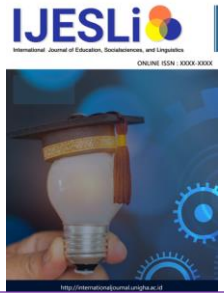
JOURNAL ON EDUCATION, SOCIAL SCIENCES AND LINGUISTICS

IJESLi



http://internationaljournal.unigha.ac.id/ - ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)

Breast Cancer Prediction In Virtue Of Big Data Analytics

[1]G S Pradeep Ghantasala, [2]AnuRadha Reddy, [3]Subbarao Peyyala, [4]D. Nageswara Rao
[1,4]Chitkara University Institute of Engineering & Technology, Chitkara University, Punjab, India
[2,3]Malla Reddy Institute of Technology and Science, Hyderabad, Telangana, India

[1]ggs.pradeep@chitkara.edu.in, [2]anuradhareddy.anu@gmail.com, [3]psubbarao1406@gmail.com, [4]nageswara.rao@chitkara.edu.in

***Corresponding Author :** *ggs.pradeep@chitkara.edu.in*

Doi :

Keywords :

Healthcare, Big Data, Breast Cancer, KNN, Dataset, Machine Learning

Abstract

The term big data is used to collect information, which is enormous and still emergent and increases exponentially concerning time. Big data covers both arranged and rearranged data. In the present scenario, big data is playing an essential vital role in the healthcare industry for the forecasting of the diseases. In the healthcare industry, Breast cancer is one of the crest cancers occurring in women of various age groups. Breast cancer is another reason for most of the deaths of women across the nation. The only solution to this is the detection of the disease in the early stages, which gives us chances for finding a better solution and cures of the illness. This problem can be implemented by taking the dataset using the K nearest neighbor (KNN) algorithm to find the classification of the accuracy (i.e., Percentage) based on the machine learning for dealing with the problem analyze the medical issues for further treatments.

Volume 1, No.1, February 2021, Pages : 130-136

Volume 1, No.1, February 2021, Page: 3119-3125
ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)
http://internationaljournal.unigha.ac.id/

Now a day's often, a therapeutic action doesn't have that much to do with a person exclusively. It is the same that a private clinic would supply to effectively anyone based on the past condition. Since because medication as we knew that it revolve round "values of care", the finest courses of avoidance of action for the all-purpose human beings around the nation. By means of breast cancer, these principles mean self-examination and mammograms later a certain period and standard chemotherapy to extravagance a cancer if it is originated. While if the primary action doesn't work, hospital and patients shift on to one another and the subsequently best hospital for the treatment. It is a trial and error method, through life on the procession. Hence a more specific approach towards the patient's healing is necessary [1].

Big data analytics provides the finest resolution for managing, storing, and analyzing a huge quantity of mammographic images. The life cycle has a most important forecast construction which improves the medical harsh conditions approach by exciting forecasting construction, estimating statistical improvement and shaping a variety of virus pattern. It also provides enhanced solution for managing healthcare records [2].

