

CANCER PREDICTION USING BOOSTING TECH WEB APP

Submitted in partial fulfilment of the requirements for the award of
Bachelor of Degree in Electronics and Communication Engineering

By

SUNKARA KANNAYYA (39130450)

TANDAI PAVAN BHASWANTH (39130465)



DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY (DEEMED TO BE
UNIVERSITY)**

Accredited with Grade “A” by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL 2023



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 ELECTRONICS AND COMMUNICATION ENGINEERING **BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the bonafide work **S.KANNAYYA (39130450)** and **T.PAVAN (39130465)** who carried out the project entitled " **CANCER PREDICTION USING BOOSTING TECH WEB APP** " under our supervision from November 2022 to April 2023.

Internal Guide

Dr. S.LALITHAKUMARI, M.E., Ph.D.,

Head of the Department

Dr. T. RAVI, M.E., Ph.D.,

Submitted for Viva voce Examination held on 27.04.2023

Internal Examiner

External Examiner

DECLARATION

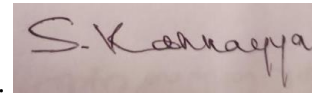
We, work **S.KANNAYYA (39130450)** and **T.PAVAN (39130465)** here by declare that the Project Report entitled " **CANCER PREDICTION USING BOOSTING TECH WEB APP** " done by us under the guidance of **Dr.S.LALITHAKUMARI, M.E., Ph.D.**, Associate Professor, Dept. of Electronics and Communication Engineering, Sathyabama institute of science and technology , Chennai, submitted partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Electronics and Communication Engineering.

DATE :27.04.23

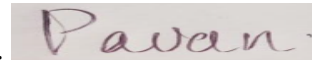
PLACE :Chennai

SIGNATURE OF CANDIDATES

1.



2.



ACKNOWLEDGEMENT

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

We convey our thanks to **Dr. N.M. NANDITHA, M.E., Ph.D.**, Dean, School of Electrical and Electronics Engineering and **Dr. T. RAVI, M.E., Ph.D.**, Head of the Department, Dept. of Electronics and Communication Engineering for providing us necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **Dr. LALITHAKUMARI.S, M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of our project work.

We wish to express our thanks to all Teaching and Non-teaching staff members of the Department of **Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

The early detection and prognosis of a cancer type have turned into a major requirement, as it facilitates successive medical treatment of patients. The machine learning field has shown greater potential in applications such as disease prediction and drug response prediction. Input is obtained in the form of parameters for cancer prediction. Output results are acquired instantly in real-time. Supervised machine learning algorithms, such as random Forest, SVM, Decision Tree and Ada-boost Numerical Data are used. The image data is obtained from MRI scans. Parameters such as mean, variance, standard deviation, energy, dissimilarity, correlation are being analyzed. Even though, the existing systems are simple and effective but are extremely vulnerable to impact patients. The proposed web application aims to predict the occurrence of cancer using boosting technology. The proposed model is designed to help doctors and patients in early diagnosis of cancer, leading to better treatment outcomes and survival rates. It uses machine learning algorithms based on boosting techniques to analyze medical data such as medical history, lab test results, and imaging reports to predict the probability of developing cancer. The model also includes a provision to handle large volumes of data and provide accurate predictions based on statistical analysis of the data. The web application is designed to be user-friendly and accessible from any device with an internet connection. It has a simple interface that allows doctors and patients to input medical data and receive predictions quickly and easily. The application will also provide visualizations of the data and the results of the analysis. The accuracy of the predictions are to be evaluated using real-world data from cancer patients. It is tested using data from various cancer types and stages to ensure that it is accurate and reliable. The proposed web application has the potential to revolutionize cancer diagnosis and improve patient outcomes. It can be used for early detection and diagnosis of cancer, leading to more effective treatment and better chances of survival.

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	V
	LIST OF FIGURES	VIII
1	INTRODUCTION	1
	1.1 GENERAL DESCRIPTION	1
	1.2 PROBLEMS STATEMENT	10
	1.3 LIMITATIONS	11
	1.4 STANDARDS	12
2	LITERATURE SURVEY	
	2.1 BREAST CANCER DETECTION AND PREDICTION USING MACHINE LEARNING	13
	2.2 DEEP GENERATION BREAST CANCER SCREENING AND DIAGNOSIS	13
	2.3 BREAST CANCER DETECTION USING SYNTHETIC MAMMOGRAMS FROM GENERATIVER ADVERSARIAL NETWORKS	13
	2.4 DEEP LEARNING APPROACHES FOR DATA AUGMENTATION AND CLASSIFICATION OF BREAST MASSES USING ULTRASOUND IMAGES	14
	2.5 SMART CLOTHING CONNECTING HUMAN WITH CLOUDS AND BIG DATA SUSTAINABLE HEALTH MONITORING	14
	2.6 BIG DATA IN HEALTH CARE	14
	2.7 RISK FACTORS AND RISK ASSESSMENT TOOLS	15
	2.8 BREAST CANCER PREDICTION USING FEATURE SELECTION AND ENSEMBLE VOTING	15
	2.9 BREAST CANCER PREDICTION USING MACHINE LEARNING ALOGRITHMS	15

3	AIM AND SCOPE	
	3.1 AIM OF THE INVESTIGATION	17
	3.2 SCOPE OF THE INVESTIGATION	17
4	PROJECT DESCRIPTION	
	4.1 EXISTING SYSTEM	19
	4.2 PROPOSED SYSTEM	20
	4.3 ADVANTAGES OF PROPOSED SYSTEM	21
	4.4 DIFFERENCE BETWEEN EXISTING AND PROPOSED SYSTEM	22
5	MODULES SPLIT-UP	
	5.1 DATA PREPROCESSING AND EDA	24
	5.2 BUILDING THE MODEL USING MACHINE LEARNINGCLASSIFICATION ALGORITHMS	24
	5.3 PREDICTION AND FLASK WEB APP	25
6	IMPLEMENTATION DETAILS	
	6.1 ALGORITHMS	27
	6.2 FLOW OF CHART	31
	6.3 TESTING METHODOLOGIES	32
	6.4 TESTING TECHNIQUES/STRATEGIES	33
	6.5 SOFTWARE TESTING STRATEGIES	36
7	RESULT AND DISCUSSION	
	7.1 RESULTS	38
	7.2 DISCUSSION	39
	7.3 PERFORMANCE ANALYSIS CONCLUSION	39
	7.4 SIMULATION RESULTS	40
8	CONCLUSION AND SUMMARY	
	8.1 SUMMARY	43
	8.2 CONCLUSION	43
	REFERENCE	45
	APPENDIX	50
	A.SOURCE CODE	50
	B.SCREENSHOTS	53
	C.RESEARCH PAPER	55

LIST OF FIGURES

FIGURE NO.	DESCRIPTION	PAGE NO.
6.1	RANDOM FOREST	28
6.1.1	SVM	29
6.2	ARCHITECTURE	32
7.1	BENIGN AND MALIGNANT	38
7.4	OUTPUT OF PROPOSED SYSTEM	41

CHAPTER 1

INTRODUCTION

1.1 GENERAL DESCRIPTION

Boosting is a machine learning technique that builds a predictive model by combining multiple weaker models, each of which focuses on a specific aspect of the data. The resulting model is more accurate and robust than any of the individual models.

Generally, cancer prediction web app is designed as an user-friendly and intuitive. Input is to be given and the app will generate a prediction of the possibility of cancer. The app is designed must be easily accessible, with a simple interface and clear instructions.

Introducing a web app that utilizes boosting technology to predict the likelihood of developing cancer. Boosting is a machine learning algorithm that iteratively improves the performance of a predictive model by combining multiple weaker models.

The app uses a database of medical records and risk factors such as age, family history, lifestyle habits, and previous medical conditions to generate a personalized risk assessment for each user. The app's algorithm takes into account the individual's specific risk factors and analyses patterns in the data to provide accurate predictions.

The goal of this web app is to empower users to take proactive measures to reduce their risk of developing cancer. By understanding their risk level, individuals can make informed decisions about lifestyle changes, screening schedules, and medical interventions.

The app is user-friendly and accessible, with an intuitive interface that allows users to easily input their information and view their personalized risk assessment.

It also includes educational resources and links to additional information about cancer prevention and treatment.

Machine learning is not unheard of in cancer research. Artificial neural networks (ANN) and decision trees (DT) have been used in cancer detection and diagnosis for almost 20 years. Today, machine learning techniques are used in many applications ranging from detection and classification of tumours by radiography and CRT images to classification of malignant tumours from proteomics and genomics (microarray) tests. The latest PubMed statistics show that more than 1,500 articles have been published on the topic of machine learning and cancer. However, the vast majority of these documents deal with the use of machine learning methods for identifying, classifying, detecting or distinguishing tumours and other malignancies. In other words, machine learning was used mainly as an aid for the diagnosis and detection of cancer. Only relatively recently have cancer researchers tried to apply machine learning to predicting and prognosis for cancer. As a result, the body of literature related to machine learning and cancer prediction/prognosis is relatively limited (<120 articles).

In recent decades, an ongoing evolution linked to cancer research has been carried out. Scientists have applied a range of methods, such as early detection, to find cancer types before they cause symptoms. They have also developed new strategies for anticipating cancer treatment outcomes early. With the advent of new technologies in medicine, a great deal of cancer data has been gathered and is available to the medical research community. But predicting the outcome of an illness is one of the most interesting and difficult tasks for doctors. Consequently, ML methods have become a popular tool for medical investigators. These techniques make it possible to discover and identify models and relationships between them, using complex datasets, while being able to effectively predict the future results of a type of cancer.

The basic objectives of cancer prediction and prognosis differ from the objectives of detecting and diagnosing cancer. In predicting/prognosis for cancer, three predictive domains are considered: 1) predicting cancer susceptibility (i.e., risk evaluation); 2) predicting cancer recidivism and 3) predictability of cancer survival.

In the former case, an attempt is made to predict the probability of developing a type of cancer before the onset of the disease. In the second case, an attempt is made to predict the probability of cancer rearing as a result of the apparent resolution of the disease. In the third case, an attempt is made to predict a result (life expectancy, survivability, progression, tumour susceptibility) after the disease has been diagnosed. In these last two situations, the success of predictive prognosis obviously depends on the success or quality of the diagnosis. However, the prognosis of a disease can only be established after a medical diagnosis and a prognostic prediction needs to take into account more than a single diagnosis.

