

Logistic and Softmax Regression for Prognosis & Diagnosis of Breast Cancer Using Multilayered Neural Networks

¹Deepa B G, ²S.Senthil, ³Aniruddha Prasad

¹Research Scholar, School of CSA, REVA University, Bengaluru, India

¹deepabg03@gmail.com

²Professor & Director, School of CSA, REVA University, Bengaluru, India

²dir.csa@reva.edu.in

³BSc (Research) in CS, Cloud Computing & Big Data, School of CSA, REVA University, Bengaluru, India

³theanidaniel@gmail.com

Article Info

Volume 83

Page Number: 3535-3542

Publication Issue:

May-June 2020

Abstract

Presently Breast Cancer is spiking as 3rd most dangerous cancer in the world. Breast Cancer has become the wide known epidemic as it occurs in all ages of women from adult to old ones too. The malignant tumor is the main cause, which is very hard to even identify with mammography & human discretion. Thus, advancement in technological researches have become a winning factor as Machine Learning and Artificial Neural Networks are helping researchers and doctors worldwide. The valuation of prediction is being done by using algorithms like Logistic Regression and SoftMax Regression to train models with data set Wisconsin Breast Cancer Data of UCI Machine Learning Repository. To perform all of these operations we have used Google Colab (Jupyter notebook backed by Python 3.x). The proposed work results into trained neural network model which is having much better accuracy when compares to the other pre-existing models.

Article History

Article Received: 19 August 2019

Revised: 27 November 2019

Accepted: 29 January 2020

Publication: 12 May 2020

Keywords: Artificial Neural Networks, Machine Learning, Logistic Regression, SoftMax Regression, Google Colab

1. Introduction

Breast cancer [1] is a sort of cancer that begins inside the breast. It is by and by third most risky cancer inside the world. Cancer begins once cells start to develop out of the breast. Breast cancer cells normally type a tumor which will commonly be seen on the x-ray or felt as an irregularity. Breast cancer happens almost totally in ladies; in any case, men can get breast cancer, as well.

Breast cancer [2] has become the leading cancer affecting most females around the world, early discovery and analysis of malignant tumour of Breast cancer can save many lives. Different diagnostic techniques are available for early detection as it plays a major role in survival rate for people. The general treatment accessible for breast cancer are medical

procedure, chemotherapy, radiation treatment, hormone treatment, organic treatment and so on.

It's important to get a handle on that practically all breast irregularities are benign and not cancer (malignant). Non-cancerous breast tumors makes anomalous developments, anyway they won't spread outside of the breast. They're not hazardous, anyway a few types of benign breast protuberances will build a lady's danger of getting breast cancer. Any breast bump or alteration must be checked by a medicinal services talented to work out if it's benign or malignant (cancer).

Researchers [3] have recognized way of life, ecological and hormonal elements that may build your danger of breast cancer. It's as yet not cleared why a few people with chance factors never create cancer, yet individuals who have no hazard factors does. It's

majorily that breast cancer are brought about by the perplexing cooperation of an individual's hereditary cosmetics and their condition.

In last years AI and neural systems are most tirelessly utilized in clinical field and determination. In these domains breast cancer prediction is one of the classification problem to determine the tumours in to benign or malignant. Both machine learning and neural networks uses the preset algorithms that processes the input and helps in predicting the output, under them your model optimizes and improvises itself on accordance to the fed data and algorithms used in it.

Our proposed research work discusses about the model's classifier accuracy and statistical measures. We are using dataset available under UCI repository. Many of the researchers will use WBCD to test their model's accuracy. The proposed model has been trained under 3 different ratios of train-test to determine and get best proposed model for our work.

The organization of research paper can be envisioned in different sections that is section 2 previews the reviews of literature on Breast Cancer Prediction, Section 3 explains our selected dataset, Section 4 briefs about existing ML classifiers, Section 5 briefs about neural network methods, Section 6 presents the proposed work, Section 7 discusses about experimental results & accuracy and finally section 8 concludes our proposed work with few contributions.

