ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

The Mechanism in Identifying and Predicting Brest Cancer using Machine Learning Algorithms

Raghu T Mylavarapu, Bharadwaja Krishnadev Mylavarapu

Manager, Regional Development Center, KPMG LLP, Montvale, NJ , USA

Business Analytics Lead, Analytics; PLM COE; PDM, John Deere Co, Moline, IL, US

ABSTRACT: Machine learning is a advanced technology playing major role in all the domains such as Medical, Industrial, etc., the main advantage of using Machine learning is helps to come up with conclusions and makes easy to take the decisions. Firstly the ML started in cancer detection as a positive response so it is started using in detecting the cancer diagnosis. Based on the condition the user can analyze the health condition and can diagnose the health of the disease. For designing of this project we are using Artificial Neural Networks (ANN) mechanism for data collection and user alert generation system. The machine learning mechanisms proved more morality and vulnerability manner and the usage of ML algorithms were increased up to 25%. Here at the basic level it is giving the understanding and developing in the most basic level.

KEYWORDS: Artificial Neural Networks, Cancer, Support Vector Machine, Machine Learning, Artificial Intelligence and SVM classifier, Prediction, identification.

I. INTRODUCTION

Day by day the technology is increasing as a result the resistance power of the person is decreeing due to the certain issues due to several reasons like sound pollution, noise pollution, air pollution, sudden changes in the climate etc., everyday new kind of viruses were developing due to keeping stock of unfreeze meat, beef etc.,. By eating such type of items and outside items many new diseases were affecting human body. There are different types of cancer such as Blood cancer, Brest cancer, Lung Cancer etc., Some diseases can see effect immediately where as some diseases takes time. One of the kinds of disease which affects more is “Cancer”. Generally cancer has different stages based on person to person the disease symptoms can see at different levels. If the effect was seen at last level the curing of the disease is highly difficult and vice versa. Based on this cancer disease the Artificial Intelligence (AI) mechanism came into existence. The AI detects the person condition from time to time and based on that the alerts will give to the person regarding diet, fitness etc., [1]

The using of machine learning the development was started almost 20 years ago. By the help of Artificial Neural Networks (ANN) and Decision Trees the mechanism came into existence. As per the recent surveys of PubMeb survey more than 1800 research papers were published based on the concept of Cancer detection with Machine Learning. All these papers were based on Identifying, Detecting, Classifying, defining the cancer disease. The complete agenda in using Ml in this concept is because of 3 reasons namely

- The prediction of suspecting disease.
- The risk estimation.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

- The risk of surveillance

Generally the cancer is based on the different factors like size of the tumor, location of place etc., Comparable difficulties likewise exist for the two doctors and patients the same with regards to the issues of disease counteraction and malignant growth vulnerability forecast. Family ancestry, age, diet, weight, high-hazard propensities smoking, substantial drinking, and introduction to ecological cancer-causing agents all assume a job in anticipating a person's hazard for creating disease Unfortunately these ordinary "large scale" clinical, natural and social parameters, for the most part, don't give enough data to make strong forecasts or guesses. In a perfect world what is required is some unmistakable atomic insights concerning either the tumor or the patient's own hereditary make-up. [2]

With the quick improvement of genomic proteomic and imaging innovations, this sort of atomic-scale data about patients or tumors would now be able to be promptly obtained. Sub-atomic biomarkers, for example, physical transformations in specific qualities, the appearance or articulation of certain tumor proteins or the concoction condition of the tumor have been appeared to fill in as incredible prognostic or prescient markers. All the more as of late, mixes or examples of numerous atomic biomarkers have been seen as significantly more prescient as single-part tests or readouts if these sub-atomic examples are joined with full-scale clinical information, [3] the power and exactness of disease anticipations and expectations improve much more. Be that as it may, as the quantity of parameters we measure develops, so too does the test of attempting to understand this data.

