

DIABETIC RETINOPATHY DETECTION USING DEEP LEARNING TECHNIQUES

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

by

SHAIK SAMEER (Reg. No. 37110703)

SHAIK FARUKHE AZAM (Reg. No. 37110704)



**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**

MARCH - 2021



SATHYABAMA



INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited with "A" grade by NAAC
Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119
www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of **SHAIK SAMEER (Reg. No. 37110703)** and **SHAIK FARUKHE AZAM (Reg. No.37110704)** who carried out the project entitled "**DIABETIC RETINOPATHY DETECTION USING DEEP LEARNING TECHNIQUES**" under my supervision from **November 2021** to **March 2021**.

Internal Guide
Dr.A.SIVASANGARI,M.E.,Ph.D.,

Head of the Department

Submitted for Viva voce Examination held on_____

Internal Examiner

External Examiner

DECLARATION

I **SHAIK SAMEER** hereby declare that the Project Report entitled “**DIABETIC RETINOPATHY DETECTION USING DEEP LEARNING TECHNIQUES**” is done by me under the guidance of **Dr.A.SIVASANGARI.,M.E.,Ph.D.**, Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: CHENNAI

SIGNATURE OF THECANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management of SATHYABAMAINSTITUTE OF SCIENCE AND TECHNOLOGY** 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. Vigneswari, M.E., Ph.D.,** and **Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.A.SIVASANGARI,M.E.,Ph.D., Assistant Professor**, 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

There are several machine learning techniques that are used to perform predictive analytics over big data in various fields. Predictive analytics in healthcare is a challenging task but ultimately can help practitioners make big data-informed timely decisions about patient's health and treatment. This paper discusses the predictive analytics in healthcare, six different machine learning algorithms are used in this research work. For experiment purpose, a dataset of patient's medical record is obtained and six different machine learning algorithms are applied on the dataset. In this project we are using Convolutional Neural Network algorithm. Performance and accuracy of the applied algorithms is discussed and compared. Comparison of the different machine learning techniques used in this study reveals which algorithm is best suited for prediction of diabetes. This project aims to help doctors and practitioners in early prediction of diabetes using deep learning techniques.

TABLE OF CONTENTS

	ABSTRACT	v
	LIST OF FIGURES	viii
CHAPTER No	TITLE	PAGE No.
1	INTRODUCTION	1
	1.1 INTRODUCTION	1
	1.2 ONTRODUTION TO COMPUTER AIDED DIAGNOSIS	2
2	AIM AND SCOPE	5
	2.1 AIM OF THE PROJECT	5
	2.2.1 OBJECTIVES	5
	2.2 SCOPE	5
	2.3 ADVANTAGES	5
3	SYSTEM DESIGN & METHODOLOGY	6
	3.1 EXISTING SYSTEM	6
	3.2 PROPOSED SYSTEM	6
	3.3 REQUIREMENT SPECIFICATION	6
	3.4 TECHONOLOGIES USED	7
	3.4.1 INTRODUCTION TO PYTHON	7
	3.4.2 INTRODUCTION TO RANDOM FOREST	10
4	SOFTWARE DESIGN AND IMPLEMENTATION	24
	4.1 DESIGN AND IMPLEMENTATION CONSTRAINTS	24
	4.2 OTHER NONFUNCTIONAL REQUIREMENTS	24
	4.3 ARCHITECTURE DIAGRAM	25
	4.4 SEQUENCE DIAGRAM	26
	4.5 USE CASE DIAGRAM	26
	4.6 ACTIVITY DIAGRAM	27
	4.7 COLLABRATION DIAGRAM	28
	4.8 MODULES	29
	4.9 CODING AND TESTING	31
	4.10 TEST DATA AND OUTPUT	32
	4.11 TESTING TECHNIQUES	34
5	RESULT AND DISCUSSION	38
	5.1 RESULT	38

6	CONCLUSION AND FUTURE WORK	44
	6.1 CONCLUSION	44
	6.2 FUTURE WORK	44
	REFERENCES	45
	APPENDIX	47
	A. PAPER WORK	47
	B. REPORT	54
	C. SOURCE CODE	57

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE No.
3.1	RANDOM FOREST	10
3.2	SUPPORT VECTOR MACHINE	14
3.3	CLASSIFICATION OF ANALYSIS	15
3.4	SVM FOR CASE IN HAND	16
3.5	DISTRIBUTION	17
3.6	TRANSFORMATION	17
3.7	SCENARIO-1	18
3.8	SCENARIO-2	19
3.9	MARGIN	19
3.10	SCENARIO-3	20
3.11	SCENARIO-4	20
3.12	CLASSIFY TWO CLASSES	21
3.13	SCENARIO-5	21
3.14	SEGREGATE TWO CLASSES	22
3.15	HYPER-PLANE	23
4.1	ARCHITECTURE DIAGRAM	25
4.2	SEQUENCE DIAGRAM	26
4.3	USECASE DIAGRAM	27
4.4	ACTIVITY DIAGRAM	28
4.5	COLLABORATION DIAGRAM	29

5.1	WELCOME PAGE	38
5.2	REGISTRATION PAGE	39
5.3	LOG-IN PAGE	39
5.4	HOME PAGE	40
5.5	FILE UPLOAD PAGE	41
5.6	RESULT PAGE	41
5.7	STAGES OF DR	42
5.8	COMPARISON OF DR RESULTS	43

CHAPTER 1

INTRODUCTION

N

1.1 INTRODUCTION

Healthcare industry contains very large and sensitive data and needs to be handled very carefully. Diabetes Mellitus is one of the growing extremely fatal diseases all over the world. Medical professionals want a reliable prediction system to diagnose Diabetes. Different machine learning techniques are useful for examining the data from diverse perspectives and synopsisizing it into valuable information. The accessibility and availability of huge amounts of data will be able to provide us useful knowledge if certain data mining techniques are applied to it. The main goal is to determine new patterns and then to interpret these patterns to deliver significant and useful information for the users. Diabetes contributes to heart disease, kidney disease, nerve damage, and blindness. Mining the diabetes data in an efficient way is a crucial concern. The data mining techniques and methods will be discovered to find the appropriate approaches and techniques for efficient classification of Diabetes dataset and in extracting valuable patterns. In this study, medical bioinformatics analyses have been accomplished to predict diabetes. The WEKA software was employed as a mining tool for diagnosing diabetes. The Pima Indian diabetes database was acquired from UCI repository used for analysis. The dataset was studied and analyzed to build an effective model that predicts and diagnoses diabetes disease. In this study, we aim to apply the bootstrapping resembling technique to enhance the accuracy and then applying Naïve Bayes, Decision Trees and (KNN) and compare their performance.

With an increase in the computing and processing power of computers and the development of image processing techniques in recent years, the idea of using a computer to analyze medical images and automatically diagnose diseases has attracted the attention of many medical and computer specialists. This idea of analyzing retina images and diagnosing ocular and vascular diseases is very efficient and inexpensive. Retinal vascular structures contain a great deal of medical information. Diabetes, hypertension, cardiovascular diseases, and some others can only be detected through the examination of retinal fundus images. Diabetic Retinopathy (DR) The associate

editor coordinating the review of this manuscript and approving it for publication was Tao Zhou . monitoring is one of the acknowledged methods in examining patients. Annual screening of the eye fundus of diabetic patients is one of the essential parts of

their treatment. However, the analysis of retinal fundus images demands an ophthalmologist expert.

The requirements for DR monitoring and limitation in resources inspired researchers to develop an automated system to analyze retinal fundus images. Generally, there are five different DR categories: No DR, Mild DR, Moderate DR, Severe DR, and Proliferative DR. An automated detection of DR, classifies each patient into one of these classes through the examination of their retinal fundus images. The early models of such systems were limited to computer-aided systems that provided vessel segmentation of fundus images. Some computer vision DR detection models are also proposed, but their accuracy rates are rather low. Convolutional Neural Networks (CNNs) and deep learning introduced a new stage for this problem. They helped researchers to design a reliable end to end DR detection system.

1.2 INTRODUCTION TO COMPUTER AIDED DIAGNOSIS

Diabetes mellitus (DM) is specified by a group of unifying metamorphoses caused by insulin resistance . Insulin resistance leads to a high ratio of glucose in the blood (hyperglycemia), which may eventually ensue to macro (stroke, heart attack) and micro (neuropathy, retinopathy, and nephropathy) vascular pathologies . According to recent statistics , there are about 382 million people diagnosed as diabetic patients and this number is expected to increase to reach 592 million by 2030. DM has two main clinical types, including type1 diabetes (T1D) and type 2 diabetes (T2D). About 90% of diabetic patients across the globe have T2D diabetes, which is stimulated by the resistance of insulin. Diabetic retinopathy (DR) is one of the significant ramifications of DM and creates a progressive visual illness in the adult population characterized by the deterioration of tiny retinal vessels . However, early screening and treatment through computer-aided diagnostic (CAD) programs can assist patients to prevent complete blindness occurring.

Several clinical studies show that the presence of microaneurysms (MAs), soft/hard exudates (EXs), and hemorrhages (HMs) in the fundus image are pathognomonic signs of diabetic retinopathy. MAs are bulges of thin vessels and appear as small and sharp borders of red spots on the retinal surface whereas EXs exist in the retinal region due to the leakage of proteins from vessels and appear as white/yellowish-white spots of a small size. HMs look like red deposits with non-uniform margins and are caused by the leaking of weak capillaries. Furthermore, DR is broadly classified into two stages based

on vessel degeneration and ischemic change, including non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). NPDR is further

subdivided in to mild (presence of MAs), moderate (advance than mid-level), and severe (venous beading in at least two quadrants, MA, HM in four retina quadrants) stages, while PDR is an advanced stage of DR. Moreover, NPDR is clinically recommended for the detection of the severity level of DR.

Innovation in the field of science and technology has made human life healthier, secure, and more livable. For instance, automated medical imaging analysis is a novel area of research that identifies serious diseases by utilizing imaging technology. Those automated diagnosis systems (ADSs) are called computer-aided systems (CADs), which provide services for the ease of humanity. The role of CAD systems has been recognized in the diverse range of medical diagnoses, such as brain tumor detection, glaucoma detection, diabetic retinopathy detection, gastrointestinal detection, macular edema, cardiac grading, Alzheimer's and Parkinson's assessment, breast cancer diagnosis, and many more. Those CAD systems have produced similar results with human observations. The development of CAD devices relies on the image features extracted from images using computational methods. Manual inspection of retinal anatomic (e.g., optic disc, cup, fovea, macula, and vascular structure) features and DR lesions is a tedious and repetitive task for clinical personnel. Moreover, human analysis of DR is subjective and requires technical domain knowledge to observe inter- or intra-changes in retinal components. As a result, computer-aided diagnostic (CAD) systems are required to execute early DR screening programs using color fundus images. These CAD systems may reduce the time, cost, and efforts of clinicians in the manual analysis of retinal images.

Diabetic retinopathy (DR) is a complication of diabetes that exists throughout the world. DR occurs due to a high ratio of glucose in the blood, which causes alterations in the retinal microvasculature. Without preemptive symptoms of DR, it leads to complete vision loss. However, early screening through computer-assisted diagnosis (CAD) tools and proper treatment have the ability to control the prevalence of DR. Manual inspection of morphological changes in retinal anatomic parts are tedious and challenging tasks. Therefore, many CAD systems were developed in the past to assist ophthalmologists for observing inter- and intra-variations. In this paper, a recent review of state-of-the-art CAD systems for diagnosis of DR is presented. We describe all those CAD systems that have been developed by various computational intelligence and image processing techniques. The limitations and future trends of current CAD systems are also described

in detail to help researchers. Moreover, potential CAD systems are also compared in terms of statistical parameters to quantitatively evaluate them. The comparison results

indicate that there is still a need for accurate development of CAD systems to assist in the clinical diagnosis of diabetic retinopathy.

Diabetic Retinopathy (DR) is the leading cause of blindness in the workingage group. Among 23 million Americans, 59 million Europeans, and as many as 50 million Indians suffering from diabetes.

Regular eye examination among these vulnerable groups is necessary to diagnose DR at an early stage, when it can be treated with the best prognosis. Currently, detecting DR is a time-consuming and manual process that requires a trained clinician to examine and evaluate digital color fundus photographs of the retina. The clinical grading process consists of detection certain subtle features, such as microaneurysms, exudates, intra-retinal hemorrhages and sometimes their position relative to each other on images of the eye. We make use of deep Convolutional Neural Networks (CNNs), which have proven revolutionary in multiple fields of computer vision including medical imaging, and we bring their power to the diagnosis of eye fundus images. With continuous development of our Deep Learning models we expect to further increase the accuracy of the method and expand it to cataract and glaucoma diagnostics.

CHAPTER 2

AIM AND SCOPE

2.1 AIM OF PROJECT

The primary aim of this proposed system is to help doctors and practitioners in early prediction of diabetes using machine learning techniques.

2.1.1 OBJECTIVES

Diabetic retinopathy (DR) is a common complication of diabetes associated with retinal vascular damage caused by long standing diabetes. Furthermore, the diagnosis of DR mostly depends on the observation and evaluation to fundus photographs of which procedure can be time-consuming even for experienced experts. Therefore computer aided automated diagnosis approaches have great potential in clinical to accurately detect DR in a short time which can further help to improve the screening rate of DR and reduce the blindness. For a deep learning model, the most important parts that should be focused on are data set, network architecture and training method. Before being used to train our model, fundus images data set obtained from public resources is preprocessed and augmented. The model accepts two fundus images corresponding to the left eye and right eye as inputs and then transmits them into the Siamese like blocks. The information from two eyes is gathered into the fully connected layer and finally the model will output the diagnosis result of each eye respectively.

2.2 SCOPE

A dataset of patient's medical record is obtained and three different machine learning algorithms are applied on the dataset. Performance and accuracy of the applied algorithms is discussed and compared. Comparison of the different machine learning techniques used in this study reveals which algorithm is best suited for prediction of diabetes

2.3 ADVANTAGES

- Time saving and efficient, it is operated online.
- This system has database which helps in retrieving images.
- It provides authentication and it is user friendly.

CHAPTER 3

SYSTEM DESIGN & METHODOLOGY

3.1 EXISTING SYSTEM

The obtained results are compared with the results of existing models within the same domain and found to be improved. The data of diabetic patients collected from the UCI laboratory is used to discover patterns with are K Nearest Neighbours (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR) and Random Forest (RF). The results are compared for performance and accuracy with these algorithms. The proposed hybrid method returns results of 78.5%, competing with the other existing methods.

3.2 PROPOSED SYSTEM

In this paper, six machine learning algorithms are used to predict diabetes disease. These six algorithms are K Nearest Neighbours (KNN), Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR) and Random Forest (RF). Comparison of the different machine learning techniques used in this study reveals which algorithm is best suited for prediction of diabetes.

3.3 REQUIREMENTS SPECIFICATION

This proposed software runs effectively on a computing system that has the minimum requirements.

