

# Use of Artificial Intelligence (AI): A Developing Assessment Techniques for Study of Tumor Diversity from Gene Expression

Rajesh Kumar Maurya<sup>1</sup>, Sanjay Kumar Yadav<sup>2</sup>, Pragya Tiwari<sup>3</sup>

<sup>1</sup> Research Scholar, <sup>2</sup> Associate Professor,

<sup>1,2</sup> Department of Computer Science and Information Technology, Sam Higginbottom University of Agriculture, Technology and Sciences (SHUATS) Allahabad UP – (INDIA)

<sup>3</sup> Assistant Professor, Department of Computer Applications, ABES Engineering College, Ghaziabad UP – (INDIA)

## ABSTRACT

Cancer is one of the commonly occurring diseases in the human that causes the greatest number of deaths globally per year. Usually, the classification of tumor diversity has been based on the consensus of study of the signs of the disease using the microscopic examination, with very limited consideration of molecular pathology. Tissue helps us to decide the prognostic and predictive aspects of breast cancer. Tumor and clinical multiformity are one of the key causes of the letdown of proper medical treatment. The Classification of tumor multiformity using evolving tools definitely improves prognosis and later helps to propose proper treatment plans. AI is one of the evolving methods for classifying the heterogeneous data available, including diverse risk factors. Moreover, the application of AI seems one of the effective tools for the analysis of raw feature data for all the genes cancer. In this article, we review the distinct AI techniques for grouping of data and extraction of key features

## Keywords

Cancer, Bioinformatics, ML application, UCI machine learning repository, Diseases Prognosis

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

## Introduction

Cancer is basically an abandoned increase of cells in any part of the human body, which develop as a result of cytological changes. The pathogeny and proteomics techniques offer an efficient way to categorize Malignant Tumors (MT) and Benign Tumors (BT). Basically a tumor is any irregular propagation of cells, it could be either BT or MT. A BT, like a usual protuberance, remains constricted to its creative place, neither attacking neighboring usual tissue nor advancing to distant body locations.. Solely MT are called cancers, moreover it is their capability to invade and refine that drive the cancer so dangerous. Although BT can usually be evacuated surgically, the spread of MT to remote body sites regularly makes them contrary to such localized medication. [1]. In contrast, cells constituting a MT show more or less proteins peculiarity of the cell nature from which it originates, and a high fragment of the cells raise and divide much more promptly than normal. A few MT stand localized and enclose, at least for a time; for example carcinoma in situ in the breast or ovary [2]. Cell growth and initiation of minor regions of extension is named metastasis; but malignant cells by and by acquiring the capacity to metastasize. Hence the major property that distinguishes the MT from benign ones are their spread and presumptuous [3]. Because of the unique blend of metabolic adaptation and genetic variability a cancer cell results in break out resistant surveillance. So, cancer cells are all the time developing to adapt to new surroundings [4]. Researchers presume that a quintessential lifestyle induced stress and that is the vital cause of developing breast cancer. In inclusion physical opinions like abnormal discharge, temperature and nipple sizes also often use working for cancer prediction. It has been suggested that the accumulation of genetic modifications and epigenetic in

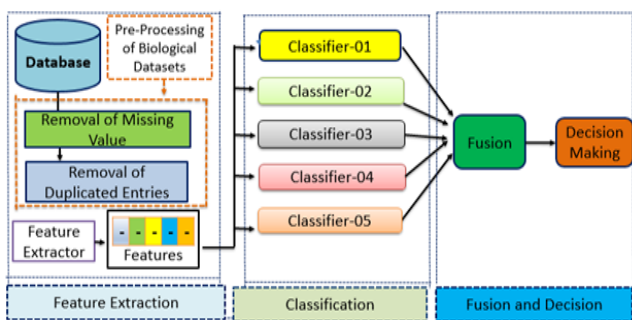
tissue inhabitant normal breast stem cells can result in the existence of irregular cells inside a milk duct in the breast. If we assemble simultaneously, offers a complicated dataset. Hence, innovative tools are necessary in deeper understanding the complex dataset and their relationship for identification of tumor diversity. [5][6][7].