2. Related Work on Breast Cancer Prediction

Murat Karabatak, M. Cevdet Ince [4] presents partner in nursing programmed ID framework for sleuthing breast cancer upheld affiliation rules (AR) and neural system (NN). during this examination, AR is utilized for decreasing the component of breast cancer information and NN is utilized for clever arrangement. The arranged AR + NN framework execution is contrasted and NN model. The element of info include house is diminished from 9 to four by misuse AR. Under control arrange, 3-crease cross approval strategy was applied to the Wisconsin breast cancer information to pass judgment on the arranged framework exhibitions. the best possible arrangement pace of arranged framework is 95.6%. This examination incontestable that the AR is utilized for decreasing the element of highlight house and arranged AR + NN model is acclimated get fast programmed indicative frameworks for various illnesses.

C E Floyd, Jr, J A Baker, J Y Lo, P J Komguth, M E Williford, [5] decide whether artificial neural system (ANN) to sort benign and malignant breast sores might be institutionalized to be utilized by all radiologists. MATERIALS AND METHODS: an ANN was made bolstered the institutionalized dictionary of the Breast Imaging Recording and framework (BI-RADS) of the yank workforce of Radiology. Eighteen contributions to the system encased ten BI-RADS injury descriptors and eight

information esteems from the patient's anamnesis. The system was prepared and tried on 206 cases (133 benign, seventy three malignant cases). Recipient in activity trademark bends for the system and radiologists were thought about. RESULTS: At a predetermined yield limit, the ANN would have improved the positive prognosticative worth (PPV) of indicative test from thirty fifth to sixty one with a general affectability of 100 percent. At an immovable affectability of ninety fifth, the explicitness of the ANN (62%) was extensively bigger than the particularity of radiologists (30%) ($P < .01$). End: The BI-RADS vocabulary gives a normal language in between mammographers partner degreed an ANN this might improvise PPV for breast symptomatic test.

N. Albert Singh, Dheeba, J., and S. Tamil Selvi. [6] Computer aided identification (CAD) helps the radiotherapist in recognition of the variations from the norm in the capable methodology. This paper examines a fresh out of the box new grouping approach for recognition of breast irregularities in computerized mammograms misuse Particle Swarm Optimized wave Neural Network (PSOWNN). The arranged anomaly recognition recipe is predicated on extricating Laws Texture Energy Measures from the mammograms and characterizing the suspicious locales by applying an example classifier. the technique is applied to genuine clinical data of 216 X-beam pictures gathered from mammogram screening focuses. The discovery execution of the CAD framework is dissected abuse Receiver in activity Characteristic (ROC) bend. This bend demonstrates the exchange offs among affectability and explicitness that is out there from a symptomatic framework, and along these lines portrays the characteristic segregation ability of the arranged framework. The outcome shows that the domain underneath the legendary animal bend of the arranged equation is zero.96853 with an affectability 94.167% of and explicitness of 92.105%.

A. R. Razavi, A. Poorebrahimi, A. T. Eshlaghy, , M. Ebrahimi, and Ahmad, L. Gh. [7] utilizing information preparing methods, creators created models to foresee the redundancy of breast cancer by investigating data gathered from ICBC register. Continuous segments of this paper audit associated work, depict foundation of this investigation, evaluate 3 characterization models (C4.5 DT, SVM, and ANN), put forth a defense for the approach acclimated direct the forecast, exhibited exploratory outcomes, and in this way the last area of the paper covers the end. To figure the approval everything being equal, affectability, particularity and precision have been utilized as alternative and were broke down with one another.

Polat, Kemal, and Salih Güneş. [8] Right now, cancer analysis was directed utilizing least square help vector machine (LS-SVM) classifier calculation. The vigor of the LS-SVM is inspected utilizing characterization precision, examination of

affectability and explicitness, k-overlay cross-approval strategy and disarray network. The got order exactness is 98.53% and it is exceptionally encouraging contrasted with the recently announced grouping procedures. Subsequently, by LS-SVM, the got outcomes show that the pre-owned technique can make a successful understanding and point out the capacity of structure of another astute help analysis framework.

