and get yourself checked sweetie!!!!

</p>

<a

href="https://www.google.com/maps/search/gynecologist+near+me/@13.093350
1,80.2657557,15z/data=!3m1!4b1" class="btn-learn-more">Click Here

<p>

What is food to eat for polycystic ovarian syndrome?

</p>

<a href="https://www.mariongluckclinic.com/blog/foods-to-eat-with-pcos.html"

class="btn-learn-more">Click Here

<p>

YOGA for polycystic ovarian syndrome?

</p>

Click Here

<p>

Approaches and strategies for polycystic ovarian syndrome?

</p>

Click Here

<p>

What is the main cause of polycystic ovarian syndrome?

</p>

Click Here

<p>

polycystic ovarian syndrome Q&A?

</p>

<a href="https://www.cedars-sinai.org/blog/polycystic-ovary-syndrome.html"

class="btn-learn-more">Click Here

<p>

Myths about polycystic ovarian syndrome?

</p>

```
<a
href="https://www.pennmedicine.org/updates/blogs/fertility-blog/2020/march/five-
myths-about-pcos" class="btn-learn-more">Click Here</a>
</div>
</div>
```

APP.PY

```
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle

app = Flask(__name__)
model = pickle.load(open('model.pkl', 'rb'))

@app.route('/')
def home():
    return render_template('index.html')

#####
#####

@app.route('/about')
def about():
    #CHANGES MADE ONLY HERE
    return render_template('about.html')

#####
#####

@app.route("/practice")
def practice():
    """ return the rendered template """
    return render_template("practice.html")

#####
```

```

#####

def check_user_input(input):
    try:
        # Convert it into integer
        val = int(input)

    except ValueError:
        try:
            #BLOCK CAN BE REMOVED
            # Convert it into float
            val = float(input)

        except ValueError:
            val=str(input)
    return val

#####

#####

@app.route('/predict',methods=['POST'])
def predict():

    l=[]

    for x in request.form.values() :
        z=x
        if isinstance(check_user_input(x),int)== True:
            #REMOVE CHECK_USER_INPUT()
            x=z
        elif isinstance(check_user_input(x),str)== True:
            #REMOVE CHECK_USER_INPUT()
            if x == 'yes':
                x=1
            elif x=='no':
                x=0

```

```

        elif x=='regular':
            x=2
        elif x == 'irregular':
            x=4
#####
#####
        else:
            return render_template('practice.html', prediction_text='Kindly enter valid
input ')
        else:
            #BLOCK CAN
            BE REMOVED
            return render_template('practice.html', prediction_text='Kindly enter valid
input ')
#####
#####
        l.append(int(x))

final_features = [np.array(l)]
prediction = model.predict(final_features)

output = round(prediction[0], 2)
if output == 1:
    output ='prone'

elif output == 0:
    output =' not quite prone'

return render_template('practice.html', prediction_text='After hearing you out, we
feel that you are {} to PCOS. Although, we strongly encourage you to adapt to
healthy lifestyle and stick with it!'.format(output))

