iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** IX **Month of publication:** September 2020

DOI: https://doi.org/10.22214/ijraset.2020.31726

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IX Sep 2020- Available at www.ijraset.com

Software for Disease Prediction

Megha Vyas[1], Priya Kunwar[2], Vinod Joshi[3], Mihir Maheshwari[4], Mr. Girish Ameta[5]
[1, 2, 3, 4]*UG Student,* [5]*Assistant Professor , Department of Computer Science & Engineering, Geetanjali Institute of Technical Studies Udaipur, Rajasthan 313001, India*

Abstract: ***Diseases are one of the most serious challenges in both developing and developed countries. According to the International Diabetes Federation, there are 285 million ill people worldwide. This total is expected to rise to 380 million within 20 years.***
Due to its importance, a design of classifier for the detection of symptoms of disease with optimal cost and better performance is the need of the age.
The Pima Indian database at the UCI machine learning laboratory has become a standard for testing data mining algorithms to see their prediction accuracy in data classification. The proposed method uses Support Vector Machine (SVM), a machine learning method as the classifier for diagnosis of disease. The machine learning method focus on classifying disease from high dimensional medical data set. The experimental results obtained show that support vector machine can be successfully used for diagnosing disease.

I. INTRODUCTION

A. Disease Prediction Using Machine learning Software Project Report

Diabetes is one of the common and rapidly increasing diseases in the world. It is a major health problem in most of the countries. Diabetes is a condition in which your body is unable to produce the required amount of insulin needed to regulate the amount of sugar in the body.

This leads to various diseases including heart disease, kidney disease, blindness, nerve damage and blood vessels damage. There are two general reasons for diabetes: The pancreas does not make enough insulin or the body does not produce enough insulin. Only 5-10 % of people with this have this form of the disease (Type-1). Cells do not respond to the insulin that is produced (Type-2). Insulin is the principle hormone that regulates uptake of glucose from the blood into most cells (muscle and fat cells). If the amount of insulin available is insufficient, then glucose will not have its usual effect so that glucose will not be absorbed by the body cells that require it.

Disease being one of the major contributors to the mortality rate. Detection and diagnosis of diabetes at an early stage is the need of the day. Disease diagnosis and interpretation of the data is an important classification problem. A classifier is required and to be designed that is cost efficient, convenient and accurate. Artificial intelligence and Soft Computing Techniques provide a great deal of human ideologies and are involved in human related fields of application. These systems find a place in the medical diagnosis. A medical diagnosis is a classification process. A physician has to analyze lot of factors before diagnosing the diabetes which makes physician‟s job difficult.

In recent times, machine learning and data mining techniques have been considered to design automatic diagnosis system for diseases. Recently, there are many methods and algorithms used to mine biomedical datasets for hidden information including Neural networks (NNs).

These algorithms decrease the time spent for processing symptoms and producing diagnoses, making them more precise at the same time. There is a great variety of methods related to diagnosis and classification of diabetes disease in the literature. Polat etal used principal component analysis and neuro fuzzy inference for diabetes data classification. Deng and kasabov obtained 78.4% classification accuracy with 10-fold cross-validation (FC) using ESOM. Yu et al. combined Quantum Particle Swarm Optimization (QPSO) and Weighted Least Square (WLS) Support Vector Machine to diagnose Type-2 diabetes. Smith et al. proposed a neural network ADAP algorithm to build associative models. 576 randomly selected data are used for training and 192 test cases showed an accuracy of 76%. Quinlan applied c4.5 algorithm and the classification accuracy was 77.1%. Sahan et al used Attribute Weighted Artificial Immune System with 10- fold cross validation method. The Support Vector Machine (SVM) is a novel learning machine introduced first by Vapnik and has been applied in several financial applications recently, mainly in the area of time series prediction and classification.