The requirements are split into two categories, namely:

Software Requirements

The basic software requirements to run the program are:

- ✓ Operating System : Windows 7 , 8, 10 (64 bit)
- ✓ Software : Python
- ✓ Tools : Anaconda (Jupyter Note Book IDE)

Hardware Requirements

The basic hardware required to run the program are:

- Hard Disk : 500GB and Above
- RAM : 4GB and Above

- Processor : I3 and Above

3.4 TECHNOLOGIES USED

Python

Random Forest and Support Vector

3.4.1 Introduction to Python

Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently.

It is used for:

- web development (server-side),
- software development,
- mathematics,
- System scripting.

What can Python do?

- Python can be used on a server to create web applications.
- Python can be used alongside software to create workflows.
- Python can connect to database systems. It can also read and modify files.
- Python can be used to handle big data and perform complex mathematics.
- Python can be used for rapid prototyping, or for production-ready software development.

Why Python?

- Python works on different platforms (Windows, Mac, Linux, Raspberry Pi, etc).
- Python has a simple syntax similar to the English language.
- Python has syntax that allows developers to write programs with fewer lines than some other programming languages.
- Python runs on an interpreter system, meaning that code can be executed as soon as it is written. This means that prototyping can be very quick.
- Python can be treated in a procedural way, an object-orientated way or a

functional way.

Good to know

- The most recent major version of Python is Python 3, which we shall be using in this tutorial. However, Python 2, although not being updated with anything other than security updates, is still quite popular.
- Python 2.0 was released in 2000, and the 2.x versions were the prevalent releases until December 2008. At that time, the development team made the decision to release version 3.0, which contained a few relatively small but significant changes that were not backward compatible with the 2.x versions. Python 2 and 3 are very similar, and some features of Python 3 have been backported to Python 2. But in general, they remain not quite compatible.
- Both Python 2 and 3 have continued to be maintained and developed, with periodic release updates for both. As of this writing, the most recent versions available are 2.7.15 and 3.6.5. However, an official End Of Life date of January 1, 2020 has been established for Python 2, after which time it will no longer be maintained.
- Python is still maintained by a core development team at the Institute, and Guido is still in charge, having been given the title of BDFL (Benevolent Dictator For Life) by the Python community. The name Python, by the way, derives not from the snake, but from the British comedy troupe Monty Python's Flying Circus, of which Guido was, and presumably still is, a fan. It is common to find references to Monty Python sketches and movies scattered throughout the Python documentation.
- It is possible to write Python in an Integrated Development Environment, such as Thonny, Pycharm, Netbeans or Eclipse which are particularly useful when managing larger collections of Python files.

Python Syntax compared to other programming languages

- Python was designed to for readability, and has some similarities to the English language with influence from mathematics.
- Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.

- Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

Python is Interpreted

- Many languages are compiled, meaning the source code you create needs to be translated into machine code, the language of your computer's processor, before it can be run. Programs written in an interpreted language are passed straight to an interpreter that runs them directly.
- This makes for a quicker development cycle because you just type in your code and run it, without the intermediate compilation step.
- One potential downside to interpreted languages is execution speed. Programs that are compiled into the native language of the computer processor tend to run more quickly than interpreted programs. For some applications that are particularly computationally intensive, like graphics processing or intense number crunching, this can be limiting.
- In practice, however, for most programs, the difference in execution speed is measured in milliseconds, or seconds at most, and not appreciably noticeable to a human user. The expediency of coding in an interpreted language is typically worth it for most applications.
- For all its syntactical simplicity, Python supports most constructs that would be expected in a very high-level language, including complex dynamic data types, structured and functional programming, and object-oriented programming.
- Additionally, a very extensive library of classes and functions is available that provides capability well beyond what is built into the language, such as database manipulation or GUI programming.
- Python accomplishes what many programming languages don't: the language itself is simply designed, but it is very versatile in terms of what you can accomplish with it.

3.4.2 Introduction to Random Forest

- With increase in computational power, we can now choose algorithms which perform very intensive calculations. One such algorithm is “Random Forest”, which we will discuss in this article. While the algorithm is very popular in various competitions (e.g. like the ones running on Kaggle), the end output of the model is like a black box and hence should be used judiciously.
- Before going any further, here is an example on the importance of choosing the best algorithm.

Random Forest Case Study

- Following is a distribution of Annual income Gini Coefficients across different countries:

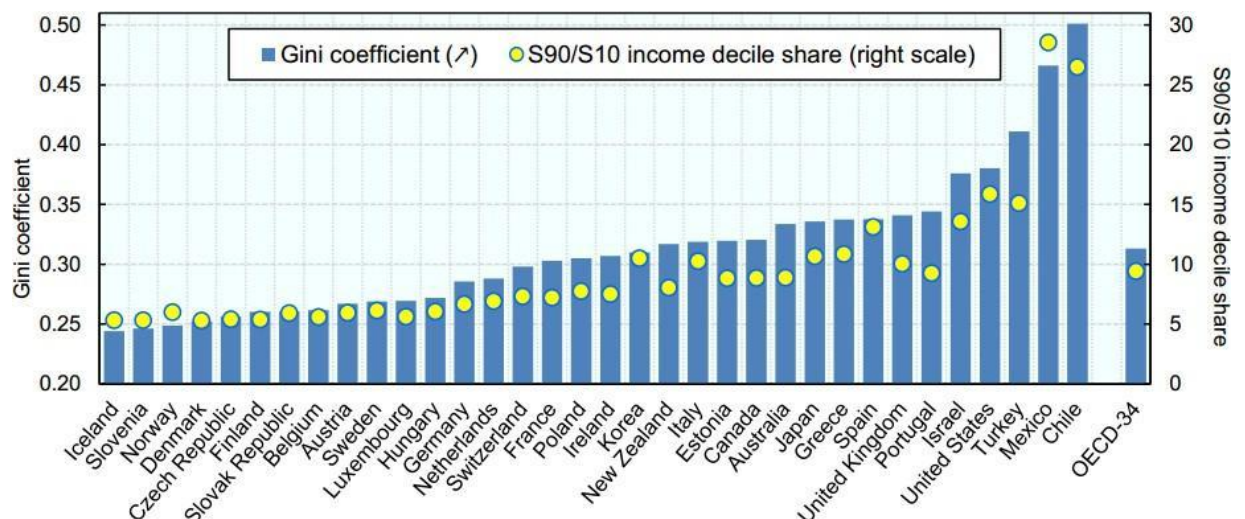


FIG 3.1: RANDOM FOREST

- Mexico has the second highest Gini coefficient and hence has a very high segregation in annual income of rich and poor. Our task is to come up with an accurate predictive algorithm to estimate annual income bracket of each individual in Mexico. The brackets of income are as follows:
 - 1. below \$40,000
 - 2. \$40,000 – 150,000
 - 3. More than \$150,000
- Following are the information available for each individual:
 - 1. Age, 2. Gender, 3. Highest educational qualification, 4. Working in Industry, 5. Residence in Metro/Non-metro

- We need to come up with an algorithm to give an accurate prediction for an individual who has following traits:
- 1. Age: 35 years, 2. Gender: Male, 3. Highest Educational Qualification: Diploma holder, 4. Industry: Manufacturing, 5. Residence: Metro
- We will only talk about random forest to make this prediction in this article.
- **The algorithm of Random Forest**
- Random forest is like bootstrapping algorithm with Decision tree (CART) model. Say, we have 1000 observation in the complete population with 10 variables. Random forest tries to build multiple CART model with different sample and different initial variables. For instance, it will take a random sample of 100 observation and 5 randomly chosen initial variables to build a CART model. It will repeat the process (say) 10 times and then make a final prediction on each observation. Final prediction is a function of each prediction. This final prediction can simply be the mean of each prediction.
- **Back to Case study**
- Disclaimer: The numbers in this article are illustrative
- Mexico has a population of 118 MM. Say, the algorithm Random forest picks up 10k observation with only one variable (for simplicity) to build each CART model. In total, we are looking at 5 CART models being built with different variables. In a real life problem, you will have more number of population sample and different combinations of input variables.
- Salary bands:
- Band 1: Below \$40,000
- Band 2: \$40,000 – 150,000
- Band 3: More than \$150,000
- Following are the outputs of the 5 different CART model.
- **CART 1: Variable Age**

	Salary Band	1	2	3
Age	Below 18	90%	10%	0%
	19-27	85%	14%	1%
	28-40	70%	23%	7%
	40-55	60%	35%	5%
	More than 55	70%	25%	5%

- **CART 2: Variable Gender**

	Salary Band	1	2	3
Gender	Male	70%	27%	3%
	Female	75%	24%	1%

- **CART 3: Variable Education**

	Salary Band	1	2	3
Education	<=High School	85%	10%	5%
	Diploma	80%	14%	6%
	Bachelors	77%	23%	0%
	Master	62%	35%	3%

- **CART 4: Variable Residence**

	Salary Band	1	2	3
Residence	Metro	70%	20%	10%
	Non-Metro	65%	20%	15%

- **CART 5: Variable Industry**

	Salary Band	1	2	3
Industry	Finance	65%	30%	5%
	Manufacturing	60%	35%	5%
	Others	75%	20%	5%

- Using these 5 CART models, we need to come up with single set of probability to belong to each of the salary classes. For simplicity, we will just take a mean of probabilities in this case study. Other than simple mean, we also consider vote method to come up with the final prediction. To come up with the final prediction let's locate the following profile in each CART model:
- 1. Age: 35 years, 2, Gender: Male, 3. Highest Educational Qualification: Diploma holder, 4. Industry: Manufacturing, 5. Residence: Metro
- For each of these CART model, following is the distribution across salary bands:

CART	Band	1	2	3
Age	28-40	70%	23%	7%
Gender	Male	70%	27%	3%
Education	Diploma	80%	14%	6%
Industry	Manufacturing	60%	35%	5%
Residence	Metro	70%	20%	10%
Final probability		70%	24%	6%

- The final probability is simply the average of the probability in the same salary bands in different CART models. As you can see from this analysis, that there is 70% chance of this individual falling in class 1 (less than \$40,000) and around 24% chance of the individual falling in class 2.
- **Conclusion**
- Random forest gives much more accurate predictions when compared to simple CART/CHAID or regression models in many scenarios. These cases generally have high number of predictive variables and huge sample size. This is because it captures the variance of several input variables at the same time and enables high number of observations to participate in the prediction.

Support Vector Machine

(SVM) Introduction:

Mastering machine learning algorithms isn't a myth at all. Most of the beginners start by learning regression. It is simple to learn and use, but does that solve our purpose? Of course not! Because, you can do so much more than just Regression!

Think of machine learning algorithms as an armory packed with axes, sword, blades, bow, dagger etc. You have various tools, but you ought to learn to use them at the right time. As an analogy, think of 'Regression' as a sword capable of slicing and dicing data efficiently, but incapable of dealing with highly complex data. On the contrary, 'Support Vector Machines' is like a sharp knife – it works on smaller datasets, but on them, it can be much stronger and powerful in building models.

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for either classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature

being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiate the two classes very well (look at the below snapshot).

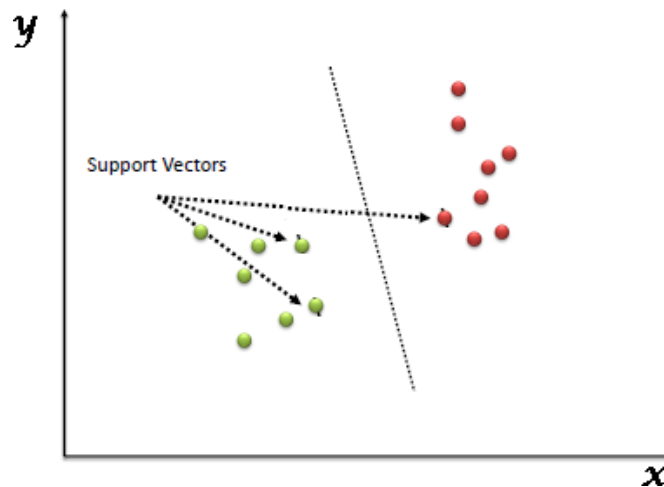


FIG 3.2: SUPPORT VECTOR MACHINE

Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

You can look at definition of support vectors and a few examples of its working [here](#).

Let's talk about how support vector machine works. This is suitable for readers who do not know much about this algorithm and have a curiosity to learn a new technique. In following explore the technique in detail and analyze cases where such techniques are stronger than other techniques.

Classification analysis

Let's consider an example to understand these concepts. We have a population composed of 50%-50% Males and Females. Using a sample of this population, you want to create some set of rules which will guide us the gender class for rest of the population. Using this algorithm, we intend to build a robot which can identify whether a person is a Male or a Female. This is a sample problem of classification analysis. Using some set of rules, we will try to classify the population into two possible segments. For simplicity, let's assume that the two differentiating factors identified are: Height of the individual and Hair Length. Following is a scatter plot of the sample.

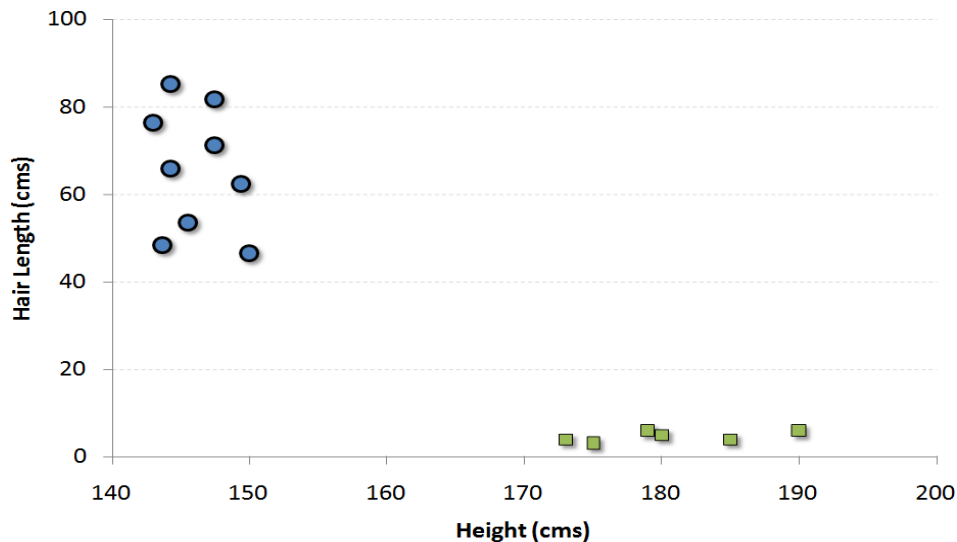


FIG 3.3: CLASSIFICATION ANALYSIS

The blue circles in the plot represent females and green squares represents male. A few expected insights from the graph are:

1. Males in our population have a higher average height.
2. Females in our population have longer scalp hairs.

If we were to see an individual with height 180 cm's and hair length 4 cm's, our best guess will be to classify this individual as a male. This is how we do a classification analysis.

A Support Vector and SVM

Support Vectors are simply the co-ordinates of individual observation. For instance, (145,150) is a support vector which corresponds to a female. Support Vector Machine is a frontier which best segregates the Male from the Females. In this case, the two classes are well separated from each other; hence it is easier to find a SVM.

How to find the Support Vector Machine for case in hand?

There are many possible frontiers which can classify the problem in hand. Following are the three possible frontiers.