Osareh, Alireza, and Bita Shadgar [9] KNN, SVMs, and probabilistic neural systems classifiers are joined with signal-to-commotion proportion include positioning, successive forward choice based element determination and head segment investigation highlight extraction to recognize the benign and malignant tumors of breast. The best by and large exactness for breast cancer analysis is accomplished equivalent to 98.80% and 96.33% individually utilizing help vector machines classifier models against two broadly utilized breast cancer benchmark datasets.

Alyami, Reem, Jinan Alhajjaj [10] to have an exact analysis, Support Vector Machine (SVM) and Artificial Neural Network (ANN) have been chosen in many research papers to take care of this issue with high grouping exactness. Right now breast cancer analysis is tended to utilizing SVM and ANN joined with highlight choice. The element choice depends on the relationship coefficient of each component against the objective class where distinctive element subsets are utilized. The model is tried on the well known Wisconsin Diagnosis Breast Cancer (WDBC) dataset to lead the analyses. 10-Fold Cross approval has been utilized for information parceling while at the same time building up the model and the result demonstrates better grouping precision. Concerning correlation among SVM and ANN, exact examinations result showed that SVM beat ANN with characterization exactness of 97.13 and 96.71 individually.

Saeid Eslami Alireza Rowhanimesh, Hadi Shahraki, and Aalaei, Shokoufeh, [11] it is seen that include determination improved the exactness of all classifiers expect of ANN and the best precision with highlight choice accomplished by PS-classifier. For WDBC and WPBC, results show highlight determination improved precision of every one of the three classifiers and the best exactness with include choice accomplished by ANN. Additionally, explicitness and affectability improved after component choice.

Peter M., Gary M. Clark and Ravdin [12] Neural Network was prepared, tried, and approved utilizing controlled endurance information from a gathering of 1373 patients with hub positive breast cancer. The Neural Network technique anticipated patient result as precisely as Cox Regression demonstrating. The last

Neural Network model can be given a patient's prognostic data and make a progression of expectations about likelihood of backslide at various occasions of development, permitting one to draw endurance likelihood bends for singular patients.

3. Data Set

The dataset used for this experiment, its model obtained from the university of Wisconsin made by Dr. William H Wolberg, Clinical Sciences Center Madison, WI 53792. It is available for download in UCI repository. [13] Features have been taken from a digitized picture of fine needle suction (FNA) from breast mass. They portrayed highlights of the cell cores given in the pictures. 10 genuine esteemed features are then assessed for every cell core: range (mean of good ways from focus to focuses on the edge), surface (standard deviation of dim scale esteems), border, region, smoothness (neighborhood variety in span lengths), conservativeness ($\text{perimeter}^2/\text{territory} - 1.0$), concavity (seriousness of sunken bits of the shape), inward focuses (number of curved bits of the form), evenness, fractal measurement ("coastline guess" - 1) Thus 32 highlights have been pulled from it.

Heat Map of Dataset: A heat map is 2-dimensional representation of data within all values have been through colors codes. Here we've got shown all features correlation with one another through this heatmap.

Pair Plot: Pair Plot also known as the scatterplot, is graphical plot which represents matching and correlation of one variable/feature value to another variable/feature value in the same dataset.

