Prediction of Breast Cancer Application By Using Machine Learning And Data Mining Techniques

Kogali Rohit[1], Dr.H.Girisha[2]
Department of CS&E Engineering
[1] M.Tech, RYMEC, BALLARI
[2] M.Tech, Ph.d Professor & Co-ordinator, RYMEC, BALLARI.

Abstract- *Breast Cancer (BC) is one in all the foremost normal tumors among women's around the globe, addressing the primary piece of latest disease cases and threatening development related going with respect to all inclusive estimations, making it an essential open debilitated prosperity in the present society. The essential conspicuous confirmation of B.C. will improve the estimate and likelihood of survival astonishingly, in light of the way that it will lift promising clinical treatment to patients. Dynamically right plan of great tumors will ruin patients encountering pointless drugs. Thusly, the right distinctive verification of B.C. likewise, request of patients into Cancer or generous gatherings is that the subject of a lot of examination. By virtue of its specific supports in essential options distinguishing proof from ensnared B.C. informational collections, AI (ML) is wide seen because the strategy of alternative in B.C. configuration request and guess illustrating. In the midst of this paper, we will in general plan to review mil frameworks and their applications in B.C. conspicuous verification and expectation. At first, we offer a structure of mil methodologies together with phony neural frameworks (ANNs), reinforce vector machines (SVMs), decision tree (DTs), and k-nearest neighbors (k-NNs).we will in general research their applications in B.C; our basic information is drawn from the Wisconsin Diagnostic Breast Cancer (WDBC) that can't avoid being that the benchmark database for relationship result through extremely astonishing computations. Finally a thought system model of our progressing the work is in addition showed up. The prognosticative models referenced here are supported vacillated controlled mil frameworks relatively as on altogether unforeseen information options and information tests. Given the creating design on the applying of mil courses in dangerous development examination, we tend to gift here the superior late conveyances that use these techniques as accomplice hope to show infection peril or patient outcomes.*

I. INTRODUCTION

Over previous year's, ceaseless development related with disease examination has been performed. Researchers connected totally unique ways, such as Screening in beginning period, in order to look out sorts of Cancer growth before they cause indications. They need grew new ways for the principal forecast of cancer treatment result. With the presence of most recent advances inside the fields of medicine, goliath measures of cancer growth data are gathered and are available to the restorative investigation network. It may, the right forecast of a disease result is one among the chief eye catching and troublesome undertakings for doctors. Accordingly, cubic centimeter ways turned into an all around preferred device for restorative specialists. These systems will find and set up examples and connections between them, from cutting edge datasets, while they're ready to adequately anticipate future results of a Cancer sort.

ML methods are wide utilized in insightful consideration diseases, especially for carcinoma (BC) assignment and anticipation. Verifiably the symptomatic precision of a patient relies upon a doctor's skill be that as it may, this experience is developed up over a couple of long periods of perceptions of different patients' side effects and affirmed analyze. And, after it's all said and done the precision still can't be verified. With the presence of figuring innovations, it's as of now relatively clear to aggregate and store heaps of learning, for instance the devoted data of electronic patient records and furthermore the European Friedreichs engine ataxia pool for movement Studies (EFACTS) persistent database while not the assistance of PCs it's unrealistic for wellbeing experts to investigations these progressed datasets altogether once attempt complex cross examinations of the information.

As the most widely recognized Cancer development in Women's, BC has reliably had a high recurrence rate and passing rate. As shown by the latest Cancer development bits of knowledge, BC alone is depended upon to speak to 25% of all new illness decisions and 15% of all ailment passings among Women's around the globe. Specialists have thought about the dangers of BC from in all regards at a beginning period, as such much early research has quite recently been realized in the treatment of BC. As a result of the undertakings of researchers and early recognizable proof methods, the demise rate has seemed unwavering and declining design over the earlier decades. As demonstrated by the bits of knowledge of Cancer Research UK, the five years survival rate for BC is basically 100% at whatever point recognized at its most

prompt stage, anyway can be as low as 15% when recognized at the latest stage. Recently, ML techniques are accepting an imperative employment in assurance and speculation of BC by applying request systems to perceive people with BC, perceive great from destructive tumors and to predict expectation. Accurate request can moreover assist clinicians with recommending the most fitting treatment schedule. Portrayal is a kind of complex improvement issue.

