

Discern breast cancer vetting Jupyter Notebook

K. Kusuma¹ Y. Sruthi² G.S.V. Revanth³ B. Tharun Pradeep⁴ Er. K. Tulasi Krishna Kumar⁵
⁵Guide

^{1,2,3,4,5}Department of Computer Science and Engineering
^{1,2,3,4,5}SVPEC, Vizag, India

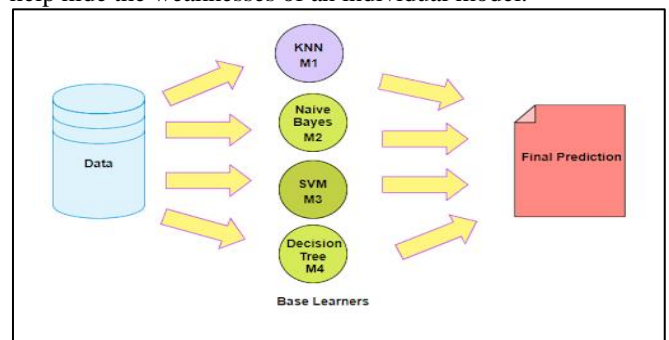
Abstract— This is a paper which is built on the idea of generating a quick diagnosis report prior to the symptoms and tests due to the ever-lingering chance of any female of developing a Breast cancer. This can also be implemented for any type of cancer provided the required datasets. In order to generate a quick diagnosis report for any patient the system uses Machine learning algorithms which extract the knowledge required from previous cases which have been diagnosed with cancer and also without cancer. The end goal is to eliminate the time constraint which is very crucial in determining a person's life as the ratio of doctors to patient is very less in the present-day world, and also reduces the chance of any human error. Specialists are not constantly precise in diagnosing to start with phases of breast cancer. Standard registration is endorsed to each lady in the wake of intersection particular age breaking point and in this manner presented to radiation as a symptom of it, expanding malignant growth hazard. Therefore, there is a need of a substitute arrangement other than exorbitant and hazardous medicinal tests. This paper introduces the utilization of AI calculations for simple acknowledgment of breast malignant growth. Mammograms, Breast ultrasound, and so on are a portion of the therapeutic test, regularly recommended by the specialists, for the finding of breast malignancy. We present a model using different kernels of Ensemble learning for the prediction of breast cancer.

Keywords: Breast Cancer, Machine Learning, Human Error, Diagnosis Report, Ensemble Learning, Algorithms, Mammograms, Breast Ultrasound

I. INTRODUCTION

The structure squares of our body are cells. These cells join to shape various organs and tissues having various capacities. Every one of these cells fix and partitions themselves in a controlled way. Be that as it may, at times, because of some explanation, these cells begin developing in an anomalous manner and structures protuberances which is known as tumour. The abundance of cells around the breast channel prompts structure breast tumour. The breast tumour can additionally arrange into amiable or harmful. Favourable tumours are not malignant growths and can be evacuated effectively with no hazard. In any case, threatening tumour, whenever left untreated the malignant growth cells can spread further into the breast channel and lobules and subsequently cause breast cancer. Breast malignant growth is reparable if the manifestations are business-like in a beginning time however late phases of it are not treatable. Among a wide range of malignancy happen in ladies, breast cancer is the most well-known one... "Breast Cancer influences one out of eight ladies in their lifetime", a general saying about a harmful tumour that starts in cells of the breast and gets into the encompassing tissues also. This infection is normally apparent in ladies yet men can likewise get influenced from it.