II. WORKING MODULES

Fig. 1: Welcome page

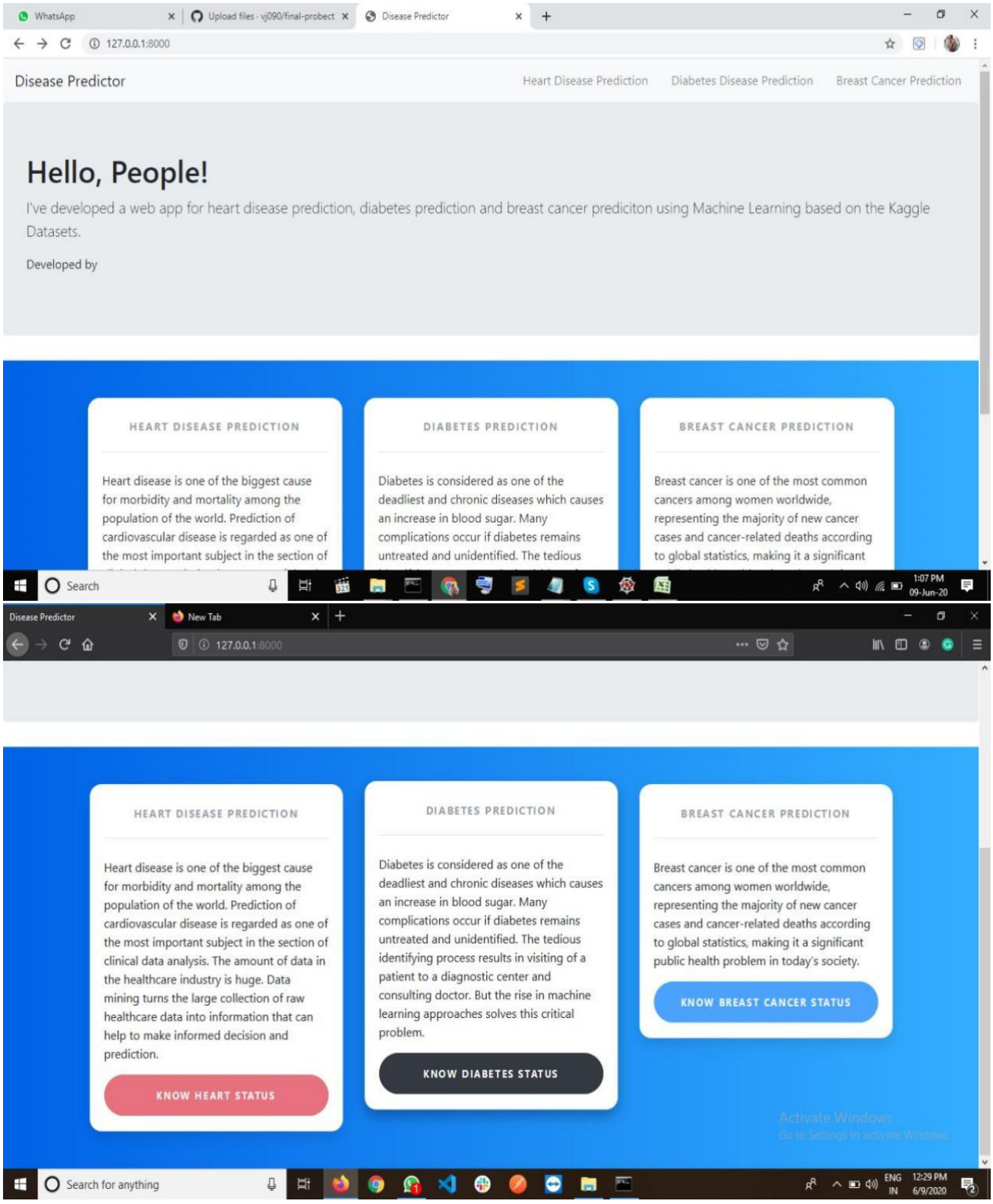


Fig. 2: Select option to check health.

III. DATA-SET

The Pima are a group of Native Americans living in Arizona. A genetic predisposition allowed this group to survive normally to a diet poor of carbohydrates for years. In the recent years, because of a sudden shift from traditional agricultural crops to processed foods, together with a decline in physical activity, made them develop the highest prevalence of disease and for this reason they have been subject of many studies.

A. Dataset

The dataset includes data from 768 women with 8 characteristics, in particular:

1) Number of times pregnant (opction in this project)
2) Plasma glucose concentration a 2 hours in an oral glucose tolerance test
3) Blood pressure (mm Hg)
4) Triceps skin fold thickness (mm)
5) 2-Hour serum insulin (mu U/ml)
6) Body mass index (weight in kg/(height in m)^2)
7) Diabetes pedigree function
8) Age (years)

The last column of the dataset indicates if the person has been diagnosed with diabetes (1) or not (0)
Source: The original dataset is available at UCI Machine Learning Repository and can be downloaded from this address:
http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

B. The Problem

The type of dataset and problem is a classic supervised binary classification. Given a number of elements all with certain characteristics (features), we want to build a machine learning model to identify people affected by type 2 diabetes.
To solve the problem we will have to analyse the data, do any required transformation and normalisation, apply a machine learning algorithm, train a model, check the performance of the trained model and iterate with other algorithms until we find the most performant for our type of dataset.