Various ML methodologies have been associated by experts in dealing with this request issue. In the going with zones, a broad elucidation of different portrayal techniques associated with BC will be given. We center around the counterfeit neural system (ANNs), bolster vector machine (SVMs), choice tree (DTs) and k-closest neighbor (k-NNs) procedures as they are the rule strategies as they are the fundamental techniques utilized in BC finding and forecast. Scientists attempt to find the best calculation to achieve the most precise request result; in any case, data of variable quality will in like manner sway the gathering result. Further, the extraordinary quality of data will affect the amount of calculation applications too. All things considered, most ML methodology are first attempted in open source databases. After some time, Over time, a benchmark dataset has emerged in the writing: Wisconsin bosom disease analysis (WBCD).There is additionally numerous other BC benchmark informational indexes, for example Wisconsin Prognostic Breast Cancer Chemotherapy (WPBCC), Wisconsin Diagnostic Breast Cancer (WDBC, etc. ML procedures that have been utilized on the WBCD database in BC conclusion and visualization show diverse dimensions of exactness that extended somewhere in the range of 94.36% and 99.90%. Correspondingly, there are results with particularly modified Algorithms relating to BC databases. This review attempts to give per clients the essential segments of BC examination and speculation using ML techniques on WBCD. By using ML diseases to examinations the WBCD database, BC can be analyzed precisely base on 9 traits as can be seen from Table (Sample Cancer Attribute). In the principle body of the survey area we will focus on how the WBCD has been utilized to represent the incredible guarantee of ML Algorithms.

Table- Sample Cancer Attribute

Number	Attribute	Domain
1	Sample code number	Id number
2	Clump Thickness	1-10
3	Uniformity of cell size	1-10
4	Uniformity of cell shape	1-10
5	Marginal Adhesion	1-10
6	Single Epithelial cell size	1-10
7	Bare Nuclei	1-10
8	Bland chromatin	1-10
9	Normal nucleoli	1-10
10	Mitoses	1-10
11	Class	2 for benign 4 for cancer

II. RELATED WORK: SURVIVAL PAPER

We chose these investigations that utilize conspicuous ML strategies and coordinated information from heterogeneous sources so as to anticipate the alluring result. Cancer growth type, ML technique, number of patients, sort of information just as the general exactness accomplished by each proposed strategy are introduced. Each sub-table compares to examine with respect to a particular situation (for example Cancer growth defenselessness forecast, disease repeat expectation and disease survival expectation). It ought to be noticed that in articles that more than one ML strategies are connected for forecast, we chose to introduce here the most precise predictive model.

I) Prediction of cancer survival Method

In a predictive model is produced for the assessment of survival in ladies that have been determined to have bosom Cancer growth, while they tended to the significance of vigor under the model's parameter variety. They analyzed three order models specifically SVM, ANN and SSL dependent on the SEER Cancer growth database. The dataset is made out of 162,500 records with 16 key highlights. A class variable was likewise considered, to be specific survivability, alluding to patients that had not endured and those that had endure. Among the most instructive highlights are: (I) the tumor estimate, (ii) the quantity of hubs and (iii) the age at determination. By looking at the best execution of every one of the three models they found that the determined precision for ANN, SVM and SSL was 65%, 51% and 71% individually. Five-overlap cross approval was utilized for assessing the execution of the prescient models. The little boxes of the SSL display uncovered its preferred exactness over different models.

II) Machine learning models in breast cancer survival predictions.