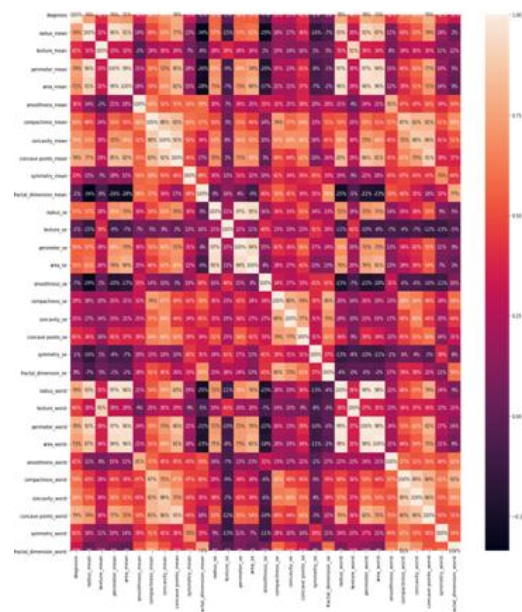


Figure 1: Heatmap chart for correlation of all features in the dataset

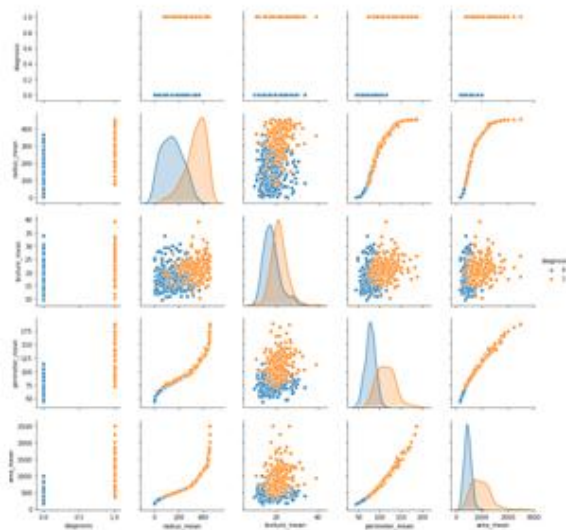


Figure 2: Pair plot graph of few features from a sample of our dataset

4. Machine Learning Classifiers

Machine Learning is science of making computers to program and taught and act like humans do, and improve their learning over time in autonomous fashion, by feeding them information and knowledge within the variety of observations and real-world interactions. [14] it's prognostic analytics helps in scientific study of applied math models helps in accuracy prediction. it's separate in three parts: (i) supervised learning (ii) unattended / unsupervised learning & (iii) Semi-supervised learning. the major goal of machine learning is to develop models that helps in classification, prediction, automation & so on [15].

Here in this proposed research we have used Logistic Regression, KNN, SVM Linear, SVM RBF, Gaussian Naïve Bayes, Decision Tree Classifier, Random Forest Classifier.

Logistic Regression (Lr)

Logistic regression is defined for the function used at the core of its method, the logistic function. The logistic function, additionally known as the sigmoid function was developed by statisticians to explain properties of population increase in ecology, rising quickly and maxing out at the carrying capability of the surroundings. It's a curved from S-shaped curve which will take any real-valued range and map it into a worth between zero and one, however ne'er specifically at those limits.

$$1 / (1 + e^{-\text{value}})$$

Where e is base of all natural logarithms (Euler's values or the EXP() function in your excel spreadsheet) & value is a actual numerical value that

you want to transform. Below is a plot of all numbers in between -5 and 5 transforms into a range 0 and 1 using the logistic function. [16]

K Nearest Neighbor (KNN)

K-Nearest Neighbor(KNN) is additionally lazy algorithmic program (as opposition of an eager algorithm). will it that mean that KNN will do nothing, like these polar bears imply? Not really. What this suggests is that it doesn't utilize the training data focuses to attempt to do any speculation. In elective words, there's no particular instructing area or it's awfully base. This moreover implies the preparation segment is really brisk . Absence of speculation implies KNN keeps all the preparation data. To be a ton of exact, all (or most) the preparation data is required all through the testing section.[17] KNN algorithmic program is predicated on include closeness: anyway intently out-of-test alternatives compare our training set decides anyway we will in general order a given information focuses.

Support Vector Machine(S) (Svm) Linear & Rbf

The goal of the support vector machine calculation is to discover a hyperplane in a N dimensional space (N — the quantity of highlights) that particularly orders the information focuses. To isolate the two classes of information focuses, there are numerous conceivable hyperplanes that could be picked. Our goal is to locate a plane that has the most extreme edge, i.e the greatest separation between information purposes of the two classes. Boosting the edge separation gives some support so future information focuses can be ordered with more certainty. [18]

As per the SVM calculation we discover the focuses nearest to the line from both the classes. These focuses are called bolster vectors. Presently, we process the separation between the line and the help vectors. This separation is known as the edge. We will likely boost the edge. The hyperplane for which the edge is most extreme is the ideal hyperplane.