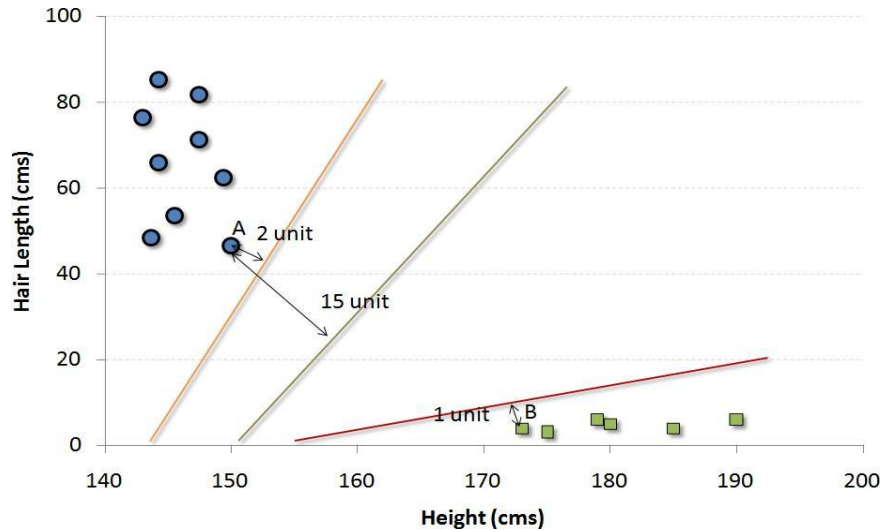


FIG 3.4: SVM FOR CASE IN HAND

How do we decide which is the best frontier for this particular problem statement?

The easiest way to interpret the objective function in a SVM is to find the minimum distance of the frontier from closest support vector (this can belong to any class). For instance, orange frontier is closest to blue circles. And the closest blue circle is 2 units away from the frontier. Once we have these distances for all the frontiers, we simply choose the frontier with the maximum distance (from the closest support vector). Out of the three shown frontiers, we see the black frontier is farthest from nearest support vector (i.e. 15 units).

What if we do not find a clean frontier which segregates the classes?

Our job was relatively easier finding the SVM in this business case. What if the distribution looked something like as follows?

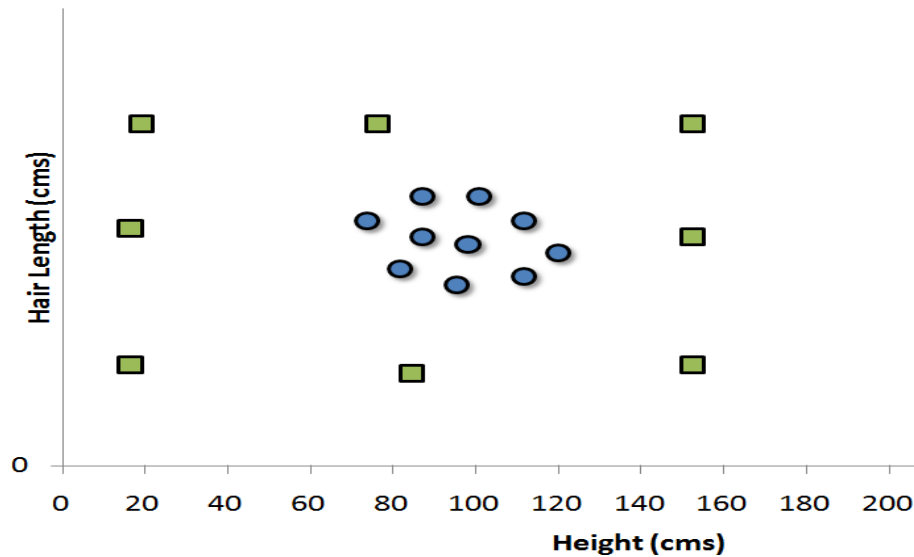


FIG 3.5: DISTRIBUTION

In such cases, we do not see a straight line frontier directly in current plane which can serve as the SVM. In such cases, we need to map these vectors to a higher dimension plane so that they get segregated from each other. Such cases will be covered once we start with the formulation of SVM. For now, you can visualize that such transformation will result into following type of SVM.

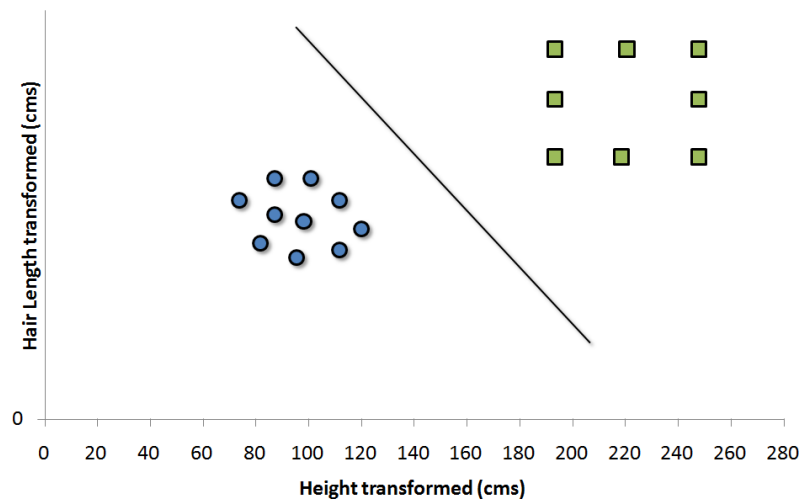


FIG 3.6: TRANSFORMATION

Each of the green squares in original distribution is mapped on a transformed scale. And transformed scale has clearly segregated classes. Many algorithms have been proposed to

make these transformations.

How does it work?

Above, we got accustomed to the process of segregating the two classes with a hyper-plane. Now the burning question is “How can we identify the right hyper-plane?” Don’t worry; it’s not as hard as you think!

Let’s understand:

- **Identify the right hyper-plane (Scenario-1):** Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane to classify star and circle.

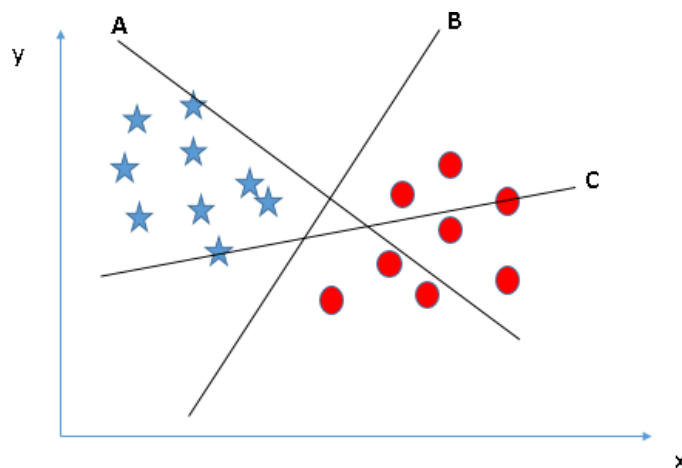


FIG 3.7: SCENARIO-1

You need to remember a thumb rule to identify the right hyper-plane: “Select the hyper-plane which segregates the two classes better”. In this scenario, hyper-plane “B” has excellently performed this job.

- **Identify the right hyper-plane (Scenario-2):** Here, we have three hyper-planes (A, B and C) and all are segregating the classes well. Now, how can we identify the right hyper-plane?

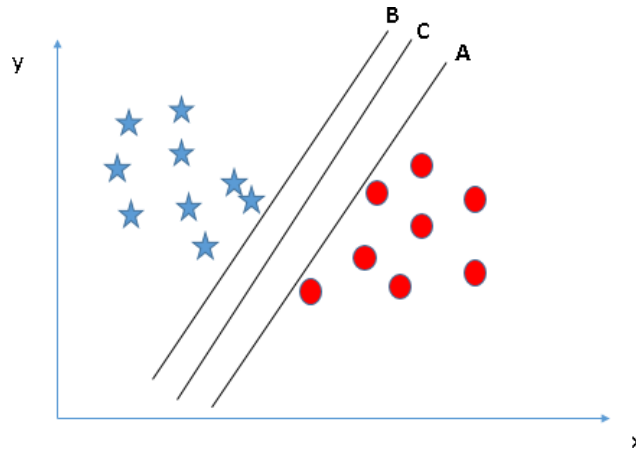


FIG 3.8: SCENARIO-2

Here, maximizing the distances between nearest data points (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as **Margin**. Let's look at the below snapshot:

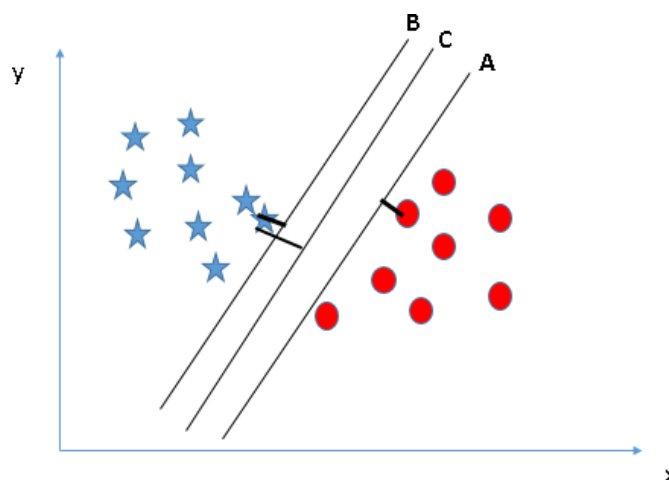


FIG 3.9: MARGIN

Above, you can see that the margin for hyper-plane C is high as compared to both A and B. Hence, we name the right hyper-plane as C. Another lightning reason for selecting the hyper-plane with higher margin is robustness. If we select a hyper-plane having low margin then there is high chance of miss-classification.

- **Identify the right hyper-plane (Scenario-3):**Hint: Use the rules as discussed in previous section to identify the right hyper-plane

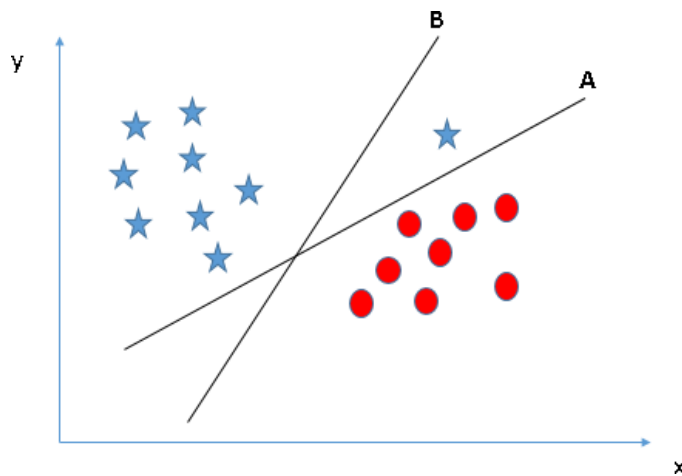


FIG 3.10: SCENARIO-3

Some of you may have selected the hyper-plane **B** as it has higher margin compared to **A**. But, here is the catch; SVM selects the hyper-plane which classifies the classes accurately prior to maximizing margin. Here, hyper-plane B has a classification error and A has classified all correctly. Therefore, the right hyper-plane is **A**.

- **Can we classify two classes (Scenario-4)?:** Below, I am unable to segregate the two classes using a straight line, as one of star lies in the territory of other(circle) class as an outlier.

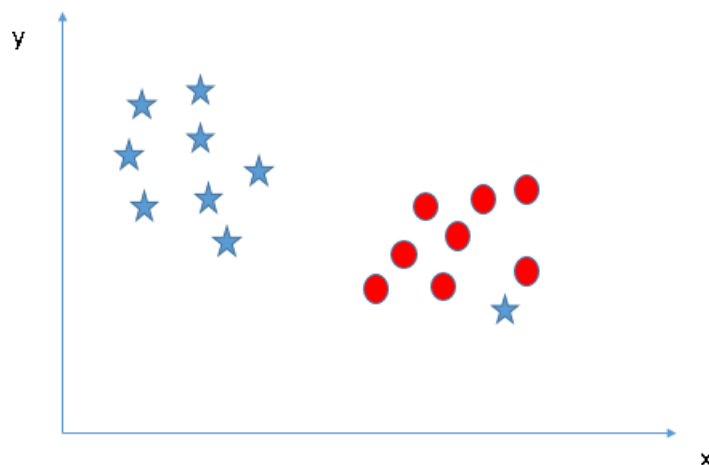


FIG 3.11: SCENARIO-4

As I have already mentioned, one star at other end is like an outlier for star class. SVM has a feature to ignore outliers and find the hyper-plane that has maximum margin. Hence, we can say, SVM is robust to outliers.

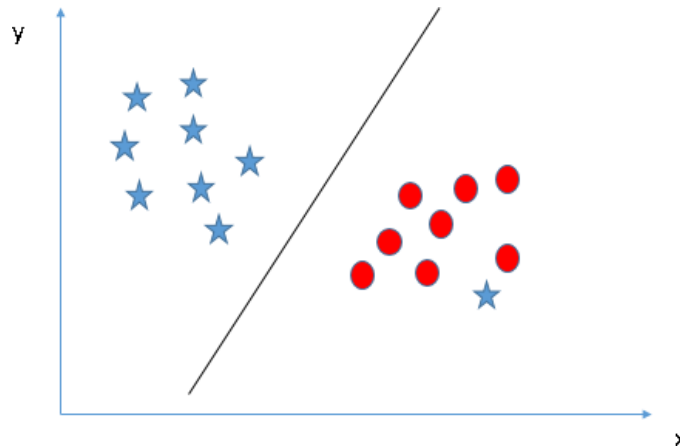


FIG 3.12: CLASSIFY TWO CLASSES

- **Find the hyper-plane to segregate two classes (Scenario-5):** In the scenario below, we can't have linear hyper-plane between the two classes, so how does SVM classify these two classes? Till now, we have only looked at the linear hyper-plane.