Previously, our reliance on large scale data by and large kept the quantities of factors sufficiently little with the goal that standard measurable techniques or even a doctor's own instinct could be utilized to anticipate malignant growth dangers and results. Be that as it may, with the present high-throughput indicative and imaging innovations we currently end up overpowered with handfuls or even many sub-atomic, cell and clinical parameters. In these circumstances, human instinct and standard insights don't for the most part work. Rather, we should progressively depend on nontraditional, seriously computational methodologies, for example, AI. The utilization of PC sickness expectation and forecast is a piece of a developing pattern towards customized, prescient medication. This development towards prescient medication is significant, for patients as well as for doctors just as wellbeing business analysts and strategy organizers. Given the developing significance of prescient drug and the developing dependence on AI to make forecasts, we trusted it would bear some significance with direct a nitty gritty audit of distributed examinations utilizing AI techniques in disease expectation and visualization. The expectation is to distinguish key patterns regarding the sorts of AI strategies being utilized, the kinds of preparing information is incorporated, [4] the sorts of endpoint forecasts being made, the kinds of malignant growths being considered and the general execution of these techniques in anticipating disease powerlessness understanding results. We additionally found that practically all forecasts are made utilizing only four sorts of info information: genomic information proteomic information, clinical information or mixes of these three. In looking at and assessing the current examinations various general patterns were noted and various regular issues identified. A portion of the more clear patterns remember quickly developing utilization of AI strategies for malignancy expectation and visualization (Figure 1), a developing dependence on protein markers and microarray information, a pattern towards utilizing blended information, a solid predisposition towards applications in prostate and bosom disease, and an unforeseen reliance on more established innovations, for example, fake neural systems. Among the more generally noted problems was an awkwardness of prescient events with parameters, overtraining, and an absence of external validation or testing. In any case, among the better planned and better-approved examinations it was clear that AI techniques, relative to simple factual strategies, could considerably (15-25%) improve the exactness of malignant growth susceptibility and disease result forecast. [5] As it were, AI has a significant task to carry out in malignant growth forecast and guess.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

II. EXISTING WORK AND PROPOSED WORK

Based on the research of cancer detection many of the scholars proposed many algorithms to make the task simple and to give user the ease of understand in usage. Some of the proposed algorithms were Naïve byes, Decision Tree, K-Nearest Neighborhood, Neural Network, Support Vector Machine (SVM) and Genetic Algorithm. [6] These are the few algorithms used to identify the cancer risk and the prediction analysis. In the existing works many of the mechanisms came with taking the input of Diagnostic Brest Cancer (DBC) dataset. [7] But the results were generated was not significant.

We have studied the same topic and proposed in another manner by the help of GRU-SVM classifier. The multiple algorithms were there in ML which was mentioned in WDBC. For doing the task we have taken Machine Intelligence Library and we have taken Wisconsin Diagnostic Breast Cancer (WDBC) dataset. [8] Totally we have taken 1200 images in the dataset in that each and every image follows with any of the below properties such as [9]

- Radius
- Texture
- Perimeter
- Area
- Smoothness
- Compactness
- Connectivity
- Fractal Dimension

For designing of the project different case studies were studied namely:

Case Study 1: Cancer Risk Estimation. [10]

Case Study 2: Estimating of cancer Survivability [11]

Case Study 3: Prediction of Cancer Recurrence [12]

For the cancer detection the raw data is collected and the raw data goes through the following steps for making the data filteration.

2.1 Data Pre-Processing [13]

To avoid conflicts the dataset was utilized with term

$$Z=X-\mu/\alpha$$

2.2 Algorithm

For the designing of the project we have designed GRU-SVM classifier. We have proposed Neural Network with GRU-SVM classifier [14] for the dataset and project implementation processing.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

$$z = \sigma (Wz \cdot [ht-1, xt])$$

Here the terms

Z= Update Gate and

R=Rest Gate of GRU-SVM Classifier.