Gaussian Naïve Bayes

Naive Bayes classifiers are assortment of characterization calculations dependent on Bayes' Theorem. It is anything but a solitary calculation however a group of calculations where every one of them share a typical rule, for example each pair of the highlights being arranged is autonomous of one another. In Gaussian Naive Bayes, nonstop worth related with each component are thought to be conveyed on as indicated by a Gaussian dispersion. A Gaussian appropriation is otherwise called Normal dispersion graph.

Decision Tree

A decision tree is flowchart-like structure in which each internal node represents a "test" on an attribute (e.g. whether coin flip comes up heads or tails), each

branch represent the outcome of test, and each leaf node represent class label (decision taken after computing all attributes)

Random Forest (RF)

Random forest, similar to its name infers, comprises of countless individual choice trees that work as a group. Every individual tree in the random forest lets out a class expectation and the class with the most votes turns into our model's forecast [19].

5. Neural Network Methods

Logistic Regression

Logistic regression are the binary classification methodology. It is often modelled as a fun() which will pull into given any range of inputs and constrain the output to be between zero and one. This means, we are able to consider Logistic Regression as a one-layer neural network. For a binary output, if verity label is y (y = zero or y = one) and y_hat is that the foretold output – then y_hat represents the chance that y = 1 - entered inputs x & w. Therefore, the chance that y = zero given inputs w and x is (1 - y_hat) [20].

$$P(y = 1|w, x) = \hat{y}$$

$$P(y = 0|w, x) = 1 - \hat{y}$$

SoftMax Regression

Softmax regression (or multinomial logistic regression) are generalization of logistic regression to the case where we want to handle multiple classes. In logistic regression we assumed that the labels were the binary : $y(i) \in \{0,1\}$ $y(i) \in \{0,1\}$. We used such a classifier to distinguish between two kinds of given numbers. SoftMax regression allows us to handle $y(i) \in \{1, \dots, K\}$ $y(i) \in \{1, \dots, K\}$ where KK is the number of classes. [21]

$$J(\theta) = -\left[\sum_{i=1}^m y(i) \log h(\theta(x(i))) + (1-y(i)) \log (1-h(\theta(x(i))))\right]$$

6. Proposed Work

Our proposed work defines a neural network model based on logistic regression and SoftMax regression to check the accuracy of the model defined as per our dataset given.



Figure 3: Flowchart of the workflow

Above flowchart starts with loading of the UCI dataset, then pre-processing it by segregating it in two parts i.e., X table consisting of diagnosis column and Y table with all the features available in it. Then creating the model using the dataset for training and testing. And running the model using either logistic regression algorithm or SoftMax regression algorithm to get required results and calculating the accuracy of the created model.

Algorithm 1 – Logistic Regression Method

Input: UCI repository dataset segregated as X & Y dataset(s)

Output: Predicts and give accuracy of the model.

Step 1: Load the dataset as X (containing only diagnosis – Malign or Benign) & Y (All other features from the UCI dataset)

Step 2: Create a model while using X_Train, Y_Train divided in 3 ratios 75-25, 80-20, 60-40 to analyse the model.

Step 3: Create a single neural network using shallow logistic regression model, with input layers=20, hidden layers=30, output layer = 1, activation = sigmoid, loss = binary_crossentropy, optimizer learning rate = 0.001 with metric = accuracy.

model = Sequential()

model.add(Dense(20,input_shape=(30,), activation='relu'))

model.add(Dense(1, activation='sigmoid'))

model.compile(loss='binary_crossentropy', optimizer=Adam(lr=0.001), metrics = ['accuracy'])

Step 4: Create an earlystopping() function using epochs with threshold=2000 and monitor validation loss over it.

earlystopper = EarlyStopping(monitor='validation_loss', verbose=1, patience=15, min_delta=0, mode='auto')

history = model.fit(x_train, y_train, epochs = 2000, validation_split = 0.15, verbose = 0, callbacks = [earlystopper])

Step 5: Run the epochs to generate the plot of training loss and validation training loss, also plot accuracy over epochs.