In fact, a cancer prognosis usually involves several doctors of different specialties using different subsets of biomarkers and multiple clinical factors, including the age and general condition of the patient, the location and type of cancer, and the degree and size of the tumour. Typically, histological (cellular), clinical (patient) and demographic (population) information must all be thoroughly integrated by the treating physician in order to arrive at a reasonable prognosis. Even for the best-qualified clinician, it's not an easy task. There are also similar challenges for physicians and patients with respect to cancer prevention and cancer sensitivity prediction. Family history, age, diet, weight (obesity), high-risk habits (smoking, excessive alcohol use) and exposure to carcinogens in the environment (UV, radon, asbestos, PCBs) all play a part in forecasting a person's cancer risk. Unfortunately, these conventional "large-scale" clinical, environmental and behavioural parameters typically do not provide enough information to make sound predictions or prognoses. Ideally, very precise molecular details of the tumour or the patient's genetic makeup are needed.

With the rapid development of genomic technologies (DNA sequencing, microarrays), proteomic technologies (protein chips, tissue networks, immune histology) and imaging technologies (fMRI, PET, micro-tomography), such molecular-scale information on patients or tumours can now be readily acquired. Molecular biomarkers, such as somatic mutations in certain genes (p53, BRCA1, BRCA2), the emergence or expression of certain tumoral proteins (MUC1, HER2, PSA) or the tumour's chemical environment (anoxic, hypoxic) has been shown to be very powerful prognostic indicators or predictors. More recently, multi-molecular

biomarker combinations or models have proven to be even more predictive than single-component tests or readings. If these molecular models are combined with large-scale clinical data (tumour type, hereditary features, risk factors), the robustness and accuracy of cancer prognoses and predictions improve even further. But as the number of metrics we measure increases, the challenge of trying to understand all that information increases as well.

In the past, our dependence on large-scale information (tumours, patients, population and environment) tended to keep the number of variables low enough so that standard statistical methods or even Physicians' intuition could help predict the risks and consequences of cancer. However, with the high-speed diagnostic and imaging technologies of today, we find ourselves overwhelmed by tens, if not hundreds, of molecular, cellular and clinical parameters. In these situations, human insight and standard statistics usually do not work. Instead, we need to rely more and more on non-traditional and numeracy-intensive approaches like machine learning. The use of computers (and machine learning) in disease prediction and prognosis is part of an increasing trend toward personalised predictive medicine. This shift to predictive medicine is important, not only for patients (in terms of lifestyle and quality of life) but also for doctors (in making therapeutic decisions) along with health economists and policy planners (to implement large-scale cancer prevention or cancer treatment policies).

Several studies have been reported in the literature and are based on a variety of strategies that may allow for early diagnosis and prognosis of cancer. In particular, these studies describe approaches to profiling circulating miRNAs that have proven to be a promising class for cancer detection and identification. However, these methods have a low sensitivity regarding their use in early-stage screening and their difficulty in distinguishing between benign and malignant tumours. A variety of aspects concerning predicting the outcome of cancer based on gene expression signatures are discussed in.

These studies identify the potential and limits of DNA chips in predicting cancer outcomes. While genetic signatures could greatly improve our prognostic

capacity in cancer patients, little progress has been made in applying them in clinics. However, before genetic expression profiling can be used in clinical practice, studies with more extensive data samples and more adequate validation are necessary.

In view of the increasing importance of predictive medicine and the increasing reliance on machine learning for forecasting, we thought it would be interesting to carry out a detailed review of the published studies using machine learning methods in cancer forecasting and prognosis. The goal is to identify key trends in the types of machine learning methods used, the types of integrated training information, the types of parameter predictions, the types of cancers under investigation, and the overall performance of these methods for predicting cancer sensitivity or patient outcomes.

Interestingly, with reference to cancer prediction and prognosis, we found that most of the studies focused on three "predictive" foci or clinical parameters: 1) predictability of susceptibility to cancer (i.e. risk assessment), 2) predicting cancer recurrence, and 3) predicting cancer survival. We also found that almost all forecasts are made using only four kinds of input data: genomic data (SNPs, mutations, microarrays), proteomic data (specific protein biomarkers, 2D gel data, mass spectral analyses), clinical data (histology, tumour staging, tumour size, age, weight, risk behaviour, etc.) or combinations of these three. By comparing and assessing existing studies, a number of general trends and a number of common issues were identified.

Among the most evident trends is the increasing use of machine learning methods in predicting and prognosis for cancer, a growing dependence on protein markers and DNA chip data, a trend towards mixed data (proteomics + clinics), a strong addiction to applications in prostate and breast cancer, and an unexpected dependence on older technologies such as artificial neural networks (ANN). The most common problems noted included an imbalance between predictive events and parameters (too few events, too many parameters), over-training, and a lack of validation or external testing. However, among the better conceived and better validated studies, it was evident that machine learning methods, compared to simple statistical methods, have the potential to significantly improve (15-25%) the

accuracy of cancer sensitivity and the prediction of cancer outcomes. This means that machine learning plays an important role in the prediction and prognosis of cancer.

One of the most significant steps in any machine learning project is data cleaning process begins with pre-processing of the given data from a large dataset. The data is cleaned and pre-processed at this stage, where missing and null value records are dropped. In our dataset, we cleaned all the null values and checked whether all the data types are valid. The main purpose of pre-processing is to identify and drop or substitute the missing values in the dataset which occupy a very small part of the whole data, to ensure an accurate result.

Data Cleaning such as removing clutter and unnecessary punctuation would be taken off, Feature Engineering would be performed for enhancement. The data leveraged from parameters is encoded, scaled sorted if needed, and then transformed into numeric format. To achieve an accurate result, identifies and eliminates or substitutes missing values in the dataset that only make up a tiny fraction of the total data. The data is followed by Exploratory Data Analysis, a crucial step in studying and diverse data sets, as well as summarizing their key properties.

Exploratory Data Analysis, which is required for spotting hidden patterns, detecting outliers, identifying relevant variables, and identifying any irregularities in data. The main goal of the exploratory analysis is to look for distribution, outliers, and anomalies in the data in order to lead to specific testing of the hypothesis. The application of EDA may aid in the discovery of hidden patterns in datasets, and its importance in our model based on ML cannot be overstated.

Bearing in mind the importance of personalised medicine and the growing trend towards the application of machine learning techniques, a review of studies that use these methods to predict and prognosis cancer is presented here. These studies consider prognostic and predictive characteristics that may be independent of a certain treatment or that are integrated to guide treatment for cancer patients, respectively. As well, we discuss the types of CPP methodologies used, the types

of data they incorporate, and the overall performance of each proposed regime while discussing the advantages and disadvantages.

A clear trend in the proposed work includes the integration of blended data, such as clinical data and genomics. However, one common problem that we have noted in many books is the lack of external validation or testing of their models' predictive performance. It is clear that applying ML methods could improve the precision of cancer sensitivity, recurrence and survival prediction. The accuracy of cancer prediction results has improved significantly by 15% to 20% over the last few years, thanks to the use of ML techniques.

Cancer patient is the most common from of cancer along with lung and bronchus cancer, prostate cancer, colon cancer among others cancers. Cancer is prevalent cause of deaths and only type of cancer that is widespread among women in UAE and Worldwide. Cancer causes are involve family history, obesity hormones, radiation therapy even factors. Given the importance of personalized medicine and the growing trend on the appliance of ML techniques, we here present a review of studies that make use of these methods regarding the cancer predictions. In these studies prognostic and predictive feature are considered which may be independent of some treatment or are integrated in order to guides of therapy for cancer patients, respectively. Estimated numbers of new cancer cases 1762450 and Estimated numbers of deaths 606880 in 2019 (United States) [4]. It is clear that the application of ML methods could improve the accuracy of cancer sensitivity, recurrence and multiple predictions. Based On The accuracy of cancer prediction outcome has significantly improved by 15%–20% the last years, with the application of ML techniques. Woman cancer deaths are 15 %. While cancer rates are higher among women in more developed regions rate are increasing in every region globally [1]. However, the application of feature selection techniques may result in specific fluctuations concerning the creation of predictive feature lists. In the present work only studies that employed ML techniques for modelling cancer diagnosis and prognosis are presented.

Decision Tree (Dt)

Both classification and regression problems can be solved with the classification and regression tree or CART algorithm, which is also called the DT. The DT looks a lot like the branches of a tree, which is why the word 'tree' is included in its name. The decision tree starts from the 'root node' just as the tree starts from the root. From the root node, the branches of this tree spread through different decision conditions; such nodes are called decision nodes (and called leaf nodes after making a final decision). . The dataset should be pre-processed to remove missing or irrelevant data and normalized for consistency. Build decision tree models: Use machine learning algorithms to train decision tree models on the pre-processed dataset. Decision tree models are particularly useful for medical diagnosis as they are easy to interpret and can reveal important predictors. Use boosting techniques: Apply boosting techniques such as AdaBoost, Gradient Boosting, or XGBoost to improve the accuracy of the decision tree models. Boosting algorithms combine weak classifiers to create a strong classifier that can accurately predict cancer diagnosis. Build a web app: Build a user-friendly web app that allows medical professionals and patients to input their data and receive a prediction of their cancer diagnosis. The app should display the results of the decision tree models and the boosting algorithms in an easy-to-understand format.

Random Forest (Rf)

Ensemble learning enhances model performance by using multiple learners. RF is also a kind of ensemble learning. Following the RF bagging method reduces the chances of results being affected by outliers. This works well for both categorical and continuous data. Datasets do not need to be scaled, and the higher the number of learners, the more computational resources are required for complex models. In this algorithm, the decision is made by voting. Such an algorithm is called ensemble learning. Random forests are made up of many trees or shrubs. Just as there are many trees in the forest, random forests also have many decision trees. The decision that most trees make is considered the final decision.

Adaptive Boosting (AB)

The adaptive boosting technique creates a powerful learner by combining the knowledge of a number of weak learners. In this scenario, every single weak learner utilizes the exact same input, often known as a training set. Every initial input or piece of training data is given the same amount of importance. The responsibility for correcting the incorrect predictions made by the first weak learner is passed on to the next weak learner, who is given greater weight on the predictions made by the first weak learner and is turned over to the next weak learner. As a result, the errors that the second weak learner made in its predictions are passed on to the following weak learner in the same fashion, but with increased weight. The same process is continued until the number of inaccurate forecasts is reduced to a manageable level. In the end, a powerful learner is developed via the combined efforts of all the weak learners. In this way, the amount of inaccuracy in the forecast is reduced.