## Artificial Intelligence for cancer Prediction

AI is very helpful to understand how cancer cells get resistant to anticancer drugs, which can quickly help to improve drug use and drug development. Artificial intelligence offers a number of arithmetical, optimization, and probabilistic methods that enable workstations to get “learn” from previous instances and to identify patterns from noisy, erroneous biological datasets. This kind of capability is compatible to biological applications, in particular mainly for those that rely on the complex genomic and proteomic measurements. As an outcome, AI is very often used in cancer detection and diagnosis .AI makes usage of mining procedures to identify patterns in enormous datasets [10] [11][12]. For analyzing gene expression data using AI and ML based feature extraction techniques are like Artificial Neural Networks (ANN), Decision Trees (DT), Naive Bayes Classifier (NBC), Random Forest (RF), Support Vector Machines (SVM), Logistic Regression (LR), and Nearest Neighbor (KNN) will undeniably give accurate outcomes and that will be supportive to cut the cancer deaths. [13][14]. The right study of the cancer using ML method will definitely support the clinician in redeemable the life of a patient. Thus it provides a great hope to the cancer patients. AI and its application are very obliging to cancer specialists to trail of the drug beside cancer. Moreover it can also enable for personalized cell therapy and drug design. AI and its applications are now

effectively employ to build suitable models based on transcript analysis [15][16][17]. A proposed model discussed in Figure -01. Steps of proposed model to classify gene expression data are

- i. The Gathering of gene expression records.
- ii. Pre-processing of gene expression records.
- iii. The Selection of appropriate classifiers for removing the redundant information.
- iv. Classification of tumor data using ensemble classifiers.
- v. Combine the performances of different classifier using fusion strategy.
- vi. Combined Performance of different classifier are evaluated.



**Figure-01:** An Example of ML Classifiers to classify features from gene expression using fusion strategy.

### Logistic Regression (LR)

This technique often used in ML to address the classification. LR method is a traditional statistical and supervised ML tool for prediction. This model uses a logistic function to model values 0 or 1 at each observation. LR model is like to other model and used to display the relationship between known and unknown variables [18].

**Table 1:** Linked Research Articles using LR

S.N	Datasets	Authors	Year	Accuracy	References
1	Bladder cancer, Breast cancer, Cervical cancer	More et al.,	2020	83.33%	[19]
2	Breast Cancer and Prostate Cancer	Agrawal et al.,	2020	90%	[20]
3	Breast Cancer	Islam and Poly	2019	80%	[21]
4	Breast Cancer	Parhusip et al.,	2020	91%	[22]
5	Breast Cancer	Mohammed et al.,	2019	74.7%	[23]

### Support Vector Machines (SVM)

SVM is a type of supervised ML approach which achieves a better efficiency and accuracy classifier for classification. Normally SVM method is to discover an ideal separating hyperplane in a t-dimensional space (t—the number of features) that clearly categories the experimental datasets. The dimensional space and most distance between close data points of boundary [24].

**Table 2:** Linked Research Articles using SVM

S.N	Dataset	Authors	Year	Accuracy	References
1	Breast Cancer	Panwar et al.,	2020	99.12%	[25]
2	Breast cancer	Ude, A. A	2019	97%	[26]
3	Breast cancer	Obaid et al.,	2019	97.9%	[27]
4	Breast cancer	Liu et al.,	2018	98.04%	[28]
5	Breast Cancer	Abdarjet al.,	2019	89.70%	[39]

### Artificial Neural Networks (ANN)

ANN are system motivated ML classifiers inspired by structure and the function of biological Nervous systems. This model is a set of connected neural networks. The main components of the ANN that are used for the transformation are Input, output and intermediate layer. These neurons are then linked via edges and every single edges is linked with their vertex called weight [29][30].

**Table 3:** Linked Research Articles using ANN

S.N	Dataset	Authors	Year	Accuracy	References
1	Breast Cancer	Kaya, Y.	2015	98.54%	[31]
2	Breast cancer	Abdel-Zaher et al.,	2016	99.68%	[32]
3	Breast Cancer	Spanhol et al.,	2017	77.80%	[33]
4	Breast Cancer	Alickovic et al.,	2019	99.27%	[34]
5	Breast Cancer	Bhardwaj & Tiwari	2015	98.24%	[40]

### k-Nearest Neighbors (k-NN)

Particularly k-NN is a non-parametric ML method. This approach is commonly used for regression and classification tasks based on discovering the k- nearest neighbors. In the case of cancer dataset, majority symptoms of tumor among the boundary of this k. Euclidean distance techniques is a capable technique between two vectors and Ki and Li. [35].

$$\text{Euclidean Distance Function (E DF)} = \sqrt{\sum_{i=1}^k (Li - Ki)^2} \tag{4}$$

**Table 4:** Linked Research Articles using KNN

S. N	Dataset	Authors	Year	Accuracy	References
1	Breast cancer	Liantoni et al.,	2020	90.00%	[38]
2	Breast Cancer	El-Baz, A. H.	2015	97.62%	[44]
3	Breast Cancer	MajdahAlshammari	2020	98.00%	[27]
4	Breast Cancer	Meriem et al.,	2018	97.51%	[36]
5	Breast Cancer	Odajima et al.,	2014	97.60%	[43]