Step 6: Now we'll calculate the accuracy and loss of the testing data, generate AUC score and ROC curve for the testing data to analyse the model.

Algorithm 2 – SoftMax Regression Method

Input: UCI repository dataset segregated as X & Y dataset(s)

Output: Predicts and give accuracy of the model.

Step 1: Load the dataset as X (containing only diagnosis – Malign or Benign) & Y (All other features from the UCI database)

Step 2: Create a model while using X_Train, Y_Train divided in 3 ratios 75-25, 80-20, 60-40 to analyse the model. Also we'll change the Y_Train into categorical data values for deep softmax algorithm model.

Step 3: Create a multi-layered neural network using deep softmax regression model, with multiple input layers, hidden layers=30, output layer =2, activation=softmax, loss=categorical_crossentropy, optimizer learning rate = 0.0001 with metric=accuracy.

```
model = Sequential()
model.add(Dense(20,input_shape=(30,), activation='softmax'))
model.add(Dense(20, activation='softmax'))
model.add(Dense(20, activation='softmax'))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer=Adam(lr=0.001), metrics = ['accuracy'])
```

Step 4: Create an earlystopping() function using epochs with threshold=2000 and monitor validation loss over it..

```
earlystopper = EarlyStopping(monitor='validation_loss', verbose=1, patience=15, min_delta=0, mode='auto')
history = model.fit(x_train, y_train, epochs = 2000, validation_split = 0.15, verbose = 0, callbacks = [earlystopper])
```

Step 5: Run the epochs to generate the plot of training loss and validation training loss, also plot accuracy over epochs

Table 1: General ML algorithms accuracy

ML ALGORITHM	CONFUSION MATRIX	CLASSIFICATION ACCURACY
Logistic Regression	0.9440	0.9440
K Nearest Neighbour	0.9510	0.9510
Support Vector Machine – Linear Classifier	0.9650	0.9650
Support Vector Machine – RBF Classifier	0.9650	0.9650
Gaussian Naïve Bayes	0.9300	0.9300
Decision Tree Classifier	0.9510	0.9510
Random Forest Classifier	0.9650	0.9650

Resultant accuracy using Neural Networks algorithms

Table 2: Neural Network model score when ratio is 60:40 train-test dataset.

RATIO 60:40	LOGISTIC REGRESSION	SOFTMAX REGRESSION
TEST ACCURACY	0.9649	0.6228
TEST LOSS	0.1934	0.0614
AUC SCORE	0.9913	0.9880
ROC CURVE SCORE	0.991	0.988

Step 6: Now we'll calculate the accuracy and loss of the testing data, generate AUC score and ROC curve for the testing data to analyse the model.

7. Experimental Results & Accuracy

To estimate the proficiency of the proposed research work, experiments have been conducted with WBCD. Neural Network models has been created using sklearn and keras libraries and resultants has been portrayed using matplotlib libraries. Here we are using logistic regression and SoftMax regression for our models. The experiment has been carried out on 3 ratios of the dataset i.e., 60:40, 75:25 and 80:20 training-test data splits.

Google Colab (based on Python Jupyter notebook) has been used as IDE to develop this program and model for our research work. The best results are coming with ratio 75:25 for our dataset. The results obtained over on neural networks are much better in comparison to machine learning models on the same dataset with same ratio.

Resultant accuracy chart using machine learning algorithms

Table 3: Neural Network model score when ratio is 75-25 train-test dataset.

RATIO 75:25	LOGISTIC REGRESSION	SOFTMAX REGRESSION
TEST ACCURACY	0.9859	0.9929
TEST LOSS	0.0523	0.0455
AUC SCORE	0.997	0.9982
ROC CURVE SCORE	0.999	0.998

Table 4: Neural Network model score when ratio is 80-20 train-test dataset.