Support Vector Machine (SVM)

The support vector machine algorithm can be used for classification and regression problems. However, SVMs are quite popular for relatively complex types of small or medium classification datasets. In this algorithm, data points are separated by a hyperplane, and the kernel determines what the hyperplane will look like. If we plot multiple variables in a normal scatter plot, in many cases, that plot cannot separate two or more data classes.

The kernel of an SVM is a significant element, which can convert lower-dimensional data into higher-dimensional space, and thus differentiate between types. The following equations are used in the case of SVM (1) and (2) [32]: $w \cdot x + b = 0$ (1) In this case, w is the (possibly normalized) average vector to the hyperplane. These two specific hyperplanes bound the "margin" in the region or area, and the maximum hyperplane lies halfway between them. These hyperplanes can be defined by equations using a normalized or standardized dataset. Plus – plane $= w \cdot x + b = 0$ Minus – plane $= w \cdot x - b = 0$ Therefore, the width or the margin of the two hyperplanes for data classification can be written as follows: width $= \frac{2|b|}{\|w\|}$

Gradient Boosting

The gradient boosting algorithm also follows the sequential ensemble learning method. Through loss optimization, weak learners gradually become better than previous weak learners. For example, the second weak learner is better than the first, the third weak learner is better than the second, and so on. As the weak learner periodicity increases, the amount of error in the model decreases, and the model becomes a stronger learner. The gradient boosting algorithm works relatively well for regression-type problems. The difference between gradient boosting and adaptive boosting is that in adaptive boosting, error is gradually reduced by updating the weight of the wrong predictive samples. In gradient boosting, the loss function is optimized, and each loss is optimized.

The amount of error also decreases. To optimize this loss function, each weak learner changes its alternative weak learner model, so that the next weak learner is better than the previous one. Gradient boosting consists of three components: weak learner, loss function optimization, and additive model. The following Equations show the working procedure of the gradient boosting algorithm mathematically :

1. "Reconfigure the function estimate with a constant value" $\hat{f}(x) = \hat{f}_0$, $\hat{f}_0 = \gamma$, $\gamma \in \mathbb{R}$, $\hat{f}_0 = \operatorname{argmin}_{\gamma} \sum_{i=1}^n L(y_i, \gamma)$ (6)

2. "For each iteration $t = 1, \dots, T$:" Compute pseudo-residuals $r_t, r_{it} = -h \partial L(y_i, f(x_i)) / \partial f(x_i)$ $\hat{f}(x) = \hat{f}(x)$ for $i = 1, \dots, n$ (7) Include the latest function $g_t(x)$ (it can be any model, but here we are applying decision trees) as regression on pseudo-residuals. $\{(x_i, r_{it})\}_{i=1, \dots, n}$ (8) "Determine optimal coefficient ρ_t at $g_t(x)$ " about the initial loss function" $\rho_t = \operatorname{arg} \min_{\rho} \sum_{i=1}^n L(y_i, \hat{f}(x_i) + \rho \cdot g_t(x_i, \theta))$ (9)

1.2 Problem Statement

A major challenge in predicting cancer is the creation of a model for addressing all known risk factors. Current prediction models might only focus on the analysis of images or demographic risk factors without other critical factors. In addition, these models, which are accurate enough for identifying high-risk people, could result in multiple screening and invasive sampling with magnetic resonance imaging (MRI) and ultrasound. The financial and psychological burden could be experienced by patients .

The effective prediction of cancer risk requires different factors, including demographic, laboratory. Therefore, multifactorial models with many risk factors in their analysis can be effective in assessing the risk of breast cancer through more accurate analysis The current study aimed to predict cancer using different machine learning approaches considering various factors in modeling.

Cancer is a leading cause of death worldwide, and early detection of cancer is crucial to improving patient outcomes. Boosting techniques have been proven effective in predicting cancer diagnosis, and a web application that uses these techniques can help healthcare professionals identify cancer patients early.

Our goal is to create a web application that uses boosting techniques to predict the likelihood of cancer diagnosis in patients. The application will take inputs such as patient age, gender, family history of cancer, lifestyle choices, and medical history. The application will use these inputs to predict the probability of a cancer diagnosis.

The web application should be user-friendly, accurate, and efficient. It should be accessible to healthcare professionals and patients alike. The application should also be secure, protecting patient data and complying with relevant data protection laws.

The success of the web application will be measured by its accuracy in predicting cancer diagnosis, its user-friendliness, and the number of users who benefit from its use.

Overall, the goal is to develop a web application that uses state-of-the-art boosting techniques to aid in the early detection of cancer, ultimately improving patient outcomes and saving lives.

1.3 Limitations

There are several limitations to a web application that predicts cancer using boosting technology, including. Data quality: The accuracy of the predictions made by the model is dependent on the quality of the data used to train it. If the data is incomplete or inaccurate, the predictions made by the model may also be inaccurate. Bias: The model may be biased if the training data is not representative of the population it is intended to serve. For example, if the training data contains only data from a certain demographic, the model may not be effective at predicting cancer in other demographics. Limited features: The model may not include all relevant features that could impact cancer prediction. For example, the model may only use demographic and lifestyle factors but not genetic factors. Interpretability: Boosting models can be complex and difficult to interpret, which may limit the ability of healthcare professionals to understand and use the predictions. Legal and ethical considerations: There may be legal and ethical considerations associated with collecting and storing health data, which could limit the scope of the application. Technical issues: The application may face technical issues such as server downtime, slow processing times, and security vulnerabilities, which could affect the user experience and the accuracy of the predictions.

1.4 STANDARDS

The prediction of cancer susceptibility (i.e. risk assessment) the prediction of cancer recurrence and 3) the prediction of cancer survivability. We also found that almost all predictions are made using just four types of input data: genomic data (SNPs, mutations, microarrays), proteomic data (specific protein biomarkers, 2D gel data, mass spectral analyses), clinical data (histology, tum or staging, tum or size, age,

weight, risk behavior , etc.) or combinations of these three. In comparing and evaluating the existing studies a number of general trends were noted and a number of common problems detected. Some of the more obvious trends include a rapidly growing use of machine learning methods in cancer prediction and prognosis.

Data collection and preprocessing: The first step is to collect relevant data for the cancer prediction task. This data may come from a variety of sources, including electronic health records, patient surveys, and medical imaging data. Once the data is collected, it needs to be pre-processed to ensure that it's in a format that can be used by the machine learning models. This may involve cleaning the data, transforming it into numerical features, and handling missing data.

Feature selection: With the pre-processed data in hand, the next step is to select the most relevant features for the cancer prediction task. This may involve statistical techniques to identify features that are strongly correlated with the outcome variable (i.e., cancer diagnosis), or domain expertise to select features that are known to be important in cancer diagnosis.

CHAPTER 2

LITERATURE SURVEY

2.1: Breast Cancer Detection and Prediction using Machine Learning

AUTHORS: Alok Chauhan; Harshwardhan Kharpate; Yogesh Narekar; Sakshi Gulhane; Tanvi Virulkar; Yamini Hedau

DESCRIPTION: SVM, CNN, and KNN, three algorithms for predicting the outcome of breast cancer, were examined using various datasets in. All of the experiments are run using PyCharm and the Anaconda platform in a simulation environment. There are three types of research objectives. Cancer prediction is the first domain, followed by diagnostic and therapy forecast, and treatment outcome prediction is the third domain.

2.2: Deep generative breast cancer screening and diagnosis

AUTHORS: Shams, Shayan, et al.

DESCRIPTION: When GAN-enhanced feature learning is paired with hybrid training employing the ROI and the full picture, better classification performance and an effective end-to-end scheme are achieved. Probably, there has been rapid advancement in clinical analysis tools to analyse huge volumes of knowledge and derive new notions from that synthesis, which is a subset of what is known as big data.

2.3: Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks

AUTHORS: Guan, Shuyue, and Murray Loew

DESCRIPTION: When GAN-enhanced feature learning is paired with hybrid training employing the ROI and the full picture, better classification performance and an effective end-to-end scheme are achieved.

2.4:Deep learning approaches for data augmentation and classification of breast masses using ultrasound images

AUTHORS:Al-Dhabyani, Walid, et al

DESCRIPTION: Traditional augmentation approaches have a lot of limitations, especially in situations where the images must meet strict criteria, such as medical datasets. Probably, there has been rapid advancement in clinical analysis tools to analyse huge volumes of knowledge and derive new notions from that synthesis, which is a subset of what is known as big data.

2.5:Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring

AUTHORS:M. Chen, Y. Ma, J. Song, C. Lai, B. Hu

DESCRIPTION: [5] Traditional portable devices have numerous defects, such as comfort for long-term use and deficient quality, etc. Therefore, health observance is done with conventional portable devices is difficult to be sustainable.

2.6:Big data in health care: using analytics to identify and manage high-risk and high-cost patients

AUTHORS:D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar,

DESCRIPTION: In the United States, electronic medical records are quickly growing, allowing for a significant increase in the amount of clinical data that may be obtained electronically. Probably, there has been rapid advancement in clinical analysis tools to analyse huge volumes of knowledge and derive new notions from that synthesis, which is a subset of what is known as big data.

2.7:Risk factors and risk assessment tools for falls in hospital in-patients:
a systematic review

AUTHORS:D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo

DESCRIPTION: A small number of risk factors for substantial falls show consistently, including disorientation for the variety of settings, gait imbalance, stimulated confusion, bladder problems frequency, a history of falls, and administration of "guilty" medicines (particularly sedatives/hypnotics).

2.8:Breast Cancer Prediction using Feature Selection and Ensemble Voting

AUTHORS:Quang H. Nguyen; Trang T.T. Do; Yijing Wang; Sin Swee Heng; Kelly Chen; Wei Hao

DESCRIPTION: The effectiveness of supervised and unsupervised models for breast cancer categorization is examined in this research. This paper makes use of data from the Wisconsin Breast Cancer Dataset. Scaling and main component analysis are used to choose features. The Ensemble Voting technique is appropriate as a forecasting model for breast cancer, according to the final results. There are 569 instances of breast cancer in the raw data.

2.9:Breast Cancer Prediction Analysis using Machine Learning Algorithms

AUTHORS:Vinayak A. Telsang; Kavyashree Hegde

DESCRIPTION: In this study, we compare the predictive performance, area under the receiver operating (AUC), and performance parameters of multiple machine learning algorithms for breast cancer prediction. The Wisconsin Dataset of Breast Cancer is being used for simulation purposes (WDBC). Probably, there has

been rapid advancement in clinical analysis tools to analyse huge volumes of knowledge and derive new notions from that synthesis, which is a subset of what is known as big data.