```

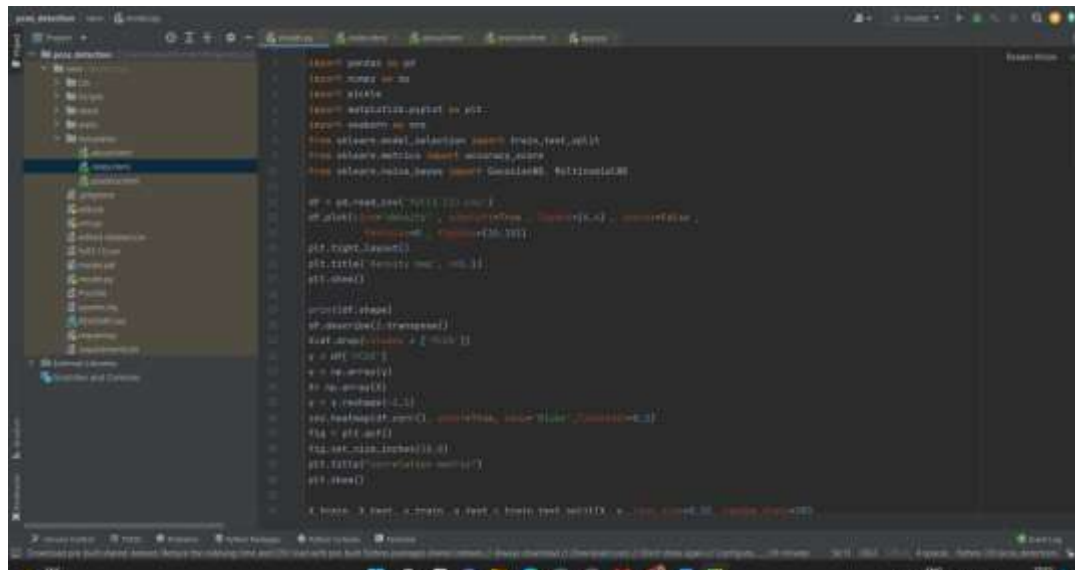
```
@app.route('/predict_api',methods=['POST'])
def predict_api():

    data = request.get_json(force=True)
    prediction = model.predict([np.array(list(data.values()))])

    output = prediction[0]
    return jsonify(output)

if __name__ == "__main__":
    app.run(debug=True)
```