### Naive Bayes Classifier (NBC)

A NBC is said to be a probabilistic ML model which is used for classification purposes. The core of the NBC classifier is the basis of Bayes theorem. The root of this method is based on Bayes Theorem (BT). By BT one can always get the probability of event S1 that is happening, when S2 has already occurred. Now S1 is the hypothesis and S2 is the evidence shown in equation-01.

$$P(S1/S2) = \frac{P\left(\frac{S2}{S1}\right) * P(S2)}{P(S1)} \tag{1}$$

The assumption built here is that the features are mutually independent. Means to say that the presence of one feature does not influence the other feature. Therefore it is termed as naive. A probabilistic perspective of classification is basically the relationship between the input features and class demonstrated as probabilities [37].

**Table 5:** Linked Research Articles using NBC

S.N	Dataset	Authors	Year	Accuracy	References
1	Breast cancer	Salama et al.,	2012	95.99%	[41]
2	Breast cancer	Ude, A. A	2019	95%	[26]
3	Breast Cancer	Meriem et.al.,	2018	96.19%	[35]
4	Breast Cancer	Asri et al.,	2016	0.97%	[42]
5	Breast Cancer	Liantoni et al.,	2020	90%	[45]

## Conclusion

This review article talks about the use of AI and its application to accurately analyze the various types of cancer and their sub types. Better classification of tumor variety using AI based application and its utility in the prediction, prognosis can decrease the casualty rate per year. The extended study of research emphasizes on growth of right risk assessment models using AI based approaches. While several efficient AI techniques have achieved very good accuracy in biological dataset, there is still the requirement of better-quality ways for saving the life of a cancer patient. Application and usability of different AI methodologies for the study of complex data integration, feature selection and extraction is an excellent way for better human understanding about cancer and its predictions and prognosis.

## References

- [1] Cooper, G. M., & Hausman, R. E. (2000). The development and causes of cancer. The cell: A molecular approach, 2.
- [2] Afify, S. (2007). Drug targeting delivery systems for treatment of Raf-1 induced lung tumors in mice.
- [3] Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. (2000). Tumor cells and the onset of cancer. Molecular cell biology.
- [4] Gonzalez, H., Hagerling, C., & Werb, Z. (2018). Roles of the immune system in cancer: from tumor initiation to metastatic progression. Genes & development, 32(19-20), 1267-1284.
- [5] Key, T. J., Verkasalo, P. K., & Banks, E. (2001). Epidemiology of breast cancer. The lancet oncology, 2(3), 133-140.
- [6] Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. Proceedings of the National Academy of Sciences, 100(7), 3983-3988.
- [7] Interagency Breast Cancer and Environmental Research Coordinating Committee (IBCERCC). (2013). Breast cancer and the environment: prioritizing prevention. National Institute of Environmental Health Sciences.
- [8] Yue, W., Wang, Z., Chen, H., Payne, A., & Liu, X. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. Designs, 2(2), 13.
- [9] Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013). Using three machine learning techniques for predicting breast cancer recurrence. J Health Med Inform, 4(124), 3.
- [10] Sbissi, S., & Gattoufi, S. (2017, August). Contextual Decision Making for Cancer Diagnosis. In International Conference on Information Technology in Bio-and Medical Informatics (pp. 55-65). Springer, Cham.
- [11] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463-477.
- [12] Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon) (pp. 35-39). IEEE.
- [13] Alpaydin, E. (2016). Machine learning: the new AI. MIT press.
- [14] Manogaran, G., & Lopez, D. (2017). A survey of big data architectures and machine learning algorithms in healthcare. International Journal of Biomedical Engineering and Technology, 25(2-4), 182-211.
- [15] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