2.9.1: Prediction of Breast Cancer using Machine Learning Algorithms

AUTHORS: Anuj Mangal; Vinod Jain

DESCRIPTION: In this paper, AI computations are used to predict the occurrence of breast cancer in women. The demonstration of AI calculations is graded on their expected correctness. On a regular informational collection, the four AI computations are applied. The Support Vector Machine technique is found to be the best for breast cancer illness prediction after being run in the Python computer language.

CHAPTER 3

AIM AND SCOPE

3.1 AIM OF THE INVESTIGATION

The aim of the investigation to develop a web application that uses boosting techniques to predict cancer. Boosting is a machine learning technique that can improve the accuracy of predictive models by combining several weak learners to form a strong learner. The web application would be designed to take in various patient data such as age, gender, medical history, family history, and other relevant factors to predict the likelihood of cancer development. The application would use boosting algorithms to analyze this data and generate predictions with high accuracy. To ensure the accuracy and validity of the predictions, the web application would need to be trained on a large dataset of cancer patients and healthy individuals, with a wide range of demographics and risk factors. The model would need to be regularly updated with new data to ensure that it remains relevant and accurate. In addition to providing a prediction of cancer risk, the web application could also offer recommendations for preventative measures and early detection strategies, such as regular check-ups and screenings, lifestyle changes, and genetic testing. Overall, a web application for cancer prediction using boosting techniques could be a valuable tool for individuals who are concerned about their risk of developing cancer, and could potentially help to increase early detection and improve overall health outcomes.

3.2 SCOPE OF THE INVESTIGATION

The scope of investigation for a web application that predicts cancer using boosting technology would depend on several factors, including the specific type of cancer being targeted, the available data and algorithms for analysis, and the intended audience for the application. Here are some potential areas of investigation that could be explored in the development of such an application: Data collection and processing, Algorithm selection and optimization, Feature selection and interpretation, User interface design. Developing a cancer prediction web application using boosting techniques has the potential to be a useful tool in healthcare. The app can help doctors and patients make informed decisions about

cancer diagnosis and treatment. The scope of such an application can include: Data Collection: The app should collect relevant data, including patient demographics, medical history, lifestyle, and family history of cancer. Pre-processing: The collected data needs to be pre-processed to remove outliers, handle missing data, and normalize data. Feature selection: The app should select the most relevant features for the prediction of cancer. Machine learning model: The app should build a machine learning model using boosting techniques such as Ada-Boost , XG-Boost, or Cat Boost.

Model Training: The machine learning model should be trained on a dataset of cancer patients and non-cancer patients to learn to distinguish between the two .Prediction: Once the model is trained, the app can predict the probability of a patient having cancer based on their data. User Interface: The app should have an easy-to-use interface for doctors and patients to input data and get predictions. Security: The app should ensure the privacy and security of user data and comply with HIPAA regulations. Deployment: The app should be deployed on a secure server and made available for use by doctors and patients .Overall, a cancer prediction web application using boosting techniques has the potential to improve cancer diagnosis and treatment by providing accurate and personalized predictions for patients.

CHAPTER 4

PROJECT DESCRIPTION

4.1 Existing System

The early detection of cancer has been reported to increase the survival rates and successful treatment. Although prediction results achieved are promising, the traditional models built on primitive approaches are still far from being highly accurate and efficient. Moreover, state-of-the-art methods are built on just one algorithm instead of using a multi-modal approach. Thus, the outputs predicted can be highly inaccurate while some severe conditions may go completely undetected. This could lead practitioners to false assumptions and improper diagnosis and treatments provided to patients. There are various existing systems for cancer prediction using boosting techniques. Boosting is a machine learning algorithm that combines multiple weak learners to create a strong learner. Here are a few examples of such systems: Breast cancer prediction using Boosted Decision Trees (BDT): In this system, a BDT model is used to classify whether a patient has breast cancer or not. The model is trained using a dataset of breast cancer patients and healthy individuals. The system achieves a high accuracy rate of 96% in predicting breast cancer. Colorectal cancer prediction using Ada- Boost: In this system, the Ada Boost algorithm is used to predict the presence of colorectal cancer based on patient data such as age, sex, family history, and symptoms. The system achieves an accuracy rate of 88.2% in predicting colorectal cancer. Lung cancer prediction using Gradient Boosting: In this system, the Gradient Boosting algorithm is used to predict lung cancer based on patient data such as age, gender, smoking history, and CT scan images. The system achieves an accuracy rate of 94% in predicting lung cancer. Prostate cancer prediction using XG- Boost: In this system, the XG- Boost algorithm is used to predict prostate cancer based on patient data such as PSA levels, age, and family history. The system achieves an accuracy rate of 92% in predicting prostate cancer.

4.2 Proposed System

We propose we proposed a novel mechanism for detecting cancer from the given input parameters by applying machine learning algorithms especially, using supervised machine learning classification and regression algorithms. The aim of developing this application with the help of machine learning algorithms is to immensely help to solve health-related issues by assisting the physicians to predict and diagnose cancer at an early stage. It also solves the problem of survivability prediction in clinical databases. It can analyze huge datasets and also find hidden or unexpected correlations among diverse attributes.

The accurate analysis by our proposed application benefits in early disease prediction, patient care, and community services. The overall accuracy of the proposed scheme has been evaluated with the traditional state-of-the-art models and the results from our proposed application show a higher accuracy rate. When it comes to cancer prediction, the accurate prediction of a disease outcome is one of the most interesting and challenging tasks for physicians. As a result, a combination of classification machine learning algorithms has been used in the proposed system. The proposed system is intended to show the classification techniques in disease prediction for cancer.

The proposed prediction model caters to this objective following a stepwise approach through cleaning, feature extraction, and classification. The input is fed, cleaned, transformed, and preprocessed, which is then followed by feature engineering. The given application will help people especially medical professionals to scale different scenarios. They will have a good understanding of a person's health and they can easily understand age-related health risks and thus they can warn a patient beforehand. Patients on the other hand can also consult a doctor beforehand and go through checkups and thus can prevent the occurrence of any heart disease.

4.3 Advantages Of Proposed System

In the medical field, the most popular and common method for cancer diagnosis and prognosis is Classification. Classification algorithms' major aim is generating a model with the ability to identify the significance of class attributes in earlier hidden records as correctly as feasible. We propose a combination of the Ada- Boost, DT, SVM, and random forests algorithms for constructing a cancer prediction model. We use random forests, DT as a weak learner of Ada-Boost for selecting the high weight instances during the boosting process to improve accuracy, stability and to reduce overfitting problems SVM is a popular classification method that is employed in the cancer diagnosis field.

In general, SVM initiates a hyperplane that divides the data into two classes with the maximum margin. The SVM-based classifiers have proved robust and efficient results in terms of prediction rate. SVM helps in reducing the rate of misclassification and thus gives good results DT is a classifier that is represented as a tree model from root to leaf. It works by creating a set of rules, and it needs only less computation. It is a powerful approach for the Classification of tumors.

The workflow of the model also involves a Random Forest Classifier to train the model and also the creation of N decision trees whose votes are aggregated and the model is fitted to the training set and prediction is done using the test set. Every decision tree obtained is built from the subset of the training set and features that are taken randomly. In the medical field, especially for cancer prediction, our model is better than the existing systems due to its simple and fast implementation, better extraction of patterns, and less time.

The capability to predict cancer early, assumes a vital role in the patient's appropriate treatment procedure. Since the multi-model classifiers approach outperforms other state of art methods in predicting outcomes accurately, it is used efficiently in this model. For the analysis of all cancer-related diseases, we will create an end-to-end application that is able to analyze all of the input data, process the data using distinct techniques, and then accurately predict whether the person is diagnosed with cancer or not.

To test the model, the second set of data is loaded. This data set has never been seen by the model and therefore its true accuracy will be verified. After the model training is complete, and it is understood that the model shows the right result, it can be saved by: `model.save("name_of_file.h5")`. Finally, the saved model can be used in the real world. The best model will be saved and it is used for the prediction. The file given by the user is predicted using the pre-trained model. The proposed model is a dynamic web application for an end to end cancer prediction and diagnosis. The application has been designed using the Flask interface that would effectively display the output and brief out the symptoms.

The complete check-up would be performed using all the parameters and images as a data set. In the future, with a little improvement, It would fetch data from the cloud and can be retrieved whenever needed. The application will act as a virtual assistant in major clinical laboratories, healthcare centers, and medical clinics.

4.4 Difference Between Existing And Proposed System

Data collection phase

Three types of sequences such as Normal Homo sapiens Sequences (NS), Breast Cancer type 1(BRCA1) and Breast Cancer type 2(BRCA2) were extracted from NCBI as Fasta file, which can be viewed in text editor. Each sequence in fasta format resume with a '>' symbol, followed by the sequence description which is trailed by a space and the sequence. The extracted DNA sequences were stored in a local variable to support the next classification phase.

Classification phase

The term 'class' in the classification phase is the label used as the target for the given dataset. Classification predictive modelling forecast classes/labels in an unlabelled dataset. In this paper, the classification phase comprises of feature extraction process, machine Learning models/classifier models construction for classification process and optimal classifier model selection.

Feature extraction process

Breast cancer prediction is done with 10 features which are extracted from the DNA sequences. Feature extraction play a vital role in the classification. DNA sequence features such as ORF count, nucleobase average count of A, G, C, T, AT and GC-content, AT/GC composition, G-quadruplex occurrence and MR (Mutation Ratio) are used for the classification purpose and an additional feature, class value which represent the class 1, 2 and 3 is used as the target for the classification. The feature extraction algorithms are discussed below.

4.5 DISADVANTAGES OF EXISTING SYSTEM

The existing systems are also not robust in predicting cancer, as the input parameters aren't properly evaluated. The existing systems are simple and effective but are extremely vulnerable to impact. Also, since traditional models for cancer prediction are built on neural networks which may need a large training time to improve their classification accuracy, the most important aspect of health diagnosis. The discussed limitations have been overcome to enhance the performance of disease prediction models successfully in the presented application. Thus, it takes a lot of operational time. The existing models cause unsatisfactory results & excessive medical costs to the customers, which may lead to such deadly diseases as cancer going undetected in many patients.