B) SCREENSHOTS



```
import pandas as pd
import numpy as np
import xgboost
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix, classification_report

# Load the data
df = pd.read_csv('data.csv')

# Check the data
df.info()
df.describe()

# Split the data
X = df[['LH', 'FSH', 'TSH', 'PRL', 'Prog', 'Test', 'FreeT4', 'FreeT3', 'FreeT4I', 'FreeT3I', 'FreeT4I2', 'FreeT3I2', 'FreeT4I3', 'FreeT3I3', 'FreeT4I4', 'FreeT3I4', 'FreeT4I5', 'FreeT3I5', 'FreeT4I6', 'FreeT3I6', 'FreeT4I7', 'FreeT3I7', 'FreeT4I8', 'FreeT3I8', 'FreeT4I9', 'FreeT3I9', 'FreeT4I10', 'FreeT3I10', 'FreeT4I11', 'FreeT3I11', 'FreeT4I12', 'FreeT3I12', 'FreeT4I13', 'FreeT3I13', 'FreeT4I14', 'FreeT3I14', 'FreeT4I15', 'FreeT3I15', 'FreeT4I16', 'FreeT3I16', 'FreeT4I17', 'FreeT3I17', 'FreeT4I18', 'FreeT3I18', 'FreeT4I19', 'FreeT3I19', 'FreeT4I20', 'FreeT3I20', 'FreeT4I21', 'FreeT3I21', 'FreeT4I22', 'FreeT3I22', 'FreeT4I23', 'FreeT3I23', 'FreeT4I24', 'FreeT3I24', 'FreeT4I25', 'FreeT3I25', 'FreeT4I26', 'FreeT3I26', 'FreeT4I27', 'FreeT3I27', 'FreeT4I28', 'FreeT3I28', 'FreeT4I29', 'FreeT3I29', 'FreeT4I30', 'FreeT3I30', 'FreeT4I31', 'FreeT3I31', 'FreeT4I32', 'FreeT3I32', 'FreeT4I33', 'FreeT3I33', 'FreeT4I34', 'FreeT3I34', 'FreeT4I35', 'FreeT3I35', 'FreeT4I36', 'FreeT3I36', 'FreeT4I37', 'FreeT3I37', 'FreeT4I38', 'FreeT3I38', 'FreeT4I39', 'FreeT3I39', 'FreeT4I40', 'FreeT3I40', 'FreeT4I41', 'FreeT3I41', 'FreeT4I42', 'FreeT3I42', 'FreeT4I43', 'FreeT3I43', 'FreeT4I44', 'FreeT3I44', 'FreeT4I45', 'FreeT3I45', 'FreeT4I46', 'FreeT3I46', 'FreeT4I47', 'FreeT3I47', 'FreeT4I48', 'FreeT3I48', 'FreeT4I49', 'FreeT3I49', 'FreeT4I50', 'FreeT3I50', 'FreeT4I51', 'FreeT3I51', 'FreeT4I52', 'FreeT3I52', 'FreeT4I53', 'FreeT3I53', 'FreeT4I54', 'FreeT3I54', 'FreeT4I55', 'FreeT3I55', 'FreeT4I56', 'FreeT3I56', 'FreeT4I57', 'FreeT3I57', 'FreeT4I58', 'FreeT3I58', 'FreeT4I59', 'FreeT3I59', 'FreeT4I60', 'FreeT3I60', 'FreeT4I61', 'FreeT3I61', 'FreeT4I62', 'FreeT3I62', 'FreeT4I63', 'FreeT3I63', 'FreeT4I64', 'FreeT3I64', 'FreeT4I65', 'FreeT3I65', 'FreeT4I66', 'FreeT3I66', 'FreeT4I67', 'FreeT3I67', 'FreeT4I68', 'FreeT3I68', 'FreeT4I69', 'FreeT3I69', 'FreeT4I70', 'FreeT3I70', 'FreeT4I71', 'FreeT3I71', 'FreeT4I72', 'FreeT3I72', 'FreeT4I73', 'FreeT3I73', 'FreeT4I74', 'FreeT3I74', 'FreeT4I75', 'FreeT3I75', 'FreeT4I76', 'FreeT3I76', 'FreeT4I77', 'FreeT3I77', 'FreeT4I78', 'FreeT3I78', 'FreeT4I79', 'FreeT3I79', 'FreeT4I80', 'FreeT3I80', 'FreeT4I81', 'FreeT3I81', 'FreeT4I82', 'FreeT3I82', 'FreeT4I83', 'FreeT3I83', 'FreeT4I84', 'FreeT3I84', 'FreeT4I85', 'FreeT3I85', 'FreeT4I86', 'FreeT3I86', 'FreeT4I87', 'FreeT3I87', 'FreeT4I88', 'FreeT3I88', 'FreeT4I89', 'FreeT3I89', 'FreeT4I90', 'FreeT3I90', 'FreeT4I91', 'FreeT3I91', 'FreeT4I92', 'FreeT3I92', 'FreeT4I93', 'FreeT3I93', 'FreeT4I94', 'FreeT3I94', 'FreeT4I95', 'FreeT3I95', 'FreeT4I96', 'FreeT3I96', 'FreeT4I97', 'FreeT3I97', 'FreeT4I98', 'FreeT3I98', 'FreeT4I99', 'FreeT3I99']]
y = df['PCOS']

# Train the model
model = xgboost.XGBClassifier()
model.fit(X, y)

# Predict the results
y_pred = model.predict(X)

# Evaluate the model
accuracy = accuracy_score(y, y_pred)
confusion = confusion_matrix(y, y_pred)
report = classification_report(y, y_pred)
```