The measurements demonstrated that this ailment is regular in ladies and there is a lot of work being done and to be done right now deal with such a deadliest infection. Restorative research focusing on breast cancer isn't new and its underlying foundations return into sixteenth century. Because of the absence of correspondence and progression in restorative field this malady continued taking edge on people and still viewed as probably the deadliest infection of the considerable number of times. Late headway in medicinal field and all the more correctly the contribution of data innovation in the therapeutic field presents another analysis system called Medical Image Processing. Medicinal Image Processing isn't just restricted to malignancy sickness, rather it has helped significantly in the finding of various types of infections and it is clear through measurements. With the assistance of picture handling procedures, it has gotten simpler to recognize tumour from a contaminated breast and analyse breast malignancy. Early location can help in appropriate determination and treatment bringing about limiting the danger of most undesirable result of this infection (demise). Ensemble learning uses multiple machine learning models to try to make better predictions on a dataset. An ensemble model works by training different models on a dataset and having each model make predictions individually. The predictions of these models are then combined in the ensemble model to make a final prediction. Every model has its strengths and weaknesses. Ensemble models can be beneficial by combining individual models to help hide the weaknesses of an individual model.



Ensemble models can also be used for regression problems, where the ensemble model will use either the mean output of the different models or weighted averages for its final prediction.

II. EXISTING METHODOLOGY AND ADVANCEMENTS

Data mining and Machine Learning is comprising of numerous systems which have been utilizing in the clinical field from long back. Heaps of research has been led utilizing distinctive dataset to anticipate disease. Hiba Asri analyzes the presentation of various arrangement ML calculation. In the investigation, he looks at SVM, NB, KNN and c4.5 as far as exactness, exactness, affectability and explicitness to locate the best precise outcome and

closed SVM as a best entertainer. Abdullah-Al Nahid and Yanan Kong, intricately clarifies contribution of AI in the characterization of picture for bosom malignancy. They put forth an attempt to show detail conversation of Convolutional Neural Network (CNN) technique for breast picture order. Abdel-ilah L. furthermore, sahinbegovic H. utilizes Artificial Neural Network so as to structure a model which and finished up, ANN exhibitions with number of neurons in shrouded layers, with best system configuration can be designed with 3 concealed layers and 21 neurons in the shrouded layer with TANSIG actuation work. In a paper introduced by Kihan Park and javdev p. desai breast malignant growth is recognized utilizing delicate edge SVM. Delicate edge SVM separate nonlinear divisible dataset. They utilize exclusively fabricated magnifying instrument perfect small-scale space framework where every indentation is name as typical or dangerous as per the picture confirmed by pathologists on which SVM is applied to sort among these two focuses. The trial gives better outcome when multiple parameters is mulled over. Aside from these there are numerous papers where we can discover effect of AI for the arrangement of breast malignancy. Ensemble methods helps improve machine learning results by combining multiple models. Using ensemble methods allows to produce better predictions compared to a single model. Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging), bias (boosting), or improve predictions (stacking).

III. MACHINE LEARNING

Machine learning is a field of computer science and was coined in 1959 by Arthur Samuel. Machine learning algorithms are very powerful because its concept is drawn from statistics, mathematics and computer science. Machine learning algorithm learns the rules by applying the concept of statistics and mathematics without any written program for it only through the data provided to it. Machine Learning is further divided into Supervised and Unsupervised.

IV. ENSEMBLE LEARNING

In ensemble learning, an agent takes a number of learners and combines their predictions to make a prediction for the ensemble. The algorithms being combined are called base-level algorithms. Random forests are an example of an ensemble method, where the base-level algorithm is decision tree, and the predictions from the individual trees are averaged or are used to vote for a prediction.

In boosting, there is a sequence of learners in which each one learns from the errors of the previous ones. The features of a boosting algorithm are:

There is a sequence of base learners (that can be different from each other or the same as each other), such as small decision trees, or (squashed) linear functions. Each learner is trained to fit the examples that the previous learners did not fit well. The final prediction is a mix (e.g., sum, weighted average or mode) of the predictions of each learner.

The base learners can be weak learners, in that they do not need to be very good; they just need to be better than random. These weak learners are then boosted to be components in the ensemble that performs better than any of them. A simple boosting algorithm is functional gradient boosting which can be used for regression, as follows. The final prediction, as a function of the inputs, is the sum

$$p_0(X) + d_1(X) + \dots + d_k(X)$$

where $p_0(X)$ is an initial prediction, say the mean, and each $d_i(X)$ is the difference from the previous prediction.