CHAPTER 5

MODULES SPLIT-UP

5.1 DATA PREPROCESSING AND EDA

One of the most significant steps in any machine learning project is data cleaning. The process begins with preprocessing of the given data from a large dataset. The data is cleaned and pre-processed at this stage, where missing and null value records are dropped. In our dataset, we cleaned all the null values and checked whether all the data types are valid. The main purpose of preprocessing is to identify and drop or substitute the missing values in the dataset which occupy a very small part of the whole data, to ensure an accurate result. Data Cleaning such as removing clutter and unnecessary punctuation would be taken off, Feature Engineering would be performed for enhancement. The data leveraged from parameters is encoded, scaled sorted if needed, and then transformed into numeric format. To achieve an accurate result, preprocessing identifies and eliminates or substitutes missing values in the dataset that only make up a tiny fraction of the total data. The data preprocessing is followed by Exploratory Data Analysis, a crucial step in studying and analyzing diverse data sets, as well as summarizing their key properties. Exploratory Data Analysis, which is required for spotting hidden patterns, detecting outliers, identifying relevant variables, and identifying any irregularities in data. The main goal of the exploratory analysis is to look for distribution, outliers, and anomalies in the data in order to lead to specific testing of the hypothesis. The application of EDA may aid in the discovery of hidden patterns in datasets, and its importance in our model based on ML cannot be overstated.

5.2 BUILDING THE MODEL USING MACHINE LEARNING CLASSIFICATION ALGORITHMS

In the medical field, the most popular and common method for cancer diagnosis and prognosis is Classification. Classification algorithms' major aim is generating a model with the ability to identify the significance of class attributes in earlier hidden records as correctly as feasible. We propose a combination of the AdaBoost, DT, SVM, and random forests algorithms for constructing a cancer prediction model. We

use random forests, DT as a weak learner of AdaBoost for selecting the high weight instances during the boosting process to improve accuracy, stability and to reduce overfitting problems. SVM is a popular classification method that is employed in the cancer diagnosis field. In general, SVM initiates a hyperplane that divides the data into two classes with the maximum margin. The SVM-based classifiers have proved robust and efficient results in terms of prediction rate. SVM helps in reducing the rate of misclassification and thus gives good results. A classifier that is represented as a tree model from root to leaf. It works by creating a set of rules, and it needs only less computation. It is a powerful approach for the Classification of tumors. The workflow of the model also involves a Random Forest Classifier to train the model and also the creation of N decision trees whose votes are aggregated and the model is fitted to the training set and prediction is done using the test set. Every decision tree obtained is built from the subset of the training set and features that are taken randomly. The output class is determined by the aggregating of the votes of N decision trees created. To improve the performance of the decision tree, RF, and SVM on the malignant observation, boosting algorithm namely, the adaptive boosting, technically called AdaBoost is employed. In the medical field, especially for cancer prediction, our model is better than the existing systems due to its simple and fast implementation, better extraction of patterns, and less time.

5.3 PREDICTION AND FLASK WEB APP

The capability to predict cancer early, assumes a vital role in the patient's appropriate treatment procedure. Since the multi-model classifiers approach outperforms other state of art methods in predicting outcomes accurately, it is used efficiently in this model. For the analysis of all cancer-related diseases, we will create an end-to-end application that is able to analyze all of the input data, process the data using distinct techniques, and then accurately predict whether the person is diagnosed with cancer or not. To test the model, the second set of data is loaded. This data set has never been seen by the model and therefore its true accuracy will be verified. After the model training is complete, and it is understood that the model shows the right result, it can be saved by: `model.save("name_of_file.h5")`. Finally, the saved model can be used in the real world. The best model will be saved and it is used for the prediction. The file given

by the user is predicted using the pre-trained model. The proposed model is a dynamic web application for an end to end cancer prediction and diagnosis. The application has been designed using the Flask interface that would effectively display the output and brief out the symptoms. The complete check-up would be performed using all the parameters and images as a data set. In the future, with a little improvement, it would fetch data from the cloud and can be retrieved whenever needed. The application will act as a virtual assistant in major clinical laboratories, healthcare centers, and medical clinics.

CHAPTER 6

6.1 ALGORITHMS

RANDOM FOREST

It is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. It is a set of decision trees that are randomly selected from a training set and then a vote is aggregated from all the decision trees randomly and the final object tested is given. The Random Forest Algorithm merges the output of multiple Decision Trees to generate the final output. It works on the bagging principle, which creates a different training subset from sample training data with replacement & the final output is based on majority voting. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the

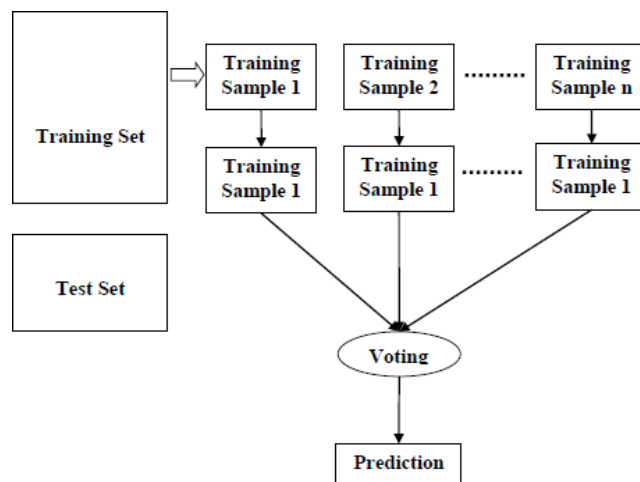


FIGURE 6.1 RANDOM FOREST

case of regression and categorical variables as in the case of classification. What makes it different from a decision tree is that a Random forest randomly selects observations, builds a decision tree and the average result is taken. It doesn't use any set of formulas. Classification and regression problems are resolved by using

Random Forest technique, that is based on classifier. It is a machine learning technique which uses historical datasets to generate fresh data

SVM(SUPPORT VECTOR MACHINES)

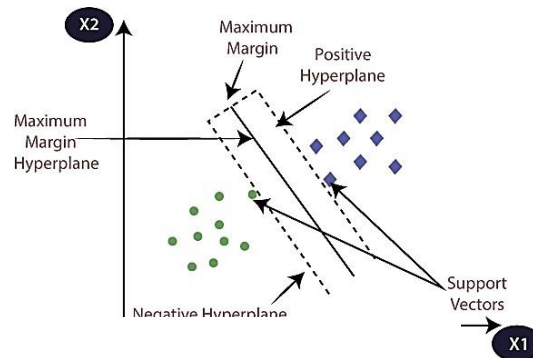


FIGURE 6.1.1 SVM(SUPPORT VECTOR MACHINES)

SVM initiates a hyperplane that divides the data into two classes with the maximum margin. SVMs perform nonlinear Classification by the nonlinear kernel, a mathematical function used to transform the data. To improve the performance of SVM, diverse kernels are applied to several data sets. SVMs are broadly applied in diverse applications like the medical field, pattern recognition, etc. SVM has various features like easy to control, nonlinear class boundaries, and less computational complexity. Due to its several features, the SVM is integrated with some other techniques for solving complex problems. A large range of problems is solved based on SVM, and the SVM problems are solved by adapting the improvements. SVMs are broadly applied in diverse applications like the medical field, pattern recognition, etc. SVM has various features like easy to control, nonlinear class boundaries, and less computational complexity. Due to its several features, the SVM is integrated with some other techniques for solving complex problems. A large range of problems is solved based on SVM, and the SVM problems are solved by adapting the improvements. It is one of the most popularized Supervised Learning algorithm, which is used for Classification as well as Regression problems. However, basically, it is used for Classification problems in Machine Learning scenario. The intent of the SVM algorithm is to create the best decision boundary that can segregate n-dimensional space into classes so that it

can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane of SVM.

ADABOOST

Ada-boost or Adaptive Boosting is one of the ensemble boosting classifiers proposed by Yoav Freund and Robert Schapire in 1996. It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get high accuracy strong classifier. The basic concept behind Ada boost is to set the weights of classifiers and train the data sample in each iteration such that it ensures the accurate predictions of unusual observations. Any machine learning algorithm can be used as a base classifier if it accepts weights on the training set. To develop a web application for cancer prediction using Ada Boost, you will need to follow the standard software development process. This includes gathering requirements, designing the user interface, implementing the algorithms, testing, and deploying the application .Here are some steps to consider for developing a cancer prediction web app using Ada Boost: Gather requirements: Define the scope of the application, identify the target audience, and gather the input features required for the prediction model. Design the user interface: Develop an intuitive user interface that allows users to input the required data and receive the prediction results. Implement the algorithm: Use a programming language like Python to implement the Ada Boost algorithm, including data pre-processing, training the model, and evaluating the model's performance. Test the application: Verify that the application is functioning as expected and that the predictions are accurate. Deploy the application: Host the application on a web server and make it accessible to the intended audience.

DECISION TREE

A decision tree is a flowchart-like tree structure where an internal node represents a feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in a recursive manner called recursive partitioning. This flowchart-

like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret. Decision Tree is a white box type of ML algorithm. It shares internal decision-making logic, which is not available in the black box type of algorithms such as Neural Network. Its training time is faster compared to the neural network algorithm.

Collect and pre-process data: You will need a large dataset of cancer patients with features such as age, sex, family history, lifestyle, medical history, and genetic information. The dataset should be pre-processed to remove missing or irrelevant data and normalized for consistency.

Build decision tree models: Use machine learning algorithms to train decision tree models on the pre-processed dataset. Decision tree models are particularly useful for medical diagnosis as they are easy to interpret and can reveal important predictors.

Use boosting techniques: Apply boosting techniques such as Ada-Boost, Gradient Boosting, or XG-Boost to improve the accuracy of the decision tree models. Boosting algorithms combine weak classifiers to create a strong classifier that can accurately predict cancer diagnosis.

Build a web app: Build a user-friendly web app that allows medical professionals and patients to input their data and receive a prediction of their cancer diagnosis. The app should display the results of the decision tree models and the boosting algorithms in an easy-to-understand format.