We utilize a dataset with eight properties that incorporate the records of 900 patients in which 876 patients (97.3%) and 24 (2.7%) patients were females and guys separately. Innocent Bayes (NB), Trees Random Forest (TRF), 1-Nearest Neighbor (1NN), AdaBoost (AD), Support Vector Machine (SVM), RBF Network (RBFN), and Multilayer Perceptron (MLP) AI systems with 10-cross overlay strategy were utilized with the proposed model for the forecast of bosom Cancer growth survival. The execution of AI methods was assessed with exactness, accuracy, affectability, particularity, and territory under ROC bend. Out of 900 patients, 803 patients and 97 patients were alive and dead, individually. In this investigation, Trees Random Forest (TRF) system demonstrated better outcomes in contrast with different procedures (NB, 1NN, AD, SVM and RBFN, MLP). The exactness, affectability and the zone under ROC bend of TRF are 96%, 96%, 93%, individually. Be that as it may, 1NN AI method gave poor execution (exactness 91%, affectability 91% and region under ROC bend 78%).

III) Prediction of cancer recurrences

In based of our overview, we here present the most significant and late distributions that proposed the utilization of ML methods for disease repeat expectation. A work which thinks about the repeat forecast of oral squalors cell carcinoma (OSCC) is proposed in. They recommended a multi parametric Decision Support System so as to break down the premise of OSCC evolvement after absolute reduction of Cancer growth patients. They misused heterogeneous wellsprings of information (clinical, imaging and genomic) so as to anticipate a possible relapse of OSCC and along these lines a resulting repeat.

A particular component choice technique was pursued with the work of two element determination algorithms, specifically CFS and wrapper calculation. Therefore, any inclination could be dodged while choosing the most useful highlights of their reference heterogeneous dataset. At that point they chose critical factors could be utilized as info vectors to explicit classifiers. Before the work of the component determination procedures the all out number of the clinical, imaging and genomic highlights was 65, 17 and 40 in every class.

III. EXISTING SYSTEM

Cancer is one of the serious issue today, diagnosing Cancer in prior stage is as yet trying for specialists. Bosom Cancer growth is one of the significant passing causing illnesses of the ladies today everywhere throughout the world. This makes us to take up this issue and to actualize the ML based Cancer growth forecast System. Past ML algorithms are not demonstrated exactness and forecast on Breast Cancer growth utilizing ANN and Feature determination calculation. To defeat foreseeing and guess of Breast Cancer growth precision and dependability in our proposed work Every year in excess of million ladies are determined to have bosom disease worldwide over portion of them will kick the bucket on account of the late diagnosing of the infection. Such a significant number of examines have experienced for recognizing the Cancer dependent on ML innovation each methodology has its very own constraints. This makes us to take up this issue and to actualize the ML based Cancer expectation System.

Disadvantages of Existing System

- Cancer prediction is certainly very complex.
- The system is expensive.
- The system accuracy is less.

IV. PROPOSED SYSTEM

In this disease, we've anticipated a model for partner early prediction of Cancer growth ailment by abuse practical AI strategies. The arrangement of errands which will be completed in our anticipated work is investigated, planned, authorized and tested abuse one among the apparatuses like Python by altering the present AI algorithms. These have collection of AI algorithms for information preparing errands. The algorithmic program is connected on to the data sets or alluded to as from our own fake language code. Instruments contain methods for data pre-preparing, order, relapse, affiliation rules, and visual picture. For capacity necessities, MySQL information server is utilized. Partner configuration AI procedure basically based disease forecast disease consolidating the expectation disease with metric limit unit innovation was utilized. Amid this model we've utilized one among the grouping algorithms alluded to as choice tree. When the client goes into the Cancer expectation disease, they need to answer the inquiries, related with hereditary and non-hereditary elements. At that point the expectation disease doles out the opportunity worth to each question upheld the client reactions. When the shot worth is normal, the shift of the opportunity is controlled by the forecast disease. Its four dimensions of hazard like low dimension, middle of the road level, abnormal state and incredibly abnormal state. Bolstered the normal hazard esteems the fluctuate of hazard are dispensed. On the off chance that the peril cost is a littler sum than or sufficient to thirty five the individual probably won't have a Cancer the tests or educated. In the event that the threat cost is bigger than or sufficient to forty, at that point the individual could have a disease, the Cancer kind and tests are told. In the event that the peril cost is bigger than or sufficient

