A MINI PROJECT REPORT
ON

"CLASSIFICATION OF ORTHOPEDIC PATIENTS USING ML"

FOR
Laboratory Practice 2
(B.E. Computer Engineering)

BY

Abhijit Shendage (405C033)
Saurabh Gore (405C008)
Yash Patil (405C025)
Samyak Vaidya (405C030)

Under the guidance of

Prof. Kirti Satpute

Sinhgad Institutes

Department of Computer Engineering

Sinhgad College of Engineering

Vadgaon (Bk.), Pune - 41

Affiliated to Savitribai Phule Pune University, Pune.

2020-2021

Sinhgad Technical Education Society,
Department of Computer Engineering
Sinhgad College of Engineering , Pune-41

Sinhgad Institutes

Date: / /

CERTIFICATE

This is to certify that,

Abhijit Shendage
Saurabh Gore
Yash Patil
Samyak Vaidya

have successfully completed their mini project work titled "CLASSIFICATION OF ORTHOPEDIC PATIENTS USING ML'' subject Laboratory Practice 2(LP2) at Department of Computer Engineering, SCOE, Pune for the partial fulfillment of the Bachelor Degree of Computer Engineering , Savitribai Phule Pune University, in semester-I, academic Year 2020-2021.

Prof. K.B. Satpute
Internal Guide
Department of Computer Engineering

Prof. M. P. Wankhade
Head of Dept.
Department of Computer Engineering

Dr. S.D. Lokhande
Principal
Sinhgad College Of Engineering, Pune

ACKNOWLEDGEMENT

We feel great pleasure in expressing our deepest sense of gratitude and sincere thanks to our guide **Prof. Kirti Satpute** for her valuable guidance during the Project work, without which it would have been very difficult task. I have no words to express my sincere thanks for valuable guidance, extreme assistance and cooperation extended to all the **Staff Members** of our Department.

This acknowledgement would be incomplete without expressing our special thanks to **Prof.M. P. Wankhade**, Head of the Department (Computer Engineering) for his support during the work.

We would also like to extend our heartfelt gratitude to our Principal**, Dr. S. D. Lokhande** who provided a lot of valuable support, mostly being behind the veils of college bureaucracy.

Last but not least we would like to thank all the Teaching, Non- Teaching staff members of our Department, our parents and colleagues those who helped us directly or indirectly for completing this Project successfully.

Abhijit S. Shendage
Saurabh H. Gore
Yash P. Patil
Samyak Vaidya

ABSTRACT

A person's orthopedic health condition can be detected from his biomechanical features. Now a days, disease prediction can be done automatically. Application of machine learning algorithms in medical science is not new. Different algorithms are applied to detect diseases and classify patients accordingly. Use of machine learning is spreading widely with the growth of medical data in medical field to improve medical service and diagnosis of diseases.

The proposed system aims to assist specialists to predict the type of orthopedic disease. In this paper we have applied various machine learning algorithms to find out how each algorithm performs to detect and classify orthopedic patients. Each of the patients in the dataset is represented by six biomechanical attributes derived from the shape and orientation of pelvis and lumbar spine. We performed our operation in two stages and got an average accuracy of more than 90 percent for most of the algorithms.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

Chapter 1

Introduction

Machine learning has been implemented in various medical fields and proven to be very accurate in classifying and predicting diseases. Use of machine learning is spreading widely with the growth of medical data in medical field to improve medical service and diagnosis of diseases. Bio-mechanics is the study of the movement of living beings using the science of mechanics. According to the science of mechanics, motion is created by force. Living beings create motion using force. Bio-mechanics is essential to understand how living things make moves.

The condition when the gel like material (Nucleus Pulposus) get squeezed out through fractures in outer wall of inter-vertebral disc is known as Lumbar Disc Herniation Fig. 1.1 shows an example of Disk herniation.

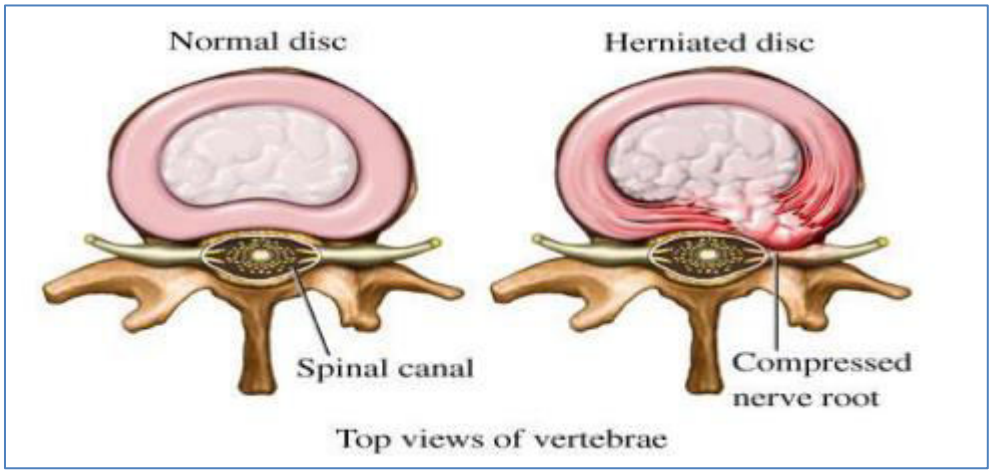

Fig 1.1 Disk herniation

Spondylolisthesis is a medical condition in which one of the bones of a person's back (vertebrata) slides forward over the bone below it. Most of the time it occurs in lower spine or lumbosacral area. Fig. 1.2 shows an image of Spondylolisthesis, where it can be seen that upper disk has moved forward with respect to lower disk.