Let the i th prediction be $\pi_i(X) = p_0(X) + d_1(X) + \dots + d_i(X)$

Then $\pi_i(X) = \pi_{i-1}(X) + d_i(X)$. Each $d_i(X)$ is constructed so that the error of π_i is minimal, given that π_{i-1} is fixed. At each stage, the base learner learns d_i to minimize

$$\sum_{e \in E} \text{error}(Y_i(e) - \pi_i(e)) = \sum_{e \in E} \text{error}(Y_i(e) - \pi_{i-1}(e) - d_i(e))$$

The i th learner can learn $d_i(e)$ to best fit $Y_i(e) - \pi_{i-1}(e)$. This is equivalent to learning from a modified data set, where the previous prediction is subtracted from the actual value of the training set. In this way, each learner is made to correct the errors of the previous prediction.

- 1) procedure Boosting learner (X_s, Y, E_s, L, k)
- 2) Inputs
- 3) X_s : set of input features
- 4) Y : target feature
- 5) E_s : set of examples from which to learn
- 6) L : base learner
- 7) k : number of components in the ensemble
- 8) Output
- 9) function to make prediction on examples
- 10) mean: $\text{mean}(Y) = \sum_{e \in E} Y(e) / |E|$
- 11) define $p_0(e) = \text{mean}(Y)$
- 12) for each i from 1 to k do
- 13) let $E_i = \{ \langle X_s(e), Y(e) - \pi_{i-1}(e) \rangle \mid e \in E_s \}$
- 14) let $d_i = L(E_i)$
- 15) define $\pi_i(e) = \pi_{i-1}(e) + d_i(e)$
- 16) return π_k

Functional gradient boosting regression learner

V. K- NEAREST NEIGHBOURS (KNN)

K-nearest neighbours (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry. The following two properties would define KNN well –

- Lazy learning algorithm – KNN is a lazy learning algorithm because it does not have a specialized training phase and uses all the data for training while classification.
- Non-parametric learning algorithm – KNN is also a non-parametric learning algorithm because it doesn't assume anything about the underlying data.

A. Working of KNN Algorithm:

K-nearest neighbours (KNN) algorithm uses 'feature similarity' to predict the values of new datapoints which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps –

- Step 1 – For implementing any algorithm, we need dataset. So, during the first step of KNN, we must load the training as well as test data.
- Step 2 – Next, we need to choose the value of K i.e. the nearest data points. K can be any integer.
- Step 3 – For each point in the test data do the following
 - Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most commonly used method to calculate distance is Euclidean.
 - Now, based on the distance value, sort them in ascending order.
 - Next, it will choose the top K rows from the sorted array.
 - Now, it will assign a class to the test point based on most frequent class of these rows.
- Step 4 – End

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

VI. NAÏVE BAYES

Naive Bayes Classifiers rely on the Bayes' Theorem, which is based on conditional probability or in simple terms, the likelihood that an event (A) will happen given that another event (B) has already happened. Essentially, the theorem allows a hypothesis to be updated each time new evidence is introduced. The equation below expresses Bayes' Theorem in the language of probability:

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

Let's explain what each of these terms means.

- "P" is the symbol to denote probability.
- P (A | B) = The probability of event A (hypothesis) occurring given that B (evidence) has occurred.
- P (B | A) = The probability of the event B (evidence) occurring given that A (hypothesis) has occurred.
- P(A) = The probability of event B (hypothesis) occurring.
- P(B) = The probability of event A (evidence) occurring.

VII. SUPPORT VECTOR MACHINE (SVM)

Support vector machines is one of the most noteworthy performing order calculations which was presented in 1992. It is famous on the grounds that it conveys superior with small tuning's classifier separate classes by making hyper plane. Hyperplane is a line that parts a plane into two sections where every one of two classes lies in either side.