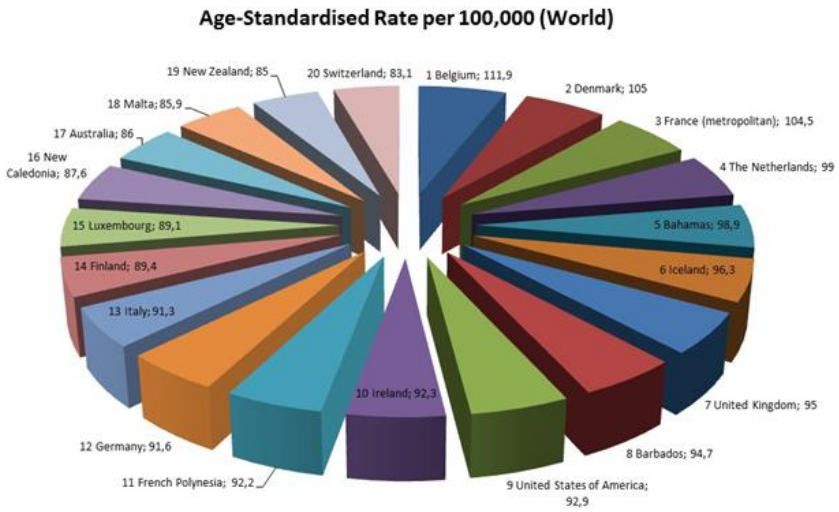


Figure 1: Prevention of Breast Cancer statistics as per WCRFI

Breast Cancer: Nowadays Breast cancer is one of the majorly widespread type of malignancy diagnose in women across the globe. It's one of the foremost causes of death in women's universal. World Health Organization, International agency for cancer investigation and American cancer society statement that 15 million new-fangled disease cases were record in 2013 internationally across the world [3]. Usually, breast tumor cells are alienated based on shape of the figure like micro classification, ample and designed distortion. In majority of the cases the tissue might be either ample or micro classification [4]. Premature recognition and healing can

Volume 1, No.1, February 2021, Page: 3119-3125
ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)
http://internationaljournal.unigha.ac.id/

decrease the breast cancer death rate quiet significantly. Regrettably, the symptoms of the breast cancer are much understated and differ in emergence at early stages [5].

Signs & Symptoms:

Various types of symptoms for breast cancer are:

- Lumps present in the breast with any size, uniformity, lineament, and with soft or hard edges.
- Ache in the breast.
- The whole or a few portion of the breast being suffered from inflammation and redness.
- Present may be some change take place in the nipple i.e., nipple renunciation, ulceration, impatient.
- Hemorrhage from the nipple that may be in some color.

2. K-Nearest Neighbor (KNN)

KNN algorithm is also known as Instance-Based Learning. It is the simplest method for categorization of samples. At this point various expanse procedures are used for categorization of samples. K-nearest neighbor find the numerals of samples from training facts which is close to the investigation samples and assigns to the most common class label. In this algorithm, instruction samples produce the classification rules without taking into account of additional information. It has elevated probability when associated instance belongs to the same class. Based on K training sample KNN algorithms identify the test samples. For all the instances, K value is always a positive integer. In our dataset (WDBC) all the instances are there in amid one (1) and ten (10), thus no need to compute normalization connecting those in-stances. Because not any characteristic will persuade the others in the detachment computation of KNN. It is extensively used in a variety of resolution support scheme for biomedical applications [6-9].

Volume 1, No.1, February 2021, Page: 3119-3125
ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)
http://internationaljournal.unigha.ac.id/

3. Proposed Work

For this experimentation, we use KNN algorithm for the prediction of breast cancer risk. We had used breast cancer Wisconsin (original) dataset that is measured from the UCI machine learning algorithm [10,16]. The dataset include 699 instances and 10 attributes next to with the category tag and it contain missing values (?) which are replaced by the mean value of the attribute. The allocation of class will be 459 (65.6%) instances fit in to the benign class and other 245(35.5%) instances fit in to the spiteful class [11].

ATTRIBUTE	DOMAIN
Sample code number	id number
Clump Thickness	1 - 10
Uniformity of Cell Size	1 - 10
Uniformity of Cell Shape	1 - 10
Marginal Adhesion	1 - 10
Single Epithelial Cell Size	1 - 10
Bare Nuclei	1 - 10
Bland Chromatin	1 - 10
Normal Nucleoli	1 - 10
Mitoses	1 - 10
Class:	2 for benign 4 for malignant

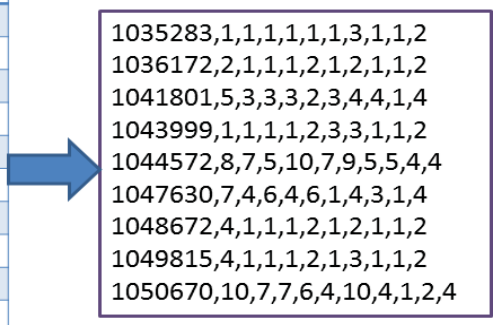


Figure 2: Representation of Breast cancer using Wisconsin Dataset

Volume 1, No.1, February 2021, Page: 3119-3125
ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)
http://internationaljournal.unigha.ac.id/