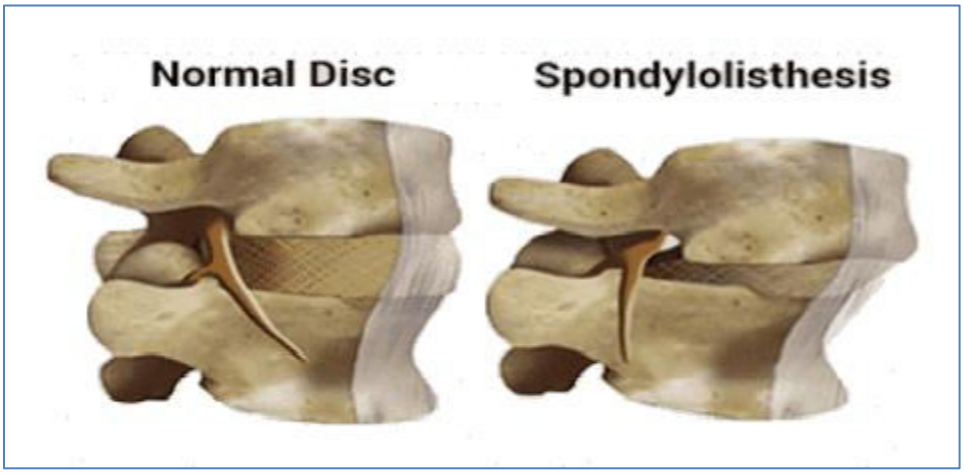

Fig. 1.2 Spondylolisthesis

Both conditions can squeeze the spinal cord or nerve roots and cause pain. Both of these diseases may cause similar types of pain but they are different. Depending on the changes they make in a normal person's bio-mechanical characteristics, the disease can be predicted. The proposed system shows the classification of orthopedic patients based on these bio-mechanical features which are represented in the datasets we used. The classification is done in two stages where firstly, the system checks if the patients are in normal condition or not. If they are not in normal condition, then the system classifies them in two different categories of diseases. The objective of this paper is to find a suitable and accurate algorithm to enhance medical diagnosis by predicting diseases in an automated way.

1.1 Motivation

Machine learning has been implemented in various medical fields and proven to be very accurate in classifying and predicting diseases.Use of machine learning is spreading widely with the growth of medical data in medical field to improve medical service and diagnosis of diseases.Since there has not been enough application of machine learning in orthopaedic field, we had a thought the biomechanical features can also be used in machine learning and thus can p0rovide a good assist to specialist and reduce their task of remembering all the values.

1.2Timeline/Evolution

For the past few years, different algorithms of machine learning has been applied to predict and classify different sorts of diseases. Algorithms like K-Nearest Neighbor (kNN), Logistic Regression (LR), Random Forest (RF) and Support Vector Machine (SVM) were used to classify cardiac arrhythmia where SVM out performed others. Again, Random Forest was used to classify Pulmonary Tuberculosis and Sarcoidosis where it provided a very good accuracy. SVM has been used to classify the types of Leukemia. Machine learning techniques are being used extensively to detect diseases in easier way. Algorithms like Artificial Neural Network (ANN) and Probabilistic Neural Network (PNN) have been comparing to predict osteoporosis where ANN provided better results than PNN. Application of ANN can be found in many medical fields including radiology, oncology, urology, and cardiology. ANN has been proven to be very useful to develop existing medical technique.

1.3 Organization of Report

The entire report is divided into five chapters. It starts with Introduction followed by Literature Review, Methodology, Result and Conclusion. The details of each section are as explained in brief as following.

- **Introduction**: This portion include why this system has been chosen for DMW mini project along with some background information on the topic.
- **Literature Review**: It discusses the existing work done so far using various methodologies and at the end there is the comparative analysis of all the considered solutions.
- **Methodology**: This chapter is the actual core of the report and all the technical details of the system are discussed in this section.
- **Result**: This chapter includes comparison of performance result discussed in the report.
- **Conclusion and Future Work:**This is the last chapter which derives conclusion and suggests future work to be done, ie discusses the future scope of the project.

Chapter 2

Literature Review

Before finalizing the proposed system as our project for LP2, we did some survey. In that we searched and studied various research papers. Along with the papers, we've also studies various books and online web resources. We found that there are plot of systems which are being developed to assists the medical field stakeholders and can become revolutionary in upcoming years. In this section we have put down the survey along with the seed idea that we did before continuing with the proposed system.

2.1 Literature Survey

Sr. No.	Title	Author	Year Of Publication	Publication	Seed Idea
1.	Disease Prediction using Machine Learning Algorithms	Sneha Grampurohit , Chetan Sagarnal	2020	Institute Of Electrical and Electronics Engineers (IEEE)	Presents comparative study of the results obtained by algorithms such as Decision tree classifier & Random Forest classifier
2.	Disease Symptom Analysis Based Department Selection Using Machine Learning for Medical Treatment	Md. Latifur Rahman, Rahad Arman Nabid , Md. Farhad Hossain	2020	Institute Of Electrical and Electronics Engineers (IEEE)	By using ML algorithms, K-Nearest neighbour and logistic regression provides analysis of easily recognized diseases

3.	An Implementation of Naïve Bayes Classifier	Feng-Jen Yang	2018	Institute Of Electrical and Electronics Engineers (IEEE)	Ensure the correctness of all probabilistic computations involved in Naïve-Bayes classification.
4.	Analysis of Symptoms Wise Disease Inference System Using Data Mining Technique	Tejal P. Burange, Dr. P. N. Chatur	2018	Institute Of Electrical and Electronics Engineers (IEEE)	Implement automatic question-answer system in which user will ask any question and system will process that question using data mining algorithms to find out proper answer

Table 2.1.1 Literature Survey

2.2 Applications

- Medical Specialists
- Research and Development team of medical officers
- Intelligent Medical Infrastructures
- Medical Training
- Simulations
- Orthopedic Research

Chapter 3

Implementation

3.1 Algorithms Used

To find the best algorithms for the given datasets we did a survey. In that we searched and studied various research papers. Along with the papers, we've also studies various books and online web resources. We found that there are multiple algorithms which can be used for classification but we narrowed it down to the two best performing ones. In this section we have put down the algorithms used along with the methodologies that we did discussed before continuing with the proposed system.

3.1.1 K-Nearest Neighbors (KNN)

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other. K-Nearest Neighbor is non-parametric. It classifies object based on majority of vote it gets from its closest neighbors. It is a supervised machine learning algorithm that can be used to solve both classification and regression problems.

$$K = n^{1/2}$$

Here, n = number of features.

The value of 'K' influences result. Smaller value of k means noise will have higher influence on result. On the other hand large value of 'K' increases the computation cost.

3.1.2 Naïve Bayes

It is a probabilistic classification technique based on Bayes' Theorem with an assumption of independence among predictors. Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

Here, P(A|B)= posterior probability of class given predictor, P(B|A)= likelihood which is the probability of predictor given class , P(A)= class prior probability of the class , P(B)= predictor prior probability.

3.2 Methodology

In this section, we will discuss about how the system works. We would also see how the system uses server side using architecture diagram.