The overall computation of the algorithm was limited by the help of Adam Algorithm. The same concept was used in the Regression process. There are few other types of regression such as Linear regression, soft sonic regression etc,. [15] Here in our project we have made the regression mechanism where we have done the classification mechanism of the datasets with the available image data. The complete process takes by three stages namely:

Mutation operator

Mutation can be defined as the adding or removing the variables.[16]

Crossover operator

Cross over is defined as a 1-point selection of a pointer of a task [17]

Selection operator

Selection makes the values of Top 20 of the values and makes the values estimation from the overall data values[18]

III. RESULTS

Here for the development of the project we have taken the dataset of Brest cancer people with images for mapping one image to other image. As per the existing work they have taken the single images and made comparison with each and every image. This process is a time consuming process. So here we have taken multiple images of data items with complete images first we take images, select images then point out the images with detection of problem in it. In the feature selection process the number of methods made is equal to the number of features were tested. We can estimate out of cent percent 62% will be the unique ones. It is proved that the exactness, accuracy will be formed in the project by implementing the proposed mechanism. Here accuracy, exactness etc., were formed because of acquiring the multiple image reading features.

$$ACC = TP + TN / FP + FN + TP + TN$$

Where as TP, TN and FN are number of true values

and

FN is Number of False Values

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

The final result correlates with the help of true positives and false positives. Here the higher and maximum mean rate factor is generated with value of 81.

For the development of the project different algorithms were used such as CNN, Naïve Bays and SVM Classifier. All these algorithms were proposed with accuracy of project with 97.34, 96.38 and 98.37 respectively.

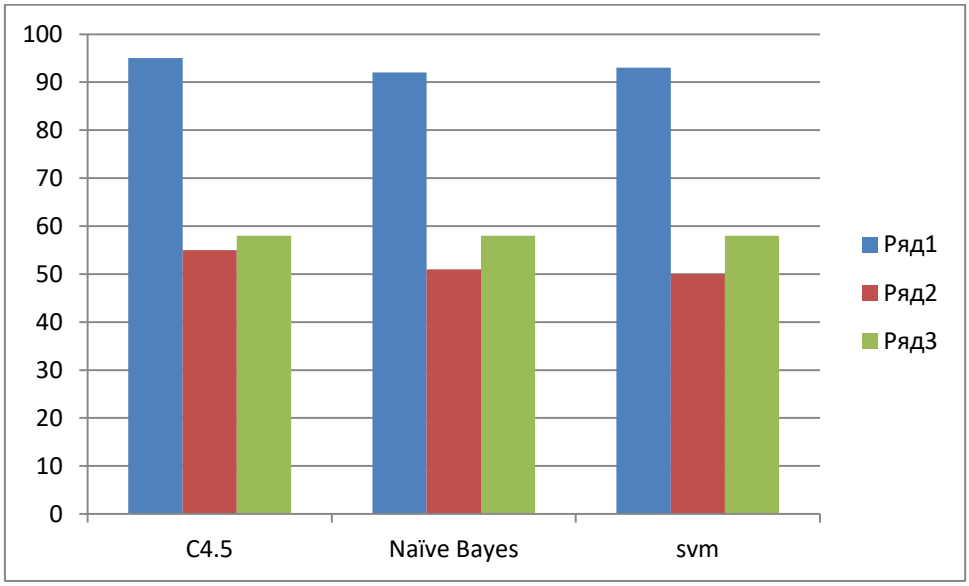

Figure 1: Risk Analysis of Brest Cancer

In the graph consider horizontal axis as the performance and vertical axis as the privacy . The graph and results concluded that compared to all graphs SVM classifier gives more efficient results compared to other relevant algorithms.

IV. CONCLUSION AND FUTURE WORK

In the proposed work Extreme Brest Cancer Detection Mechanism was developed and by identifying the importance of the work and further enhanced up to multiple image classification. The Brest cancer is the second largest disease and women are ending their life overall the world. In our paper we have proposed the dataset which having the real world data. Here we have proposed the SVM classifier mechanism framework. In the future of this work is to extend the mechanisms in understanding the all the possible solutions of the problem and all the solutions were taken into the consideration in having the diagnosis of cancer patients.