to forty five, at that point the individual could have a Cancer growth, the sort of disease is anticipated upheld the indications. On the off chance that the peril cost is bigger than or satisfactory to fifty five, at that point the individual could have a Cancer growth, at that point it's predicted by fundamental incessant examples. a call tree might be a stream diagram like tree structure, wherever every interior hub signifies an investigate partner degree property, each branch speaks to relate degree result of the investigate and each leaf hub holds a class mark. The most elevated most hub is that the root hub. The property cost of the data is tried against a call tree. A way is duplicated from root to leaf hub that holds the classification expectation for that learning. Call trees might be just conceived again into arrangement rules. This call tree is utilized to get visit designs inside the dataset. {the learning |the info| the information} and thing sets that happen oft inside the information base are alluded to as regular examples. The incessant examples that is above all related with explicit Cancer growth assortments and are valuable in anticipating the disease. Applying the decision tree for the given client qualities it follow a definitive outcome and anticipate the sort of the Cancer.

V. PROPOSED METHODOLOGY

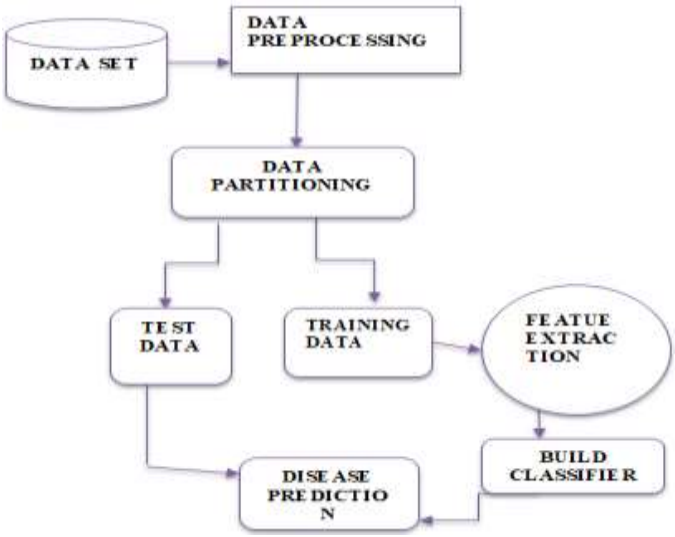

Fig 1:- Proposed method of Cancer Prediction

5.1 Machine learning work flow diagram

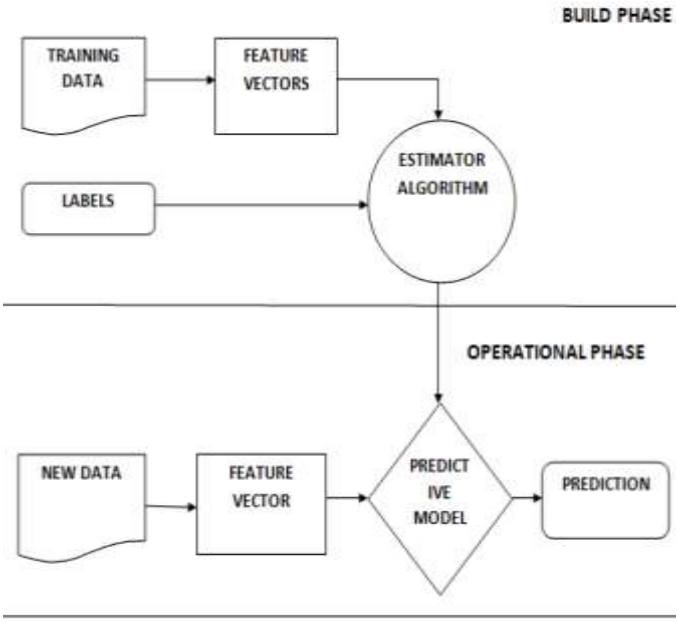

Fig2:- Work Flow diagram of ML

VI. MACHINE LEARNING APPROACHES: - REVIEW OF ML

ML algorithms have been commonly used in BC end and representation to increment different bits of learning from data tests. ML is a kind of Artificial Intelligence (AI) that uses a grouping of verifiable, probabilistic and streamlining devices to pick up and improve execution normally from new data and past experiences, without unequivocally modified rules. The two experiences and ML are used for separating data. In overseeing puzzling, colossal, high dimensional data, ML approaches are generally prepared for removing key features and potential rules which might be difficult to be found using traditional bits of knowledge. BC datasets, ML approaches have wound up being more sensible than estimations ones. Customary quantifiable methodologies for data examination to envision restorative outcomes fuse anyway are not confined to, determined backslide, direct backslide and discriminant examination. Much work differentiating particular standard quantifiable systems and ordinary ML gathering methodologies have been circulated to demonstrate the upsides of ML and its potential. Differentiated ANN and straight discriminant examination of the radiographic features of masses and patient's age in 139 suspicious masses to perceive whether they made chest ailment and demonstrated the prevalence of last referenced. Recently in any case with the advancement and improvement of the ML algorithms, and the growing sum and multifaceted nature of the data, results show ML approaches have better request exactness.

In 2004, two ML Classification strategies (DT and ANN) were contrasted and a measurable technique (straight

regression)to anticipate the bosom disease survival utilizing an extensive dataset which has more than 200, 000 cases and exhibited that ML techniques could be a promising characterization strategy for reasonable use. The outcomes demonstrated that DT was the best indicator with the exactness of 93.6% with ANN accomplishing 91.2% and both were superior to anything straight relapse accomplishing 89.2% precision ML algorithms can in like manner be used to perform nonlinear real showing indicating inclinations when stood out from standard quantifiable methods, these fuse the ability to orchestrate and perceive jumbled nonlinear associations between components, the ability to recognize each and every believable participation between pointer factors, and the openness of various readiness algorithms ML algorithms can be assembled into two general orders reliant on the way they "learn" about data; coordinated learning and unsupervised learning. Coordinated learning is so named in light of the way that the engineer empowers the learning methodology about what results should be from the arrangement data, as a manual for the algorithms. It resembles the way in which that the gatekeepers demonstrate their adolescent the names of different articles. Guided learning requires all readiness data to be named, and after that algorithms can make sense of how to envision the yield from the data. On the other hand, in unsupervised adjusting, all of the data is unlabeled, it should have the ability to find within structure or association between different information sources. Picking whether to use a directed or an unsupervised ML count generally depends upon the data types and structures. We will give a point by point presentation on the principle sorts of managed order learning algorithms that have been utilized in the BC determination and forecast, which are ANN, DT, SVM and k-NN. A counterfeit neural system (ANN) can be characterized as a model of thinking dependent on the human mind.

ANNs have managed various triumphs with extraordinary advancement in BC order and finding in the all around beginning periods the way toward preparing an ANN may include long causal chains of computational stages. Since 1986, a novel proficient inclination drop calculation got back to spread (BP) has had wide applications, particularly for restorative information. The information layer gets the information and after that transmits the information to a shrouded layer which is utilized for handling the information and giving the preparation results to the yield layer. The yield layer demonstrates the order results. An ordinary ANN show is comprised of a chain of command of layers: input, covered up and yield layers (Figure). Layers are made out of interconnected neurons which contain an actuation work for nonlinear change to reinforce the nonlinear articulation capacity. It was made by summing up the Widrow-Hoff learning guideline to numerous layer systems and nonlinear differentiable exchange function.

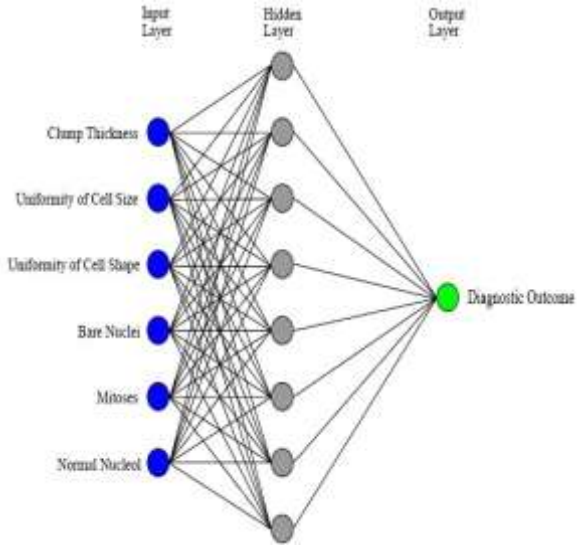

Fig:- Diagram of ANN is trained to predict the diagnostic

The idea of SVM, which was proposed by Vapnik based on the measurable learning hypothesis, has turned into a fundamental part in ML strategies. SVM was at first created for double characterization, however it very well may be effectively stretched out for multiclass issues with across the board use in fields of written by hand digit acknowledgment, speaker ID, content order and so on. The key element of a SVM classifier is to discover a streamlined choice limit that speaks to the biggest partition (greatest edge) between the classes. The standard of SVM begins from taking care of straight divisible issues at that point reaches out to manage nonlinear issues. The method for taking care of nonlinear issues is to outline tests from the first limited dimensional space to a higher dimensional space to acknowledge straight distinctness. SVM is a standout amongst the most well known methodologies in BC finding and anticipation. A case of SVM structure for BC finding is appeared in Figure.

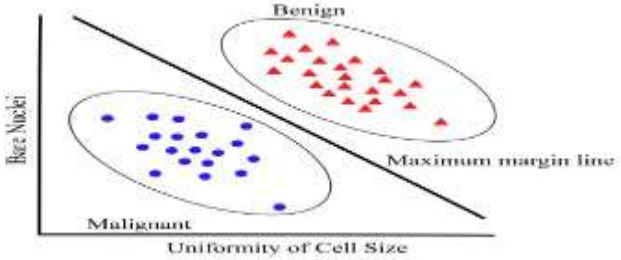

Fig:-Diagram of how an SVM might work in distinguishing between benign and malignant tumor

In ML, DT is a predictive model that represent to a mapping between object attributes and item valuesC5.0 and characterization and relapse tree (CART) which use entropy-based measures to develop the tree DT is like the flowchart, in which each non-leaf hub shows a test on a specific quality, each branch speaks to a result of that test and each leaf hub communicates an order or choice. The most well known DT strategies are Iterative Dichotomiser 3 (ID3), C4.5 A DT is a tree-like classifier that segments each conceivable result of information recursively into classes, DT works with regards to BC analysis and forecast, and there is a case of a BC determination and visualization choice tree structure.

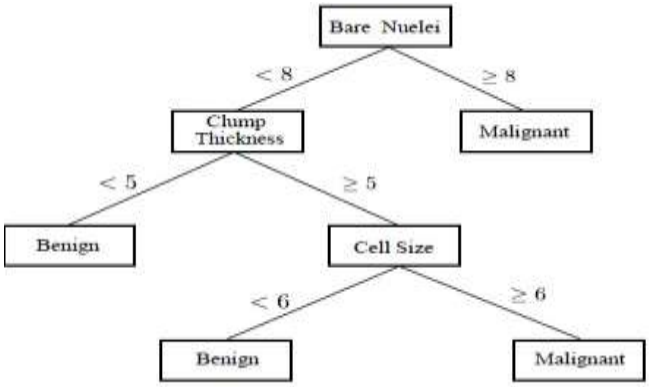

Fig:- Diagram of DT is used to solve BC diagnosis problem

k-NN is a standout amongst the most central ML methods in order, k-NN is a non-parametric lazy learning calculation utilized for characterization, which arranges the articles utilizing their "k" closest neighbors. K-NN just thinks about the neighbors around the article, not the hidden data distribution. Furthermore, there is no preparation stage with the preparation information. k-NN structure is displayed for deciding BC analysis and guess if k = 3, the test (circle) is doled out to dangerous BC (square) in light of the fact that there are 2 squares and just 1 triangle inside the inward circle. In the event that k = 5, the test is doled out to generous BC (triangle).