3.2.1 System Flow

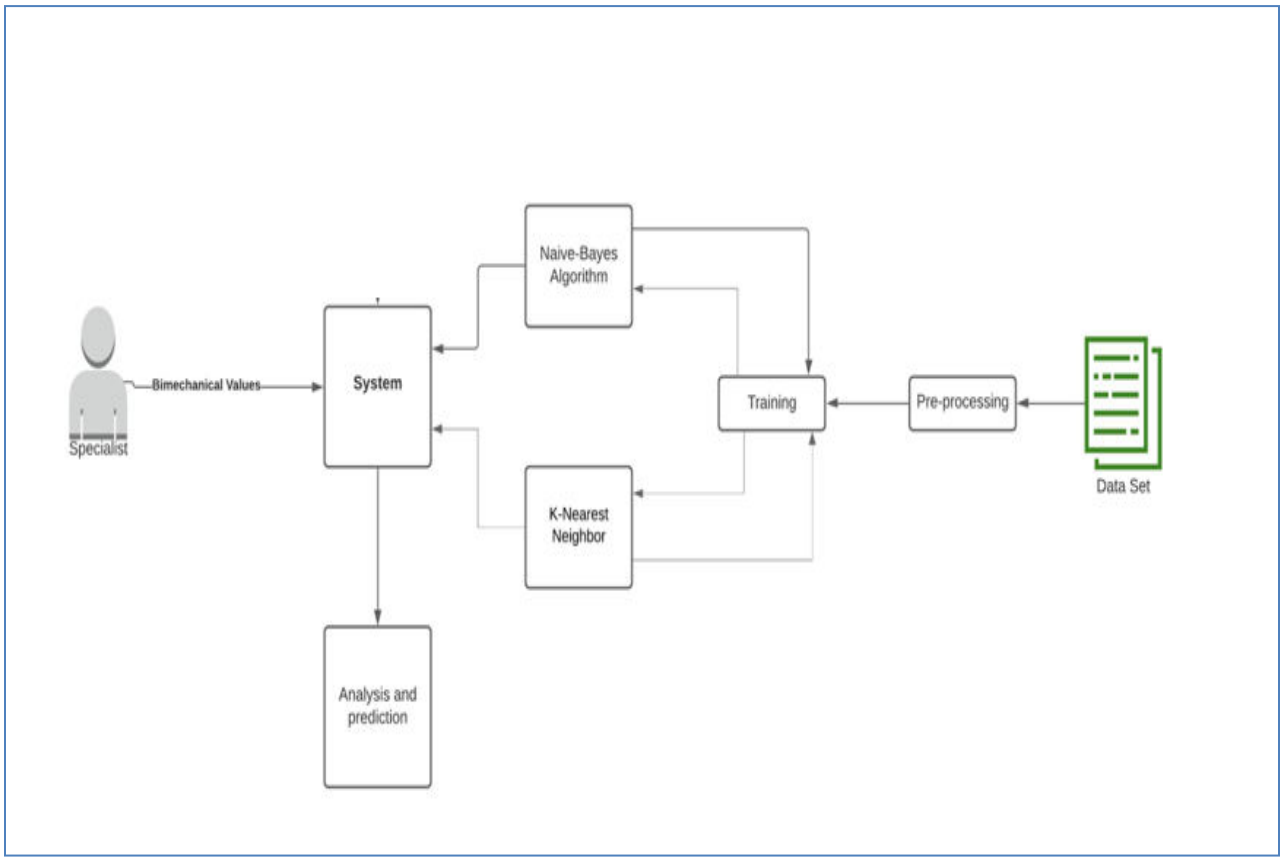

Fig. 3.2.1.1 System Flow

As shown in the above figure, the user has to be a specialist. After seeing the reports of the patient, the user will get the values of various attributes which can then be supplied to the system to classify the newly came patient with respect to his type of disease. As the system has already been trained with algorithms such as KNN and Naïve Bayes by the various' patients preprocessed and data, the system will do some magical computations and try to calculate the results. Thus the calculated results and analysis will be shown to user on the webpage. Thus, the efforts of remembering and the time taken to process the data are saved.