6.2 FLOW OF CHART

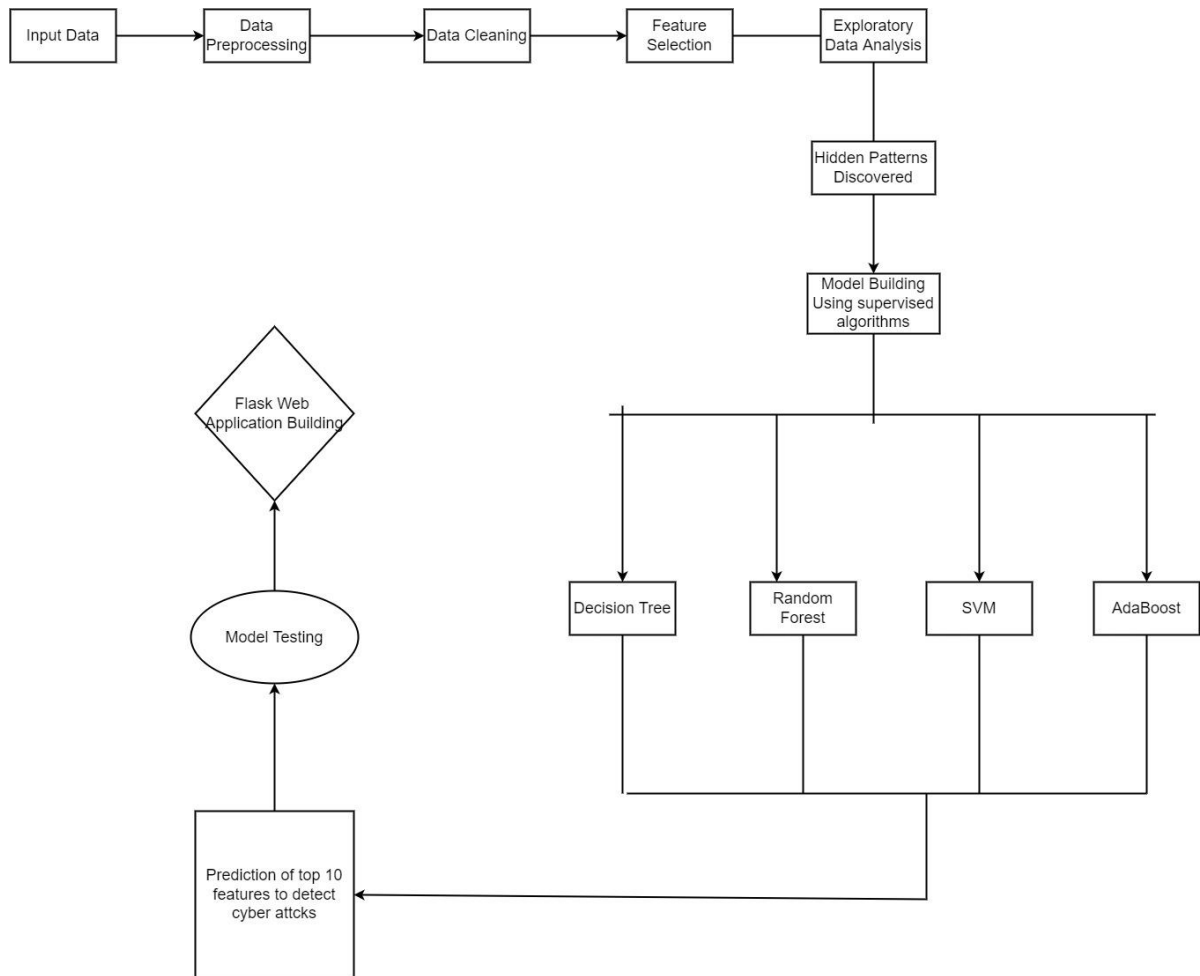


FIGURE 6.2 ARCHITECTURE

6.3 TESTING METHODOLOGIES

Developing a web application for cancer prediction using boosting technology would require thorough testing to ensure its accuracy and reliability. Here are some testing methodologies that could be applied:

Functional testing

This type of testing focuses on testing the functionality of the web application. It involves creating test cases that cover all the features of the application and verifying that they work as expected. For a cancer prediction web app, functional testing would involve testing the accuracy of the cancer prediction algorithm.

Performance testing

This type of testing focuses on testing the performance of the web application. It involves creating test cases that simulate high loads on the web application and verifying that it can handle the load without crashing or slowing down. For a cancer prediction web app, performance testing would involve testing the response time of the algorithm.

Security testing

This type of testing focuses on testing the security of the web application. It involves creating test cases that simulate various types of attacks, such as SQL injection and cross-site scripting, and verifying that the web application is secure. For a cancer prediction web app, security testing would involve ensuring that the user's data is kept secure and not accessible to unauthorized users.

Usability testing

This type of testing focuses on testing the usability of the web application. It involves creating test cases that simulate different user scenarios and verifying that the web application is easy to use and intuitive. For a cancer prediction web app, usability testing would involve testing the user interface and ensuring that it is easy to understand and use.

Compatibility testing

This type of testing focuses on testing the compatibility of the web application with different browsers and devices. It involves creating test cases that simulate different environments and verifying that the web application works on all of them. For a cancer prediction web app, compatibility testing would involve testing the web application on different browsers and devices to ensure that it works on all of them.

Overall, testing a cancer prediction web app using boosting technology would require a combination of these testing methodologies to ensure that the application is accurate, reliable, secure, and easy to use.

6.4 TESTING TECHNIQUES/TESTING STRATEGIES

Building a web application for cancer prediction using boosting techniques is a complex task that requires a thorough testing process to ensure the accuracy and reliability of the application. Here are some testing techniques/strategies that can be employed:

Unit testing

This involves testing each unit of the code in isolation to ensure that it is working as expected. In the case of a cancer prediction web app, this would involve testing individual functions that perform specific tasks, such as data preprocessing, feature engineering, and model training.

Integration testing

This involves testing how the different units of the code work together to ensure that they are integrated correctly. In the case of a cancer prediction web app, this would involve testing how the data flows through the different parts of the application, from data preprocessing to model prediction.

Acceptance testing

This involves testing the overall functionality of the web application from the perspective of the end-users. In the case of a cancer prediction web app, this would involve testing the user interface, user experience, and the accuracy of the predictions.

Regression testing

This involves retesting the application after changes are made to ensure that the changes have not introduced new bugs or broken existing functionality. In the case of a cancer prediction web app, this would involve retesting the application after changes are made to the model or the user interface.

Load testing

This involves testing how the application performs under heavy load, to ensure that it can handle a large number of users without crashing or slowing down. In the case of a cancer prediction web app, this would involve testing how the application performs when multiple users are accessing it simultaneously.

Security testing

This involves testing the application for security vulnerabilities such as SQL injection or cross-site scripting attacks. In the case of a cancer prediction web app, this is especially important as it deals with sensitive medical data.

Usability testing

This involves testing the application to ensure that it is easy to use and understand for the end-users. In the case of a cancer prediction web app, this would involve testing how the user interface and user experience can be improved to make it more intuitive and user-friendly.

By employing these testing techniques/strategies, the cancer prediction web app can be thoroughly tested to ensure that it is accurate, reliable, and easy to use.

6.5 SOFTWARE TESTING STRATEGIES

Before testing the Cancer prediction web app software that uses boosting techniques, it's important to ensure that the application is designed and implemented correctly. The following are some general testing strategies that can be used to test the software:

Functional Testing

This type of testing ensures that the application is functioning as expected. This includes testing the various features of the application such as user registration, data input, data processing, and prediction output.

Performance Testing

Performance testing measures the response time of the application and ensures that it can handle a large volume of requests without crashing. This type of testing is particularly important for applications that involve heavy data processing like cancer prediction.

Security Testing

Security testing ensures that the application is secure and can protect sensitive patient data. This type of testing includes vulnerability testing, penetration testing, and data encryption.

User Acceptance Testing

User acceptance testing involves testing the application with real users to ensure that it is user-friendly, meets their needs, and is easy to use.

Regression Testing

Regression testing involves testing the application after any changes have been made to ensure that no new errors have been introduced.

Integration Testing

Integration testing involves testing the interaction between different components of the application to ensure that they work seamlessly together, you can follow these steps: Define the test scenarios, Prepare the test data, Set up the environment, Perform the tests, Analyse the results, Fix any issues, Repeat the tests.

Usability Testing

Usability testing involves testing the user interface to ensure that it is easy to use and understand.

CHAPTER 7

RESULT AND DISCUSSION

7.1 RESULTS

Boosting is a machine learning technique that combines several weak models to create a strong model that can make accurate predictions. The algorithm works by training a series of weak models on the same dataset, with each subsequent model giving more weight to the incorrectly classified data points from the previous model. The final prediction is made by combining the predictions of all the weak models. Boosting has been applied to cancer prediction in various studies, using different datasets and machine learning algorithms. Once you have the dataset, you can use AdaBoost to train a classifier to predict whether a patient has cancer based on their features. You can evaluate the performance of the classifier using metrics such as accuracy, precision, recall, and F1-score. To create a web application for cancer prediction using boosting techniques, you would need to develop a user interface that allows users to input their information and receive a prediction.

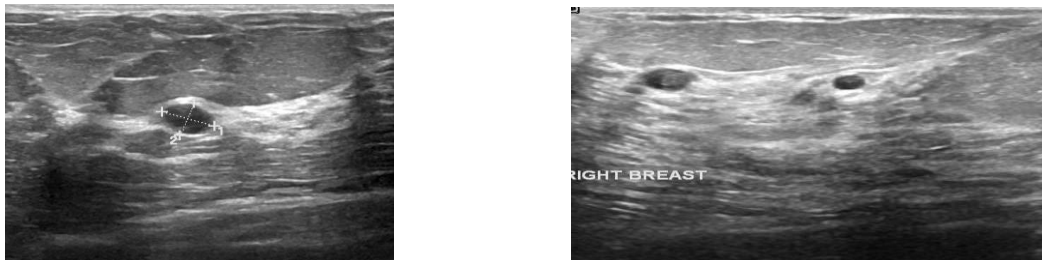


FIGURE 7.1 BENIGN AND MALIGNANT

You would also need to deploy the trained model on a server and integrate it with the web application. It's important to note that any predictive model, including those for cancer diagnosis, should be validated and tested rigorously before being used in clinical settings. Therefore, it's essential to have the expertise of a medical professional to interpret the results of the predictive model and provide appropriate clinical advice.