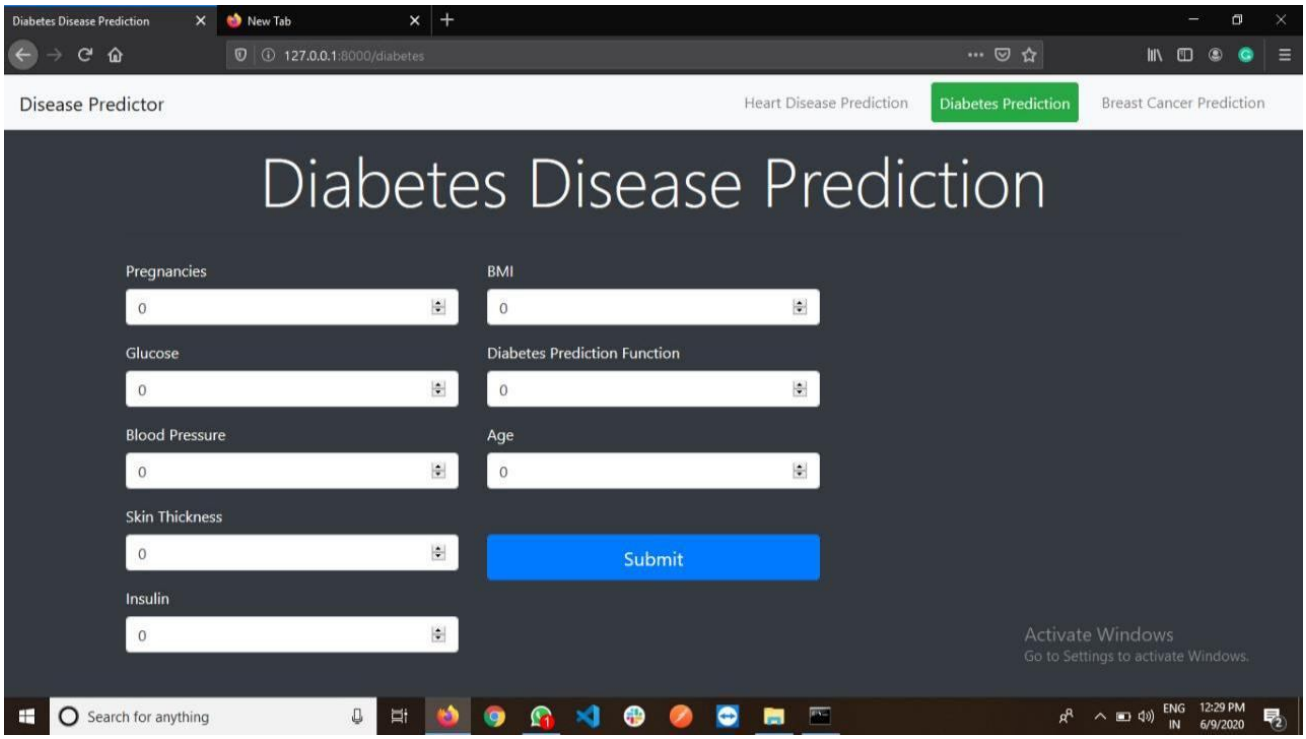


Fig 3: Symptopms for diabetes

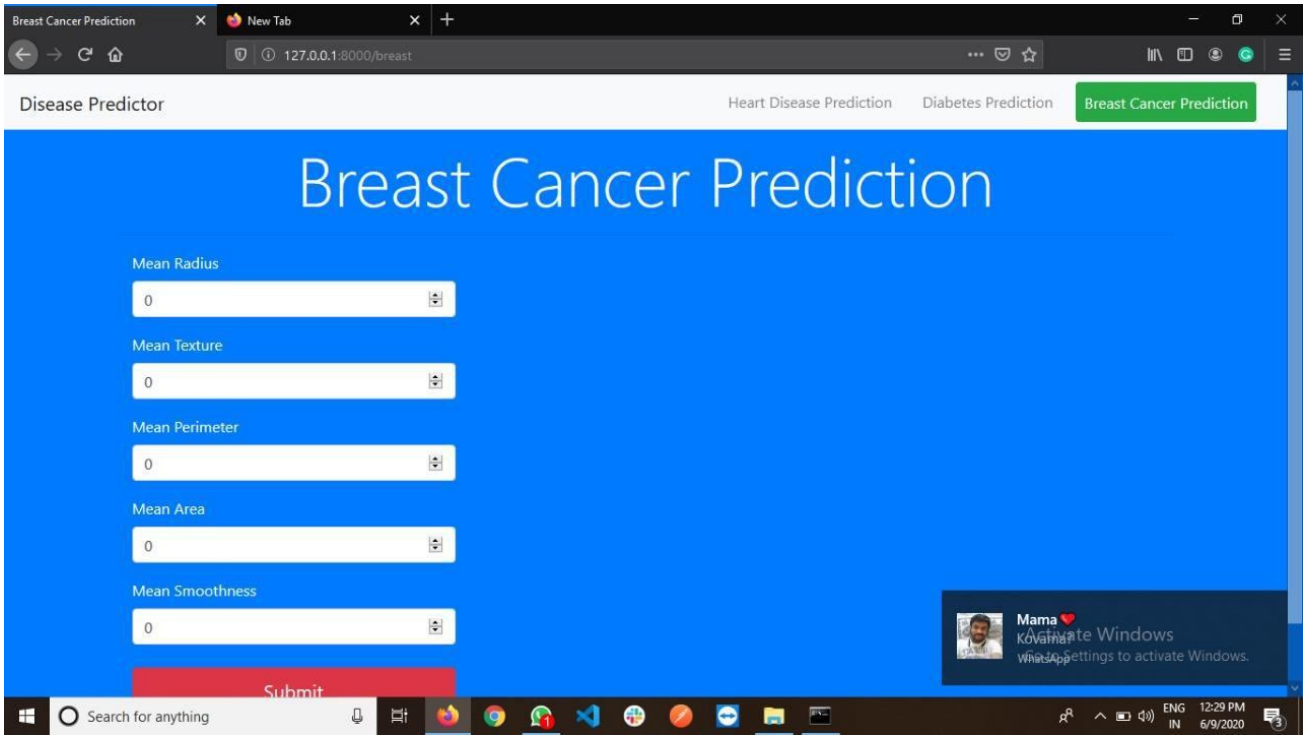


Fig 4: Symptoms for breast cancer

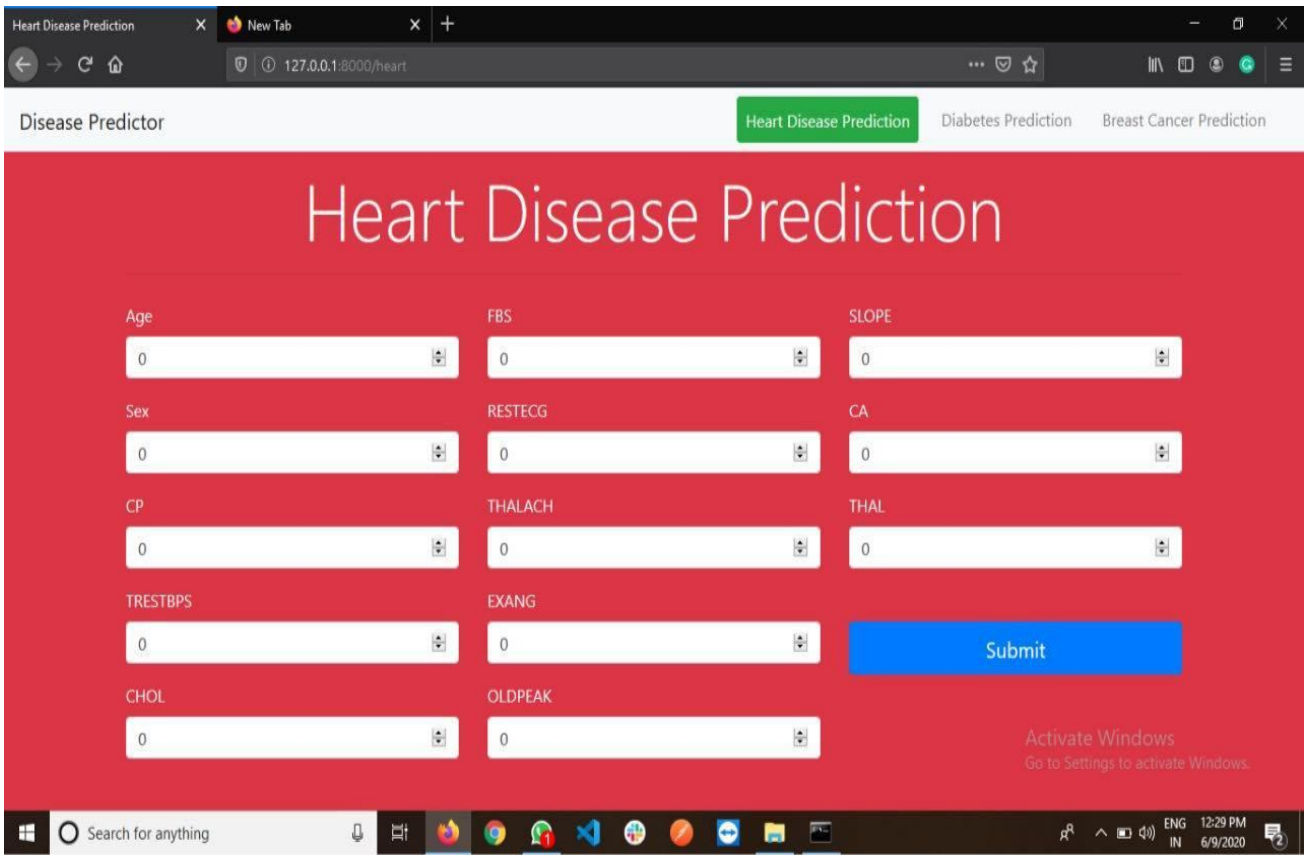


Fig 4: Symptoms for Heart Disease

IV. CONCLUSION

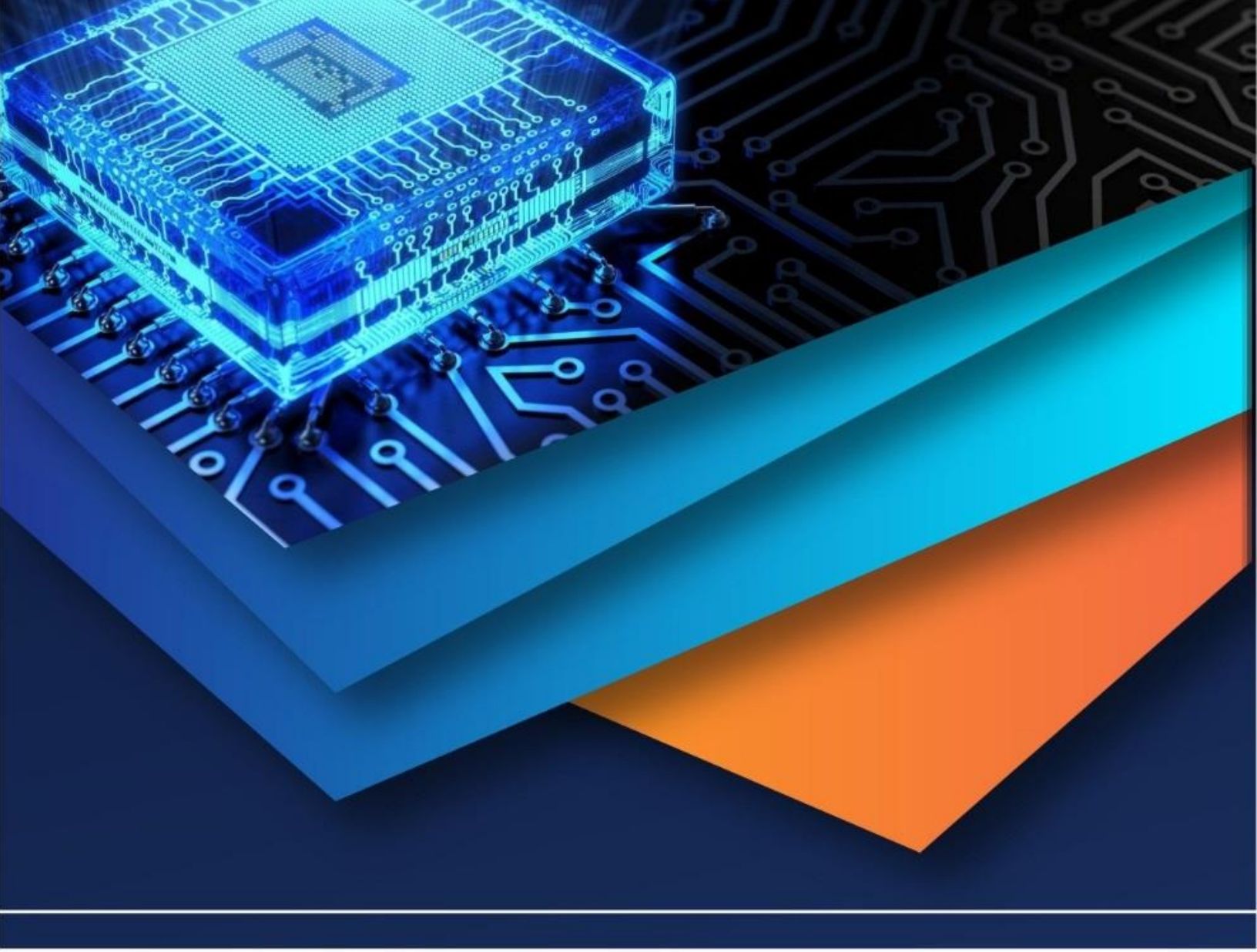
The development of software or website includes so many people like user system developer, user of system and the management, it is important to identify the system requirements by properly collecting required data to interact with supplier and customer of the system. Proper design builds upon this foundation to give a blue print, which is actually implemented by the developers.
On realizing the importance of systematic documentation all the processes are implemented using a software engineering approach. Working in a live environment enables one to appreciate the intricacies involved in the System Development Life Cycle (SDLC). We have gained a lot of practical knowledge from this project, which we think, shall make us stand in a good state in the future.