RATIO 80:20	LOGISTIC REGRESSION	SOFTMAX REGRESSION
TEST ACCURACY	0.9736	0.9736
TEST LOSS	0.0743	0.0972
AUC SCORE	0.9981	0.9975
ROC CURVE SCORE	0.998	0.998

8. Conclusion

The results shows that deep SoftMax regression multi layered neural network model with train-test data split ratio of 75-25 gives us the best result for test accuracy of the model with **99.29%** with test loss of **0.0455** only and in our logistic regression single neural network model also the model with train-test split ratio of 75-25 gives us the best result for test accuracy of the model with **98.59%** with test loss of **0.0523** only. Thus, 75-25 ratio producing the best results with followed by ratio of 80-20 as second best for both the logistic regression & softmax regression model with test accuracy of **97.36%**, whereas the ratio 60-40 produces the most unfitting data results with unpopular test accuracy for both logistic regression of **96.49%** & softmax regression bad test accuracy of **62.28%**. Therefore, the ratio 75-25 produces the best results.

References

- [1] <https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html>
- [2] Floyd Jr, Carey E., Joseph Y. Lo, A. Joon Yun, Daniel C. Sullivan, and Phyllis J. Kornguth. "Prediction of breast cancer malignancy using an artificial neural network." *Cancer: Interdisciplinary International Journal of the American Cancer Society* 74, no. 11 (1994): 2944-2948.
- [3] <https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470>
- [4] Karabatak, Murat, and M. Cevdet Ince. "An expert system for detection of breast cancer based on association rules and neural network." *Expert systems with Applications* 36, no. 2 (2009): 3465-3469.
- [5] Baker, Jay A., Phyllis J. Kornguth, Joseph Y. Lo, Margaret E. Williford, and Carey E. Floyd Jr. "Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon." *Radiology* 196, no. 3 (1995): 817-822.
- [6] Dheeba, J., N. Albert Singh, and S. Tamil Selvi. "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach." *Journal of biomedical informatics* 49 (2014): 45-52.
- [7] Ahmad, L. Gh, A. T. Eshlaghy, A. Poorebrahimi, M. Ebrahimi, and A. R. Razavi. "Using three machine learning techniques for predicting breast cancer recurrence." *J Health Med Inform* 4, no. 124 (2013): 3.
- [8] Polat, Kemal, and Salih Güneş. "Breast cancer diagnosis using least square support vector machine." *Digital signal processing* 17, no. 4 (2007): 694-701.
- [9] Osareh, Alireza, and Bitia Shadgar. "Machine learning techniques to diagnose breast cancer." In *2010 5th international symposium on health informatics and bioinformatics*, pp. 114-120. IEEE, 2010.

- [10] Alyami, Reem, Jinan Alhajjaj, Batool Alnajrani, Ilham Elaalami, Abdullah Alqahtani, Nahier Aldhafferi, Taoreed O. Owolabi, and Sunday O. Olatunji. "Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural network and support vector machines." In *2017 International Conference on Informatics, Health & Technology (ICIHT)*, pp. 1-7. IEEE, 2017.
- [11] Aalaei, Shokoufeh, Hadi Shahraki, Alireza Rowhanimanesh, and Saeid Eslami. "Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets." *Iranian journal of basic medical sciences* 19, no. 5 (2016): 476.
- [12] Ravdin, Peter M., and Gary M. Clark. "A practical application of neural network analysis for predicting outcome of individual breast cancer patients." *Breast cancer research and treatment* 22, no. 3 (1992): 285-293.
- [13] [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
- [14] <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>
- [15] Kourou, Konstantina, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, and Dimitrios I. Fotiadis. "Machine learning applications in cancer prognosis and prediction." *Computational and structural biotechnology journal* 13 (2015): 8-17.
- [16] <https://machinelearningmastery.com/logistic-regression-for-machine-learning/>
- [17] <https://blog.usejournal.com/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>
- [18] <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [19] <https://towardsdatascience.com/understanding-random-forest-58381e0602d2>
- [20] <https://www.datasciencecentral.com/profiles/blogs/logistic-regression-as-a-neural-network>
- [21] <http://deeplearning.stanford.edu/tutorial/supervised/SoftmaxRegression/>