- [16] Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2, 117693510600200030.
- [17] Subbaiah, S., & Muruganandam, M. S. Applications of Machine Learning in Cancer Prediction and Prognosis.
- [18] Zhang, Z. (2016). Model building strategy for logistic regression: purposeful selection. *Annals of translational medicine*, 4(6).
- [19] More, S. S., Lobo, V. B., Laban, R. M., Panchal, S., Patil, M., & Pathak, G. (2020, June). Cancer Prediction and Insurance Eligibility using Machine Learning Techniques. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 725-730). IEEE.
- [20] Agrawal, P., Deb, S., Shilpa, A. V., Raju, S. N., & Sreenivasa, B. C. Breast Cancer and Prostate Cancer Detection using Classification Algorithms.
- [21] Islam, M. M., & Poly, T. N. (2019). Machine Learning Models of Breast Cancer Risk Prediction. *BioRxiv*, 723304.
- [22] Kharya, S., & Soni, S. (2016). Weighted naive bayes classifier: A predictive model for breast cancer detection. *International Journal of Computer Applications*, 133(9), 32-37.
- [23] Mohammed, Y. A., & Saleh, E. G. (2020). Investigating the Applicability of Logistic Regression and Artificial Neural Networks in Predicting Breast Cancer. *Journal of Al-Qadisiyah for computer science and mathematics*, 12(2), Page-63.
- [24] Martinez-Arroyo, M., & Sucar, L. E. (2006, August). Learning an optimal naive bayes classifier. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 3, pp. 1236-1239). IEEE.
- [25] Panwar, N., Sharma, D., & Narang, N. (2020). Breast Cancer Classification with Machine Learning Classifier Techniques. Available at SSRN 3577709.
- [26] Ude, A. A. (2019). Classification of Breast Cancer Using Logistic Regression (Doctoral dissertation).
- [27] Obaid, O. I., Mohammed, M. A., Ghani, M. K. A., Mostafa, A., & Taha, F. (2018). Evaluating the Performance of Machine Learning Techniques in the Classification of Wisconsin Breast Cancer. *International Journal of Engineering & Technology*, 7(4.36), 160-166.
- [28] Liu, Na, Jiang Shen, Man Xu, Dan Gan, Er-Shi Qi, and Bo Gao. "Improved cost-sensitive support vector machine classifier for breast cancer diagnosis." *Mathematical Problems in Engineering* 2018 (2018). Bataineh, A. A. (2019). A comparative analysis of nonlinear
- [29] Sepandi, M., Taghdir, M., Rezaianzadeh, A., & Rahimikazerooni, S. (2018). Assessing breast cancer risk with an artificial neural network. *Asian Pacific journal of cancer prevention: APJCP*, 19(4), 1017.
- [30] El\_Jerjawi, N. S., & Abu-Naser, S. S. (2018). Diabetes prediction using artificial neural network.
- [31] Kaya, Y. (2015). A Hybrid Model for Breast Cancer Diagnosis Based on Expectation-Maximization and Artificial Neural Network: EM+ ANN. *Karaelmas Science & Engineering Journal*, 5(1).
- [32] Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139-144.
- [33] Araújo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., ... & Campilho, A. (2017). Classification of breast cancer histology images using convolutional neural networks. *PloS one*, 12(6), e0177544.
- [34] Alickovic, E., & Subasi, A. (2019, May). Normalized Neural Networks for Breast Cancer Classification. In *International Conference on Medical and Biological Engineering* (pp. 519-524). Springer, Cham.

- [35] Harrison, O. (2018). Machine learning basics with the k-nearest neighbors algorithm. *Towards Data Science*. September, 10.
- [36] Amrane, M., Oukid, S., Gagaoua, I., & Ensari, T. (2018, April). Breast cancer classification using machine learning. In 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT) (pp. 1-4). IEEE.
- [37] Xie, X., Ho, J. W., Murphy, C., Kaiser, G., Xu, B., & Chen, T. Y. (2011). Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software*, 84(4), 544-558.
- [38] Liantoni, F., Sukmagautama, C., & Myrtha, R. (2020). Increased Mammogram Image Contrast Using Histogram Equalization And Gaussian In The Classification Of Breast Cancer. *JITCE (Journal of Information Technology and Computer Engineering)*, 4(01), 40-44.
- [39] Abdar, M., & Makarenkov, V. (2019). CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer. *Measurement*, 146, 557-570.
- [40] Bhardwaj, A., & Tiwari, A. (2015). Breast cancer diagnosis using genetically optimized neural network model. *Expert Systems with Applications*, 42(10), 4611-4620.
- [41] Salama, G. I., Abdelhalim, M. B., & Zeid, M. A. E. (2012, November). Experimental comparison of classifiers for breast cancer diagnosis. In 2012 Seventh International Conference on Computer Engineering & Systems (ICCES) (pp. 180-185). IEEE.
- [42] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, 83, 1064-1069.
- [43] Odajima, K., & Pawlovsky, A. P. (2014, October). A detailed description of the use of the kNN method for breast cancer diagnosis. In 2014 7th International Conference on Biomedical Engineering and Informatics (pp. 688-692). IEEE.
- [44] El-Baz, A. H. (2015). Hybrid intelligent system-based rough set and ensemble classifier for breast cancer diagnosis. *Neural Computing and Applications*, 26(2), 437-446.
- [45] Liantoni, F., Sukmagautama, C., & Myrtha, R. (2020). Increased Mammogram Image Contrast Using Histogram Equalization And Gaussian In The Classification Of Breast Cancer. *JITCE (Journal of Information Technology and Computer Engineering)*, 4(01), 40-44.