The separation between the nearest information focuses and plane is known as edge. There can be numerous lines isolating the classes however the best line is that having biggest edge and known as Maximal Margin hyperplane. Hyperplane line for n-dimensional space will be:

$$B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n = 0$$

SVM order the information by augmenting the hole between the help vectors. On the off chance that the information is directly distinguishable, we can utilize Maximum Margin Classifier. In any case, progressively situation information is not in every case directly divisible. In such cases Support vector classifier (SVC) are utilized to draw hyperplane with some mistake. Support Vector Machine (SVM), which is an augmentation to SVC are utilized for making non-straight limits. Having diverse arrangement of arrangement MMC permit us to pick ideal arrangement. Ideal plane can be chosen by figuring the opposite good ways from each purpose of a dataset and afterward choosing the plane with greatest edge's outcomes by changing component space with a certain goal in mind with the utilization of bits. At the point when we manage high dimensionality information, utilization of bits is should so as to give elite. SVM classifier as a blend of help vector classifier and nonlinear bits. Among different kinds of bits, we can choose best pieces relying upon the dataset. Aside from utilizing various sorts of bits we can tune the classifier by changing the estimation of coefficient related with these parts so as to arrive at best order. Right now, will utilize distinctive SVM portion to fabricate a prescient model which can precisely foresee the bosom malignant growth tumour with no perplexity.

VIII. DECISION TREE

Decision Trees are a kind of Supervised Machine Learning (that is you clarify what the info is and what the comparing yield is in the preparation information) where the information is persistently part as indicated by a specific parameter. The tree can be clarified by two substances, in particular choice hubs and leaves. The leaves are the choices or the ultimate results. Also, the choice hubs are the place the information is part.

There are two primary kinds of Decision Trees:

- 1) Classification trees (Yes/No sorts)

What we've seen above is a case of arrangement tree, where the result was a variable like 'fit' or 'unfit'. Here the choice variable is Categorical.
- 2) Regression trees (Continuous information types)

Here the choice or the result variable is Continuous, for example a number like 123.

A. Working:

Since we comprehend what a Decision Tree is, we'll perceive how it works inside. There are numerous calculations out there which build Decision Trees, however a standout amongst other is called as ID3 Algorithm. ID3 Stands for Iterative Dichotomiser 3.

Before talking about the ID3 calculation, we'll experience barely any definitions.

B. Entropy:

Entropy, additionally called as Shannon Entropy is meant by $H(S)$ for a limited set S , is the proportion of the measure of vulnerability or arbitrariness in information.

$$H(S) = \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)}$$

Instinctively, it tells us about the predictability of a certain occasion. Example, consider a coin toss whose probability of heads is 0.5 and probability of tails is 0.5. Here the entropy is the highest possible, since it is extremely unlikely of figuring out what the result might be. Alternatively, consider a coin which has heads on both the sides, the entropy of such an occasion can be anticipated flawlessly since we know beforehand that it'll always be heads. At the end of the day, this occasion has no randomness thus its entropy is zero.

In particular, lower values suggest less uncertainty while higher values infer high uncertainty.

C. Information Gain:

Information gain is also called as Kullback-Leibler uniqueness signified by $IG(S, A)$ for a set S is the successful change in entropy after settling on a particular attribute A . It measures the relative change in entropy with respect to the free variables.

$$IG(S, A) = H(S) - H(S, A)$$

Alternatively,

$$IG(S, A) = H(S) - \sum_{i=0}^n P(x) * H(x)$$

where $IG(S, A)$ is the information gain by applying feature A . $H(S)$ is the Entropy of the entire set, while the second term calculates the Entropy after applying the feature A , where $P(x)$ is the probability of event x .

IX. ADOPTING CASE BASED REASONING (CBR) IN JUPYTER NOTEBOOK

Case-based reasoning is utilized for order and for relapse. It is additionally material when the cases are confounded, for example, in lawful cases, where the cases are mind boggling lawful decisions, and in arranging, where the cases are past answers for complex issues.

If the cases are simple, one algorithm that works well is to use the k -nearest neighbours for some given number k . Given a new example, the k training examples that have the input features closest to that example are used to predict the target value for the new example. The prediction could be the mode, average, or some interpolation between the prediction of these k training examples, perhaps weighting closer examples more than distant examples.