3.2.2 System Architecture

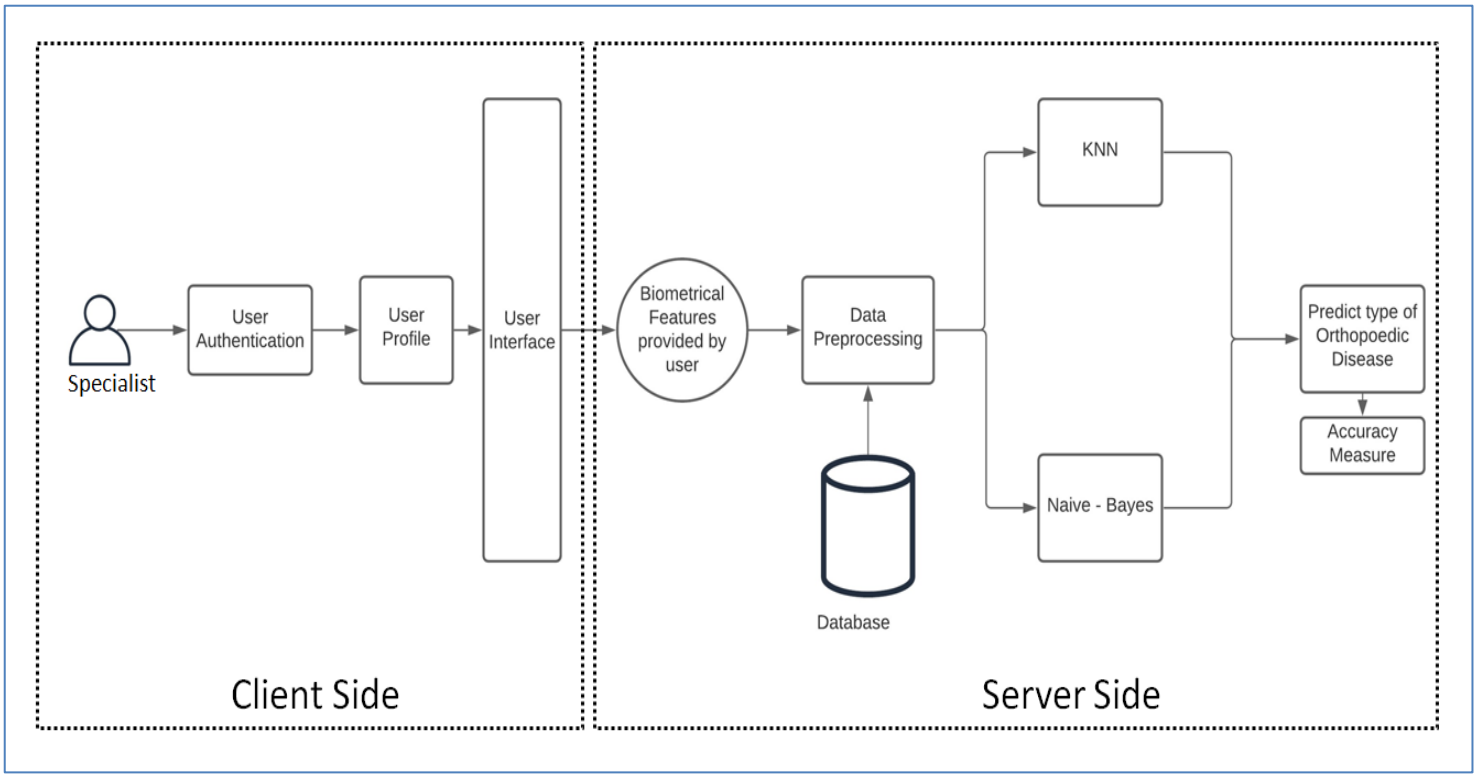

Fig. 3.2.2.1 System Architecture

As the system is a web application, it has two sides viz. server side and the client side. The system is designed by using django framework of python. The client side or the front is of web pages which are easily accessible from anywhere and easy to use. The server side has business logic and it also has the database which contains users' information and patients' database. It does all the calculations and processes which are required to produce the required output which is then shown to the end user.

3.2.3 Use Case

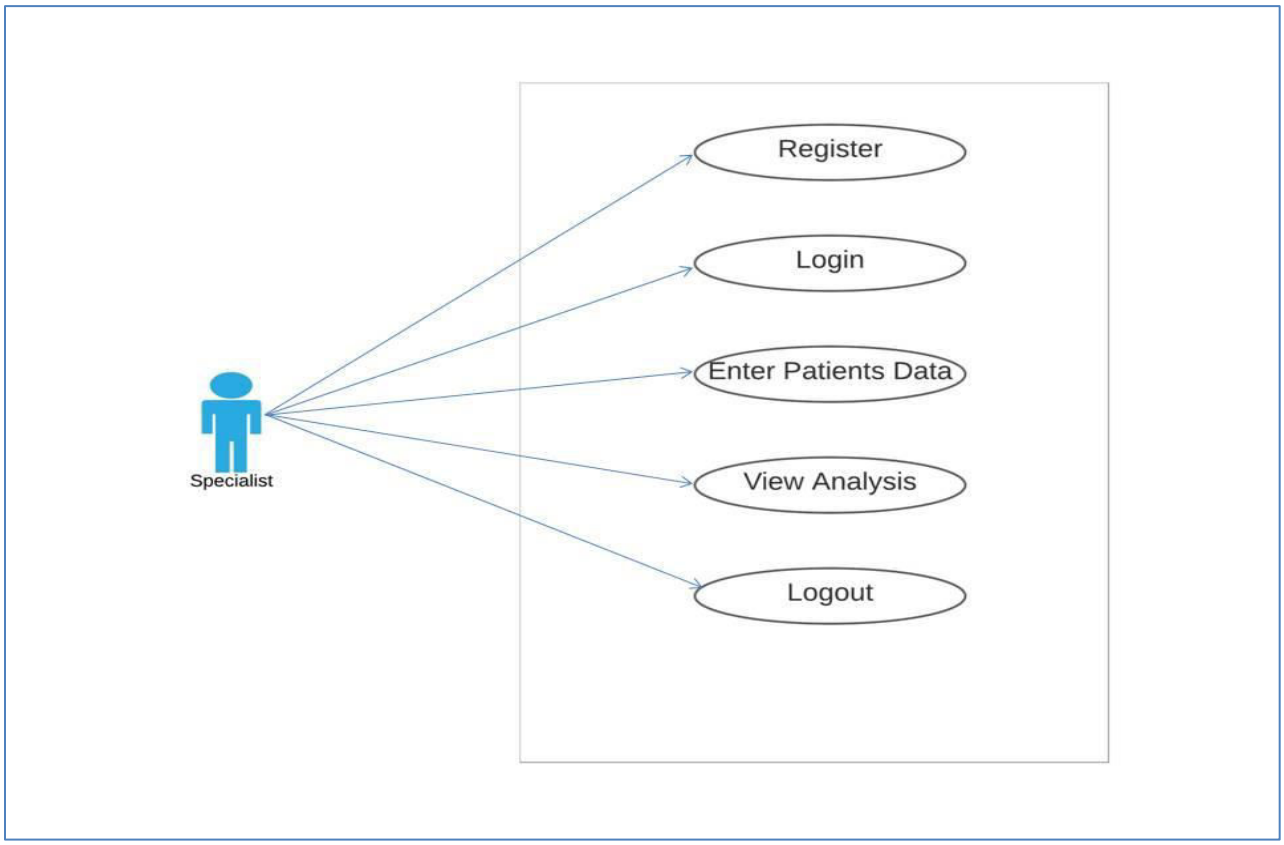

Fig. 3.2.3.1 Use Case Diagram

The main end user of the system is medical specialist. The user will be able to register, login and see the expected results after inputting the suitable values to the system.

3.2.4 Testing

3.2.4.1 Selenium Testing

0Selenium is a portable framework for testing web applications. Selenium provides a playback tool for authoring functional tests without the need to learn a test scripting language. The supporting code can be written any language such as JAVA, Python, SCALA, C++, etc. The selenium testing of the system is done using the Google Chrome browser and its driver. It uses python's pytest library. The testing can be done in both ways, in an automated testing way and in manual testing way. The proposed system is tested in both the ways to ensure its reliability.

3.2.4.2 Unit Testing

Unit testing is a method by which individual units of source code are tested to determine if they are fit for use. Library used for unit testing in python- unittest. Every unit of the proposed system has been tested, so that it ensures it's safety at very tiny level. As each and every unit of the system has been tested, the failure of any particular unit's failure does not affect the execution of other units of the system.

3.3 Implementation Overview

3.3.1 System Overview

Here are some snapshots attached of the system,

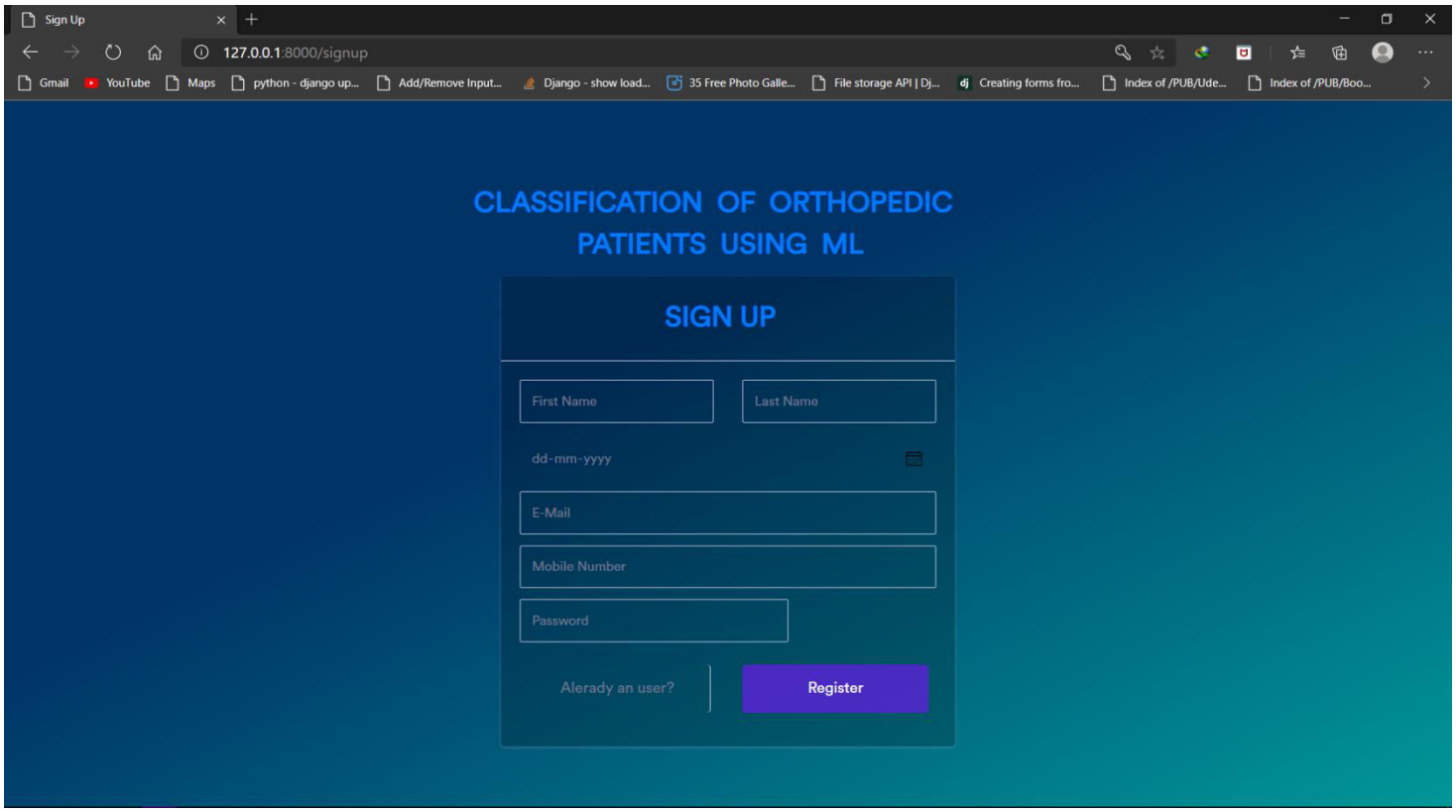

Fig 3.3.1.1 Registration Page