Fig B.1 Dataset



Fig B.2 Home Page

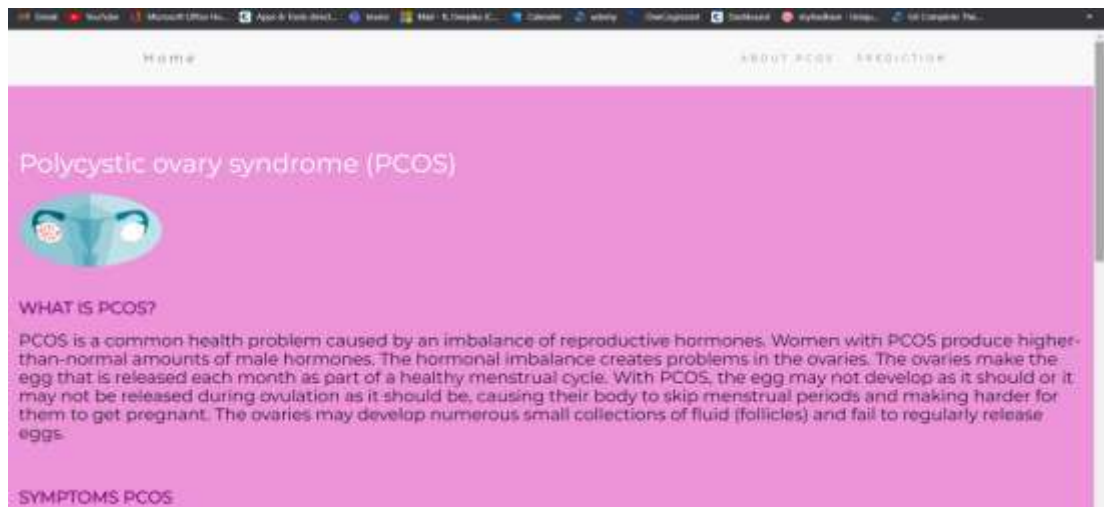


Fig B.3 About Page

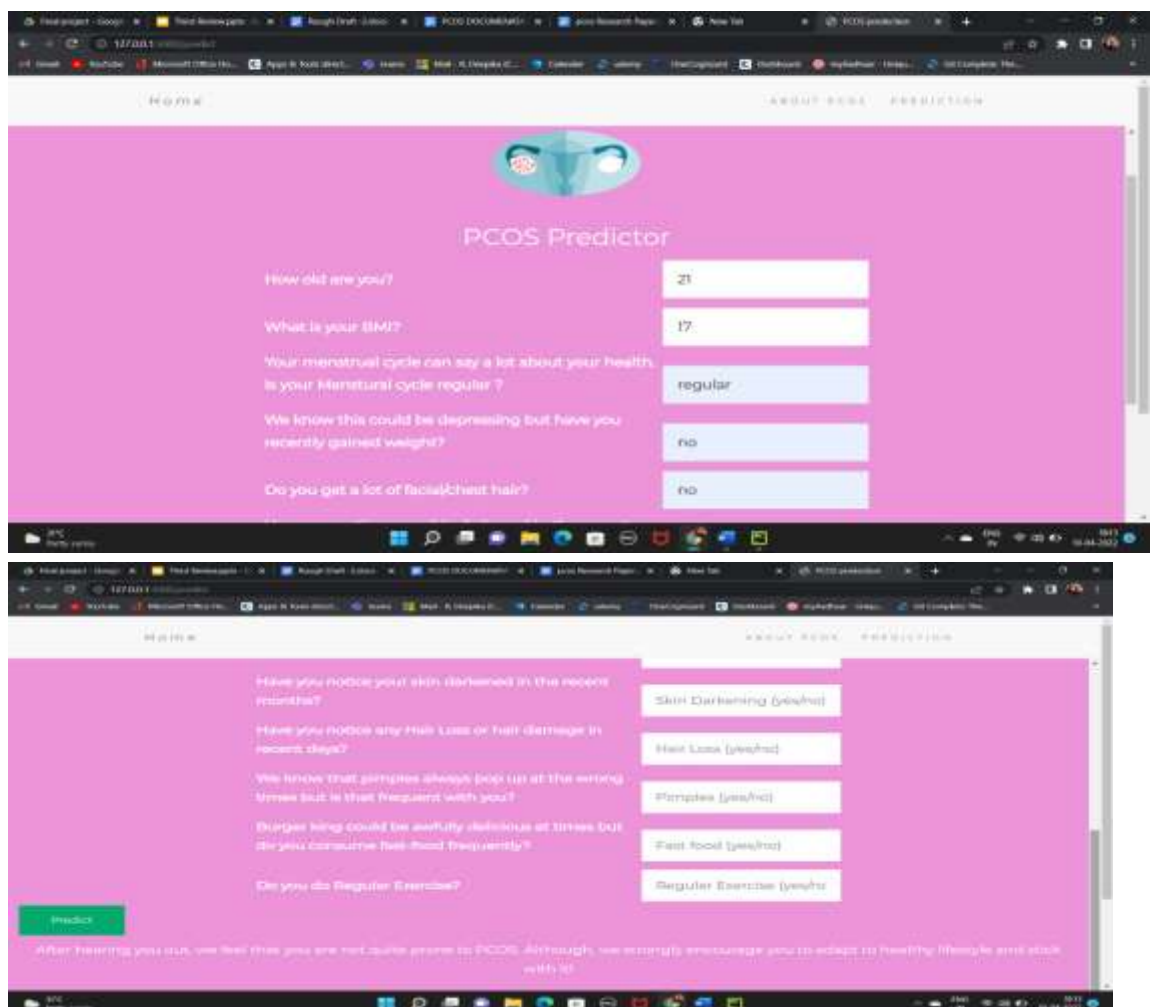


Fig B 4.Detector Page

C) PLAGIARISM REPORT

1/5

PaperPass.net

Title

YUNI - THE PCOS DETECTOR

8%

SIMILARITY INDEX

6%

ACADEMIC

10%

INTERNET

Date:	2022-01-22 14:52:48(+00:00 UTC)
Report ID:	61ec1a7ee94df538d
Word count:	1701
Character count:	9776

D) Journal Paper

YUNI-PCOS DETECTION

Dikshita Basu & Deepika R

Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamil Nadu, India
Guide: Mr. Murari Devakannan

ABSTRACT

Polycystic Ovary Syndrome or PCOS is an endocrine disorder that occurs in women of reproductive age. The condition once detected cannot be cured but treatment can help relieve its effects.

The exact cause of PCOS is still unknown but there are certain factors that portray the risk of getting PCOS. The factors that result in this syndrome are obesity, insulin resistance, blood pressure, depression, inflammation. The symptoms include hirsutism, Oligo-ovulation, acne, heavy bleeding, skin darkening. Using the causes and symptoms, a model is prepared in order to accept them as features and outputs the presence or absence of this condition. The machine learning models used for supervised classification are Logistic Regression. The reason behind building multiple models is to find out the best one for the given dataset, in the known scope of the knowledge

- web- flask server
- access to all public
- refer the nearest gynecologist

Keywords:

PCOS, Machine Learning, Multi-layer Perceptron, Logistic Regression.

INTRODUCTION

Polycystic Ovary Syndrome, or PCOS, is a health condition that affects about 10 million people in the world. The exact cause is unknown, but it is considered a hormonal problem. Genetics and environmental factors are believed to be involved in the development of PCOS. It is a leading cause of female infertility and is responsible for a number of

symptoms that can affect the body physically and emotionally. Symptoms of PCOS may begin shortly after puberty, but can also develop during the later teen years and early adulthood. Because symptoms may be attributed to other causes or go unnoticed, PCOS may go undiagnosed for some time. People with PCOS typically have irregular or missed periods as a result of not ovulating. Although some people may develop cysts on their ovaries, many people do not. Other symptoms include

Weight gain, Fatigue, Unwanted hair growth, thinning hair on the head, Infertility, Acne, Mood changes, Pelvic pain, Headaches, Sleep problems.

LITERATURE SURVEY

S. N O	AUTHOR	TITLE	DESCRIPTION	DIS-ADVANTAGE
1	Zhang, 2018	Predicting & identifying Risk Of PCOS	Hyperprolactinemia, Cushing's syndrome and non-classic congenital adrenal hyperplasia .	Low efficiency of result.
2	[P. Mehrotra, 2011]	A machine learning approach for selection of polycystic ovarian syndrome (pcos) Attributes and comparing Different classifier Performance with the help of . .	used machine learning algorithms like Bayes and Logistic Regression	Abnormal algorithm used in this system