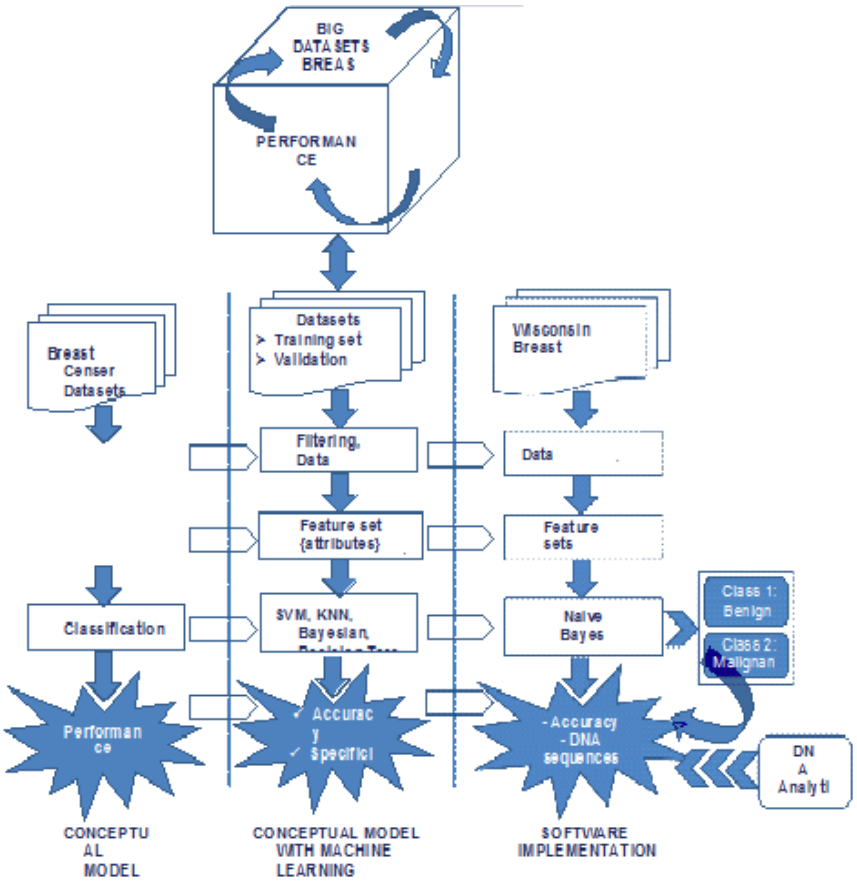


Figure 3: Theoretical representation for early recognition of Breast Cancer based on Machine Learning

4. Evaluation methods

For this experimentation, we had utilized K nearest neighbor (KNN) algorithm and it is implemented by using R tool. R is extensively used for the implementation and it is free open source software. For computing statistics, graphics it is the most commonly used software atmosphere. The initial version of R tool was implemented by the Ross Ihaka and Robert Gentleman in the 1990's, for our test we used R3.4.2 version [12]. It is incorporated software that includes several facilities they are for storing and handling the data is a very efficient tool. It also performs the array and matrix calculations. For the analysis of data it contains an enormous group of midway tools [13-15, 17, 18]. In R tool there is a wrap up available called as 'neighbr' for

Volume 1, No.1, February 2021, Page: 3119-3125
ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)
http://internationaljournal.unigha.ac.id/

KNN algorithm. By applying KNN algorithm in R tool it provides enhanced accuracy of 97.65% when compared to additional methods.

5. Conclusion

This article deal with KNN algorithm to categorize cancer tumor as any benign or malignant. We had applied characteristic assortment on the dataset to eliminate replica and inappropriate features. We applied proportioned uncertainty characteristic assessment in WEKA for facet selection. Our proposed approach is evaluated and compared using Wisconsin breast cancer dataset. The investigational result shows that accurateness, exactitude, recollect, and F-measure are enlarged by our proposed method when compared with various models. In future, we will work on feature selection technique to get better the accurateness of the model. The most vital future mission is to be crated the benchmark breast cancer datasets which will maintain the presentation evaluation and comparison of the similar algorithms consequences.