Information gain is also called as Kullback-Leibler uniqueness signified by $IG(S, A)$ for a set S is the successful change in entropy after settling on a particular attribute A . It measures the relative change in entropy with respect to the free variables.

Alternatively,

where $IG(S, A)$ is the information gain by applying feature A . $H(S)$ is the Entropy of the entire set, while the second term calculates the Entropy after applying the feature A , where $P(x)$ is the probability of event x .

For this method to work, a distance metric is required that measures the closeness of two examples. First define a metric for the domain of each feature, in which the values of the features are converted to a numerical scale that is used to compare values. Suppose $X_i(e)$ is a numerical representation of the value of feature X_i for the example e . Then $(X_i(e_1) - X_i(e_2))$ is the difference between example e_1 and e_2 on the dimension defined by feature X_i . The Euclidean distance, the square root of the sum of the squares of the dimension differences, could be used as the distance between two examples. One important issue is the relative scales of different dimensions; increasing the scale of one dimension increases the importance of that feature. Let w_i be a non-negative real-valued parameter that specifies the weight of feature X_i . The distance between examples e_1 and e_2 is then

$$d(e_1, e_2) = \sqrt{\sum_i w_i * (X_i(e_1) - X_i(e_2))^2}$$

The feature weights could be provided as input. It is also possible to learn these weights. The learning agent would try to find weights that minimize the error in predicting the value of each element of the training set, based on every other instance in the training set. This is an instance of leave-one-out cross validation.

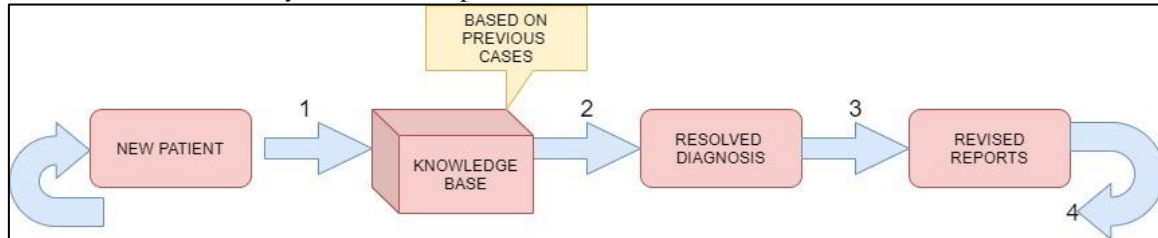
One of the problems in case-based reasoning is accessing the relevant cases. A k -d-tree is a way to index the training examples so that training examples that are close to a given example can be found quickly. Like a decision tree, a k -d-tree splits on input features, but at the leaves are subsets of the training examples. In building a k -d-tree from a set of examples, the learner tries to find an input feature that partitions the examples into set of approximately equal size and then builds k -d-trees for the examples in each partition. This division stops when there are only few examples at each leaf. A new example is filtered down the tree, as in a decision tree. The exact matches will be at the leaf found. However, the examples at the leaves of the k -d-tree could possibly be quite distant from the example to be classified; they agree on the values down the branch of the tree but could disagree on the values of all other features. The same tree is used to search for those examples that have one feature different from those tested in the tree, by allowing for one branch of a different value to be used when filtering a new case down the tree.

Case-based reasoning is also applicable when the cases are more complicated, for example, when they are legal cases or previous solutions to planning problems. In this scenario, the cases are carefully chosen and edited to be useful. Case-based reasoning consists of a cycle of the following four steps:

- 1) Retrieve
Given a new case, retrieve similar cases from the case base.
- 2) Reuse
Adapt the retrieved cases to fit to the new case.

- 3) Revise
Evaluate the solution and revise it based on how well it works.
- 4) Retain
Decide whether to retain this new case in the case base.
If the case retrieved works for the current situation, it should be used. Otherwise, it may need to be adapted. The

revision may involve other reasoning techniques, such as using the proposed solution as a starting point to search for a solution, or a human could do the adaptation in an interactive system. The new case and the solution can then be saved if retaining it will help in the future.