3	Norman, 2007	Data Analysis on the Risks of Obesity and Overweight in Women-A Study	comprehensive study on the disorder and its three diagnostic criteria in depth giving us insights on not just PCOS	abnormalities of insulin, gonadotropin and follicular genesis.
4	B. Cahyono, Adiwijaya, M.S.	Automated Detection of Polycystic Ovarian Syndrome Using Follicle Recognition.	In this paper, authors have tried To recognize PCOS based on the follicle recognition through image pre-processing to increase the quality of image. Multiscale morphological approach and top-hat transform is used to classify the image of follicles so that they can be separated according to size. Used for automated results after the screening.	image processing using it not efficient
5	Sharvari S. Deshpande, Asmita	An Implementation of Convolutional Neural Network on PCOS Classification based on Ultrasound Images.	Classifies patients into PCO and NON-PCO class by using convolutional neural network algorithms over ultrasound images. Suggests that optimizing neural networks is hard. Shows the percentage of accuracy attained by the author.	The result timing is high.
6	D. Whitley, T. Starkweather	Computational Characterization and Identification of Human Polycystic Syndrome Genes	Developed an algorithm to predict new PCOS genes. Shows out of a certain number of genes how they can be positive or Negative	Low efficiency algorithm is used
7	Honglei Jing, Xiang Xu, Jinzhu Wang	CLINICAL REVIEW: Identifying Children at risk For Polycystic Ovary Syndrom.	Provides a prospect that PCOS can be prevented at an early age. There are many risk factors that can be monitored at early get like birth age, insulin obesity in early childhood, assessment of hyperandrogenism can show the	process is not good

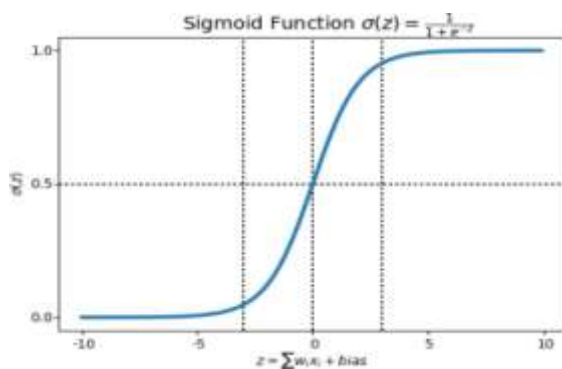
			chances of PCOS.	
8	Xiangyang Liu, Juan Bao, Yan	i-HOPE: Detection and Prediction System For Polycystic Ovary Syndrome (PCOS) Using Machine Learning Techniques.	Various machine algorithms are applied over the data obtained among which RF algorithm is found superior.	Less accuracy in result.
9	Akinrotimi, Akinyemi Omololu	Obese and Non-Obese Polycystic Ovarian Syndrome: Comparison	Patients suffering from obese PCOS are more prone to get hypertension, IR, Metabolic syndrome. Controlling obesity in patients makes them more responsive towards clomiphene.	algorithm is not supported.
10	Hector F. Escobar-Morreale	responsive towards clomiphene.	Shows how cysts get developed in the ovary and what the symptoms are and how these cysts cause hypertension, diabetes etc. Ultrasound images are one of the best methods which can be used to.	slow process.
11	S. Sheela, M. Sumathi	Segmentation of Polycystic Ovary in Ultrasound Image	ICV algorithm is developed that can give us right result in less computational time and iterations.	The icv algorithm does not give good results.
12	H Prasanna Kumar, S Srinivasan	Exploration On Polycystic Ovarian Syndrome and Data Mining Techniques.	Various data mining techniques are discussed and how they can be useful in getting right results	data mining is missing some data

PROPOSED WORK

Logistic Regression

Logistic Regression is a Machine Learning algorithm which is used for the classification problems, it is a predictive analysis algorithm and based on the concept of probability. Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving

the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1. Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems. In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.



How does Logistic Regression work ?

logistic regression can be used both in classification and regression problems but it is widely used as a classification algorithm. Regression models aim to project value based on independent features.

when the dependent variables are binary logistic regression is considered.

Multi-Layer Perceptron

Our model consists of three Multilayer Perceptron layers in a Dense layer. The first and second are identical, followed by a Rectified Linear Unit (ReLU) and Dropout activation function.

A multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses backpropagation for training the network. MLP is a deep learning method.

MLP is widely used for solving problems that require supervised learning as well as research into computational neuroscience and parallel distributed processing. Applications include speech recognition, image recognition and machine translation.