REFERENCES

[1] K. Usha Rani . (2016).Perfomance Of Synthetic Minority Over Sampling Technique On Im Balanced Breast Cancer Data. IEEE, P1-5.
[2]. Jaree Thongkam , Guandong Xu, Yanchun Zhang And Fuchun Huang. (2007). Toward Breast Cancer Survivability Prediction Models Through Improving Training Space. Expert Systems With Applications P1-10.
[3]Sourajit Behera And Rinkle Rani (2016). Comparative Analysis Of Density Based Outlier Detection Techniques On Breast Cancer Data Using Hadoop And Map Reduce. IEEE, P1-4.
[4] Uma Ojha And Dr. Savita Goel. (2017). A Study On Prediction Of Breast Censer Recurence Using Data Mining Technique . IEEE, P1-4.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 2, February 2020

[5] Swathi Rathi (2019), Estimating and Diverting the Route by Using Traffic Data over Big Data vol8,issue 9, IJIRCCE ISSN(Online): 2319-8753 ISSN (Print): 2347-6710
[6] Swathi Rathi,(2019) Attribute Evaluation of Dataset Using Chi-Square Test in Rapid Miner Studiovolume 7, issue 5,IJIRCCE SSN(Online): 2320-9801 ISSN (Print): 2320-9798
[7] Muhammad Umer Khan, Jong Pill Choi, Hyunjung Shin And Minkoo Kim . (2008). Predicting Breast Cancer Survivability Using Fuzzy Decision Trees For Personalized Healthcare. IEEE, P20-24.
[8] Dr. Medhat Mohamed, Muhamed Wael Farouq,Hala Abou And Abdel Badeeh.(). Using Data Mining For Assessing Diagnosis Of Breast Cancer. Computer Science And Information Technology,P1-7.
[9] Ahmed Iqbal Pritom, Shahed Anzarus Sabab And Md. Ahadur Rahman Munshi. (2016). Predicting Breast Cancer Recurrence Using Effective Classification And Feature Selection Technique. IEEE, P1-5.
[10] Hyunjung Shin And Yonghyun Nam. (2014). A Coupling Approach Of A Predictor And A Descriptor For Breast Cancer Prognosis. Bmc, P1-12.
[11] Kumar. Attangudi P. Perichappan, S. Sasubilli and A. Z. Khurshudyan, "Approximate analytical solution to non-linear Young-Laplace equation with an infinite boundary condition," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-5. doi: 10.1109/ICOMET.2018.8346349
[12] S. Chandrasekaran, "Contemplated Method for Predicting Disease by Deep Learning Approach Over Big Data," 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018, pp. 1-5.
doi: 10.1109/RICE.2018.8509090
[13] Gopinadh Sasubilli,Uday Shankar Sekhar, Ms.Surbhi Sharma, Ms.Swati Sharma, "A Contemplating approach for Hive and Map reduce for efficient Big Data Implementation" 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 131–135 DOI: 10.15439/2018KM20
[14]. Dursun Delen, Glenn Walker, Amit Kadam(2005). Predicting Breast Cancer Survivability:
[15] Sreenivas Sasubilli,Kumar Attangudi Perichiappan Perichappan,P. Srinivas Kumar,Abhishek Kumar,"An Approach towards economical hierarchic Search over Encrypted Cloud", 2018 Proceedings of the First International Conference on Information Technology and Knowledge Management pp. 125–129, DOI: 10.15439/2018KM38
[16] Abien Fred M. Agarap.(2018). On Breast Cancer Detection: An Application Of Machine Learning Algorithms On The Wisconsin Diagnostic Dataset.ACM, P1-5.
[17] Sara W. Dyrstad, Yan Yan , Amy M. Fowler And Graham A. Colditz. (2015). Breast Cancer Risk Associated With Benign Breast Disease: Systematic Review And Meta-Analysis. Springer, P1-7.
[18] Dana Bazazeh And Raed Shubair.(2016). Comparative Study Of Machine Learning Algorithms For Breast Cancer Detection And Diagnosis. IEEE, P1-5.