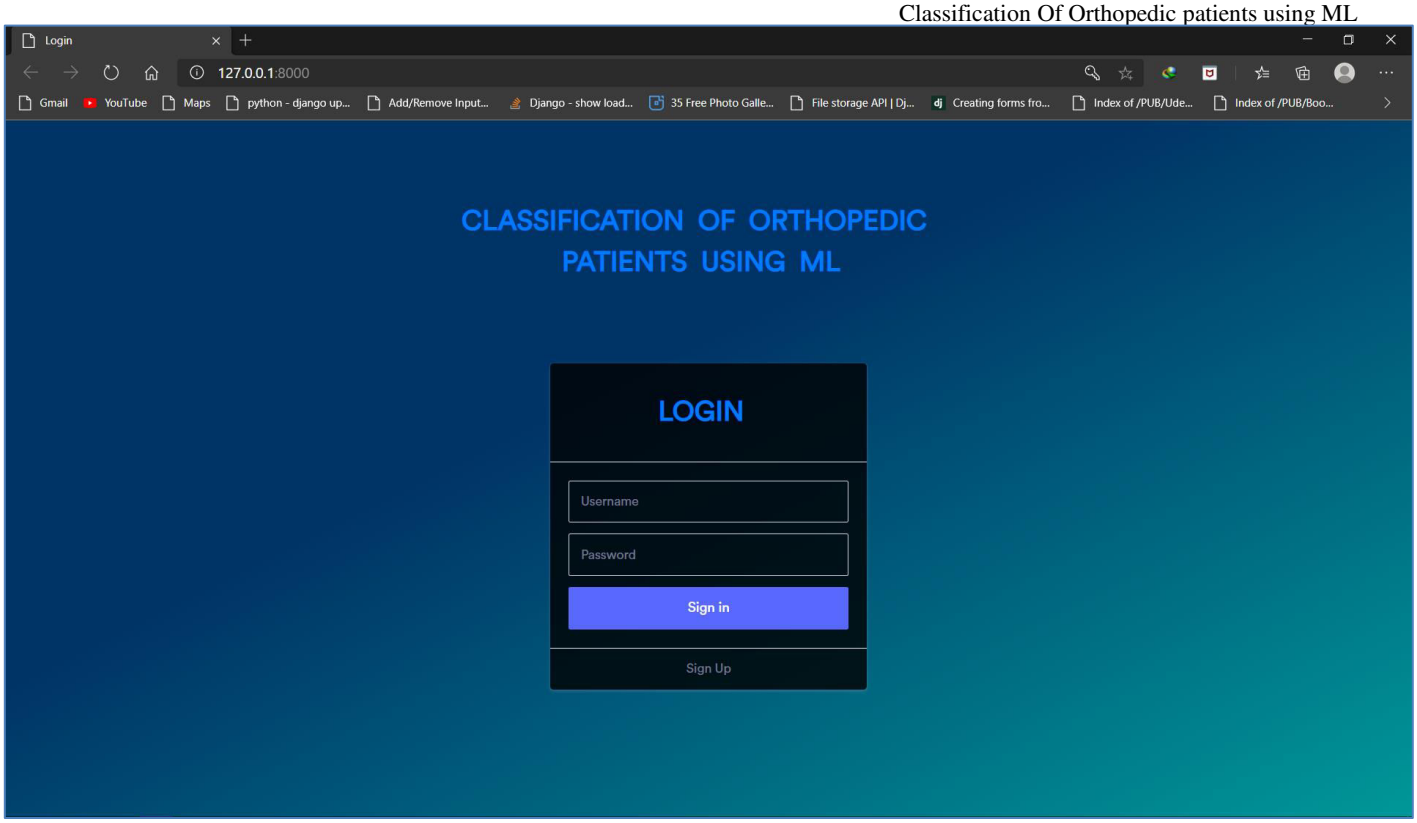

Fig 3.3.1.2 Login Page

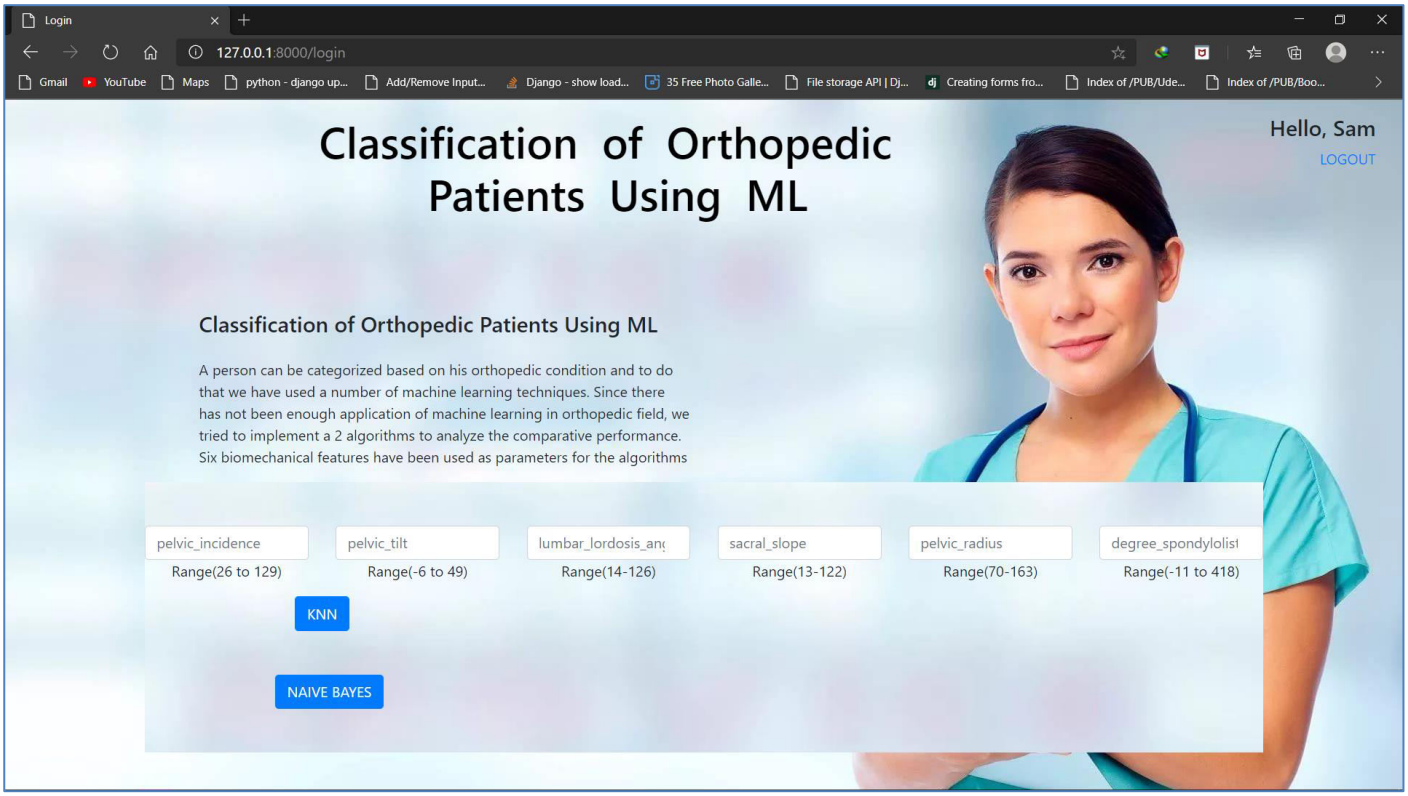

Fig 3.3.1.3 Index Page

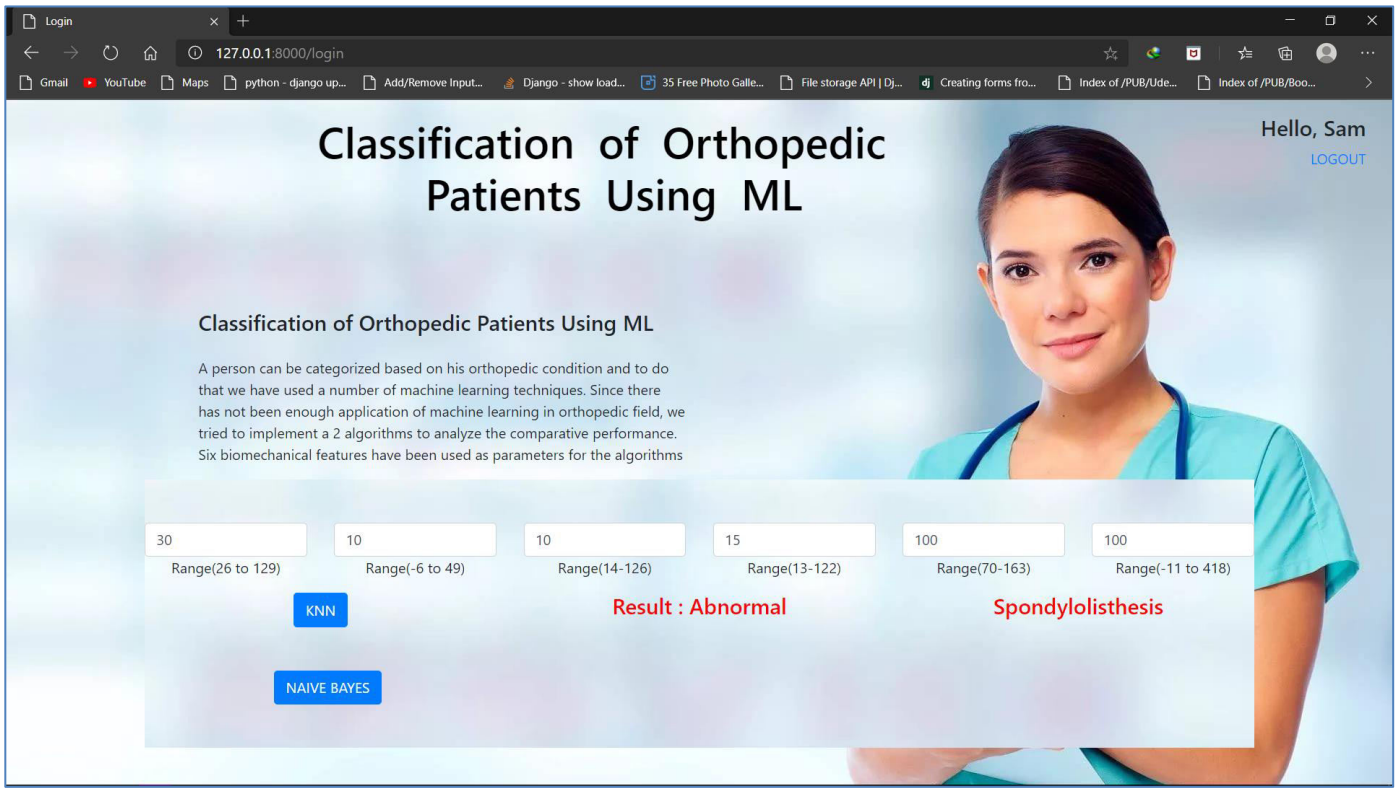

Fig 3.3.1.4 Result of KNN