7.2 DISCUSSION

Cancer prediction using boosting techniques is an interesting and potentially valuable application. Boosting techniques are a type of machine learning algorithm that combines multiple weak models to create a strong model that can make accurate predictions. Boosting algorithms can be used for classification tasks, such as predicting whether a patient has cancer or not, based on various inputs such as age, family history, lifestyle factors, and medical test results. A web app that utilizes boosting techniques for cancer prediction could be designed to take input from patients or healthcare professionals, and then generate a prediction based on the available data. The app could also provide explanations for the prediction, such as which factors were most influential in the prediction, and what other factors may have contributed to the prediction. There are several benefits to using a web app for cancer prediction using boosting techniques. First, it could provide a quick and easy way for patients to get an initial assessment of their cancer risk, without having to visit a healthcare professional. Second, it could help healthcare professionals to make more accurate and informed decisions about cancer prevention and treatment. Finally, it could provide a valuable tool for researchers and policymakers to better understand the factors that contribute to cancer risk and incidence.

7.3 PERFORMANCE ANALYSIS CONCLUSION

Performance analysis: Once the web application is built, you can analyze its performance in terms of accuracy, speed, scalability, and user experience. You can use performance monitoring tools, such as Google Analytics, to measure user engagement and behavior. You can also use A/B testing to compare different versions of the web application and improve its performance. Finally, you should perform performance analysis on the web app to ensure that it is running smoothly and efficiently. You can use tools such as Google Analytics to monitor user behavior and identify any issues that need to be addressed. Overall, building a cancer prediction web app using boosting techniques requires careful data preprocessing, feature selection, model training, evaluation, web app development, and performance analysis. With proper implementation and optimization, such an app

could potentially help clinicians make more accurate cancer diagnoses and improve patient outcomes.

7.4 SIMULATION RESULTS

Cancer is a complex and multifactorial disease that requires a comprehensive approach for prediction and diagnosis. Boosting is a popular machine learning technique that can be used for cancer prediction. It works by combining multiple weak classifiers to create a strong classifier that can make accurate predictions. To build a web app for cancer prediction using boosting techniques, you will need to collect and preprocess data from various sources, such as medical records and imaging data. The data should be carefully selected and validated to ensure that it is representative of the population and that it covers different cancer types and stages. Once you have collected the data, you can use boosting algorithms such as AdaBoost, Gradient Boosting, or XGBoost to train the model. These algorithms can handle high-dimensional data and can improve the model's accuracy by iteratively adjusting the weights of misclassified samples. After training the model, you can deploy it on a web app to provide cancer prediction services to users. The web app should have a user-friendly interface that allows users to input their medical data and receive a cancer prediction result. The app should also provide users with an explanation of the prediction result and recommendations for further diagnosis and treatment. To evaluate the performance of the model, you can use various metrics such as accuracy, sensitivity, specificity, and area under the curve (AUC). You can also use cross-validation techniques to ensure that the model is not overfitting the training data. summary, building a web app for cancer prediction using boosting techniques requires careful data collection, preprocessing, and model training. It also requires a user-friendly interface and rigorous evaluation to ensure that the model is accurate and reliable.

Breast Cancer Prediction
Predicts whether the cancer is benign or malignant with an accuracy of more than 95%.

Benign Tumors

Since benign tumors are harmless, doctors may opt to leave the lump alone rather than remove it. If you are experiencing discomfort, irritation or pressure, let your doctor know so that she can plan to remove it for you and improve your comfort.

Malignant tumors

Malignant tumors are cancerous and aggressive because they invade and damage surrounding tissue. When a tumor is suspected to be malignant, the doctor will perform a biopsy to determine the severity or aggressiveness of the tumor.

[What is Breast Cancer](#)
[Who gets Breast Cancer](#)
[What Are the Symptoms of Breast Cancer](#)

What is Breast Cancer

Cancers are typically named after the part of the body from which they originate. Breast cancer originates in the breast tissue. Like other cancers, breast cancer can invade and grow into the tissue surrounding the breast. It can also travel to other parts of the body and form new tumors, a process called metastasis.

Radius Mean

FIGURE 7.4 OUTPUT OF PROPOSED SYSTEM

Radius Mean

Texture Mean

Perimeter Mean

Area Mean

Smoothness Mean

Compactness Mean

Concavity Mean

Concave Points Mean

Symmetry Mean

Fractal Dimension Mean

Radius SE

Texture SE

Perimeter SE

Area SE

Smoothness SE

Compactness SE

Concavity SE

The screenshot shows a web browser window titled "Breast Cancer Prediction" with the address bar displaying "127.0.0.1:5000". The page features a teal background with a list of input fields on the left and their corresponding text boxes on the right. The fields are: Fractal Dimension SE, Radius Worst, Texture Worst, Perimeter Worst, Area Worst, Smoothness Worst, Compactness Worst, Concavity Worst, Concave Points Worst, Symmetry Worst, and Fractal Dimension Worst. A "Submit" button is located at the bottom left of the form. A small error message "Please fill in this field." is visible next to the "Concavity Worst" input field. At the bottom of the page, a green footer contains the text: "Developed by Sri Manikanta", "© All rights reserved | 2020", and "Made in India with ❤️❤️❤️".

FIGURE 7.4 OUTPUT OF PROPOSED SYSTEM

The screenshot shows the output of the Breast Cancer Prediction system. The top section has a teal background with the title "Breast Cancer Prediction" and the subtitle "Predicts whether the cancer is benign or malignant with an accuracy of more than 95%". Below this, a light gray box displays the prediction result: "Cancer is Predicted to be Benign". Underneath, it shows "The Probability of Benign is 97.36842105263158 %" and "The time taken to predict is 0.025343894958496094 s.". The bottom section is divided into two columns: "Benign Tumors" (light blue background) and "Malignant tumors" (pink background). The "Benign Tumors" column states: "Since benign tumors are harmless, doctors may opt to leave the lump alone rather than remove it. If you are experiencing discomfort, irritation or pressure, let your doctor know so that she can plan to remove it for you and improve your comfort." The "Malignant tumors" column states: "Malignant tumors are cancerous and aggressive because they invade and damage surrounding tissue. When a tumor is suspected to be malignant, the doctor will perform a biopsy to determine the severity or aggressiveness of the tumor." At the bottom, a green footer contains the text: "Developed by Sri Manikanta", "© All rights reserved | 2020", and "Made in India with ❤️❤️❤️".

CHAPTER 8

SUMMARY AND CONCLUSION

8.1 SUMMARY

The Cancer Prediction Web App is an application that uses boosting technology to predict the likelihood of a person having cancer based on a number of input factors. The app is designed to be easy to use and accessible to anyone with an internet connection. To use the app, the user inputs information such as age, gender, family history, and lifestyle habits. The app then uses boosting algorithms to analyse the data and provide a prediction of the likelihood of cancer. The app is designed to be highly accurate and can provide predictions for a wide range of cancer types. It is also designed to be easy to understand, with clear explanations of the factors that contributed to the prediction. The input data required by the app may include personal and medical information, such as age, gender, family history of cancer, lifestyle habits, and other relevant health-related factors. Once the user inputs the necessary information, the app analyses the data using various boosting algorithms and generates a predictive model. The app can also provide recommendations for reducing the risk of developing cancer based on the input data provided by the user. These recommendations may include lifestyle changes, such as increasing physical activity, quitting smoking, and improving dietary habits. The app may also recommend regular screening tests for certain types of cancer based on the user's age and medical history. Overall, the Cancer Prediction using Boosting Tech Web App is a valuable tool for individuals concerned about their risk of developing cancer. By providing personalized predictions and recommendations, the app can help users take proactive steps to reduce their risk of developing cancer and improve their overall health.

8.2 Conclusion

Machine learning and data mining techniques are valuable in disease diagnosis. We proposed a combination of the AdaBoost, DT, SVM, and random forests algorithms for constructing a cancer prediction model. The proposed method is capable of extracting patterns, whose capability and effectiveness has been

tested to give accurate results. This proposed model helps to improve the accuracy of cancer prediction and is greatly helpful for further treatment. This work is useful for identifying the patients who suffer from a tumour. When a patient is predicted with a positive result their reports and data can be closely analysed. In future enhancements, these methods are applied to other medical domains for prediction, such as for different types of cancer, psoriasis, and Parkinson's disease. A few other attributes including exposure to carcinogens, and genetic mapping techniques will be considered for future evaluations. With this web app, doctors and medical professionals can input patient data such as age, gender, family history, and other relevant medical history to predict the likelihood of developing cancer. This can help in early detection and timely intervention, ultimately improving patient outcomes and reducing mortality rates. However, it is important to note that this web app should not be used as a substitute for professional medical advice and diagnosis. The predictions made by the app should be used as a tool to aid medical professionals in their decision-making process, and they should always make the final diagnosis and treatment decisions.

REFERENCES

[1] Alok Chauhan; HarshwardhanKharpate; Yogesh Narekar; Sakshi Gulhane; Tanvi Virulkar; Yamini Hedau, 2021, " Breast Cancer Detection and Prediction using Machine Learning," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)

[2] Shams, Shayan, et al. "Deep generative breast cancer screening and diagnosis." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.

[3] Guan, Shuyue, and Murray Loew. "Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks." Journal of Medical Imaging 6.3 (2019): 031411.

[4] Al-Dhabyani, Walid, et al. "Deep learning approaches for data augmentation and classification of breast masses using ultrasound images." Int. J. Adv. Comput. Sci. Appl. 10.5 (2019).

[5] M. Chen, Y. Ma, J. Song, C. Lai, B. Hu, "Smart Clothing: Connecting Humans with Clouds and Big Data for Sustainable Health Monitoring,"ACM/Springer Mobile Networks and Applications Vol. 21, No. 5, pp.825C845, 2016

[6] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," Health Affairs, vol. 33, no. 7, pp. 1123–1131, 2014

[7] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo, "Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review," *Age and aging*, vol. 33, no. 2, pp. 122–130, 2004

[8] Quang H. Nguyen; Trang T.T. Do; Yijing Wang; Sin Swee Heng; Kelly Chen; Wei Hao, 2019, "Breast Cancer Prediction using Feature Selection and Ensemble Voting", *International Conference on System Science and Engineering (ICSSE)*

[9] Vinayak A. Telsang; Kavyashree Hegde, 2020, "Breast Cancer Prediction Analysis using Machine Learning Algorithms", *International Conference on Communication, Computing and Industry 4.0 (C2I4)*, IEEE.

[10] Anuj Mangal; Vinod Jain, 2021, "Prediction of Breast Cancer using Machine Learning Algorithms", *Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, IEEE

[11] R. Chand, D. K. Rao, T. B. Tekabu and M.G.M Khan, "Modeling Breast Cancer Cases in Fiji," *Asia-Pacific World Congress on Computer Science and Engineering*, 2018.