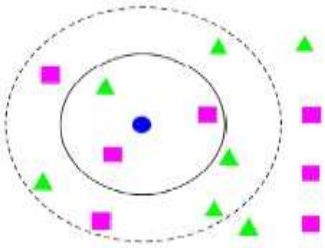

Fig:- Diagram of k-Nearest neighbor for breast cancer diagnosis. Blue circle means the test sample; green triangle means the malignant BC and Pink Square means the benign BC.

VII. MACHINE LEARNING ALGORITHMS USED

1). SVM:-SUPPORT VECTOR MACHINE

The goal of the support vector machine algorithm is to discover a hyperplane in a N-dimensional space (N—the number of highlights) that particularly groups 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 greatest edge, for example the greatest separation between information purposes of the two classes. Boosting the edge remove gives some support so future information focuses can be characterized with more certainty. In AI, support-vector machines (SVMs, additionally support-vector networks) are managed learning models with related learning calculations that dissect information utilized for arrangement and relapse examination. Given a lot of preparing precedents, each set apart as having a place with either of two classes, a SVM preparing calculation fabricates a model that doles out new guides to one classification or the other, making it a non-probabilistic paired straight classifier (despite the fact that techniques, for example, Platt scaling exist to utilize SVM in a probabilistic order setting). A SVM show is a portrayal of the precedents as focuses in space, mapped with the goal that the instances of the different classes are isolated by a reasonable hole that is as wide as could be expected under the circumstances.

2). K-NN: - K-NEAREST NEIGHBORS

K nearest neighbors is a simple algorithm that stores every accessible case and arranges new cases dependent on a closeness measure (e.g., distance functions). KNN has been utilized in factual estimation and example acknowledgment as of now in the start of 1970's as a non-parametric procedure.

Algorithm

A case is ordered by a lion's share vote of its neighbors, with the case being doled out to the class most normal among its K closest neighbors estimated by a separation work. On the off chance that K = 1, at that point the case is basically allocated to the class of its closest neighbor.

Distance functions

Euclidean $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

Manhattan $\sum_{i=1}^{k}|x_i - y_i|$

Minkowski $\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q}$

It ought to likewise be noticed that every one of the three separation measures are substantial for persistent factors. In the occasion of unmitigated factors the Hamming separation must be utilized. It additionally raises the issue of institutionalization of the numerical factors somewhere in the range of 0 and 1 when there is a blend of numerical and all out factors in the dataset.

Hamming Distance

$$D_H = \sum_{i=1}^{k}|x_i - y_i|$$

$$x = y \Rightarrow D = 0$$

$$x \neq y \rightarrow D = 1$$

X	Y	Distance
Male	Male	0
Male	Female	1

Picking the ideal incentive for K is best done by first assessing the information. When all is said in done, a substantial K esteem is increasingly exact as it diminishes the general clamor yet there is no assurance. Cross-approval is another approach to reflectively decide a decent K esteem by utilizing an autonomous dataset to approve the K esteem. Generally, the ideal K for most datasets has been between3-10. That produces much preferred outcomes over 1NN.

VIII. CONCLUSION

In this project as we discussed about the ideas of ML while we laid out their application in cancer prediction/prognosis. The majority of the examinations that have been proposed the most recent years and spotlight on the improvement of prescient models utilizing administered ML techniques and order calculations planning to anticipate substantial ailment results. Here we additionally intrigued to structure and build up a model for predicting cancer disease.