V. ACKNOWLEDGEMENT

We would like to thank the management of Geetanjali Institute of Technical Studies, Dabok, Udaipur for giving the necessary infrastructure support to smoothly conduct this research work. Also, we would like to thank Dr. Mayank Patel for his guidance at appropriate stages of the work. Finally, we would like to thank Mr. Girish Ameta and Mr. Abhishek Gupta and Ms. Nikita Somani (project Mentor) for their technical support in successful implementation of the project.

REFERENCES

[1] Kaggle.com https://www.kaggle.com/uciml/pima-indias-diabetes-dtabase
[2] www.Google.com
[3] Gill, N.S.; Mittal, P. A computational hybrid model with two level classification using SVM and neural network for predicting the diabetes disease. J. Theor. Appl. Inf. Technol. 2016, 87, 1–10
[4] Sanz, J.A.; Galar, M.; Jurio, A.; Brugos, A.; Pagola, M.; Bustince, H. Medical diagnosis of cardiovascular diseasesusinganinterval-valuedfuzzyrule-basedclassificationsystem. Appl. SoftComput. 2014,20,103–111. [CrossRef]
[5] Varma, K.V.; Rao, A.A.; Lakshmi, T.S.M.; Rao, P.N. A computational intelligence approach for a better diagnosis of diabetic patients. Comput. Electr. Eng. 2014, 40, 1758–1765. [CrossRef]
[6] Data Mining Concepts and Techniques, Third Edition: Jiawei Han, University of Illinois at Urbana–Champaign and Micheline Kamber Jian Pei, Simon Fraser University.
[7] Patrick Ernst, Cynthia Meng, Amy Siu, Gerhard Weikum : "KnowLife: a Knowledge Graph for Health and Life Sciences." 30th International Conference on Data Engineering (ICDE), 2014 IEEE, p p : 1254 - 1257.
[8] www.webmd.com [Accessed 22/10/2015] Mount Adora Hospital & Diagnostic Center, Mirboxtula, Nayashark, Sylhet-3100.
[9] http://patient.info/forums [Accessed 27/10/2015]
[10] Samaneh Moghaddam, Martin Ester: "Aspect-based opinion mining from product reviews.", The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012.
[11] Amit X. Garg, MD; Neill K. J. Adhikari, MD; Heather McDonald, MSc; M. Patricia Rosas-Arellano, MD, PhD; P. J. Devereaux, MD; Joseph Beyene, PhD; Justina Sam, BHSc; R. Brian Haynes, MD, PhD: "Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes", JAMA. 2005;293(10):1223-1238.doi:10.1001/jama.293.10.1223



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY