

MACHINE LEARNING PROJECT (CS-312b)

Breast Cancer Analysis for Diagnosis and Treatment

Presented by:

Abhay Gupta | 2016005

Prakash Nidhi Verma | 2016185

Navneet Kumar Chaurasiya | 2016167

Presented to:

Prof. Kusum Kumari Bharti



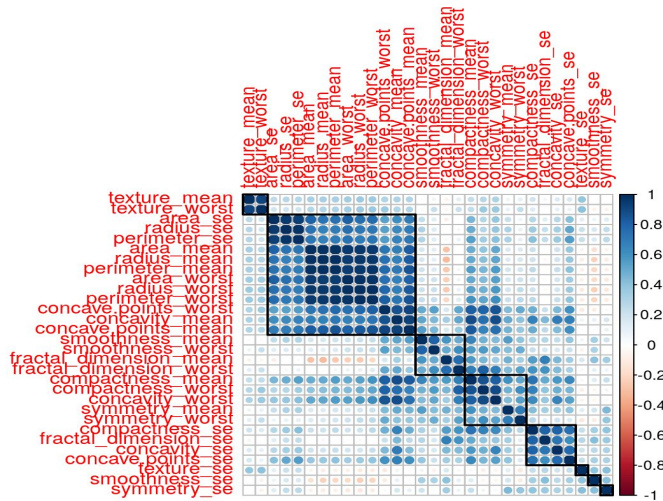
Abstract

This paper presents the use of machine learning for Breast cancer analysis as well as Diagnosis and Treatment respective to this domain. This analysis used for generating quantitative representation through prediction. The project focuses on prediction and analysis on the breast cancer data. The analysis provides all possibilities for improvement in the health area. The goal is to classify whether breast cancer is benign or malignant. To achieve this I have used machine learning classification methods to fit a function that can predict the discrete class of new input.

Introduction

Breast cancer (BC) is one of the most common cancers among women worldwide, representing the majority of new cancer cases and cancer-related deaths according to global statistics, making it a significant public health problem in today's society. Worldwide, about 2.1 million women develop breast cancer every year, and despite better treatment options, more than 627,000 die from the disease. Our ML algorithms have shown a cancer-related prediction, which is associated with breast cancer response to treatment.

Breast cancer is a cancer that develops from breast tissue. This is an analysis of the Breast Cancer Wisconsin (Diagnostic) DataSet, obtained from [Kaggle](#). We are going to analyze it and to try several machine learning classification models to compare their results. The best model found is based on a neural network and reaches a **sensibility of 0.984** with an **F1 score of 0.984**.



Problem Description

The most common cancer in women worldwide. The principal cause of death from cancer among women globally.

Early detection is the most effective way to reduce breast cancer deaths. Early diagnosis requires an accurate and reliable procedure to distinguish between benign breast tumours from malignant ones.

Breast Cancer Types - three types of breast tumours: Benign breast tumours, In-situ cancers, and Invasive cancers. The majority of breast tumours detected by mammography are benign. They are non-cancerous growths and cannot spread outside of the breast to other organs. In some cases, it is difficult to distinguish certain benign masses from malignant lesions with mammography. If the malignant cells have not gone through the basal membrane but are completely contained in the lobule or the ducts, the cancer is called in-situ or noninvasive. If cancer has broken through the basal membrane and spread into the surrounding tissue, it is called invasive. This analysis assists in differentiating between benign and malignant tumours.

Related work

A detailed analysis of more recent studies revealed that there is a growing trend in risk assessment as well as the prediction of recurrence of a cancer type regardless of the ML technique used. Many research groups have tried to predict the possibility of redeveloping cancer after remission and appeared to improve the accuracy of predictions compared to alternative statistical techniques. Moreover, the vast majority of these publications used molecular and clinical data in order to make their predictions. The use of such measurable features as input data is a growing trend based on the advent of HTTps.

In the following, we are going to discuss one case for each of the objectives of predicting (i) susceptibility, (ii) recurrence and (iii) survival, all by means of ML techniques. Each sub-section summarizes the representative studies we have selected based on their predictive outcomes. We only selected those publications that have been accepted in the last 5 years and make use of distinguishable ML methods. We provide the readers with the appropriate details of the most recent techniques used for the prediction and prognosis of most frequent cancer types.

Proposed solution

When we analysis the problem set we found that the solution is not user friendly rather, we were identify the disease by using given data set only. But we provide extra feature like user would be able check their disease condition using their real life data entry. So user can see their improvement and analyze their condition on daily basis. Our application is user friendly and easy to handle or visualize.

Evaluation study

In the present review, the recent works relevant to cancer prediction/prognosis by means of ML techniques are presented. After a brief description of the ML branch and the concepts of the data preprocessing methods, the feature selection techniques and the classification algorithms being used, we outlined three specific case studies regarding the prediction of cancer susceptibility, cancer recurrence and cancer survival based on popular ML tools. Obviously, there is a large amount of ML studies published in the last decade that provide accurate results concerning the specific predictive cancer outcomes. However, the identification of potential drawbacks including the experimental design, the collection of appropriate data samples and the validation of the classified results, is critical for the extraction of clinical decisions.

Moreover, it should be mentioned that in spite of the claims that these ML classification techniques can result in adequate and effective decision making, very few have actually penetrated the clinical practice. Recent advances in omics technologies paved the way to further improve our understanding of a variety of diseases; however more accurate validation results are needed before gene expression signatures can be useful in the clinics.

A growing trend was noted in the studies published during the last 2 years that applied semi-supervised ML techniques for modelling cancer survival. This type of algorithms employs labelled and unlabeled data for their predictions while it has been proven that they improved the estimated performance compared to existing supervised techniques. SSL can be thought as a great alternative to the other two types of ML methods (i.e. supervised learning and unsupervised learning) that use, in general, only a few labelled samples.

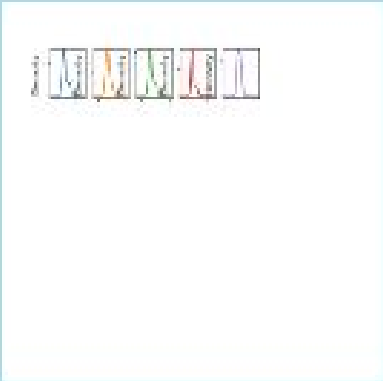
Results

Output when we doesn't enter any of the value :

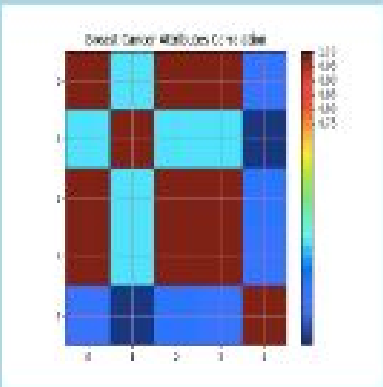


Output when we enter any of the value:

Predicted Disease based on information given	Malignant
ACCURACY	93.86 %



Density Graph



Correlation Graph

	Precision	Recall	F1-Score	Support Count
Design	0.928	0.957	0.94	75
Malignant	0.97	0.842	0.9	32

Actual/Predicted	Malignant	<u>Malignant</u>
Design	75	1
Design	5	32

<input checked="" type="checkbox"/> Radius Mean	<input type="checkbox"/> Texture mean	<input type="checkbox"/> Perimeter mean
<input checked="" type="checkbox"/> Area Mean	<input type="checkbox"/> Smoothness mean	<input type="checkbox"/> Compactness mean
<input checked="" type="checkbox"/> Concavity Mean	<input type="checkbox"/> Concave point mean	<input type="checkbox"/> Symmetry mean
<input checked="" type="checkbox"/> Fractal dimension Mean	<input type="checkbox"/> By radius	<input type="checkbox"/> By perimeter
<input type="checkbox"/> By texture	<input type="checkbox"/> By area	<input type="checkbox"/> By smoothness
<input type="checkbox"/> By Compactness	<input type="checkbox"/> By Concavity	<input type="checkbox"/> By Concave points
<input type="checkbox"/> By Symmetry	<input type="checkbox"/> By Fractal dimension	<input type="checkbox"/> Radius worst
<input type="checkbox"/> Texture worst	<input type="checkbox"/> Perimeter worst	<input type="checkbox"/> Area worst
<input type="checkbox"/> Smoothness worst	<input type="checkbox"/> Compactness worst	<input type="checkbox"/> Concavity worst
<input type="checkbox"/> Concave points worst	<input type="checkbox"/> Symmetry worst	<input type="checkbox"/> Fractal dimension worst

Optional: If you have the values of above selected checkboxes then you can test whether you have Breast cancer or not

Radius Mean:	<input type="text" value="cm"/>
Area Mean:	<input type="text" value="cm sq."/>
Concavity Mean:	<input type="text"/>
Fractal Dimension Mean:	<input type="text"/>

Submit

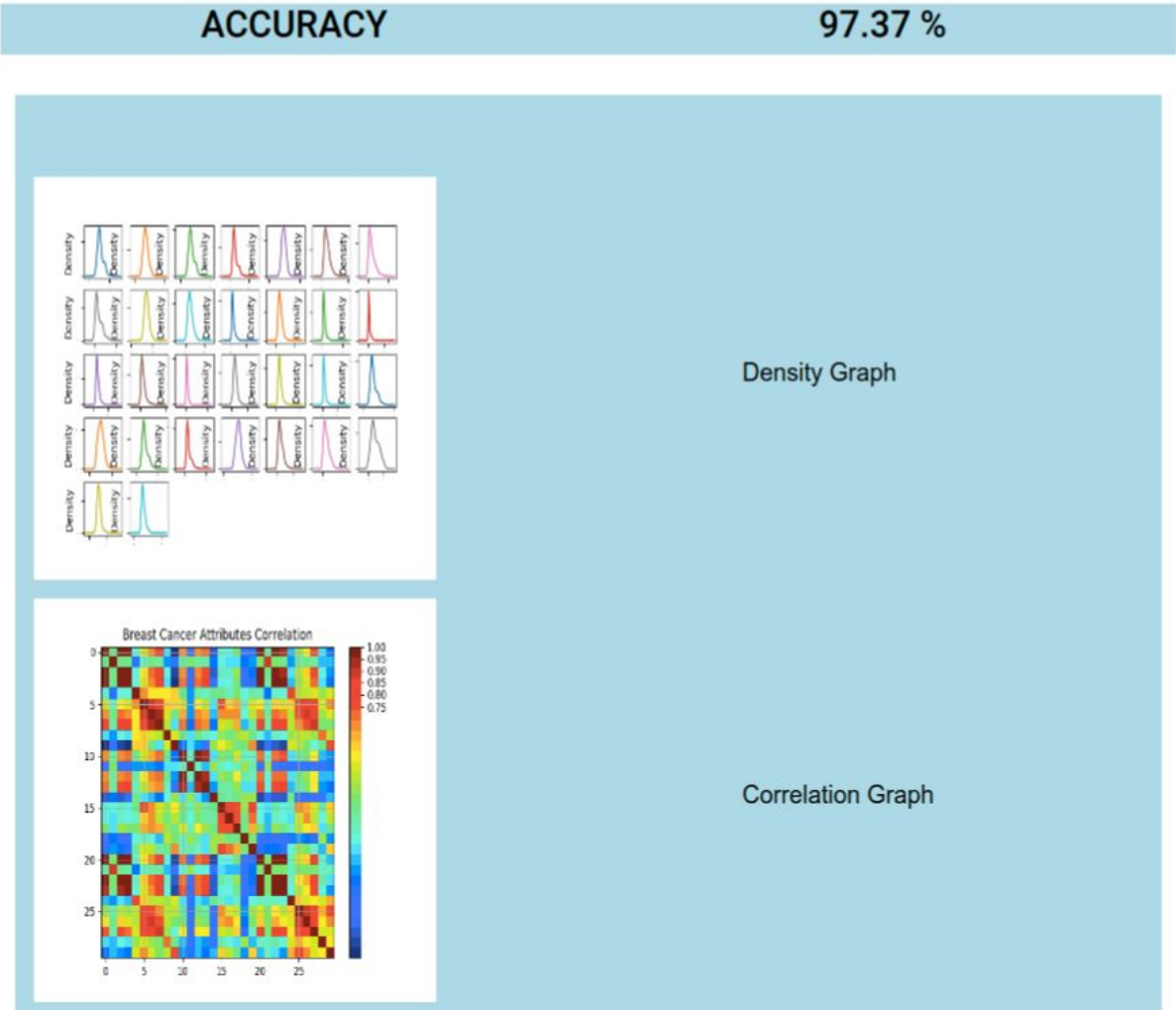
<input checked="" type="checkbox"/> Radius Mean	<input type="checkbox"/> Texture mean	<input type="checkbox"/> Perimeter mean
<input checked="" type="checkbox"/> Area Mean	<input type="checkbox"/> Smoothness mean	<input type="checkbox"/> Compactness mean
<input checked="" type="checkbox"/> Concavity Mean	<input type="checkbox"/> Concave point mean	<input type="checkbox"/> Symmetry mean
<input checked="" type="checkbox"/> Fractal dimension Mean	<input type="checkbox"/> By radius	<input type="checkbox"/> By perimeter
<input type="checkbox"/> By texture	<input type="checkbox"/> By area	<input type="checkbox"/> By smoothness
<input type="checkbox"/> By Compactness	<input type="checkbox"/> By Concavity	<input type="checkbox"/> By Concave points
<input type="checkbox"/> By Symmetry	<input type="checkbox"/> By Fractal dimension	<input type="checkbox"/> Radius worst
<input type="checkbox"/> Texture worst	<input type="checkbox"/> Perimeter worst	<input type="checkbox"/> Area worst
<input type="checkbox"/> Smoothness worst	<input type="checkbox"/> Compactness worst	<input type="checkbox"/> Concavity worst
<input type="checkbox"/> Concave points worst	<input type="checkbox"/> Symmetry worst	<input type="checkbox"/> Fractal dimension worst

Optional: If you have the values of above selected checkboxes then you can test whether you have Breast cancer or not

Radius Mean:	<input type="text" value="1.5"/>
Area Mean:	<input type="text" value="7.05"/>
Concavity Mean:	<input type="text" value=".66"/>
Fractal Dimension Mean:	<input type="text" value="2.05"/>

Submit

When we select all the entries than:



	Precision	Recall	F1-Score	Support Count
Benign	0.974	0.987	0.98	76
Malignant	0.973	0.947	0.96	38

Actual\Predicted	Malignant	Benign
Benign	75	1
Malignant	2	36

Discussion

Among the most common applied ML algorithms relevant to the prediction outcomes of cancer patients, we found that SVM and ANN classifiers were widely used. As mentioned to our introductory section, ANNs have been used extensively for nearly 30 years. In addition, SVMs constitute a more recent approach in the cancer prediction/prognosis and have been used widely due to its accurate predictive performance. However, the choice of the most appropriate algorithm depends on many parameters including the types of data collected, the size of the data samples, the time limitations as well as the type of prediction outcomes.

Concerning the future of cancer modelling, new methods should be studied for overcoming the limitations. A better statistical analysis of the heterogeneous datasets used would provide more accurate results and would give reasoning to disease outcomes. Further research is required based on the construction of more public databases that would collect valid cancer dataset of all patients that have been diagnosed with the disease. Their exploitation by the researchers would facilitate their modelling studies resulting in more valid results and integrated clinical decision making.

Conclusion

We discussed the concepts of ML while we outlined their application in cancer prediction/prognosis. Most of the studies that have been proposed in the last years and focus on the development of predictive models using supervised ML methods and classification algorithms aiming to predict valid disease outcomes. Based on the analysis of their results, it is evident that the integration of multidimensional heterogeneous data, combined with the application of different techniques for feature selection and classification can provide promising tools for inference in the cancer domain.

Future direction

Nowadays, breast cancer is a most dangerous disease. And many people not sure about they are having this disease having or not sometimes won't be able to processed their treatment and check up regularly form doctor so, they can use our application/App to identify and recognize their disease from home easily. We can hire some specialist and provide online treatment plan to use any app and this idea. This idea can change many life and give a medical support to their users.

References

1. <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>
 2. https://en.wikipedia.org/wiki/Breast_cancer
 3. [http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Original\)](http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original))
 4. ML book by Tom M Mitchell
 5. <http://science.sciencemag.org/content/244/4905/707.short>
 6. [*Journal of Women's Health*](#).
 7. [Epidemiology of soy exposures and breast cancer risk](#)".
-

Our website looks like