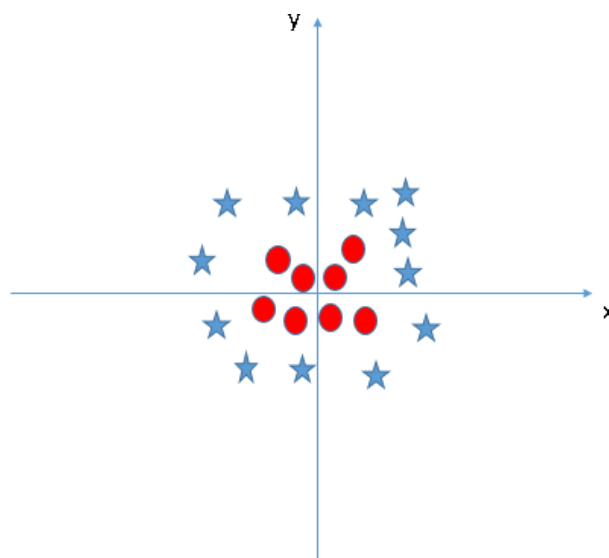


FIG 3.13: SCENARIO-5

SVM can solve this problem. Easily! It solves this problem by introducing additional

feature. Here, we will add a new feature $z=x^2+y^2$. Now, let's plot the data points

on axis x and z:

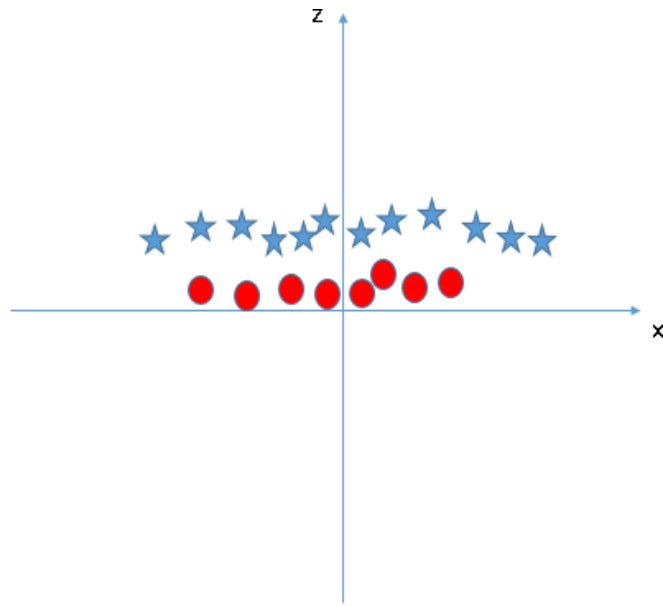


FIG 3.14: SEGREGATE TWO CLASSES

In above plot, points to consider are:

- o All values for z would be positive always because z is the squared sum of both x and y
- o In the original plot, red circles appear close to the origin of x and y axes, leading to lower value of z and star relatively away from the origin result to higher value of z .

In SVM, it is easy to have a linear hyper-plane between these two classes. But, another burning question which arises is, should we need to add this feature manually to have a hyper-plane. No, SVM has a technique called the kernel trick. These are functions which takes low dimensional input space and transform it to a higher dimensional space i.e. it converts not separable problem to separable problem, these functions are called kernels. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then find out the process to separate the data based on the labels or outputs you've defined.

When we look at the hyper-plane in original input space it looks like a circle:

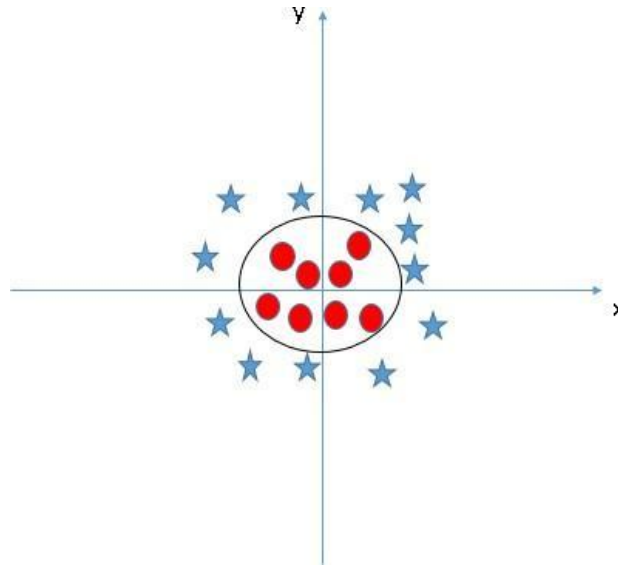


FIG 3.15 HYPER-PLANE

Pros and Cons associated with SVM

- **Pros:**
 - o It works really well with clear margin of separation
 - o It is effective in high dimensional spaces.
 - o It is effective in cases where number of dimensions is greater than the number of samples
 - o It uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.

- **Cons:**
 - o It doesn't perform well, when we have large data set because the required training time is higher
 - o It also doesn't perform very well, when the data set has more noise i.e. target classes are overlapping
 - o SVM doesn't directly provide probability estimates, these are calculated using an expensive five-fold cross-validation. It is related SVC method of Python scikit-learn library

CHAPTER 4

SOFTWARE DESIGN AND IMPLEMENTATION

4.1 DESIGN AND IMPLEMENTATION

CONSTRAINTS Constraints in Analysis

- ◆ Constraints as Informal Text
- ◆ Constraints as Operational Restrictions
- ◆ Constraints Integrated in Existing Model Concepts
- ◆ Constraints as a Separate Concept
- ◆ Constraints Implied by the Model Structure

Constraints in Design

- ◆ Determination of the Involved Classes
- ◆ Determination of the Involved Objects
- ◆ Determination of the Involved Actions
- ◆ Determination of the Require Clauses
- ◆ Global actions and Constraint Realization

Constraints in Implementation

A hierarchical structuring of relations may result in more classes and a more complicated structure to implement. Therefore it is advisable to transform the hierarchical relation structure to a simpler structure such as a classical flat one. It is rather straightforward to transform the developed hierarchical model into a bipartite, flat model, consisting of classes on the one hand and flat relations on the other. Flat relations are preferred at the design level for reasons of simplicity and implementation ease. There is no identity or functionality associated with a flat relation. A flat relation corresponds with the relation concept of entity-relationship modeling and many object oriented methods.

4.2 Other Nonfunctional

Requirements Performance

Requirements

The application at this side controls and communicates with the following three main general components.

embedded browser in charge of the navigation and accessing to the web service;

Server Tier: The server side contains the main parts of the functionality of the proposed architecture. The components at this tier are the following.

Web Server, Security Module, Server-Side Capturing Engine, Preprocessing Engine, Database System, Verification Engine, Output Module.

Safety Requirements

1. The software may be safety-critical. If so, there are issues associated with its integrity level
2. The software may not be safety-critical although it forms part of a safety-critical system. For example, software may simply log transactions.
3. If a system must be of a high integrity level and if the software is shown to be of that integrity level, then the hardware must be at least of the same integrity level.
4. There is little point in producing 'perfect' code in some language if hardware and system software (in widest sense) are not reliable.
5. If a computer system is to run software of a high integrity level then that system should not at the same time accommodate software of a lower integrity level.
6. Systems with different requirements for safety levels must be separated.
7. Otherwise, the highest level of integrity required must be applied to all systems in the same environment.

4.3 Architecture Diagram:

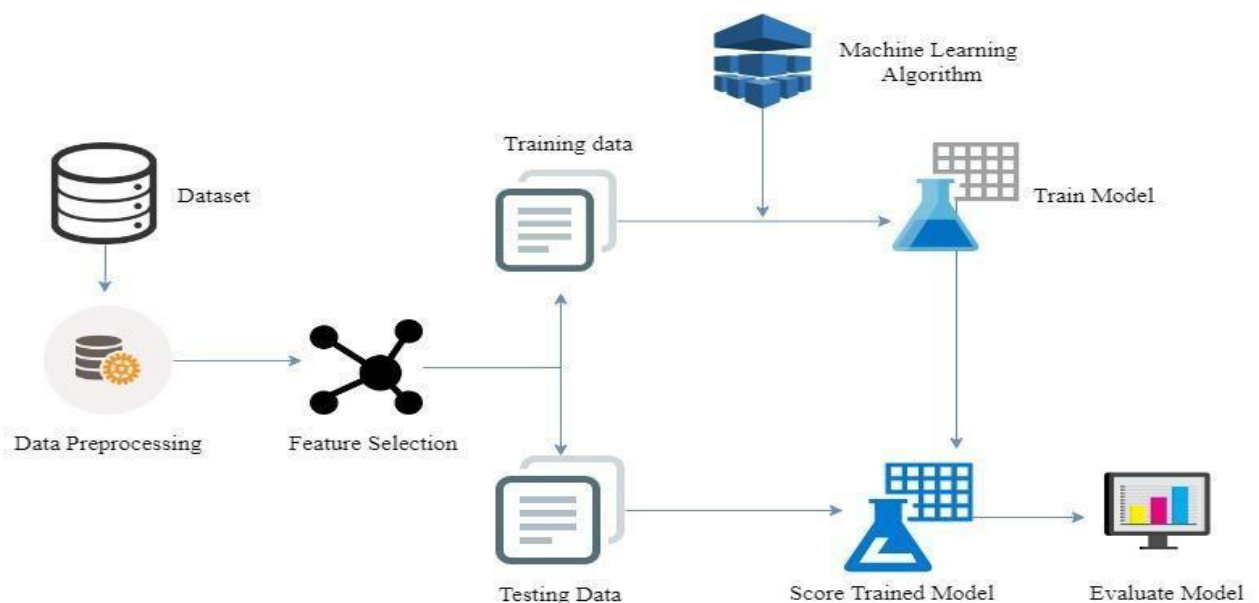


Fig: 4.1 ARCHITECTURE DIAGRAM

4.4 Sequence Diagram:

A Sequence diagram is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of Message Sequence diagrams are sometimes called event diagrams, event sceneries and timing diagram.

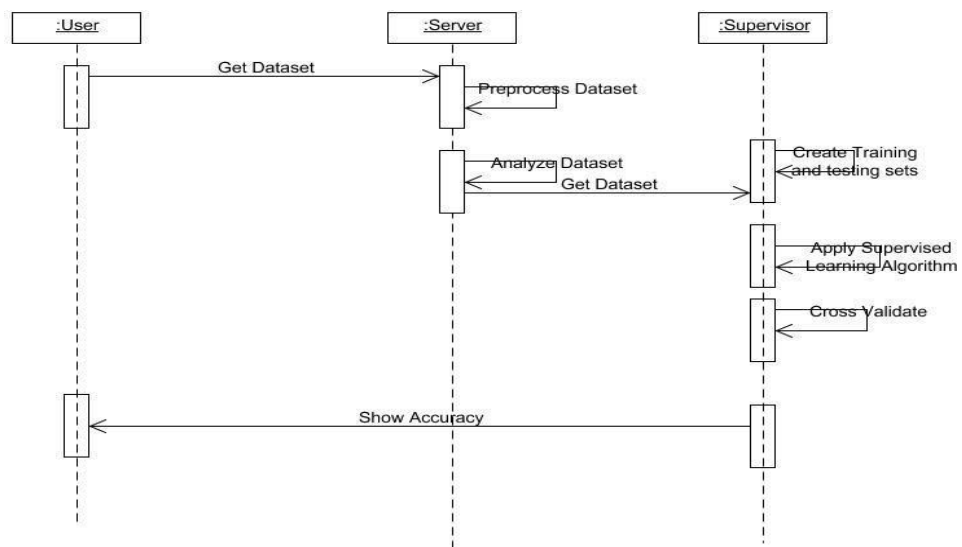


FIG 4.2 SEQUENCE DIAGRAM

4.5 USE CASE DIAGRAM:

Unified Modeling Language (UML) is a standardized general-purpose modeling language in the field of software engineering. The standard is managed and was created by the Object Management Group. UML includes a set of graphic notation techniques to create visual models of software intensive systems. This language is used to specify, visualize, modify, construct and document the artifacts of an object oriented software intensive system under development.

A Use case Diagram is used to present a graphical overview of the functionality provided by a system in terms of actors, their goals and any dependencies between those use cases. Use case diagram consists of two parts:

Use case: A use case describes a sequence of actions that provided something of measurable value to an actor and is drawn as a horizontal ellipse.

Actor: An actor is a person, organization or external system that plays a role in one or

more interaction with the system.

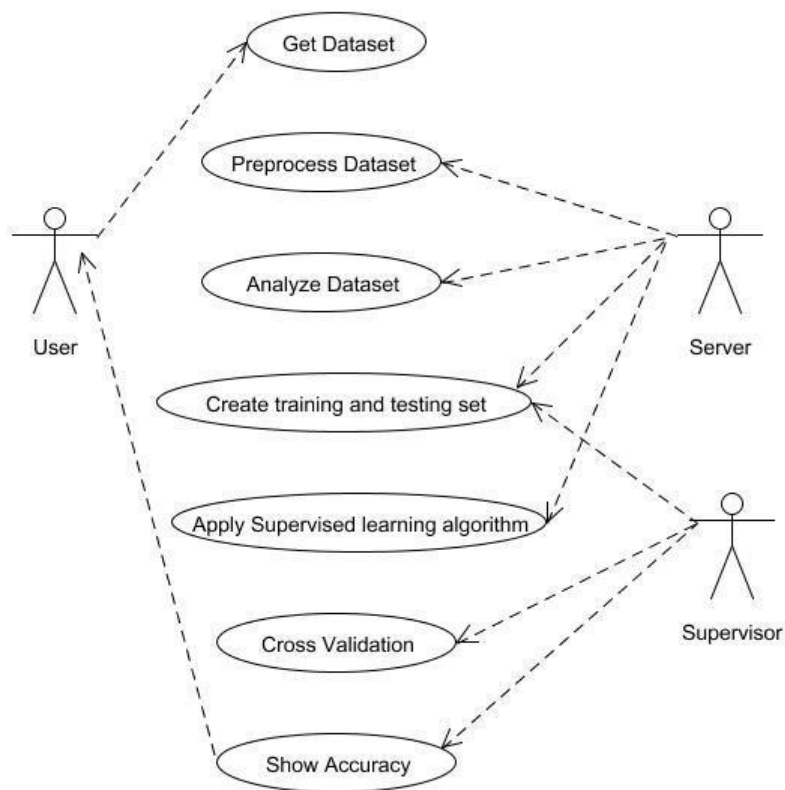


FIG 4.3: USECASE DIAGRAM

4.6 ACTIVITY DIAGRAM:

Activity diagram is a graphical representation of workflows of stepwise activities and actions with support for choice, iteration and concurrency. An activity diagram shows the overall flow of control.

The most important shape types:

- Rounded rectangles represent activities.
- Diamonds represent decisions.
- Bars represent the start or end of concurrent activities.
- A black circle represents the start of the workflow.
- An encircled circle represents the end of the workflow.

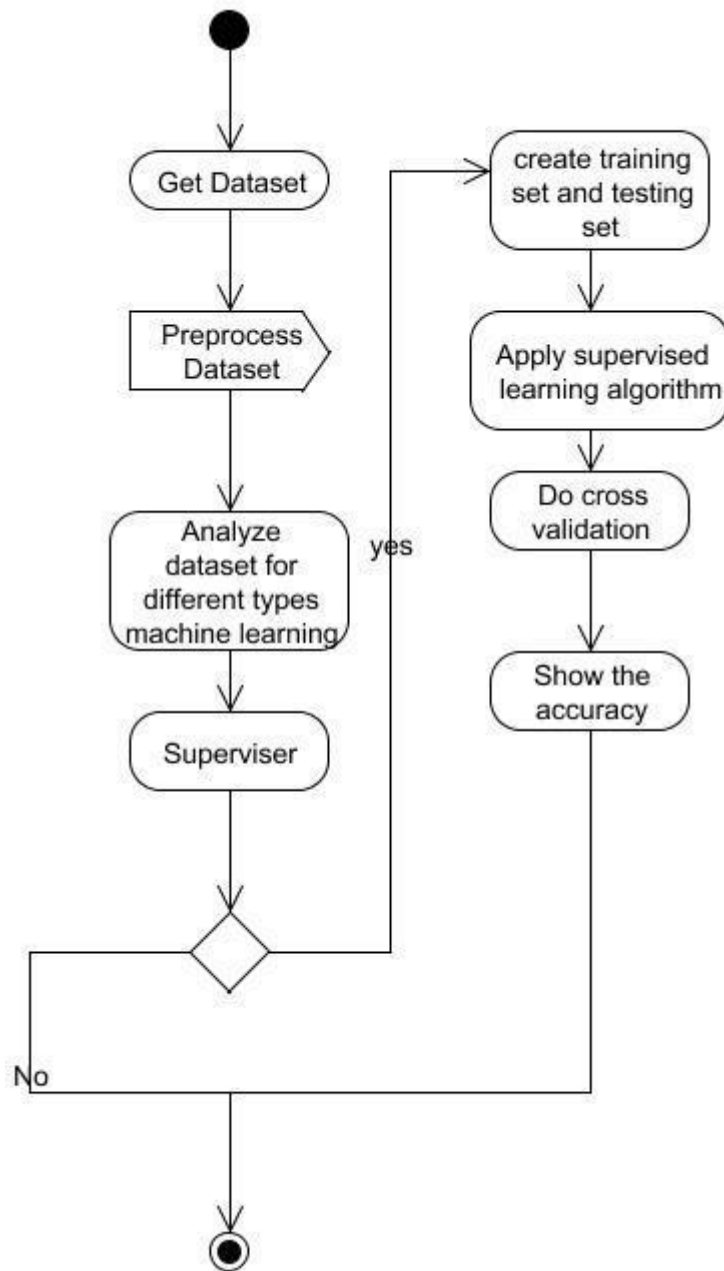


FIG 4.4: ACTIVITY DIAGRAM

4.7COLLABORATION DIAGRAM

UML Collaboration Diagrams illustrate the relationship and interaction between software objects. They require use cases, system operation contracts and domain model to already exist. The collaboration diagram illustrates messages being sent between classes and objects.

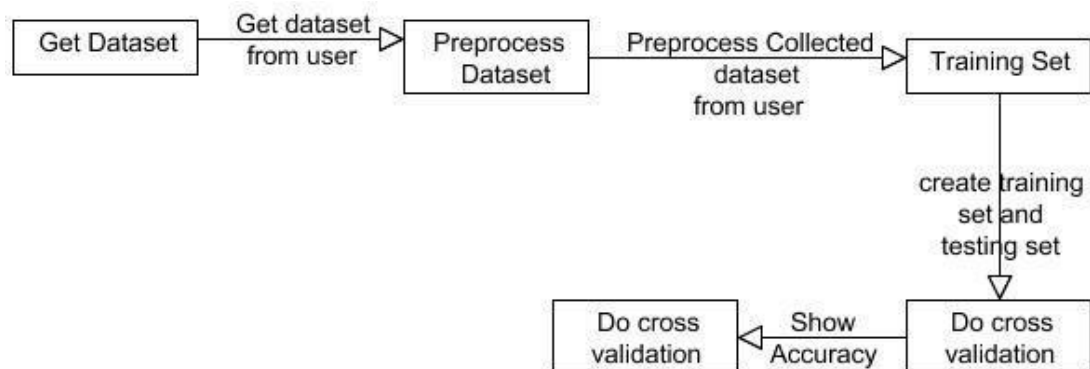


FIG 4.5: COLLABORATION DIAGRAM

4.8 MODULES

- Data Pre-Processing
- Feature Selection
- Classification Modeling
- Performance Measures

Module Description

Dataset Description:

The objective of the dataset is to predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. The datasets consists of several medical predictor variables and one target variable, Outcome. Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test

Blood Pressure: Diastolic blood pressure (mm Hg)

Skin Thickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mu U/ml)

BMI: Body mass index (weight in kg/(height in m)²)

Diabetes Pedigree Function: Diabetes pedigree function

Age: Age (years)

Outcome: Class variable (0 or 1) 268 of 768 are 1, the others are 0

Data Pre-Processing

Diabetes disease data is pre-processed after collection of various records. The dataset contains a total of 769 patient records, where 6 records are with some missing values. Those 6 records have been removed from the dataset and the remaining 763 patient records are used in pre-processing.

Feature Selection

From among the 8 attributes of the data set, one attributes pertaining to age is used to identify the personal information of the patient. The remaining 7 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of diabetes disease.

Classification Modeling

The clustering of datasets is done on the basis of the variables and criteria of Decision Tree (DT) features. Then, the classifiers are applied to each clustered dataset in order to estimate its performance. The best performing models are identified from the above results based on their low rate of error.

- Decision Trees Classifier
- Support Vector Classifier
- Random Forest Classifier
- Logistic Regression
- K Nearest neighbors
- Naive Bayes

Performance Measures:

Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficacy of this model.

Logistic Regression: 71.42857142857143

K Nearest neighbors: 78.57142857142857

Support Vector Classifier: 73.37662337662337

Naive Bayes: 71.42857142857143

Decision tree: 68.18181818181817

Random Forest: 75.97402597402598

4.9 CODING AND

TESTING CODING

Once the design aspect of the system is finalized the system enters into the coding and testing phase. The coding phase brings the actual system into action by converting the design of the system into the code in a given programming language. Therefore, a good coding style has to be taken whenever changes are required it easily screws into the system.

CODING STANDARDS

Coding standards are guidelines to programming that focuses on the physical structure and appearance of the program. They make the code easier to read, understand and maintain. This phase of the system actually implements the blueprint developed during the design phase. The coding specification should be in such a way that any programmer must be able to understand the code and can bring about changes whenever felt necessary. Some of the standards needed to achieve the above-mentioned objectives are as follows:

- Program should be simple, clear and easy to understand.

- Naming conventions

- Value conventions

- Script and comment procedure

- Message box format Exception

- and error handling

NAMING CONVENTIONS

Naming conventions of classes, data member, member functions, procedures etc., should be **self-descriptive**. One should even get the meaning and scope of the variable by its name.

The conventions are adopted for **easy understanding** of the intended message by the user.

So it is customary to follow the conventions. These conventions are as follows: **Class names**

Class names are problem domain equivalence and begin with capital letter and have mixed cases.

Member Function and Data Member name

Member function and data member name begins with a lowercase letter with each subsequent letter of the new words in uppercase and the rest of letters in lowercase.

VALUE CONVENTIONS

Value conventions ensure values for variable at any point of time. This involves the

following:

Proper default values for the variables.

Proper validation of values in the field.

Proper documentation of flag values.

SCRIPT WRITING AND COMMENTING STANDARD

Script writing is an art in which indentation is utmost important. Conditional and looping statements are to be properly aligned to facilitate easy understanding. Comments are included to minimize the number of surprises that could occur when going through the code.

MESSAGE BOX FORMAT

When something has to be prompted to the user, he must be able to understand it properly. To achieve this, a specific format has been adopted in displaying messages to the user. They are as follows:

X – User has performed illegal operation.

! – Information to the user.

TEST PROCEDURE

SYSTEM TESTING

Testing is performed to identify errors. It is used for quality assurance. Testing is an integral part of the entire development and maintenance process. The goal of the testing during phase is to verify that the specification has been accurately and completely incorporated into the design, as well as to ensure the correctness of the design itself. For example the design must not have any logic faults in the design is detected before coding commences, otherwise the cost of fixing the faults will be considerably higher as reflected. Detection of design faults can be achieved by means of inspection as well as walkthrough. Testing is one of the important steps in the software development phase. Testing checks for the errors, as a whole of the project testing involves the following test cases:

Static analysis is used to investigate the structural properties of the Source code.

Dynamic testing is used to investigate the behavior of the source code by executing the program on the test data.

4.10 TEST DATA AND

OUTPUT UNIT TESTING

Unit testing is conducted to verify the functional performance of each modular component

of the software. Unit testing focuses on the smallest unit of the software design (i.e.), the module. The white-box testing techniques were heavily employed for unit testing.

FUNCTIONAL TESTS

Functional test cases involved exercising the code with nominal input values for which the expected results are known, as well as boundary values and special values, such as logically related inputs, files of identical elements, and empty files.

Three types of tests in Functional test:

Performance Test

Stress Test

Structure Test

PERFORMANCE TEST

It determines the amount of execution time spent in various parts of the unit, program throughput, and response time and device utilization by the program unit.

STRESS TEST

Stress Test is those test designed to intentionally break the unit. A Great deal can be learned about the strength and limitations of a program by examining the manner in which a programmer in which a program unit breaks.

STRUCTURED TEST

Structure Tests are concerned with exercising the internal logic of a program and traversing particular execution paths. The way in which White-Box test strategy was employed to ensure that the test cases could Guarantee that all independent paths within a module have been have been exercised at least once.

Exercise all logical decisions on their true or false sides.

Execute all loops at their boundaries and within their operational bounds.

Exercise internal data structures to assure their validity.

Checking attributes for their correctness.

Handling end of file condition, I/O errors, buffer problems and textual errors in output information

INTEGRATION TESTING

Integration testing is a systematic technique for construction the program structure while at the same time conducting tests to uncover errors associated with interfacing. i.e., integration testing is the complete testing of the set of modules which makes up the

product. The objective is to take untested modules and build a program structure tester should identify critical modules. Critical modules should be tested as early as possible. One

approach is to wait until all the units have passed testing, and then combine them and then tested. This approach is evolved from unstructured testing of small programs. Another strategy is to construct the product in increments of tested units. A small set of modules are integrated together and tested, to which another module is added and tested in combination. And so on. The advantages of this approach are that, interface dispenses can be easily found and corrected.

The major error that was faced during the project is linking error. When all the modules are combined the link is not set properly with all support files. Then we checked out for interconnection and the links. Errors are localized to the new module and its intercommunications. The product development can be staged, and modules integrated in as they complete unit testing. Testing is completed when the last module is integrated and tested.

4.11 TESTING TECHNIQUES / TESTING

STRATEGIES TESTING

Testing is a process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an as-yet –undiscovered error. A successful test is one that uncovers an as-yet- undiscovered error. System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. It verifies that the whole set of programs hang together. System testing requires a test consists of several key activities and steps for run program, string, system and is important in adopting a successful new system. This is the last chance to detect and correct errors before the system is installed for user acceptance testing.

The software testing process commences once the program is created and the documentation and related data structures are designed. Software testing is essential for correcting errors. Otherwise the program or the project is not said to be complete. Software testing is the critical element of software quality assurance and represents the ultimate the review of specification design and coding. Testing is the process of executing the program with the intent of finding the error. A good test case design is one that as a probability of finding an yet undiscovered error. A successful test is one that uncovers an yet undiscovered error. Any engineering product can be tested in one of the two ways:

WHITE BOX TESTING

This testing is also called as Glass box testing. In this testing, by knowing the specific

functions that a product has been design to perform test can be conducted that

demonstrate each function is fully operational at the same time searching for errors in each function. It is a test case design method that uses the control structure of the procedural design to derive test cases. Basis path testing is a white box testing.

Basis path testing:

- Flow graph notation
- Cyclometric complexity
- Deriving test cases
- Graph matrices Control

BLACK BOX TESTING

In this testing by knowing the internal operation of a product, test can be conducted to ensure that “all gears mesh”, that is the internal operation performs according to specification and all internal components have been adequately exercised. It fundamentally focuses on the functional requirements of the software.

The steps involved in black box test case design are:

- Graph based testing methods
- Equivalence partitioning
- Boundary value analysis
- Comparison testing

SOFTWARE TESTING STRATEGIES:

A software testing strategy provides a road map for the software developer. Testing is a set activity that can be planned in advance and conducted systematically. For this reason a template for software testing a set of steps into which we can place specific test case design methods should be strategy should have the following characteristics:

Testing begins at the module level and works “outward” toward the integration of the entire computer based system.

Different testing techniques are appropriate at different points in time.

The developer of the software and an independent test group conducts testing.

Testing and Debugging are different activities but debugging must be accommodated in any testing strategy.

INTEGRATION TESTING:

Integration testing is a systematic technique for constructing the program structure while at

the same time conducting tests to uncover errors associated with. Individual modules, which are highly prone to interface errors, should not be assumed to work instantly when

we put them together. The problem of course, is “putting them together”- interfacing. There may be the chances of data lost across on another’s sub functions, when combined may not produce the desired major function; individually acceptable impression may be magnified to unacceptable levels; global data structures can present problems.

PROGRAM TESTING:

The logical and syntax errors have been pointed out by program testing. A syntax error is an error in a program statement that in violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted keywords are common syntax error. These errors are shown through error messages generated by the computer. A logic error on the other hand deals with the incorrect data fields, out-off-range items and invalid combinations. Since the compiler s will not deduct logical error, the programmer must examine the output. Condition testing exercises the logical conditions contained in a module. The possible types of elements in a condition include a Boolean operator, Boolean variable, a pair of Boolean parentheses A relational operator or on arithmetic expression. Condition testing method focuses on testing each condition in the program the purpose of condition test is to deduct not only errors in the condition of a program but also other a errors in the program.

SECURITY TESTING:

Security testing attempts to verify the protection mechanisms built in to a system well, in fact, protect it from improper penetration. The system security must be tested for invulnerability from frontal attack must also be tested for invulnerability from rear attack. During security, the tester places the role of individual who desires to penetrate system.

VALIDATION TESTING

At the culmination of integration testing, software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of software test-validation testing begins. Validation testing can be defined in many ways, but a simple definition is that validation succeeds when the software functions in manner that is reasonably expected by the customer. Software validation is achieved through a series of black box tests that demonstrate conformity with requirement. After validation test has been conducted, one of two conditions exists.

- * The function or performance characteristics confirm to specifications and are accepted.
- * A validation from specification is uncovered and a deficiency created.

Deviation or errors discovered at this step in this project is corrected prior to completion of

the project with the help of the user by negotiating to establish a method for resolving

deficiencies. Thus the proposed system under consideration has been tested by using validation testing and found to be working satisfactorily. Though there were deficiencies in the system they were not catastrophic

USER ACCEPTANCE TESTING

User acceptance of the system is key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system and user at the time of developing and making changes whenever required. This is done in regarding to the following points.

- Input screen design.
- Output screen design.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 RESULT

In this chapter we will discuss about the results of our system.

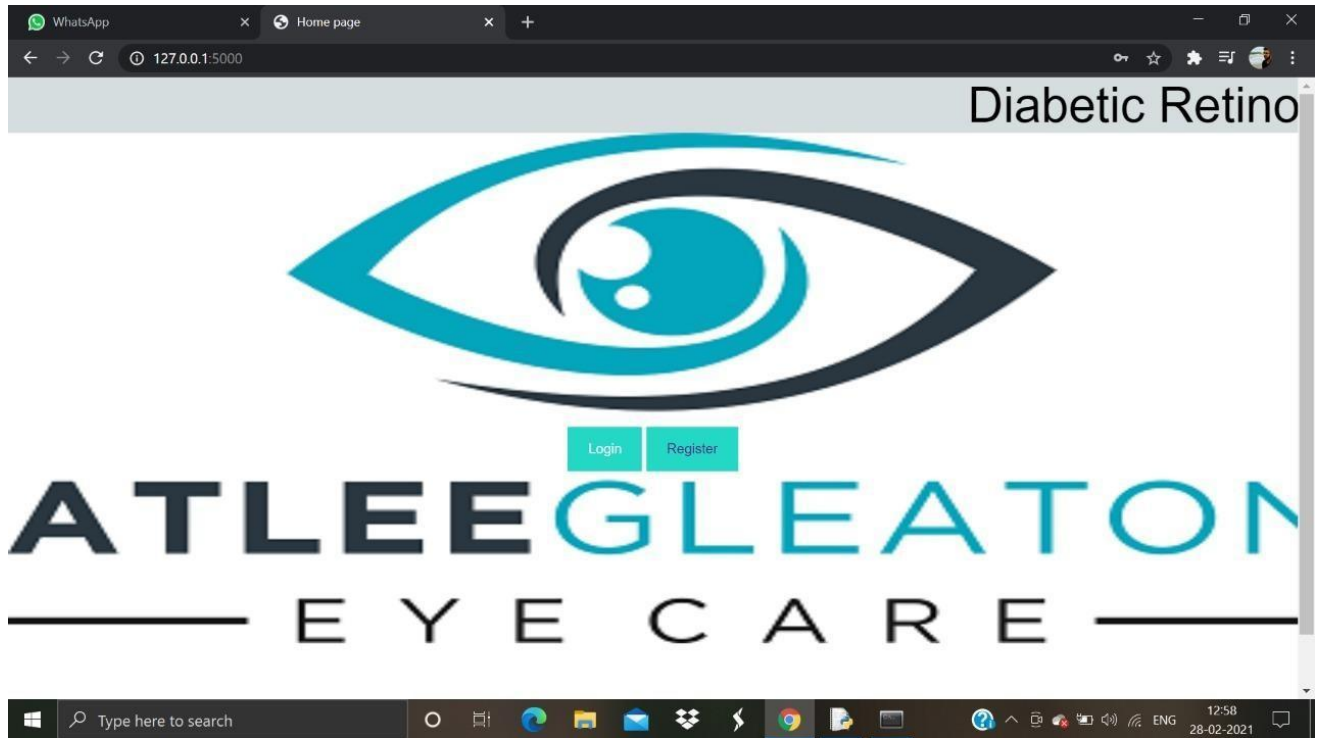


Fig 5.1: WELCOME PAGE

The Fig 5.1 is the welcome page of the project. This page will appear once the user or the admin open's the portal. It is the frontend of the project or it can also be termed as user interface. Here the user the gets the multiple options to login or register as per need.

Register

Please fill in this form to create an account.

Name

SHAIK FARUKHE AZAM

Email

farukh.sk7869@gmail.com

Password

By creating an account you agree to our [Terms & Privacy](#).

Register

Fig 5.2: REGISTRATION PAGE

The Fig 5.2 shows that the user has to register first if he or she is new in this website. After successful registration, they get their log-in credentials which are provided by the admin. They use the login credentials to log-in in the website.

Username

farukh

Password

Login

☒ Remember me

Cancel

[Forgot password?](#)

Fig 5.3: LOG-IN PAGE

The Fig 5.3 is the log-in page which generally is the entry or authentication of every secured portal. This portal can be logged in by both admin and the user. The admin can log-in to check the working of the portal and to maintain the database. The user log-in to access the portal and the admin provides the required security to the users so that the data does not get deleted or leaked. This section checks the authentication and if the log in credentials is correct then the user can log in. It is the frontend of the project or it can also be termed as user interface. Here the user gets the multiple options to execute or access their task as per need.

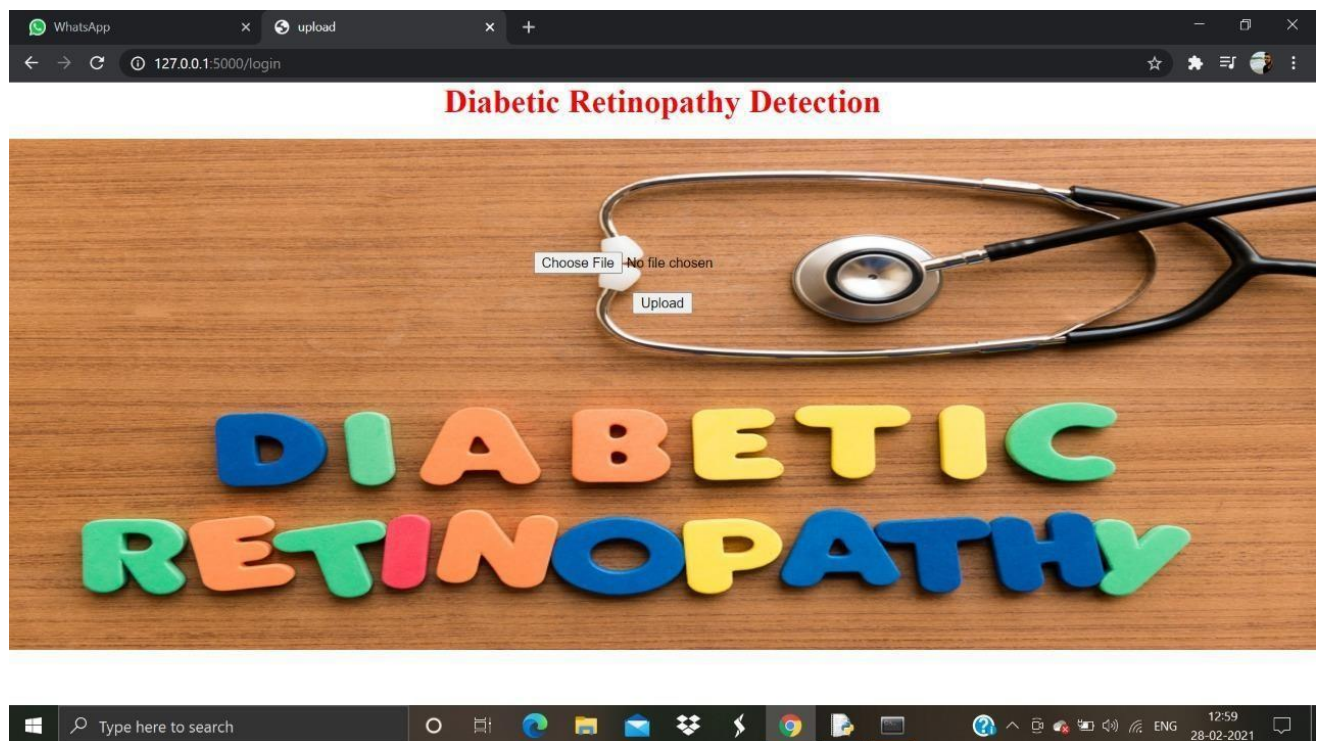


Fig 5.4: HOME PAGE

The Fig 5.4 is the home page. This page is displayed after successful log-in. Here the user gets the multiple options to choose a file and upload the file. If the user does not choose a file, it shows that no file has chosen.

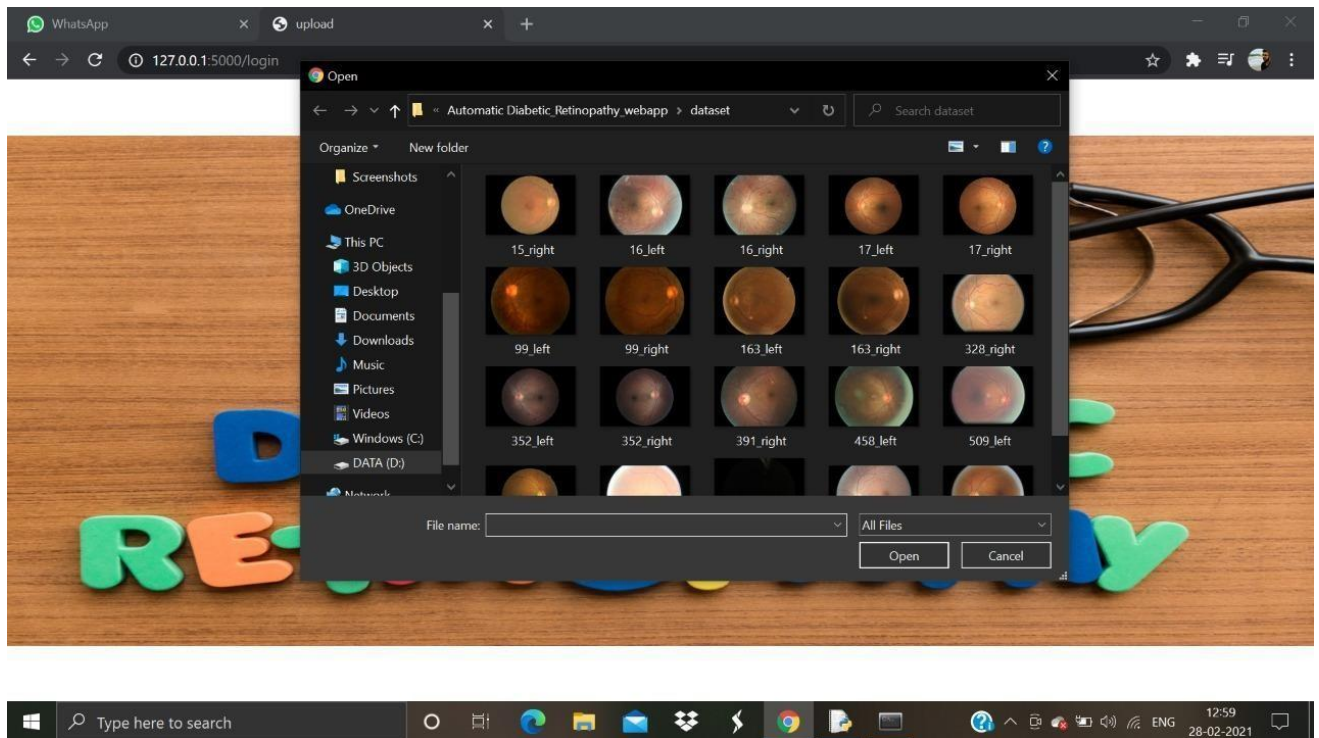


Fig 5.5: FILE UPLOAD PAGE

The Fig 5.5 is the file upload page. Here the user has to choose the file which he want to upload. In this page the user can access the image files from the data set. The user can view the images and upload the images. The admin can log-in to check the working of the portal and to maintain the database.

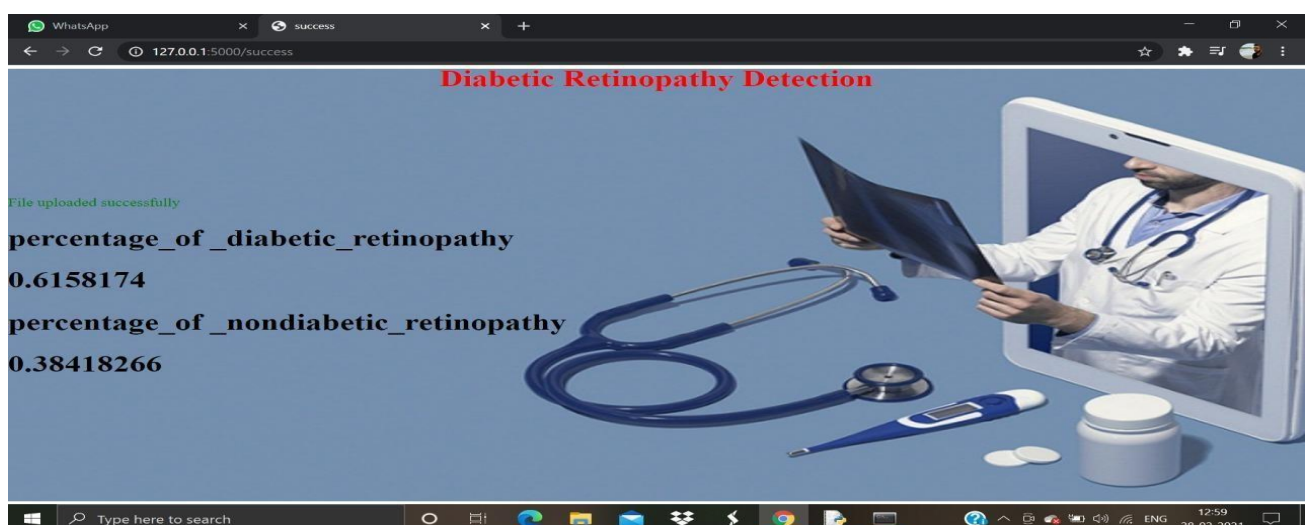


Fig 5.6: RESULT PAGE

The Fig 5.6 is the result page of our project. This page is displayed after the image file is uploaded successfully. This page shows the percentage of Diabetic Retinopathy and percentage of non-diabetic retinopathy.

If the percentage of diabetic retinopathy is higher than the percentage of non-diabetic retinopathy, then it concludes that the person is having diabetic retinopathy.

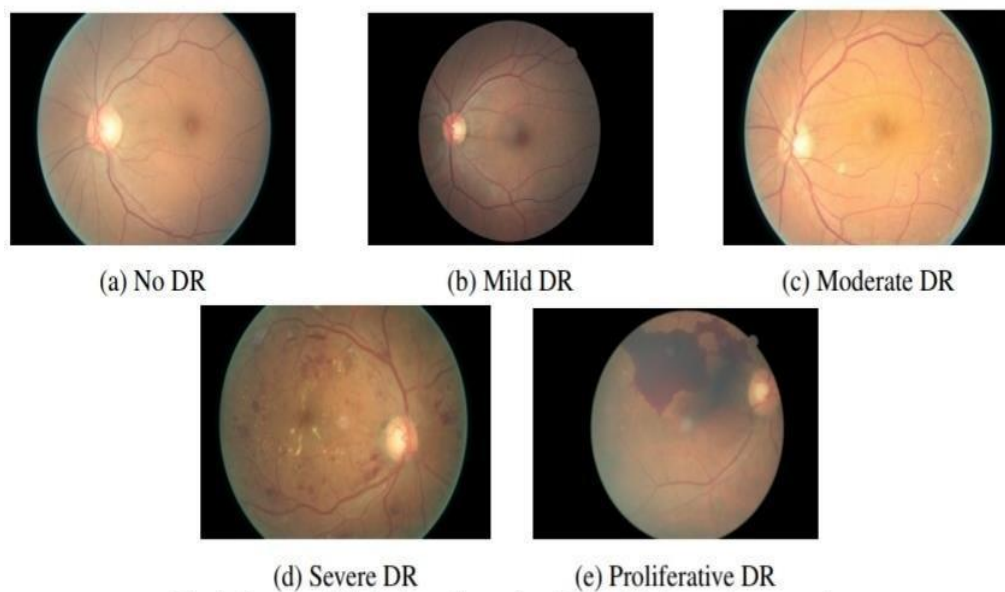


Fig 1: Stages of diabetic retinopathy (DR) with increasing severity

Fig 5.7: STAGE OF DIABETIC RETINOPHY

The Fig 5.7 shows the stages of diabetic retinopathy with increasing severity.

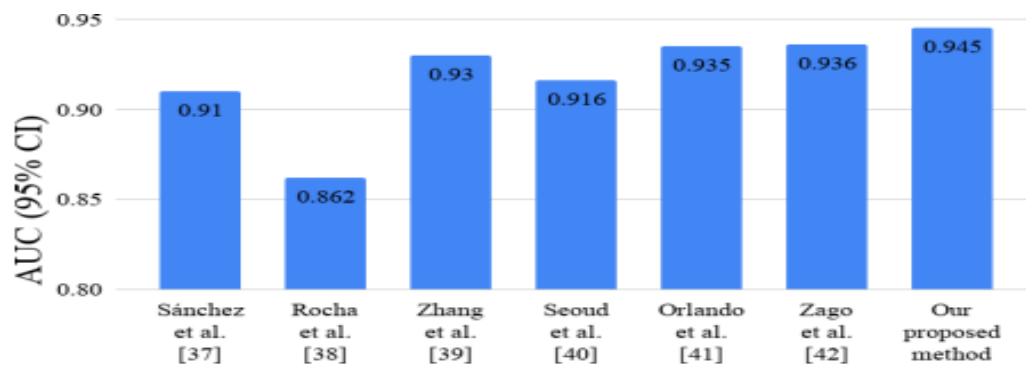


FIGURE 2. Comparison of DR results using the Messidor dataset.

Fig 5.8: Comparison of DR results

The comparison gained results with the state-of-the-art models on the Messidor dataset in fig 5.8 shows that our AUC value is the highest compared to recent work.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

This system will make the process of diabetic retinopathy detection more efficient and flexible for the users. The users no need to worry about the security and the confidentiality of the images submitted. And the user need not worry about the information to be stored because it is all maintained in a database and can be edited or retrieved whenever needed. The portal is secured, monitored and maintained by the admin time to time. Our model accepts to fundus images corresponding to the left eye and right eye as inputs and then transmits them into the Siamese – like blocks. The information from two eyes is gathered into the fully-connected layer and finally the model will output the diagnosis result of each eye respectively. The system consists of four functional modules: Image Pre-Processing, Inception, Convolution Neural network algorithm and Matching Score.

6.2 FUTURE WORK

This project summarizes the state of the art in diabetic retinopathy research and provides a perspective on opportunities for future investigations. New insights into the pathophysiology of diabetes and diabetic retinopathy will improve metabolic control. Structure-function analyses are revealing new details of diabetic retinopathy. Intraocular drug therapy provides improved visual outcomes. Together these steps will yield better means to detect and quantify vision loss, and to develop patient-specific treatments to preserve vision for persons with diabetes.

REFERENCES

- [1] XialongZeng,Haiquan Chen, Yuan Lao, Wenbin Ye,” Automated Diabetic Retinopathy Detection on Binocular Vol 09 Issue01, Jan 2020 ISSN 2456 – 5083 Page 8 Siamese-Like Convolution Neural Network” ,February 2019.
- [2] S. R. Nirmala, M. K. Nath, and S. Dandapat, “ Retinal Image Analysis: A Review,” International Journal of Computer & Communication Technology (IJCCT), vol-2, pp. 11- 15, 2011.
- [3] National Eye Institute, National Institutes of Health, “ Diabetic Retinopathy: What you should know,” Booklet, NIH Publication, no: 06-2171, 2003.
- [4] A. D. Fleming, K. A. Goatman, and J. A. Olson, “ The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy,” British Journal of Ophthalmology, vol 94, no. 6, pp. 706- 711, 2010.
- [5] Shijian Lu, et al., “ Automatic Fundus Image Classification for Computer-Aided Diagnosis” , 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009, pp.1453-1456.
- [6] Gang Luo, OpasChutatape, Huiqi Lei, Shankar. M. Krishnan, “Abnormality Detection in Automated Mass Screening System of Diabetic Retinopathy”, available online at “http://luoinnovation.tech.officelive.com/luo_research/cbms2001.pdf” (accessed on March 3, 2013).
- [7] Keerthi Ram, Gopal Datt Joshi, and JayanthiSivaswam, “A Successive ClutterRejectionBased Approach for Early Detection of Diabetic Retinopathy” , IEEE Transactions On Biomedical Engineering, Vol. 58, No. 3, March 2011, pp.664-673.
- [8] Ahmad Fadzil, et. Al, “ Analysis of Foveal Avascular Zone in Colour Fundus Images for Grading of Diabetic Retinopathy Severity” , 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010, pp. 5632-5635.

- [9] R. Sivakumar, G. Ravindran, M. Muthayya, S. Lakshminarayanan, and C. U. Velmurughendran, “Diabetic Retinopathy Analysis” , Journal of Biomedicine

and Biotechnology, 2005 pp. 20– 27. Z.

- [10] Omar, M. Hanafi, S. Mashohor, N. Mahfudz, and M. Muna'im, "Automaticdiabetic retinopathy detection and classification system," in 2017 7th IEEE International Conference on System Engineering and Technology (ICSET), pp. 162– 166, IEEE, 2017
- [11] Y. Haniza, A. Hamzah, and M. Norrima, "Edge sharpening for diabetic retinopathy detection," IEEE Conference Publications, 2010

APPENDIX

A. Paper work

DIABETIC RETINOPATHY DETECTION USING DEEP LEARNING TECHNIQUE

Submissiondate:20-Mar-202108:45AM(UTC-0700)

SubmissionID:1537788652

Filename:DIABETIC RETINOPATHY DETECTION USING DEEP LEARNING TECHNIQUES

Wordcount:2725

Charactercount:15524

DIABETIC RETINOPATHY DETECTION USING DEEP LEARNING TECHNIQUES

SHAIK SAMEER [1] , SHAIK FARUKHE AZAM [2] , SIVASANGARI A[3]

[1][2] UG Student, Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, India

[3] Associate professor, Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, India

ABSTRACT: Diabetic retinopathy (DR) is a common complication of diabetes associated with retinal vascular damage caused by long standing diabetes. Furthermore, the diagnosis of DR mostly depends on the observation and evaluation to fundus photographs of which procedure can be time-consuming even for experienced experts. Along these lines PC supported mechanized finding methods take incredible potential in clinical to precisely identify DR in a brief timeframe which can additionally assist with improving the screening pace of DR and lessen the quantity of visual impairment. The main parts of proposed method that ought to be centered around are informational collection, network engineering and preparing strategy. Prior to being utilized to prepare our model, fundus pictures informational collection got from public assets is preprocessed and increased. Our system recognizes 2 fundus fig comparing to one side eye and right eye as information sources and afterward communicates them into the Siamese like squares. The data from two eyes is accumulated into the completely associated layer and finally the model will yield the conclusion consequence of each eye separately.

Keywords: diabetic retinopathy, deep learning, eye, images

1 INTRODUCTION

Diabetic retinopathy or DR is a very severe retinal disease. This is the major cause of sightlessness in the working population of the developed nations. According to a survey, more than 93 million people all over the world are suffering from this disease. This disease makes its prey to the diabetic people. The retina as a

membrane covers the backside of the eye. This part is extremely sensitive to light. The retina does the conversion of light into signals. The human brain can interpret these signals. DR disease ruins the blood vessels in the retinal tissue. These vessels start leaking fluid and result in visions loss. There are mainly two categories of diabetic retinopathy disease. These categories are known as PDR (Proliferative diabetic retinopathy) and NPDR (Non-proliferative diabetic retinopathy) [1]. The first category refers to the most advanced DR stage. In PDR, novel, anomalous blood vessels generate in the retinal tissue. On the other hand, NPDR is the milder stage and generally does not show any symptoms. The use of machine learning algorithms is very common for diabetic retinopathy detection. A lot of algorithms have been presented by different researchers over the time. The general process of DR detection using machine learning involves image pre-processing, the classification and FE-feature extraction. The first step known as image preprocessing makes use of various image enhancement and restoration methods for the normalization of retinal funds images. In this step, the image quality gets improved. The use of different filtering techniques in this step reduces the blocking artifacts in the retinal fundus image. In this step, the conversion of RGB retinal image is carried out into many channel images. These images include gray scale, R, G and B component images. The main aim here is to separate objects from the background [2]. The use of Clahe approach enhances the contrast of local regions in gray scale images. The next step in DR detection is feature extraction. This step involves three tasks i.e. optic disc elimination, blood vessels extraction &

removal and detection of exudates and micro-aneurysms. The OD (optic disc) is a round region in the backside of the eye. In this region, retinal nerve fibers collectively make the optic nerve. Exudates contain high and similar brightness levels of OD. Hence, the removal of OD from the image of retina is essential. Region properties and region detection approaches are generally used masking and removing the high intensity OD. The use of canny edge detection is quite popular for detecting the counter. This approach preserves all local maxima and thereby improves the blurred images. The second task in feature extraction is the extraction and removal of blood vessels. In order to detect microaneurysms and exudates, it is necessary to remove blood vessels and OD from the retina image [3]. Blood vessels have same concentration level as micro-aneurysms while the concentration level of OD is similar to exudates. The blood vessels with high contrast level can be removed by applying Dilation operation on the intensity image. Then, the structuring element is used to fill the small image holes via dilation operation. Structuring elements come in different shapes such as diamond, disc, round and so on. However, the use of flat disc shaped structure is quite common for the removal of OD (optic disc) and blood vessels.

The exudates features are detected after the removal of blood vessels and optic disc from the retinal image. Exudates represent a bright lesion of retinal image. Morphological closing operation is generally used for the detection of this feature. The implementation of this closing operation is carried out on the eroded fundus image. Micro-aneurysms are other important features of diabetic retinopathy. In order to detect micro-aneurysms features, opening morphological operation are applied [4].

The last assignment in highlight extraction is the identification of exudates and miniature aneurysms. The exudates highlights

are identified after the expulsion of veins and optic circle from the retinal picture. Exudates address a splendid injury of retinal picture. Morphological shutting activity is for the most part utilized for the discovery of this component. The execution of this end activity is done on the dissolved fundus picture. Miniature aneurysms are other significant highlights of diabetic retinopathy. To identify miniature aneurysms highlights, opening morphological activity are applied [4]. In this activity, expansion follows disintegration while the miniature aneurysms show up as red spot portraying growing in retina. The extraction of highlights is completed after the recognition of exudates and microaneurysms in the shading fundus pictures. All separated are highlights are conveyed to various characterization calculations subsequent to processing. The majority of works above either depend on the factors physically estimated by specialists or put a lot of exertion into extricating handmade highlights with picture preparing approaches which bring additional intricacy and unstability. Along these lines, the profound learning technique with the capacity to gain huge highlights straightforwardly from the fundus photography has excited the consideration of scientists lately.

They embraced the CNN structure with an outfit taking in technique from the arrangement positioned second in the DR-Kaggle rivalry and got a decent discovery result with the zone Az under the FROC of 0.954 in the Kaggle dataset. Profited by the huge preparing information and all around filtered master reviewing to the fundus pictures, the work accomplished an intriguing execution with the territory under the beneficiary working bend (AUC) of 0.991/0.990 and affectability of 97.5%/96.1% on two distinctive test sets separately. [5] this work, rather than receiving fundus fig of single eye as info like most past works did we assembled a novel Siamese-like CNN model with weight-sharing layers dependent on Inception V3, which can acknowledge fundus pictures of the two eyes as

information sources and yields the grouping aftereffect of each eye simultaneously.

II RELATED WORK

Xiang Zeng et al.[1] have proposed, cnn and set up with an exchange learning framework. Here basically homogeneous pictures of binocular photographs of the fundus are dealt with, the framework simply predicts the closeness of DR burdens by seeing fundus pictures of both the eye of a solitary individual. The paper doesn't give location of the issues.

Goatman et al.[4] will when everything is said in done portray a changed openness of scaled downscale aneurysms in disguising fundus pictures, which expect a vital activity in PC helped evaluation of diabetic retinopathy, a veritable and standard eye languishing. The assessment can be removed into four phases. The basic advancement contains picture overhaul, cover rectification, and picture association of the green channel. An information base of 21 remarked on pictures has been used to set up the check. The figuring was showed up distinctively corresponding to genuinely g₁₁ researching of 94 pictures; affectability was 88.5% at an ordinary number of 2.13 phony positives per image[5].

Lu Shij et al. [6] explained the strategy for get-together fundus pictures. Here technique the fundus pictures changed over into a segment vector subject to histograms go pictures at various targets. By at that point, fundus pictures were gathered by getting from fuse vectors of an immense number of regular and strange preparing fundus pictures. The exactness of the framework is 96%. *Fleming et al. [4]* showed a technique of irregularity discovery from covering fundus pictures that usage object-based disguising capability. They demand the assortment from the norm into microaneurysms, channel, exudates, and so on as spot class, unprecedented veins, and inconsistent sound framework assessment.

Ram et al. [8] describes that procedures to see microaneurysms using a major breaking point from a preprocessed picture. By then, the insinuated and as generally as conceivable happening issue objects are exculpated. At long last, sort out, the contenders are given out on their similitude to authentic microaneurysms.

H Sivakumar et al. [10] depict a phony neural organization based technique to design diabetic retinopathy. They have recognized three-layer feed-forward back causing neural organization. In the masterminding time period, they used 6 worker ranch centers, 6 covered center centers, and 4 yield place centers. The 12th yield places wander from standard, diabetic Retinopathy, pre proliferative diabetic retinopathy, proliferative diabetic retinopathy.

III EXISTING SYSTEM

The patient influenced by Diabetic Retinopathy may not experience visual disability until the infection has advanced to a serious stage, when the treatment is less viable. Subsequently the early recognition and the standard subsequent meet-ups is important to treat diabetic retinopathy. Early conclusion is troublesome numerous people since they don't know about DR particularly in the 3rd beginning phases. As of now, the finding of DR is a sluggish and awkward cycle.

Moreover, the determination requires specific clinical visit. To beat this issue the accompanying calculations used to identify the diabetic retinopathy however shockingly it doesn't give the precision Numerous calculations are utilized for programmed diabetic location. neural network has 28 convolutional layers and after each layer there is batchnorm and ReLU nonlinear capacity aside from at the last layer. The yield from last layer is a class mark either DR or no DR. The last exactness of the model is 73.3%. The convolutional neural network calculations (CNN) utilized for diabetic retinopathy on the informational collection of 80,000 pictures utilized in this calculation and accomplishes an

affectability of 95% and an exactness of 75% on 5,000 approval pictures.

4 IV PROPOSED SYSTEM

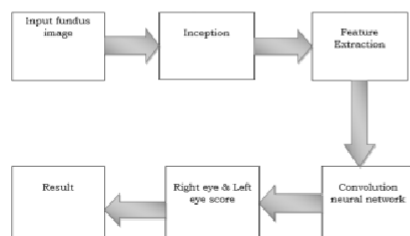


FIG 1 OVERVIEW OF THE PROPOSED SYSTEM

In the proposed system Designed for significant learning model. Principle parts that should be based on are instructive list, network plan and getting ready strategy. Just setup the model fundus pictures enlightening file got from public resources is preprocessed and amplified. The model recognizes two fundus pictures contrasting with one side & right-side eyes as data sources and subsequently conveys them into the Siamese-like squares. These kind of data is collected into the totally related coating and so that the model determination yield assurance outcome of each eye independently.

V MODULES DESCRIPTION

5.1 IMAGE PRE-PROCESSING

Dataset has fundamental part that ought to be overseen for the learning system. The dataset has immense assortment, as dis⁵epant magnificence or objective subsequently most of them are procured with different equipment in the different atmosphere. Basically, the model recognizes two fundus pictures contrasting with eyes as information sources and subsequently sends these details to Siamese like squares.

5.2 INCEPTION

The outstanding introduction of Inception is benefited by a couple of organization affiliation techniques, for instance, accepting cluster normalization, using MLP conv layers to replace direct convolution layers & factorizing huge pet lodging size. With these strategies, the amount of the organization limits similarly like computational cc-complexity has diminished.

5.3 CONVOLUTION NEURAL NETWORK ALGORITHM

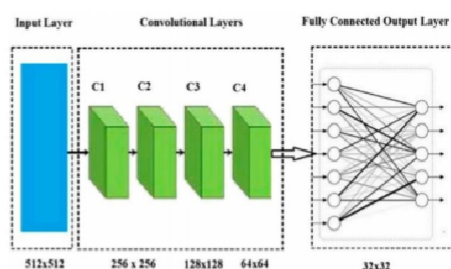


FIG 2 CNN MODEL

Move learning strategy is a generally utilized preparing technique for convolution neural network. By stacking the loads of Inception blocks pre-prepared on Image Net informational index, the model will has a superior loads instatement prior to beginning the inclination streamlining. In addition, considering the immense distinction between the fundus pictures informational index and Image Net dataset.

1 5.4 THE MATCHING SCORE

Disarray grids of forecast aftereffects of left eye, right eye, and the two eyes together. The forecast consequences of the left eye and the correct eye have fundamentally the same as appropriation designs, showing that the information parcel technique saves the first picture classes conveyance of left eyes and right eyes.

VI PERFORMANCE ANALYSIS

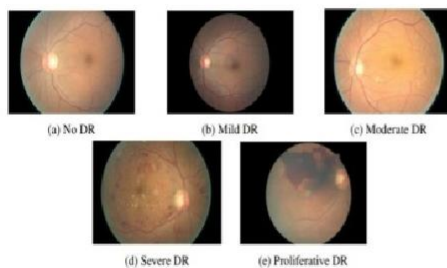


Fig 3 Stages Of Diabetic Retinopathy(Dr) With Increasing Severity

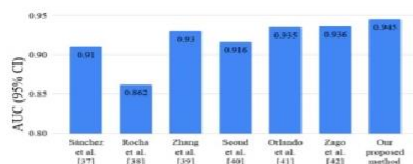


Fig 4 comparison of DR results using the messior dataset

Fig 4 shows the comparison gained results with the state-of-the-art models on the Messidor dataset in Figure2show that our AUC value is the highest compared to recent works.

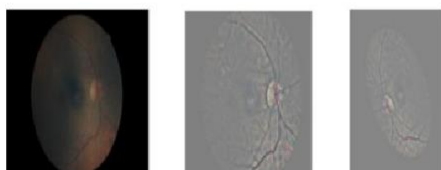


Fig 5 Illusion of the preprocessing and augmentation process

Operator	Resolution	# Channels	# Layers
Conv3 × 3	224 × 224	32	1
MBCConv1, k3 × 3	112 × 112	16	1
MBCConv6, k3 × 3	112 × 112	24	2
MBCConv6, k5 × 5	56 × 56	40	2
MBCConv6, k3 × 3	28 × 28	80	3
MBCConv6, k5 × 5	14 × 14	112	3
MBCConv6, k5 × 5	14 × 14	192	4
MBCConv6, k3 × 3	7 × 7	320	1
Conv1x1 & Pooling & FC	7 × 7	1280	1

Table 1 The efficient -BO baseline network architecture

Model	resolution
EfficientNet-B0 [2]	224 × 224
EfficientNet-B1 [2]	240 × 240
EfficientNet-B2 [2]	260 × 260
EfficientNet-B3 [2]	300 × 300
EfficientNet-B4 [2]	380 × 380
EfficientNet-B5 [2]	456 × 456
EfficientNet-B6 [2]	528 × 528
EfficientNet-B7 [2]	600 × 600

Table 2 The value of the resolution on efficient networks

VI CONCLUSION

This paper presents an investigation of different techniques for screening retinopathy and gathering its reality level. In once-over, this examination brings an overall audit of the occasion and improvement of diabetic retinopathy circulated throughout the a few years in various journals and gatherings. In our examination, we found that even a lot of made nations need data on the development of diabetic retinopathy. In like manner, data on the recurrence of diabetic retinopathy in sort 1 diabetes is insufficient. Our assessment recommends that more through and through first rate ponders subject to data characterized by sex, age, and earnestness of disease are fundamental to summarize the verification base. The adequacy of modified diabetic retinopathy discovery and characterization will profitably decrease ophthalmologist's excess weight.

REFERENCES

- [1] Xialong Zeng,Haiquan Chen, Yuan Lao, Wenbin Ye," Automated Diabetic Retinopathy Detection on Binocular Vol 09 Issue01, Jan 2020 ISSN 2456 – 5083 Page 8 Siamese-Like Convolution Neural Network" ,February 2019.

[2] S. R. Nirmala, M. K. Nath, and S. Dandapat, "Retinal Image Analysis: A Review," *International Journal of Computer & Communication Technology (IJCCT)*, vol-2, pp. 11- 15, 2011.

[3] National Eye Institute, National Institutes of Health, "Diabetic Retinopathy: What you should know," Booklet, NIH Publication, no: 06-2171, 2003.

[4] A. D. Fleming, K. A. Goatman, and J. A. Olson, "The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy," *British Journal of Ophthalmology*, vol 94, no. 6, pp. 706- 711, 2010.

[5] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp.

[6] Shijian Lu, et al., "Automatic Fundus Image Classification for Computer-Aided Diagnosis", 31st Annual International Conference of the IEEE EMBS Minneapolis, Minnesota, USA, September 2-6, 2009, pp.1453-1456.

[7] Gang Luo, Opas Chutatape, Huiqi Lei, Shankar. M. Krishnan, "Abnormality Detection in Automated Mass Screening System of Diabetic Retinopathy", available online at "http://luoinnovation.tech.officelive.com/luo_research/cbms2001.pdf" (accessed on March 3, 2013).

[8] Keerthi Ram, Gopal Datt Joshi, and Jayanthi Sivaswam, "A Successive ClutterRejectionBased Approach for Early Detection of Diabetic Retinopathy", *IEEE Transactions On Biomedical Engineering*, Vol. 58, No. 3, March 2011, pp.664-673.

[9] Ahmad Fadzil, et. Al, "Analysis of Foveal Avascular Zone in Colour Fundus Images for Grading of Diabetic Retinopathy Severity", 32nd Annual International Conference of the IEEE EMBS Buenos Aires, Argentina, August 31 - September 4, 2010, pp. 5632-5635.

[10] R. Sivakumar, G. Ravindran, M. Muthayya, S. Lakshminarayanan, and C. U. Velmurughendran, "Diabetic Retinopathy Analysis", *Journal of Biomedicine and Biotechnology*, 2005 pp. 20- 27. Z.

[11] Omar, M. Hanafi, ¹³ Mashohor, N. Mahfudz, and ¹³ M. Muna'im, "Automaticdiabetic retinopathy detection and classification system," in **2017 7th IEEE International Conference on System Engineering and Technology (ICSET)**, pp. 162- 166, **IEEE, 2017**

[12]. Y. Haniza, A. Hamzah, and M. Norrima, "Edge sharpening for diabetic retinopathy detection," *IEEE Conference Publications*, 2010

¹⁴ [13]. K. K. Palavalasa and B. Sambaturu, "Automatic diabetic retinopathy detection using digital image processing," in **2018 International Conference on Communication and Signal Processing (ICCSP)**, pp. 0072-0076, **IEEE, 2018**

[14]. D. Satyarthi, B. Raju, and S. Dandapat, "Detection of diabetic retinopathy in fundus images using vector quantization technique," in *2006 Annual IEEE India Conference*, pp. 1-4, **IEEE, 2006**

[15]. K.Ram, G. D. Joshi, and J. Sivaswamy, "A successiveclutterrejection-based approach for early detection of diabeticretinopathy,"*IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 664-673, 2010

B.Report

ORIGINALITYREPORT			
22%	6%	18%	7%
SIMILARITYINDEX	INTERNETSOURCES	PUBLICATIONS	STUDENTPAPERS
PRIMARYSOURCES			

1	SubmittedtoIndianaUniversity StudentPaper	3
	%	2%
2	AsraMomeniPour,HadiSeyedarabi,SeyedHassanAbbasiJahromi,AliRez a Javadzadeh."AutomaticDetectionandMonitoringofDiabeticRetinopathyUs i ngEfficientConvolutionalNeuralNetworksandContrastLimitedAdaptiveHist ogramEqualization",IEEEAccess,2020 Publication	
3	ShoravSuriyal,ChristopherDruzgalski,KumarGautam."Mobileassistedd iabeticretinopathydetectionusingdeepneuralnetwork",2018GlobalMedi c alEngineeringPhysicsExchanges/PanAmericanHealthCareExchanges(GMEPE/PAHCE),2018 Publication	2%

4	SubmittedtoChinoValleyUnifiedSchoolDistrict StudentPaper	%
---	---	---

5	XianglongZeng,HaiquanChen,YuanLuo,
---	------------------------------------

WenbinYe."AutomatedDiabeticRetinopathy
DetectionBasedonBinocularSiamese-Like

2

%

ConvolutionalNeuralNetwork",IEEEAccess,2019
Publication

6

2%

"NeuralNetworkTechniqueforDiabeticRetinopathyDetection",Internation
alJournalofEngineeringandAdvancedTechnology,2019
Publication

7

2%

G.Kalyani,B.Janakiramaiah,A.Karuna,L.V.NarasimhaPrasad."Diabeticre
tinopathydetectionandclassificationusingcapsulenetworks",Complex&Intel
ligentSystems,2021
Publication

SubmittedtoAmityUniversity
StudentPaper

1

%

1%

9

HaiQuanChen,XiangLongZeng,YuanLuo,WenBinYe."DetectionofDiab
eticRetinopathyusingDeepNeuralNetwork",2018IEEE23rdInternationalC
onferenceonDigitalSignalProcessing (DSP),2018
Publication

1
0

XiangLongZeng,HaiQuanChen,YuanLuo,WenBinYe."AutomatedDetection
ofDiabeticRetinopathyusingaBinocularSiamese-
LikeConvolutionalNetwork",2019IEEEInternationalSymposiumonCircuitsa
ndSystems(ISCAS),

11 InternetSource

1/0

12 www.nice.org.uk
InternetSource

1

13 www.isaacpub.org
InternetSource

Communication and Signal Processing, 2013.

Publication

14 www.warse.org
InternetSource

Manisha Manjramkar. "Survey of Diabetic Retinopathy Screening Methods", 2018 2nd

15

S. Saranya Rubini, A. Kunthavai. "Diabetic Retinopathy Detection Based on Eigenvalues of the Hessian Matrix", Procedia Computer Science, 2015
Publication

16

technicaljournalsonline.com
InternetSource

17

Gandhi, Mahendran, and R. Dhanasekaran. "Diagnosis of diabetic retinopathy using morphological process and SVM classifier", 2013 International Conference on

%

1%

1%

<1%

<1%



<1%

<1%

aykumar Lachure, A. V. Deorankar, Sagar Lachure, Swati Gupta, Romit Jadhav. "Diabetic Retinopathy using morphological operations and machine learning", 2015 IEEE International Advance Computing Conference (IACC), 2015
Publication

C. Source Code

```
from flask import *

app = Flask(__name_)

import tensorflow as tf, sys

import tensorflow.compat.v1 as tf

tf.disable_v2_behavior()


import os

import sqlite3 as sql

import base64

from flask import Flask ,render_template,request,jsonify,session from

flask import Flask,abort,render_template,request,redirect,url_for

#from werkzeug import secure_filename

#Flask Libraries

from flask import Flask, redirect, url_for, render_template, request

import requests

import urllib

@app.route('/', methods=['GET', 'POST'])

def home():

    return render_template('index.html')


def validate(username,password):

    con = sql.connect('static/chat.db')

    completion = False

    with con:

        cur = con.cursor()
```

```
cur.execute('SELECT * FROM persons')
```

```
for row in rows:

    dbuser = row[1]

    dbpass = row[2]

    if dbuser == username:

        completion = (dbpass == password)

return completion
```

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

    error = None

    if request.method == 'POST':

        username = request.form['username']

        password = request.form['password']

        completion =

        validate(username,password) if completion

        == False:

            error = 'invalid Credentials. please try again.'

        else:

            session['username'] = request.form['username']

            return render_template('file_upload_form.html')

    return render_template('file_upload_form.html', error=error)
```

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

    if request.method == 'POST':

        try:

            name = request.form['name']
```

username =

request.form['username']

```

password = request.form['password']

with sql.connect("static/chat.db") as
con:

    cur = con.cursor()

    cur.execute("INSERT INTO persons(name,username,password)
VALUES (?, ?, ?)",(name,username,password))

    con.commit()

    msg = "Record successfully added"

except:

    con.rollback()

    msg = "error in insert
operation" finally:

    return render_template("index.html",msg = msg)

    con.close()

return render_template('register.html')

```

```
@app.route('/list')
```

```
def list():
```

```
    con =
```

```
    sql.connect("static/chat.db")
```

```
    con.row_factory = sql.Row
```

```
    cur = con.cursor()
```

```
    cur.execute("select * from persons")
```

```
    rows = cur.fetchall();
```

```
return render_template("list.html",rows = rows)
```

```

def upload():
    return render_template("file_upload_form.html")

@app.route('/success', methods = ['POST'])
def success():
    if request.method == 'POST':
        f = request.files['file']

        # change this as you see fit
        #image_path = sys.argv[1]
        image_path = f.filename

        # Read in the image_data
        image_data = tf.gfile.GFile("D:/Project/Automatic
Diabetic_Retinopathy_webapp/dataset/"+str(image_path), 'rb').read()

        # Loads label file, strips off carriage return
        label_lines = [line.rstrip() for line
                        in tf.gfile.GFile("models/retrained_labels.txt")]

        # Unpersists graph from file
        with tf.gfile.GFile("models/retrained_graph.pb", 'rb') as f:
            graph_def = tf.GraphDef()
            graph_def.ParseFromString(f.read())
            _ = tf.import_graph_def(graph_def, name='')

        with tf.Session() as sess:
            # Feed the image_data as input to the graph and get first prediction

```

```
softmax_tensor = sess.graph.get_tensor_by_name('final_result:0')
```

```
predictions = sess.run(softmax_tensor,  
                        {'DecodeJpeg/contents:0': image_data})
```

```
# Sort to show labels of first prediction in order of confidence
```

```
top_k = predictions[0].argsort()[-len(predictions[0]):][::-1]
```

```
print(top_k)
```

```
m=[]
```

```
for node_id in top_k:
```

```
    human_string =
```

```
    label_lines[node_id] score =
```

```
    predictions[0][node_id]
```

```
    for j in human_string,score:
```

```
        print(j)
```

```
        m.append(j)
```

```
#print('%s (score = %.5f)' % (human_string, score))
```

```
return render_template("success.html", list2 = m)
```

```
if __name__ == '__main__':
```

```
    app.run(debug = True)
```