X. DESCRIPTION OF BREAST CANCER DATASET

In this paper we are using the dataset obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The name of this dataset is Wisconsin Breast Cancer Database which is freely available in internet in a site known as UCI. This dataset contains the record of 569 patients consisting altogether 11 attributes. Our aim is to classify Label class attribute of this dataset consisting two

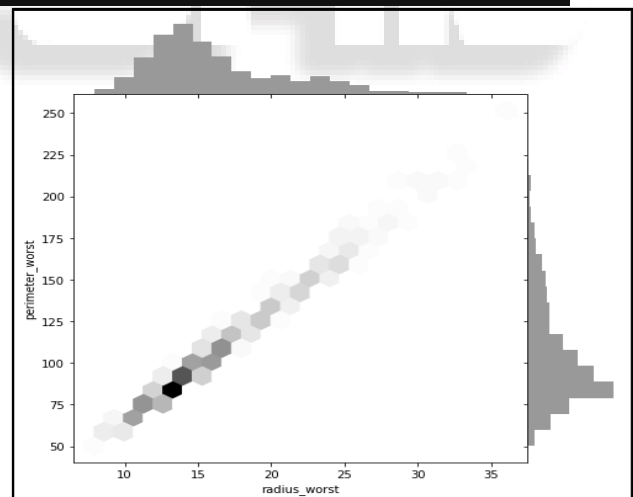
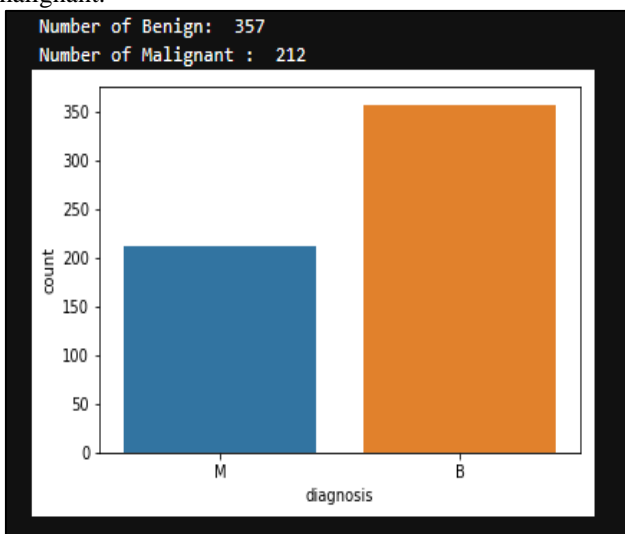
types of tumour (benign or malignant). As discussed earlier benign is non dangerous tumour while malignant can cause cancer and hence should be predictable for further investigation. This dataset uses 2 for symptomatic of benign, 4 for symptomatic of malignant. Dataset consist 16 missing values, removing them is not a good option as we can lose valuable information, we had replaced all missing values with standard number format -99999. Here's a sample of 5 rows of the dataset out of the 569 patients.

id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.147
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.070
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.127
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.105
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.104

5 rows x 33 columns

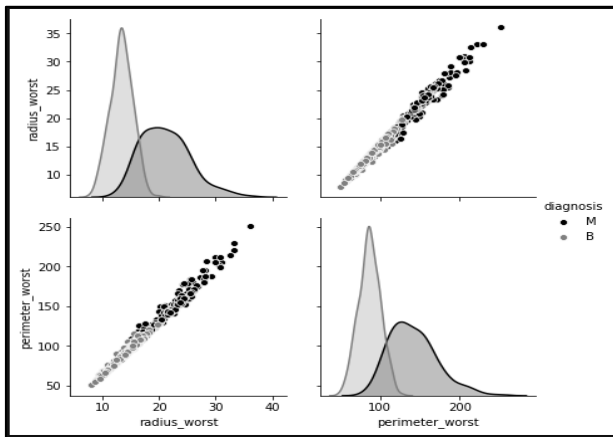
XI. DATASET OBSERVATION

The Wisconsin Breast Cancer Database is consisting of 569 records and 11 attributes. The first attribute in the dataset is id had been dropped to form a new dataset with 10 attributes. Other 9 attributes of the dataset contain random values ranging between 1 to 10. As presented in figure, we can see that we have 357 cases of benign and 212 cases of malignant.