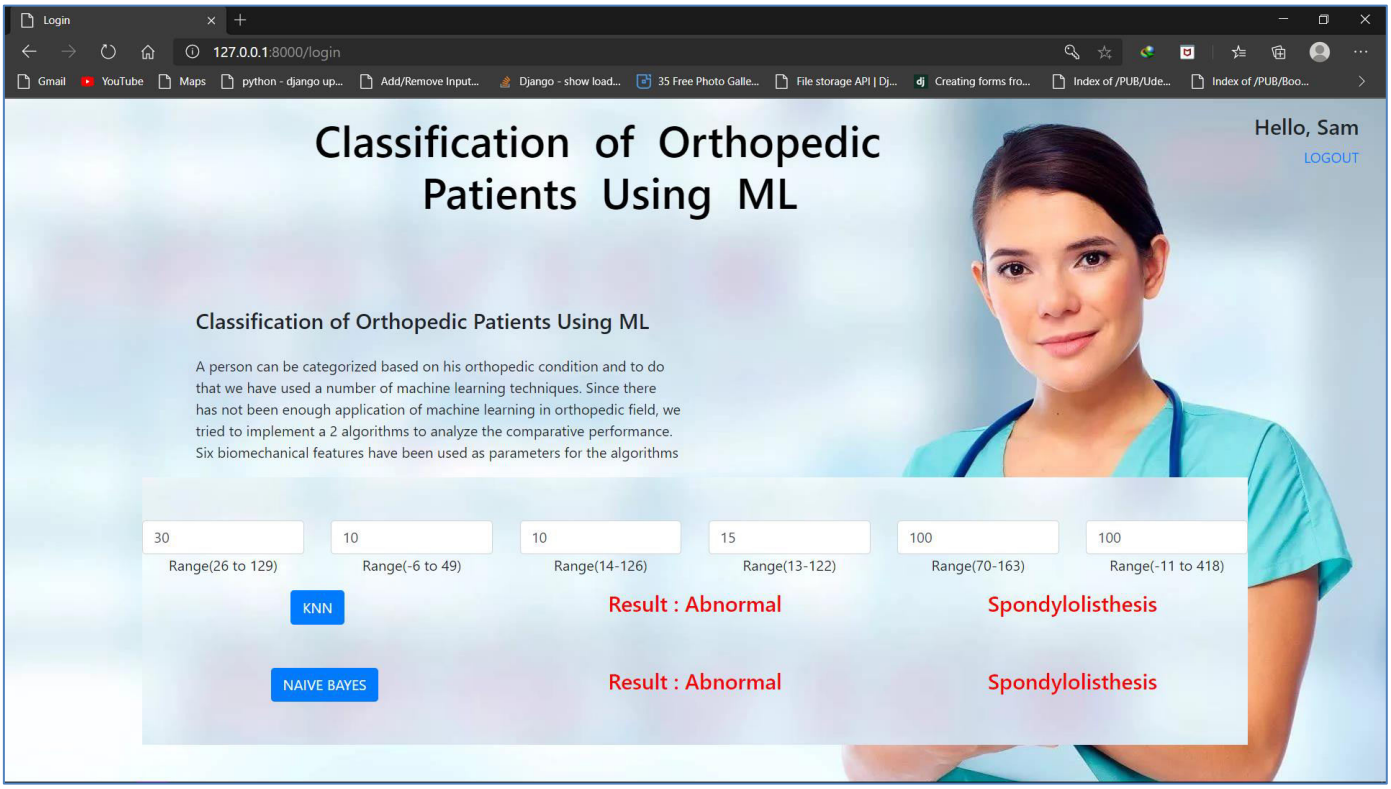

Fig. 3.3.1.5 Result of NB

3.3.2 TESTING RESULTS

Here are some snapshots of the results which were obtained during selenium and unit testing,