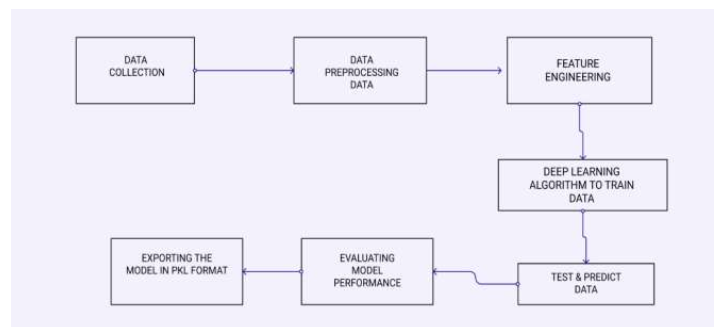
How does MLP work?

A multilayer perceptron is a neural network connecting multiple layers in a directed graph, which means that the signal path through the nodes only goes one way. Each node, apart from the input nodes, has a nonlinear activation function. An MLP uses backpropagation as a supervised learning technique. Since there are multiple layers of neurons, MLP is a deep learning technique.

Why MLP over Logistic Regression ?

We tried cleaning the data and training it in logistic regression but the accuracy was poor and we wanted to create a system that works much better than the previous system. So, we tried deep learning and when we trained the data using MLP we got an excellent accuracy of 98%. Thus our goal was achieved.

ARCHITECTURE DIAGRAM:



CONCLUSION

Polycystic Ovary Syndrome (PCOS) is a medical condition which causes hormonal disorder in women in their childbearing years.

The hormonal imbalance leads to a delayed or even absent menstrual cycle. Women with PCOS majorly suffer from excessive weight gain, facial hair growth, acne, hair loss, skin darkening and irregular periods leading to infertility in rare cases.

Our proposal seems to be better than the respective others, Hence, Multi-layer perceptron is used for our final hypothesis which will predict the result using test data with accuracy of 98%.

FUTURE SCOPE

I'd like to train our model on a larger dataset. So, the Person using this detection to check whether they have PCOS or not in a more accurate level and We would like to make a more easy and user friendly website. So the one using it finds it very reliable and helps them to detect and get a checkup in their near gynecologist, if they are prone to it.

REFERENCES

- [1] Mehrotra, P., Chatterjee, J., Chakraborty, C., Ghoshdastidar, B., & Ghoshdastidar, S.(2011).Automated screening of polycystic ovary syndrome using machine learning .
- [2] Revised 2003 consensus on diagnostic criteria and long term health risks related to polycystic ovary syndrome (PCOS) . Human Reproduction, Volume 19, Issue 1, January 2004, Pages 41-47.
- [3] Supervised Learning-classification using K-Nearest Neighbours (KNN).(2019).Python® MachineLearning,205–220.doi:10.1002/9781119557500.ch9
- [4] Bichler, Martin and Kiss, Christine, "A Comparison of Logistic Regression, k-Nearest Neighbor, and Decision Tree Induction for Campaign Management" (2004). AMCIS 2004 Proceedings. 230.
- [5] B.Cahyono, Adiwijaya,M.S. Mubarak,U.N. Wisesty "An Implementation of Convolutional Neural Network on PCO Classification",Fifth International Conference on Communication Technology,2017
- [6] "PCOS" <https://www.mayoclinic.org/diseases-conditions/pcos/symptoms-causes/syc-20353439>
- [7] Sharvari S.Deshpande,Asmita Wakankar, "Automated Detection of Polycystic Ovarian Syndrome Using Follicle Recognition",IEEE International Conference on Advanced Communication Control and Computing Technologies(ICACCCT),2014,1341-1342
- [8] "Neural Network",<https://www.investopedia.com/terms/n/neuralnetwork.asp>,Oct 2019.
- [9] Jing Li, Ji-hang Cheng, Jing Yuan Shi, Fei Huang, "Brief introduction of Back Propagation(BP) Neural Network Algorithm and Its Improvement" Springer-Verlag Berlin Heidelberg, 2012, 553-554.