6. References

[1] Song, M., Lee, K. M., & Kang, D. (2011). Breast cancer prevention based on gene–environment interaction. Molecular carcinogenesis, 50(4), 280-290.

[2] T. Stricker, D.V. Catenacci, S.Y. Siewert, (2011). Molecular profiling of cancer - the future of personalized cancer medicine: a primer on cancer biology and the tools necessary to bring molecular testing to the clinic. Semin. Oncol, 38:173–185.

[3] Mendes, E. (2015). Personalized Medicine: Redefining Cancer and Its Treatment.

[4] Vogenberg, F. R., Barash, C. I., & Pursel, M. (2010). Personalized medicine: part 1: evolution and development into theranostics. Pharmacy and Therapeutics, 35(10), 560.

[5] Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

[6] Ali, S. I., & Shahzad, W. (2012, October). A feature subset selection method based on symmetric uncertainty and ant colony optimization. In 2012 International Conference on Emerging Technologies (pp. 1-6). IEEE.

[7] Potharaju, S. P., & Sreedevi, M. (2017). A Novel M-Cluster of Feature Selection Approach Based on Symmetrical Uncertainty for Increasing Classification Accuracy of Medical Datasets. Journal of Engineering Science & Technology Review, 10(6).

[8] Peart, O. (2017). Metastatic breast cancer. Radiologic technology, 88(5), 519M-539M.

[9] Jayasinghe, U. W., Pathmanathan, N., Elder, E., & Boyages, J. (2015). Prognostic value of the lymph node ratio for lymph-node-positive breast cancer-is it just a denominator problem?. Springerplus, 4(1), 1-10.

[10] World Cancer Research Fund International: http://www.wcrf.org

[11] Park, Y. M. M., O'Brien, K. M., Zhao, S., Weinberg, C. R., Baird, D. D., & Sandler, D. P.

Volume 1, No.1, February 2021, Page: 3119-3125
ISSN 2775-4928 (Print) ISSN 2775-8893 (Online)
http://internationaljournal.unigha.ac.id/

(2017). Gestational diabetes mellitus may be associated with increased risk of breast cancer. British journal of cancer, 116(7), 960-963.

[12] Godet, I., & Gilkes, D. M. (2017). BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. Integrative cancer science and therapeutics, 4(1).

[13] Park, P. W., & Lee, S. W. (2017). Classification of Heart Disease Using K-Nearest Neighbor Imputation. In Proceedings of the Korea Information Processing Society Conference (pp. 742-745). Korea Information Processing Society.

[14] Eltalhi, S., & Kutrani, H. (2019). Breast cancer diagnosis and prediction using machine learning and data mining techniques: A review. IOSR J. Dental Med. Sci., 18(4), 85-94.

[15] D.R Umesh et al., (2016). "Big Data Analytics to Predict Breast CancerRecurrence on SEER Dataset using MapReduce Approach", Inter-national Journal of Computer Applications, 150(7), 7-11.

[16] Ghantasala, G. P., Kallam, S., Kumari, N. V., & Patan, R. (2020, March). Texture Recognization and Image Smoothing for Microcalcification and Mass Detection in Abnormal Region. In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA) (pp. 1-6). IEEE.

[17] Bhowmik, C., Ghantasala, G. P., & AnuRadha, R. (2021). A Comparison of Various Data Mining Algorithms to Distinguish Mammogram Calcification Using Computer-Aided Testing Tools. In Proceedings of the Second International Conference on Information Management and Machine Intelligence (pp. 537-546). Springer, Singapore.

[18] Patan, R., Ghantasala, G. P., Sekaran, R., Gupta, D., & Ramachandran, M. (2020). Smart healthcare and quality of service in IoT using grey filter convolutional based cyber physical system. Sustainable Cities and Society, 59, 102141.