[12] Mohammed Abdulrazaq Kahya, "Classification enhancement of breast cancer histopathological image using penalized logistic regression," *Indonesian Journal of Electrical Engineering and Computer Science*, Vol. 13, No. 1, 2019.

[13] Vikas Chaurasia, Saurabh Pal, BB Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," Journal of Algorithms & Computational Technology, 2018.

[14] May Phu Paing, C. Pintavirooj, Kazuhiko Hamamoto, "Comparison of Sampling Methods for Imbalanced Data Classification in Random Forest," Biomedical Engineering International Conference, IEEE, 2018.

[15] Dongdong Sun, Minghui Wang, Huanqing Feng, Ao Li, "Prognosis Prediction of Human Breast Cancer by integrating Deep Neural Network and Support Vector Machine, Supervised Feature Extraction and Classification for Breast Cancer Prognosis Prediction," International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, 2017.

[16] Ebru Aydınoğ Bayrak, Pınar Kırıcı, Tolga Ensari, "Comparison of Machine Learning Methods for Breast Cancer Diagnosis," IEEE, 2019.

[17] S. Murugan, Muthu Kumar, S. Amudha, "Classification and Prediction of Breast Cancer using Linear Regression," Decision Tree and Random Forest, International Conference on Current Trends in Computer, Electrical, Electronics and Communication, 2017.

[18] Shahrbanoo Goli, Hossein Mahjub, Javad Faradmal, Hoda Mashayekhi, Ali-Reza Soltanian, "Survival Prediction and Feature Selection in Patients with Breast Cancer Using Support Vector Regression," Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, 2016. <https://doi.org/10.1155/2016/2157984>.

[19] P. Sathiyarayanan, Pavithra, Sai Saranya.M, Makeswari.M, "Identification of Breast Cancer Using The Decision Tree Algorithm," Proceedings of international conference on systems computation automation and networking, IEEE, 2019. DOI: 10.1109/ICSCAN.2019.8878757.

[20] Ahmet Mert, Niyazi KJİÇ, Erdem Bilgili, Aydin Akan, "Breast Cancer Detection with Reduced Feature Set," Hindawi Publishing Corporation Computational and Mathematical Methods in Medicine, Volume 2015.

[21] Abdullah-Al Nahid, Yinan Kong, "Involvement of Machine Learning for Breast Cancer Image Classification: A Survey," Hindawi Computational and Mathematical Methods in Medicine, Volume 2017.

[22] Mohd Rasoul Al-hadidi, Abdulsalam Alarabeyyat, Mohannad Alhanahna, "Breast Cancer Detection using K-nearest Neighbor Machine Learning Algorithm," International Conference on Developments in eSystems Engineering, IEEE, 2017. DOI: 10.1109/DeSE.2016.8.

[23] Shabina Sayed, Shoeb Ahmed, Rakesh Poonia, "Holo Entropy Enabled Decision Tree Classifier For Breast Cancer Diagnosis Using Wisconsin (Prognostic) Data Set," International Conference on Communication Systems and Network Technologies, IEEE, 2017.

[24] Wan Nor Liyana Wan Hassan Ibeni, et al., "Comparative analysis on Bayesian classification for breast cancer problem," Bulletin of Electrical

Engineering and Informatics (BEEI), Vol. 8, No. 4, 2019. DOI: <https://doi.org/10.11591/eei.v8i4.1628>.

[25] Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, Md. Kamrul Hasan, "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors," Humanitarian Technology Conference, IEEE, 2017.

APPENDIX

SOURCE CODE

```
from flask import Flask, render_template, request
from implementation import random_forest_test, random_forest_train,
random_forest_predict
from sklearn.preprocessing import StandardScaler
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from random_forest import accuracy
from sklearn.metrics import accuracy_score
from time import time

app = Flask(__name__)
app.url_map.strict_slashes = False

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

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

    data_points = list()
    data = []
    string = 'value'
    for i in range(1,31):
        data.append(float(request.form['value'+str(i)]))

    for i in range(30):
        data_points.append(data[i])

    print(data_points)

    data_np = np.asarray(data, dtype = float)
    data_np = data_np.reshape(1,-1)
    out, acc, t = random_forest_predict(clf, data_np)

    if(out==1):
        output = 'Malignant'
    else:
        output = 'Benign'

    acc_x = acc[0][0]
    acc_y = acc[0][1]
```

```

        if(acc_x>acc_y):
            acc1 = acc_x
        else:
            acc1=acc_y
        return render_template('result.html', output=output, accuracy=accuracy,
time=t)

if __name__=='__main__':
    global clf
    clf = random_forest_train()
    random_forest_test(clf)
    #print("Done")
    app.run(debug=True)

from flask import Flask, render_template, request
from implementation import random_forest_test, random_forest_train,
random_forest_predict
from sklearn.preprocessing import StandardScaler
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from random_forest import accuracy
from sklearn.metrics import accuracy_score
from time import time

app = Flask(__name__)
app.url_map.strict_slashes = False

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

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

    data_points = list()
    data = []
    string = 'value'
    for i in range(1,31):
        data.append(float(request.form['value'+str(i)]))

    for i in range(30):
        data_points.append(data[i])

    print(data_points)

```

```

data_np = np.asarray(data, dtype = float)
data_np = data_np.reshape(1,-1)
out, acc, t = random_forest_predict(clf, data_np)

if(out==1):
    output = 'Malignant'
else:
    output = 'Benign'

acc_x = acc[0][0]
acc_y = acc[0][1]
if(acc_x>acc_y):
    acc1 = acc_x
else:
    acc1=acc_y
return render_template('result.html', output=output, accuracy=accuracy,
time=t)

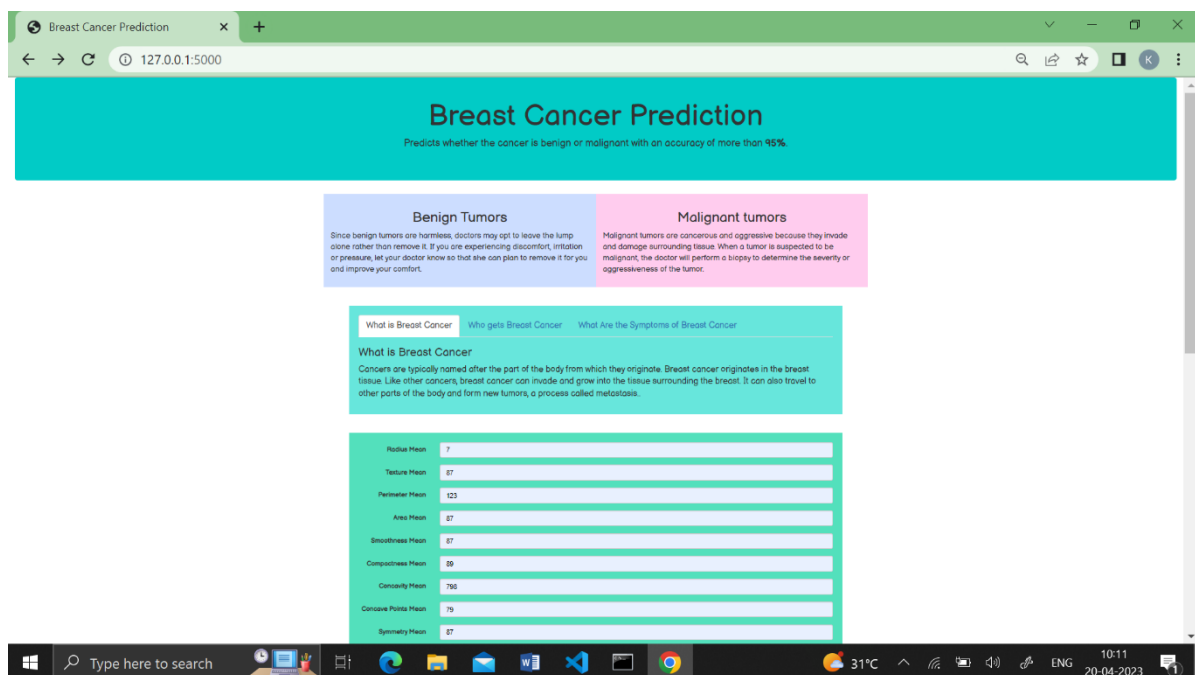
if __name__=='__main__':
    global clf
    clf = random_forest_train()
    random_forest_test(clf)
    #print("Done")
    app.run(debug=True)

```

SCREENSHOTS

```
C:\Users\THIS PC\Desktop\Breast-Cancer-Predictor-master>python app.py
Microsoft Windows [Version 10.0.19045.2728]
(c) Microsoft Corporation. All rights reserved.

C:\Users\THIS PC\Desktop\Breast-Cancer-Predictor-master>python app.py
The accuracy of testing data: 0.956140350877193
The running time: 0.28127145767211914
The accuracy of testing data: 0.9649122807017544
The running time: 0.0
* Serving Flask app 'app'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:5000
Press CTRL+C to quit
* Restarting with stat
The accuracy of testing data: 0.9649122807017544
The running time: 0.28118038177490234
The accuracy of testing data: 0.956140350877193
The running time: 0.0
* Debugger is active!
* Debugger PIN: 101-090-504
```



Breast Cancer Prediction

127.0.0.1:5000

Fractal Dimension Mean	98
Radius SE	79
Texture SE	87
Perimeter SE	387
Area SE	98
Smoothness SE	798
Compactness SE	79
Concavity SE	87
Concave Points SE	978
Symmetry SE	98
Fractal Dimension SE	79
Radius Worst	879
Texture Worst	87
Perimeter Worst	98
Area Worst	79
Smoothness Worst	879
Compactness Worst	79
Concavity Worst	7
Concave Points Worst	97
Symmetry Worst	978
Fractal Dimension Worst	9

Submit

Breast Cancer Prediction

Predicts whether the cancer is benign or malignant with an accuracy of more than 95%.

Cancer is Predicted to be Malignant

The Probability of Malignant is 96.49122807017544 %

The time taken to predict is 0.02595996856689453 s.

Benign Tumors

Since benign tumors are harmless, doctors may opt to leave the lump alone rather than remove it. If you are experiencing discomfort, irritation or pressure, let your doctor know so that she can plan to remove it for you and improve your comfort.

Malignant tumors

Malignant tumors are cancerous and aggressive because they invade and damage surrounding tissue. When a tumor is suspected to be malignant, the doctor will perform a biopsy to determine the severity or aggressiveness of the tumor.