Fig 3.3.2.1 Selenium Testing

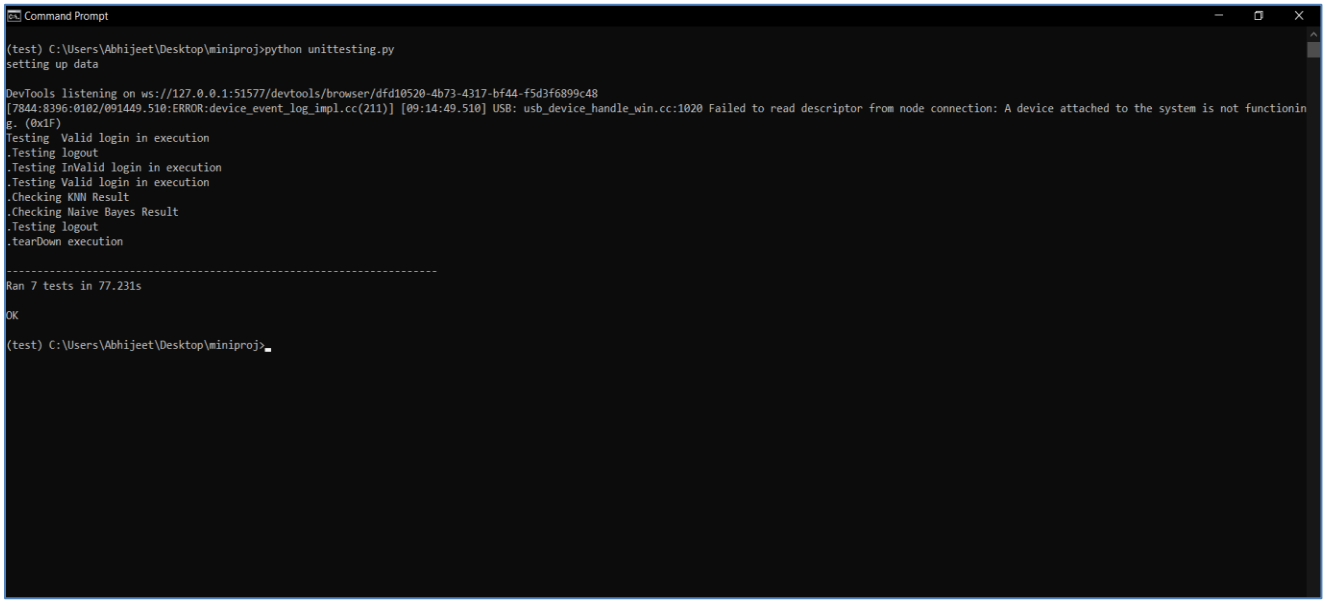

Fig 3.3.2.2 Unit Testing

Chapter 4

Result

In the proposed system, we have used two classifiers named K-Nearest Neighbor (k-NN) and Naïve Bayes (NV). The system is developed using django framework. We have used python and its related packages for classification. To train the system, we have used about 70% of the data. Accuracy score and Confusion matrix were used as metrics to calculate the performance of the algorithms.

n	**No (Prediction)**	**Yes (Prediction)**
No (Actual)	True Negatives (TN)	False Positives (FP)
Yes (Actual)	False Negatives (FN)	True Positives (TP)

TABLE 4.1 CONFUSION MATRIX

Where, True Positives (TP) = these are cases in which algorithms predicted yes (they have the disease), and they do have the disease. True Negatives (TN) = these are cases in which algorithms predicted no, and they don't have the disease. False positives (FP) = these are the cases in which algorithms predicted yes, but they don't actually have the disease. The False Negatives (FN) = these are the cases in which algorithms predicted no, but they actually do have the disease. n = total number of instances. From each algorithms confusion matrix, we used the following equation to derive the accuracy of each algorithm,

$$\text{Accuracy} = (TP + TN) / n$$

The accuracy rates for normal / abnormal of KNN is 83.87 percent and of Naive Bayes is 79.03 percent. And the accuracy rates for normal / disk hernia / spondylolisthesis of KNN is 83.8 percent and of Naive Bayes is 87.10 percent. We have used Matplotlib library to visually represent the results of each algorithm.

From the result of both of our analysis, we can state that the system provides the best result for this kind of analysis given the required data values.

Chapter5

Conclusion & Future Work

5.1 Conclusion

A person can be categorized based on his orthopedic condition and to do that we have two machine learning techniques. Since there has not been enough application of machine learning in orthopedic field, we tried to implement the two of algorithms to analyze the comparative performance. Six biomechanical features have been used as parameters for the algorithms. From the results obtained we can say that the system does stand on the expectations. Because of the higher accuracy of the system, the specialist and other end users of the system can rely on it. The system provides a cross-validation method by obtaining results from two different algorithms. The analysis can assist doctors in identifying diseases in a faster and easier way.

5.2 Future Scope

The system currently uses only two of many machine learning algorithms. In future our system can be extended with the help of deep learning and neural network. Also the end user of the system has to provide the input manually which is obtained from various medical system provides before. In future we can make the system to get its input directly from medical lab equipment so that both the time taken and the human intervention get reduced.

References

[1] S. Pouriyeh, S. Vahid, G. Sannino, G. D. Pietro, H. Arabnia, J.Gutierrez, ―A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease‖, IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, 2017.

[2] Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) "Analysis of Symptoms Wise Disease Inference System Using Data Mining Technique" by Tejal P. Burange and Dr. P. N. Chatur

[3] 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020 "Disease Prediction using Machine Learning Algorithms" by Sneha Grampurohit and Chetan Sagarnal

[4] D. Knudson, Fundamentals of Biomechanics, 2nded, New York: Springer, 2007, pp.3.

[5] K. Alawneh, M. Al-dwiekat, M. Alsmirat and M. Al-ayyoub, ―Computer-Aided Diagnosis of Lumbar Disc Herniation,‖ in 6thInternational Conference on Information and Communication System (ICICS 2013), 2013.

[6] S. Liao et al., ―Automatic Lumbar Spondylolisthesis Measurement in CT Images,‖IEEE Trans. Med. Imag., vol. 35, no. 7, pp.1658-1669, July2016.

[7] P. Shimpi, S. Shah, M. Shroff and A.Godbole, ―A Machine Learning Approach for the Classification of Cardiac Arrhythmia,‖in Proceeding of the IEEE International Conference on Computing and Communication (ICCMC), pp. 603-607, 2017.

[8] Y. Wu, H. Wang and F. Wu, ―Automatic Classification of pulmonary Tuberculosis and Sarcoidosis based on Random Forest,‖ in 10thInternational Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017.

[9] P. Jagadev and H.G. Virani, ―Detection of Leukemia and its Types using Image Processing and Machine Learning,‖ in International Conference on Trends in Electronics and Informatics (ICEI), IEEE, 2017