REFERENCES

[1] HanahanD,Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144: 646–74.

[2] Polley M-YC, Freidlin B, Korn EL, Conley BA, Abrams JS, McShane LM. Statistical and practical considerations for clinical evaluation of predictive biomarkers. J Natl Cancer Inst 2013;105:1677–83.

[3] Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. Cancer Informat 2006;2:59.

[4] Fortunato O, BoeriM, Verri C, Conte D,MensahM, Suatoni P, et al. Assessment of circulating microRNAs in plasma of lung cancer patients. Molecules 2014;19:3038–54.

[5] Heneghan HM, Miller N, Kerin MJ. MiRNAs as biomarkers and therapeutic targets in cancer. CurrOpinPharmacol 2010;10:543–50.

[6] Madhavan D, Cuk K, Burwinkel B, Yang R. Cancer diagnosis and prognosis decoded by blood-based circulating microRNA signatures. Front Genet 2013;4.

[7] Zen K, Zhang CY. CirculatingmicroRNAs: a novel class of biomarkers to diagnose and monitor human cancers. Med Res Rev 2012;32:326–48.

[8] KoscielnyS.Whymost gene expression signatures of tumors have not been useful in the clinic. SciTransl Med 2010;2 [14 ps12-14 ps12].

[9] Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 2005;365:488–92.

[10] Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.

[11] Mitchell TM. The discipline of machine learning: Carnegie Mellon University. Carnegie Mellon University, School of Computer Science, Machine Learning Department; 2006.

[12] Witten IH, Frank E. Data mining: practical machine learning tools and techniques. Morgan Kaufmann; 2005.

[13] Niknejad A, Petrovic D. Introduction to computational intelligence techniques and areas of their applications in medicine. Med ApplArtifIntell 2013;51.

[14] Pang-Ning T, Steinbach M, Kumar V. Introduction to data mining; 2006.

[15] Drier Y, Domany E. Do two machine-learning based prognostic signatures for breast cancer capture the same biological processes? PLoS One 2011;6:e17795.

[16] Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99:147–57.

[17] Reetz, K.; Dogan, I.; Costa, A.S.; Dafotakis, M.; Fedosov, K.; Giunti, P.; Parkinson, M.H.; Sweeney, M.G.; Mariotti, C.; Panzeri, M.; et al. Biological and clinical characteristics of the European Friedreich's Ataxia Consortium for Translational Studies (EFACTS) cohort: A cross-sectional analysis of baseline data. Lancet Neurol. 2015, 14, 174–182.

[18] Chen, W.; Zheng, R.; Baade, P.D.; Zhang, S.; Zeng, H.; Bray, F.; Jemal, A.; Yu, X.Q.; He, J. Cancer statistics in B China, 2015. CA Cancer J. Clin. 2016, 66, 115–132.

[19] Sharma, A.; Kulshrestha, S.; Daniel, S. Machine learning approaches for breast cancer diagnosis and prognosis. In Proceedings of the International Conference on Soft Computing and Its Engineering Applications, Changa, India, 1–2 December 2017.

[20] Howell, A.; Cuzick, J.; Baum, M.; Buzdar, A.; Dowsett, M.; Forbes, J.F.; Hoctin-Boes, G.; Houghton, J.; Locker, G.Y.; Tobias, J.S.; et al. Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. Lancet 2004, 365, 60–62.

[21] Ashraf, M.; Kim, Le.; Xu, Huang. (2010). "Information Gain and Adaptive Neuro-Fuzzy Inference System for Breast Cancer Diagnoses". Proceedings Computer Sciences Convergence Information Technology 2010 (ICCIT-2010), 30th Nov.-2nd Dec., IEEE, Seoul. pp: 911 – 915. doi:10.1109/ICCIT.2010.5711189.

[22] Senturk, Z.K.; Kara, R. Breast cancer diagnosis via data mining: Performance analysis of seven different algorithms. Comput. Sci. Eng. 2014, 4, 35–46