XII. EXPERIMENTAL RESULT

We divided the dataset randomly into two parts, training dataset (75% of dataset) and test dataset (25% of dataset). We build a model using Ensemble Learning. We check the model with different types of model kernels used in it. We further fine-tuned the model by gradient boosting classifier parameters. Ensemble model with multiple kernels with an accuracy of 99.53% (0.9953 out of 1.0) as a better model for predicting breast cancer.



```
max_boosting = GradientBoostingClassifier(max_depth=1)
max_boosting.fit(new_x_train,Y_train)

print(f"Gradient boosting training set accuracy: {format(max_boosting.score(new_x_train,Y_train), '.4f')} ")
print(f"Gradient boosting testing set accuracy: {format(max_boosting.score(new_x_train,Y_train), '.4f')} ")

Gradient boosting training set accuracy: 0.9953
Gradient boosting testing set accuracy: 0.9953
```

XIII. CONCLUSION

Breast malignancy regularly happen in ladies for the most part after the time of 35 years. Maximum level of patient experiencing this sickness passes on the off chance that it isn't known toward the start. As the hazard is high, we need an answer with best outcome of diagnosis. For this reason, we use Ensemble learning alongside direct, CBR and poly pieces to manufacture prescient model. Lastly, we conclude Ensemble model with multiple kernels with an accuracy of 99.53% (0.9953 out of 1.0) as a better model for predicting breast cancer.

XIV. FUTURE SCOPE

This can be further improved into an expert system or direct implementation into an artificial intelligence (AI). Which deal with medical advancements in technology in the future, limiting human intervention and errors with diagnosis. Wellbeing websites visit bots and online paramedical help frameworks can be created by medicinal organizations and colleges. It can likewise be utilized by master medicinal frameworks in helping the house specialists and junior specialists for recommending therapeutic guide and propelled learning. This can further be implemented for any sort of disease provided the dataset like coronavirus, infections and also predicting outbreaks of diseases.

REFERENCES

[1] Vikas Chaurasia, Saurabh Pal & etal, "Prediction of benign and malignant breast cancer using data mining techniques", Journal of Algorithm & Computational Technology,2018

[2] Prity Vijay, Bright Keshwani," Emergence of Big Data with Hadoop: A Review", IOSR Journal of Engineering (IOSRJEN)", Vol:06, Issue:03,2016, pp50-54

[3] [Machine learning: https://en.wikipedia.org/wiki/Machine_learning

[4] Ensemble learning: https://en.wikipedia.org/wiki/Ensemble_learning

[5] HibaAsria, Hajar Mousannif & etal, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis", The 6th International Symposium on Frontiers in Ambient and Mobile Systems (Elsevier), Vol:83, 2016, pp:64-69

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

[7] Layla Abdel-Ilah, Hana Šahinbegović, "Using Machine Learning Tool in Classification of Breast Cancer", Springer, Vol:62,2017

[8] Kihan Park, Jaydev P. Desai "Machine learning approach for breast cancer localization" IEEE Explorer,2017

[9] Breast Cancer Wisconsin (Original) Data Set [https://archive.ics.uci.edu/ml/datasets/Breast +Cancer+ Wisconsin+ %28Original%29](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29)

[10] Ensemble learning: <https://artint.info/2e/html/ArtInt2e.Ch7.S6.SS2.html>

[11] Case based reasoning: <https://artint.info/2e/html/ArtInt2e.Ch7.S7.html>

[12] Support Vector Machine (SVM) Kernels based approach for detection of Breast Cancer. Prity Vijay1, Bright Keswani2 1Research Scholar, Department of Computer Sciences, Suresh Gyan Vihar University, Jaipur, India 2Associate Professor, Department of Computer Applications, Suresh Gyan Vihar University, Jaipur, India.

[13] Decisiontrees:<https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html>

[14] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm

[15] NAÏVE BAYES: <https://blog.easysol.net/machine-learning-algorithms-4/>