Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777426

DETECTION AND PREVENTION OF CANCER IN EARLY STAGES USING LINEAR REGRESSION ALGORITHM

Dr. Sachin Kumar[1], Mr. Hirdesh Sharma[2], Mr. Vijay Kumar Tiwari[3], Ms. Namita Sharma[4] and Ms. Roshan Kumari[5]

[1] Associate Professor in Dept. of Computer Applications, Mangalmay Institute of Management & Technology, Greater Noida, U.P., India
ORCID ID: 0000-0002-1136-8009, Email: sachinks.78@gmail.com

[2] Assistant Professor in Dept. of CSE, JIMS Engineering Management Technical Campus (JEMTEC), Greater Noida, U.P., India
ORCID ID: 0000-0002-1278-4135, Email: hirdesharma@gmail.com

[3] Assistant Professor in Dept. of CSE, ITS Engineering College, Greater Noida, U.P., India
Email: Vijayvijay456@gmail.com

[4] Assistant Professor in Dept. of MCA, Noida Institute of Engineering & Technology, Greater Noida, U.P., India
Email: namiitasharma@gmail.com

[5] Assistant Professor in Dept. of MCA, Noida Institute of Engineering & Technology, Greater Noida, U.P., India
Email: roshan.sonu24@gmail.com

ABSTRACT: In order to boost the likelihood of successful treatment and long-term survival for cancer patients, it is important to identify and prevent cancer in its earliest stages using a linear regression method. Models that predict cancer risk factors and early symptoms can be created using the linear regression algorithm. These models can be trained using historical datasets of cancer patients' demographics, medical histories, and outcomes of diagnostic tests. Doctors may screen patients and determine which ones are more likely to have cancer or who may already have it but be in the early stages by utilizing these prediction models. This enables early detection and treatment, which can significantly raise the likelihood of positive results. On the basis of each patient's unique risk factors and medical background, these models can also be used to create individualized treatment programs for them. Better treatment outcomes for patients may arise from more focused and effective care. In general, early detection and treatment of cancer using the linear regression algorithm has the potential to save lives, enhance patient outcomes, and lessen the total toll that cancer has on people and society. Overall, a comprehensive and rigorous procedure of data collection, preprocessing, feature selection, model training, evaluation, deployment, monitoring, and updating is required for the methodology of employing the linear regression algorithm to identify and prevent cancer in its early stages and it will found 98.2 % accuracy in the model.

KEYWORDS: Cancer, Cancer detection, Cancer Prevention, Breast Cancer, Linear Regression Algorithm, Cancer Prediction, Accuracy.

INTRODUCTION:

In order to improve patient outcomes and lower mortality rates, it is essential to detect and treat cancer early on. Cancer is one of the main causes of death worldwide. Recently, machine

learning techniques like linear regression have demonstrated promise in the early detection and prevention of cancer. To create predictive models based on patient data, linear regression is a supervised learning approach that can be employed. Linear regression can uncover risk factors and early indications of cancer by examining patient variables, including demographics, medical history, and diagnostic outcomes. Improving treatment outcomes and boosting chances of long-term survival depend on the early identification of cancer. Doctors can intervene early and administer focused and efficient treatments by utilizing linear regression to identify people at high risk of acquiring cancer or those who might already be battling it in its early stages. Also, preventing cancer is essential to lessening the overall impact of the disease. Based on each patient's unique risk variables and medical background, linear regression can be used to create individualized screening and prevention regimens for them. In this situation, early cancer identification and prevention using a linear regression algorithm has the potential to save lives, enhance patient outcomes, and lessen the overall toll that cancer has on people and society [1]. Cancer has been a dangerous disease for a long time. This causes death if it is not discovered at an early stage. It is a concern that there is now no viable treatment for this sickness. The only way to save individuals from this disease is if and only if it is diagnosed at an early stage (I and II). If it is diagnosed at a later stage, the chances of survival are quite slim (III and IV). Breast tissues harboring aberrant cells that are uncontrollably growing, changing, and multiplying are just one example of the many malignancies that can result in breast cancer. It represents 25% of all cancer cases and 15% of all female cancer deaths in the expected cancer cases for 2012. According to projections, it claimed the lives of 508,000 women just in 2011. In both industrialized and less developed countries, it is the most prevalent type of cancer among women. Both richer and less economically developed nations suffer greatly from the costs of cancer. A number of factors are being blamed for the rise in cancer diagnoses, including an ageing and growing population as well as a general increase in risk-taking habits like smoking, being overweight, and being inactive. The likelihood of receiving effective therapy considerably increases with early cancer diagnosis. Delay in diagnosis leads to late presentations, which are harmful in terms of mortality, prognosis, and cost of available treatments. Cancer is a dangerous condition marked by the unregulated nature of the body's cells. Cancer is not simply one illness; it has more than 100 distinct kinds. Cancer gradually deteriorates the human body when cells start to proliferate out of control and form multiple tumors. Tumors have the potential to interact with other body components and spread. The circulatory, digestive, or nervous systems may be among them. Infected body parts have a hormonal impact that changes the body [2].

The two methods for cancer detection that are most frequently employed are magnetic resonance imaging and X-ray mammography (MRI). These modern innovations have several drawbacks, such as their high cost, broad scope, and need for truly huge healthcare facilities. The aforementioned techniques may also provide some adverse results and false-positive findings [3-4].

1. TYPES OF CANCER:

A complex and varied illness, cancer can affect different bodily parts. There are many different varieties of cancer, and each has special traits, risk factors, and treatment choices of its own [5].

The most typical cancer kinds are listed below:

1. **Breast cancer:** The cells of the breast tissue are where breast cancer typically develops, across the world.
2. **Lung cancer**: The cells of the lungs, a particular type of cancer called lung cancer can grow. Worldwide, it is the main factor in mortality from cancer.
3. **Prostate cancer:** Men who have the prostate gland acquire prostate cancer, a specific type of cancer. Among men, it is one of the most prevalent cancers.
4. **Colorectal cancer:** The colon or rectum is where colorectal cancer typically first appears.
5. **Skin cancer:** The cells of the skin are where skin cancer typically develops.
6. **Leukemia:** It is blood and bone marrow that both affected by the cancer leukemia. One of the most prevalent cancers in youngsters is this one.
7. **Lymphoma:** The immune system's lymphatic system, which is impacted by a specific type of cancer called lymphoma, is affected. The lymph nodes, spleen, and bone marrow are just a few of the organs that it may influence.
8. **Pancreatic cancer:** In the cells of the pancreas, pancreatic cancer is a particular type of cancer that can appear. With a low rate of survival, it is among the most fatal kinds of cancer.

To avoid cancer, discover it early, and cure it, it's critical to understand the many types of cancer. A patient's prognosis and chances of long-term survival can be significantly improved by early detection and treatment [4].

There are numerous various cancer types, and each has excellent symptoms and treatment methods that have recently been discovered. Breast cancer has been detrimental to women's life for a very long time. This disease has been present at various times throughout history [5]. Ancient Greeks and Egyptians were the first cultures to learn of this ailment. Breast cancer affects around 200 000 individuals nationwide each year [6-7].

Tumors are masses created when normal bladder lining cells—most often urothelial cells—change and begin to develop uncontrollably. Bladder cancer develops in this manner. Furthermore, urothelial cells line the ureters and renal pelvis. The treatment for cancer that develops in the renal pelvis and ureters, despite the fact that it is generally considered of as a type of kidney cancer, is identical to that for bladder cancer. Tumors could be either benign or malignant. In a cancerous tumor, the term "malignant" describes its ability to grow and spread to other parts of the body. The tumor, if benign, can develop but won't spread. Bladder benign tumors are quite rare. Oral and oropharyngeal cancer, which is ninth in prevalence in men, is found in the mouth cavity. Men are diagnosed with these cancers more than twice as often as women. On average, 62 years old is the diagnosis' age. A number of factors can result in cancer. Researchers are still looking into the origins of this type of cancer. Yet, there is currently no known treatment for this condition. It can be used to prevent cancers brought on by HPV. The earliest indicator of kidney cancer is a renal cortical tumor, which is a mass made up of healthy

cells that are growing abnormally in one or both kidneys. A tumor may be benign, sluggish, or malignant. Because malignant tumors are carcinogenic, they can grow and spread to other bodily places. Thyroid cancer develops in the thyroid gland. The voice box, often referred to as the gland, is located in the front of the neck, just behind the larynx. Individuals with thyroid cancer usually show few or no symptoms. Thyroid cancers are frequently found incidentally via x-rays or other imaging tests that were done for another reason or during routine neck checkups [8].

2. LITERATURE REVIEW:

An overview of studies that have used linear regression algorithms for the early cancer detection and prevention is provided below: Researchers employed linear regression methods to forecast patient risk of breast cancer recurrence in a study that was published in the Journal of Biomedical Informatics. The study discovered that increasing the algorithm's genetic and clinical input increased the algorithm's predicted precision.

A decision tree and linear regression algorithm combination was utilized in a different study that was published in the Journal of Medical Systems to estimate the likelihood that smokers will develop lung cancer. According to the study, the algorithm exhibited a high level of prediction accuracy and may be used to create individualized preventative and screening programs with populations.

This is to predict the likelihood of prostate cancer in males based on demographic and clinical characteristics. It is used a logistic regression model. According to the study, the algorithm demonstrated a high sensitivity and specificity for identifying prostate cancer.

Researchers utilized linear regression algorithms to forecast how colorectal cancer patients would respond to chemotherapy. According to the study, the algorithm's predicted accuracy was increased by adding genetic elements.

A different study that was published in the Journal of Thoracic Oncology used a linear regression approach to forecast the possibility of recurrence in different cancer patients. The algorithm is helpful in identifying individuals who might benefit from more aggressive therapy, according to the study, which revealed that it had a high predictive accuracy.

All things considered, these findings indicate that linear regression algorithms can be a useful tool. These algorithms can help with the creation of individualized prevention and treatment strategies by combining clinical and genetic data, increasing the accuracy of cancer risk prediction [9-10].

Table 1: Overview of Detection and Prevention of Cancer in Early Stages in Various Areas.

S. No.	Area	Issue	Outcome	Reference
1	Using a probabilistic and general regression neural network to	The oral malignancies that develop in India typically have advanced stages of cancer.	It is better for early detection and prevention of oral cancer because it has a classification accuracy of 80%.	Sharma, N.. et al.(2015). [30]

	find oral cancer early and prevent it			
2	Using machine learning techniques for breast cancer detection	One in two Indian women who are diagnosed with breast cancer pass away, meaning that there is a 50% risk that the case would end fatally.	The outcomes are highly competitive and useful for both diagnosis and therapy.	Sharma, S. et al.(2008). [31]
3	Working towards the best algorithm for longitudinal tumour markers-based ovarian cancer screening	The postmenopausal women older than 50 without ovarian cancer were divided at random into training and validation groups for the development of a screening test based on longitudinal marker levels.	The training set was tested using the maximum risk, which confirmed a high specificity of 99.7% .	Skates, S. J. et al. (1995).[32]
4	Automatic early stage invasive cervical cancer disease diagnosis using an appropriate machine learning model	Cells of cervical cancer develop gradually at the cervix. This malignancy is successfully treatable if caught early.	It has the highest accuracy on the top 25 characteristics in dataset splitting ratio (80:20).	Jahan, S., et al. (2021).[33]
5	use of artificial intelligence to cancer detection	Because of their lethal character, illnesses like cancer are known as chronic fatal diseases. Because cancer spreads more quickly, it is often discovered when the disease is already advanced. It has been discovered that reducing the fatality rate from cancer depends on early detection.	To apply AI and machine learning algorithms, additional technologies such as data management and image pre-processing must also be improved.	Patel, D. et al. (2020).[34]
6	employing machine learning	It is becoming more and more frequent in medical practise to use ML-based methods to	SVM, however, has been determined to have a 99% accuracy rate. The	Al Mudawi, N. et al.(2022).[35]

	methods to create a model for detecting cervical cancer	predict the early stages of major illnesses including cancer, kidney failure, and heart attacks.	effectiveness of the models is evaluated by computing the computational complexity of traditional machine learning approaches.	
7	machine learning and causal analysis	The cancer disease are correlated with the risk of human papillomavirus infection. Around 37% of cancer cases can be avoided with preventive care, which is the most expensive cancer treatment.	With the help of Python and many performance measurement criteria, such as accuracy, precision, and recall was found.	Lilhore, U. K., et al. (2022).[36]
8	Several methods of combining genetic algorithms and logistic regression for predicting hepatocellular cancer survival are contrasted.	Diagnosis in humans is challenging due to a variety of variables. Using machine learning approaches, an unique diagnostics strategy is proposed in this study.	The proposed model had a f1-score of 93.56% and a classification accuracy of 94.55%. Our algorithm demonstrates how the suggested concept's optimal machine learning techniques can introduce a fresh and accurate method for diagnosing HCC.	Książek, W., et al. (2021).[37]
9	Using a combination of feature selection and regression techniques	The liver is a necessary organ for life, but because the symptoms of liver illness are often hidden, it can be exceedingly challenging to diagnose it at an early stage. One of the most prevalent and lethally destructive liver conditions is cancer.	It is providing the highest R2-Score of 0.8923 and the lowest MSE of 0.0618 among various ML algorithms.	Mehmood, M., et al. (2022). [38]
10	To detect Corona Virus Illness, a deep learning system is used using CT	More than 26 million cases of Corona virus disease (COVID-19) were found.	Outcome is that AI techniques could be able to extract particular graphic aspects from COVID-19 .	Wang, S., et al. (2021). [39]

	images (COVID-19)			
11	Using a powerful characteristic, cervical cancer cells can be detected the CNN-SVM network.	The traditional method of detecting cervical cancer cells relies heavily on the pathologists' expertise, which also has the drawbacks of being inefficient and having low accuracy.	The outcomes suggested that the strong feature CNN-SVM will give good accuracy also.	Jia, A. D., et al. (2020).[40]
12	Bionic Glycome approach with machine learning for advanced adenoma and colorectal cancer screening and diagnosis	Most colorectal adenomas turn into colorectal cancer (CRC), one of the biggest health issues in the world. Patients are advised to have advanced adenomas removed because they are typically regarded as precancerous tumours. Fecal tests (FOBT or FIT) and colonoscopies are typically used for colorectal cancer screening, although their effectiveness is constrained by uptake and adherence.	Overall, the diagnostic model we developed here shows great promise for adenomas, and it may be able to make up for the shortcomings of the present screening techniques for these conditions.	Pan, Y. et al. (2021). [41]
13	using the deep learning Exception algorithm to identify and categorise breast cancer	The breast cancer will claim the lives of about 10 million individuals worldwide. Breast cancer is a lethal condition that is quite common among women around the world. It comes in at number four when compared to other malignancies that can be fatal, like brain tumours, cervical cancer, and colorectal cancer.	Using the pre-processed dataset, the proposed model was developed, tested, and verified. Precision was (97.60%), Recall was (97.60%), and F1-Score was (97.58%), according to the results.	Abunasser, B. S. et al. (2022). [42]
14	Multimarket panel development	The most typical gynaecologic cancer is endometrial carcinoma. Although endometrial cancer has a generally favourable	The outcome is that prolactin's capacity to successfully distinguish between cancer and control groups.	Yurkovetsky, Z. et al.(2007).[43]

		prognosis, tumours discovered in their last stage.		
15	Understanding the Impact of Internet of Things (IoT) on cancer	Effective IoT solutions, however, may link to and exchange the critical clinical data with other healthcare systems and devices operating over the extensive Internet infrastructure.	Breast temperature does somewhat rise during the development of cancer and its types.	Shukla, S. K..et al.(2022)..[44]

3. PROPOSED MODEL OF DETECTION AND PREVENTION OF CANCER IN EARLY STAGES:

Using the linear regression algorithm, the following steps can be included in a proposed model for early cancer detection and prevention [11-12]:

(1.) Data collection: Patients' demographics, medical histories, and diagnostic outcomes should be compiled into a sizable and varied dataset. Patients with both cancer and non-cancer diagnoses should be included in the dataset, which should be general population-representative.

(2.) Data preprocessing: To eliminate any discrepancies, mistakes, or missing values from the gathered data, preprocessing is necessary. Techniques for data cleansing, data transformation, and data standardization may be used in this.

(3.) Feature selection: It is important to choose from the dataset relevant elements that could predict cancer risk or cancer diagnosis. This may entail statistical analysis, feature rating formulas, or subject-matter knowledge.

(4.) Model training: A linear regression model should be trained using the chosen features so that it can predict and diagnose cancer.

(5.) Model evaluation: Using statistical criteria like as accuracy, precision, recall, and F1 score, one should assess the performance of the trained model. To confirm the model's applicability to fresh and untested data, cross-validation techniques may be applied.

(6.) Deployment: This model has been proven to be accurate, it can be used in a clinical context to screen patients for this or to detect the disease early. In order to do this, a new screening technique may need to be developed or the model may need to be integrated into already-in-use clinical operations.

(7.) Monitoring and updating: If new information becomes available or the population changes, the model should be continually checked for errors and updated. As a result, the model is guaranteed to continue being reliable and efficient throughout time.

(8.) Personalized screening and prevention plans: On the basis of the model's projections and each patient's unique risk factors and medical background, individualized screening and preventative regimens can be created.

Overall, the process of data collection, preprocessing, feature selection, model training, evaluation, deployment, monitoring, and updating that goes into this suggested model for identifying and preventing cancer in its early stages using the linear regression algorithm is extensive and rigorous. This strategy has the promise of reducing the entire toll that cancer

takes on people and society while also enhancing patient outcomes and possible lifesaving [13-16].

4. LINEAR REGRESSION ALGORITHM USED FOR FINDING RESULT IN EARLY STAGES IN PREVENTION AND DETECTION OF CANCER:

Regression analysis is done with the help of the supervised learning technique known as linear regression. Making predictions about upcoming data points entails fitting a linear equation to a dataset. Simple and multiple linear regressions can both be performed using the technique [17-18].

The steps in the procedure for linear regression are as follows:

(1.) Data collection: It gathers dependent and independent variables into a dataset.

(2.) Data preprocessing: To eliminate any discrepancies, mistakes, or missing values, the data is preprocessed. Techniques for data cleansing, data transformation, and data standardization may be used in this.

(3.) Feature selection: It is chosen from the dataset relevant features that could forecast the dependent variable.

(4.) Model training: An algorithm for linear regression is trained using the chosen features. The model finds the coefficients of a linear equation that best describe the data and fits the data to the equation.

(5.) Model evaluation: Statistical metrics like the R-squared value, mean squared error, and root mean square error are used to assess how well the trained model performed.

(6.) Prediction: After the model has been verified, predictions regarding upcoming data points can be made using it. The dependent variable's predicted value is generated by the model using the independent variable as input [19-20].

5. CANCER PREVENTION AND DETECTION TECHNIQUES IN EARLY STAGES

Cancer can be found early on, when it is most treatable, using a number of cancer prevention and detection strategies. The most typical methods are listed below [21-22]:

(1.) Self-examination: By regularly performing self-examinations, it is possible to identify several cancers, including skin and breast cancer. To check for any lumps or anomalies, for instance, women should self-examine their breasts every month.

(2.) Regular check-ups: Frequent medical exams and screenings can be done, depending on age and risk factors, the American Cancer Society advises routine screenings for breast, cervical, colorectal, and lung cancers as well as a few other types of cancer.

(3.) Genetic testing: Genetic testing can assist in identifying those who are more likely to contract cancers including breast, ovarian, or colon cancer. This knowledge can aid people in making preventative decisions and/or considering earlier or more frequent screenings, as well as helping people and their healthcare providers do so.

(4.) Lifestyle changes: The risk of cancer can be lowered by living a healthy lifestyle. This entails keeping a healthy weight, exercising frequently, eating a diet high in fruits and vegetables, abstaining from cigarette use, and not drinking excessive amounts of alcohol.

(5.) Vaccinations: The human papillomavirus, which can cause some cancers like cervical cancer, can be prevented by vaccinations (HPV). Depending on personal risk factors, additional vaccinations might also be advised.
(6.) Imaging tests: Mammograms, CT scans, and MRI scans are imaging tests that can assist identify cancer in its early stages. With testing, tumors or other abnormalities that may not be obvious through a physical examination might be found.
(7.) Biopsies: A biopsy entails removing a sample of tissue from the area that is thought to be malignant and checking it for cancer cells. By using this method, it is possible to identify the cancer's type and confirm a cancer diagnosis.
In general, lowering cancer-related morbidity and mortality requires early detection and prevention. People should discuss their personal risk factors and recommended screenings with their medical professionals [23-26].

6. TECHNIQUES USED FOR CANCER PREVENTION AND DETECTION IN EARLY STAGES:

In order to diagnose cancer and prevent it in any stages, numerous procedures are performed. The most typical ones are listed here:
(1.) Screening tests: These tests are designed to find cancer early on, before any symptoms appear. Mammography, colonoscopy, and Pap smear are a few examples of screening tests for cancers of the breast, colon, and cervix.
(2.) Genetic testing: A genetic mutation that raises the chance of acquiring a particular type of cancer will be tested for in this process. If such a mutation is discovered in a person, they may choose to get screened more frequently or take efforts to lower their risk.
(3.) Lifestyle modifications: A nutritious diet, stopping smoking, cutting back on alcohol.
(4.) Immunization: Certain cancers are preventable with certain immunizations.
(5.) Chemoprevention: To lower the likelihood of acquiring cancer, this entails using medications or other substances.
(6.) Biopsy: A biopsy may be used to confirm the cancer diagnosis if a screening test or physical examination point to the disease. The key to effectively treating cancer is, in general, early detection and prevention. To decide the appropriate course of action based on individual risk factors and medical history, it is crucial to explore these alternatives with a healthcare professional.

7. RESULTS

This is found that every machine learning algorithm will be giving very higher accuracy which is showing in the different graphs below but linear regression algorithm will give higher accuracy to the given model. The data has been taken on the link as https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data?resource=download and analyses it by the different machine algorithms.

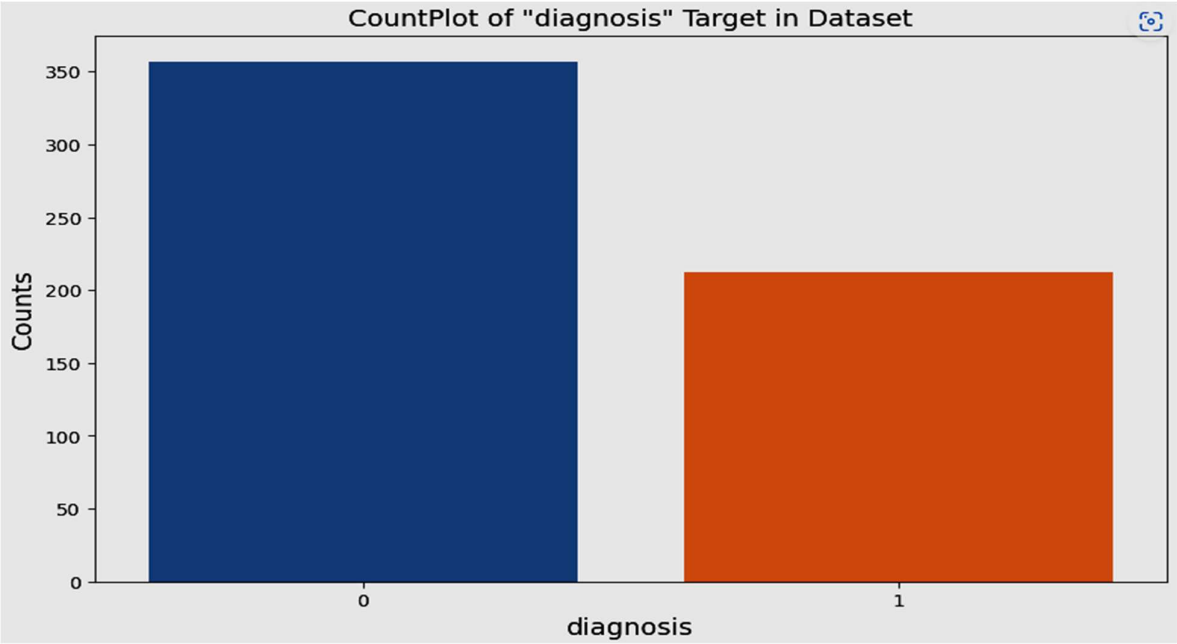


Fig. 1:- Diagnosis vs Counts

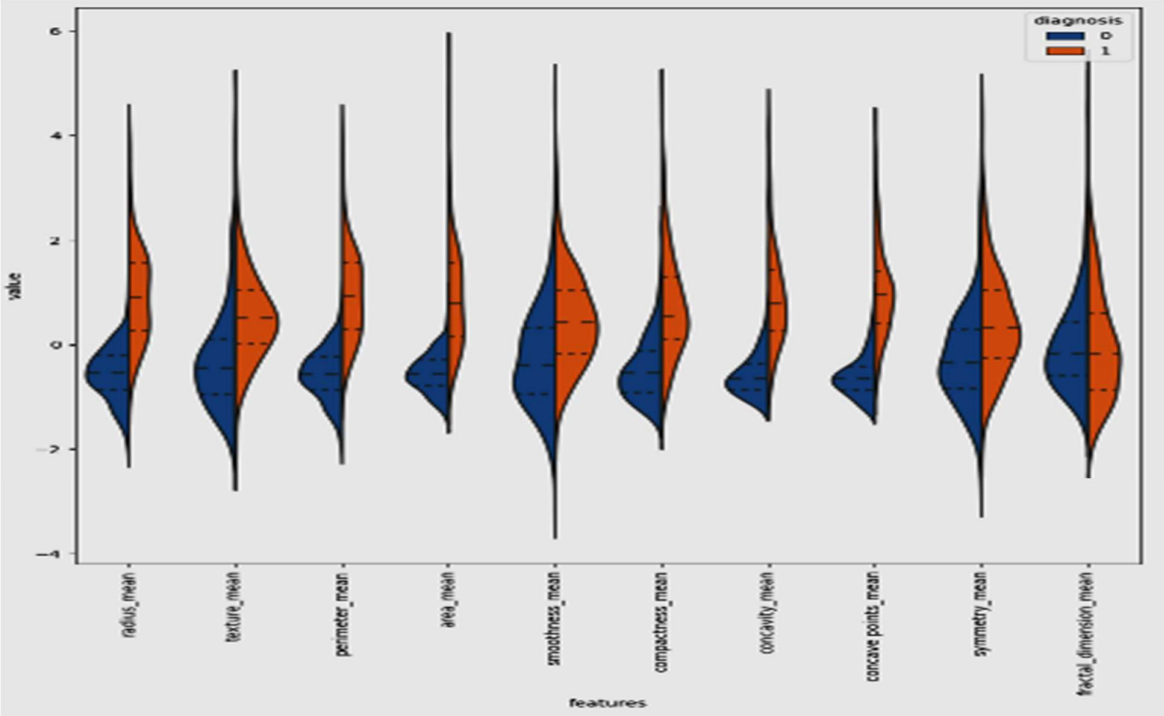


Fig 2:- Diagnosis Vs Counts on different factors

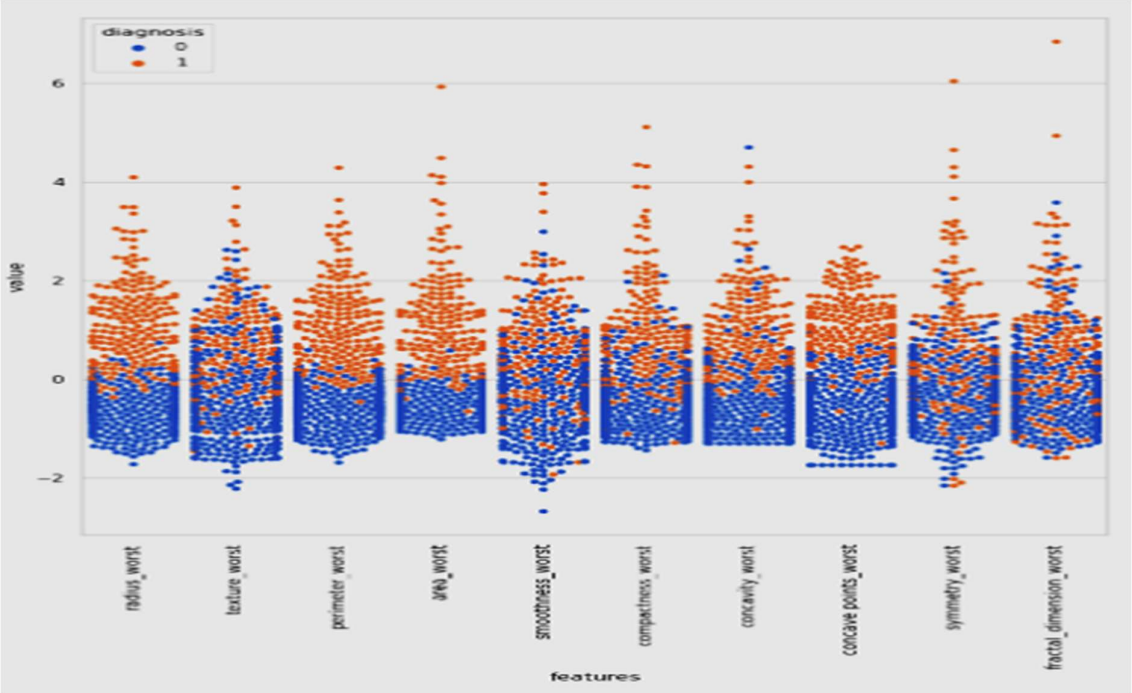
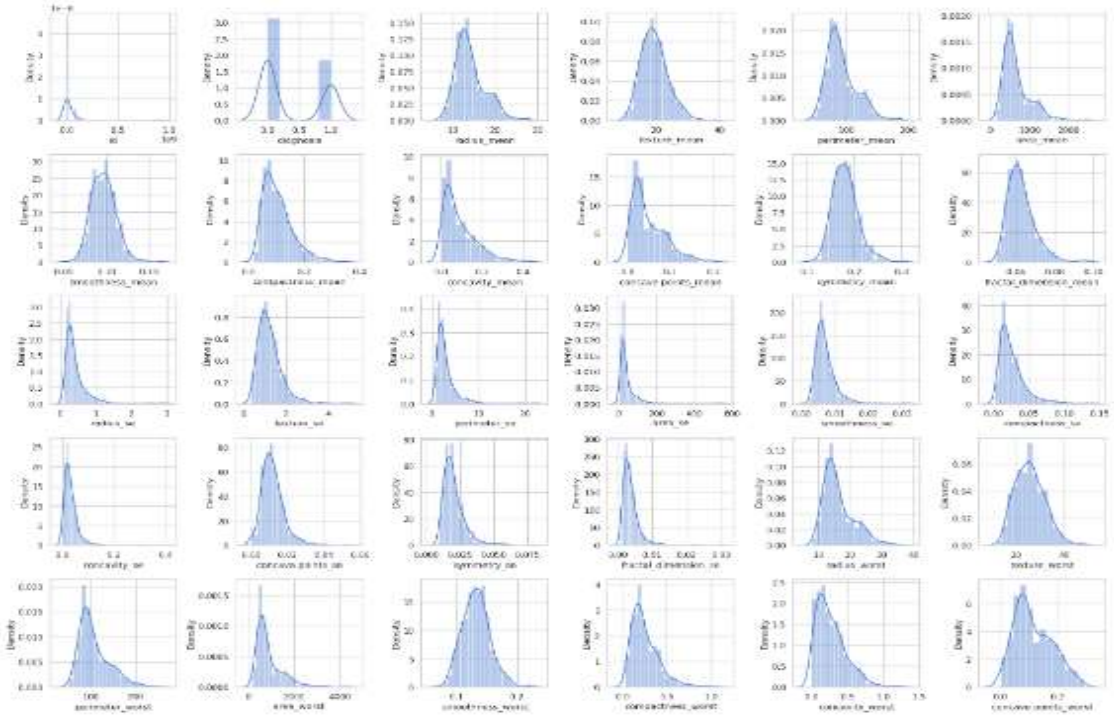


Fig 3:- Features of different factors involved in early cancer detection



Fig, 4: - Different factors on early cancer detection in graphical format

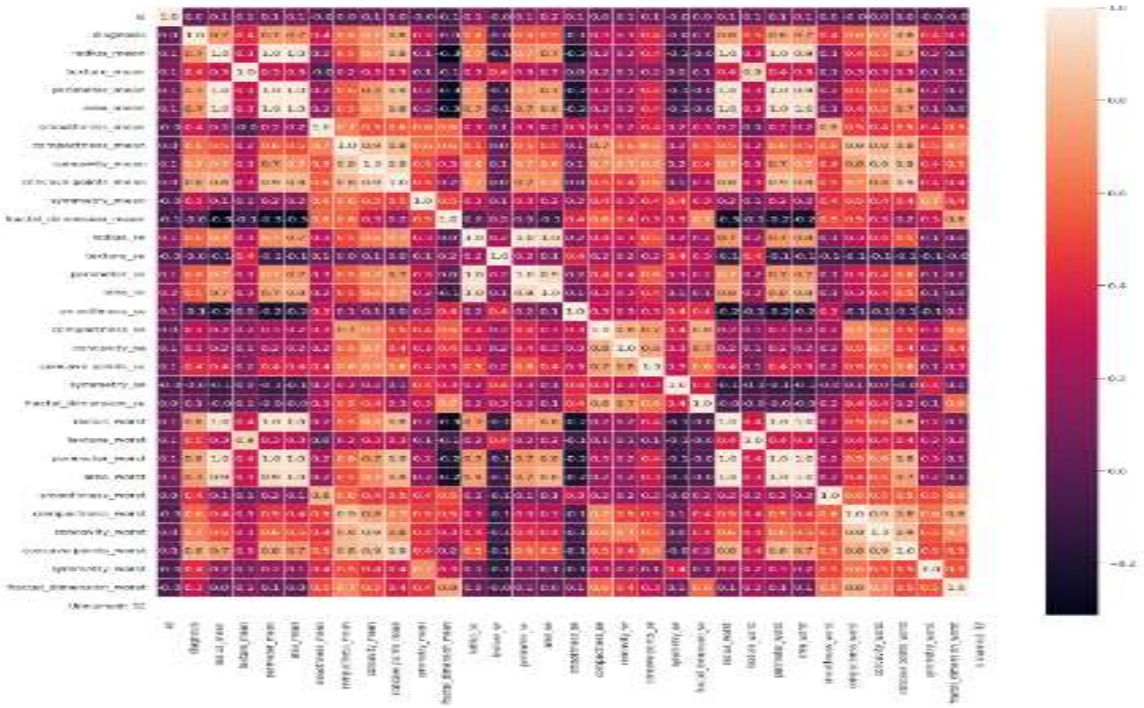


Fig 5:- Heat map of different factors involved in early cancer detection

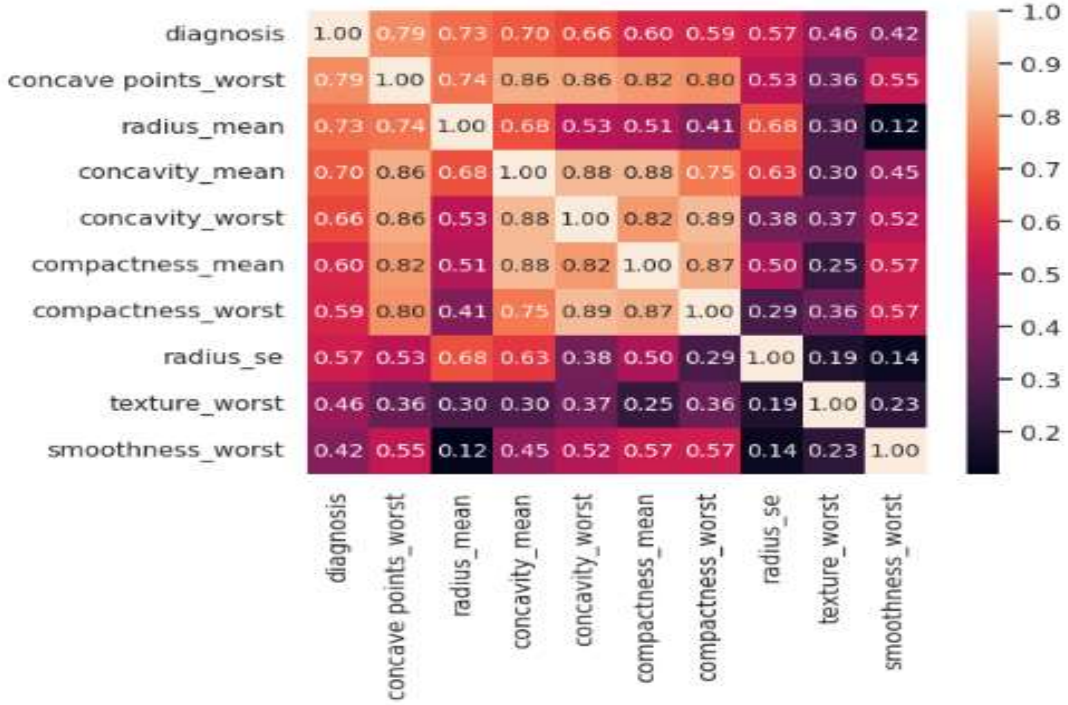


Fig. 6: - Heat map of different dependent factors involved in this research

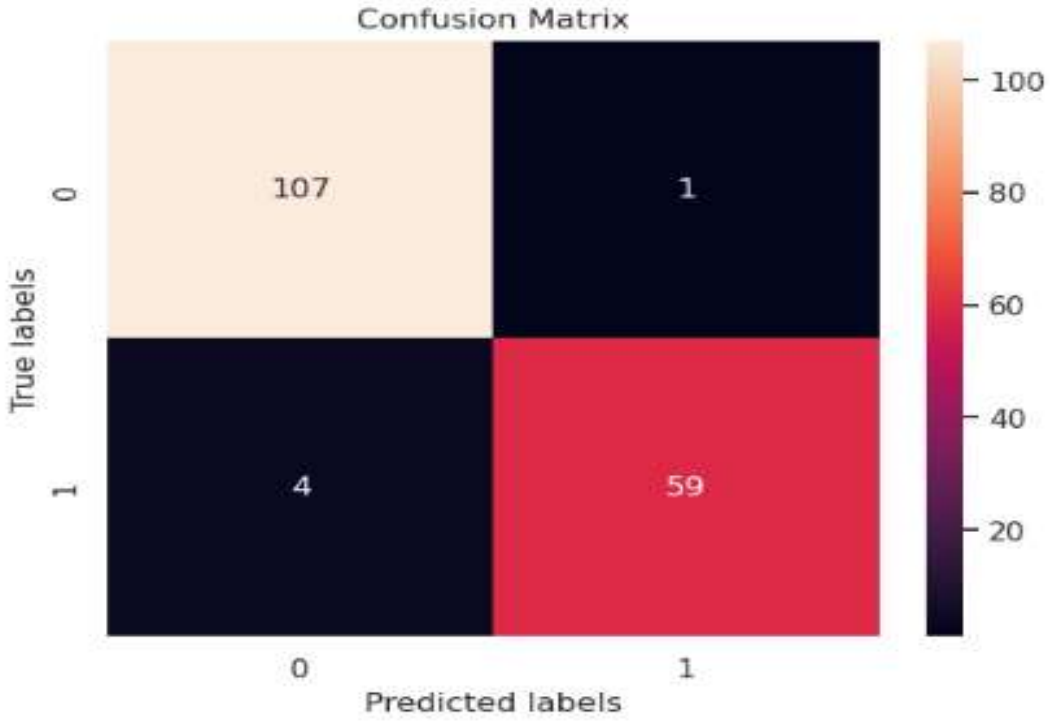


Fig 7: - Confusion matrix of predicted labels vs true labels

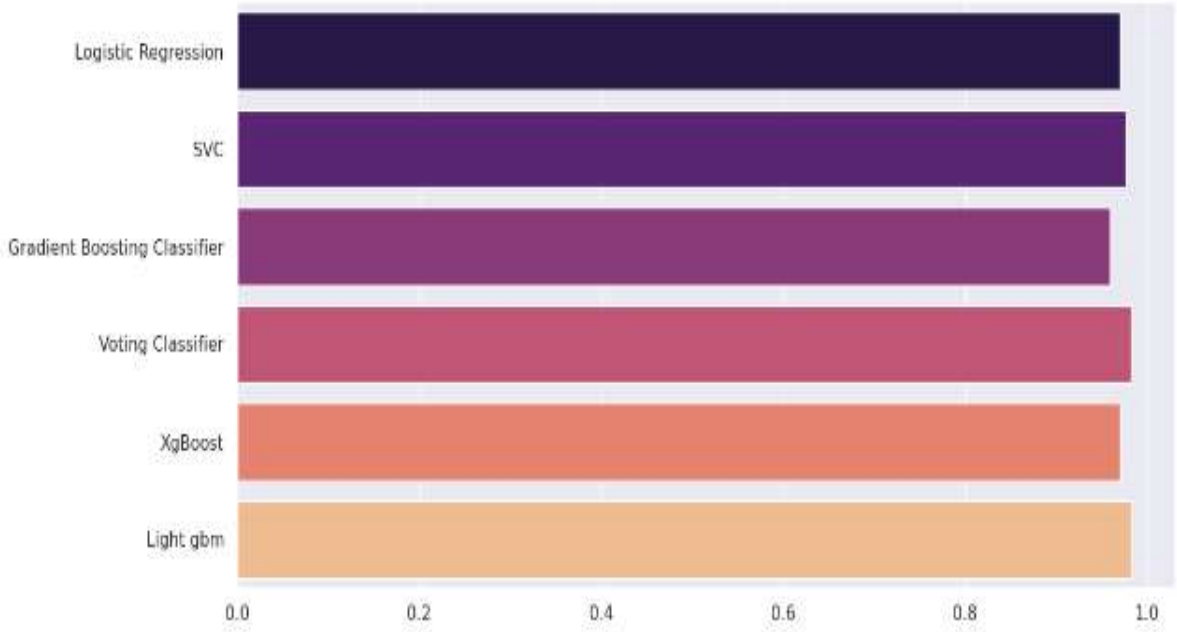


Fig. 8:- Accuracy of different algorithms with comparisons of all with linear regression algorithm.

8. CONCLUSION:

In summary, early cancer diagnosis and prevention are essential for a successful course of therapy and better patient outcomes. The application of linear regression algorithms in cancer

diagnosis and prevention can be a potent tool in determining the most efficient preventive and therapeutic approaches as well as forecasting the possibility of cancer in high-risk patients. Linear regression algorithms can find patterns and associations that might be predictive of cancer risk by gathering and evaluating data on well-known risk variables, as well as applying statistical analysis tools. It is crucial to remember that linear regression algorithms are simply one of several tools available in the battle against cancer. These should always be used with other diagnostic and preventative techniques, and the results should always be interpreted by a qualified healthcare expert. Overall, the use of linear regression algorithms to cancer detection and prevention has enormous promise to enhance patient outcomes and lessen the burden of this fatal illness. The accuracy is 98.2 % [26-30].

9. REFERENCES:

[1] Abdi, H. and Williams, L. J. (2009). Principal component analysis. *In Encyclopedia of Biometrics*.1(1). 12-23.

[2] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3). 175–185.

[3] Barandiaran, I. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence,* 20(8).11-21.

[4] Bellaachia, A. and Guven, E. (2006). Predicting breast cancer survivability using data mining techniques. *Age,* 58(13).10–110.

[5] Delen, D., Walker, G., and Kadam, A. (2005). Predicting breast cancer survivability: a comparison of three data mining methods. *Artificial intelligence in medicine*, 34(2). 113– 127.

[6] Ferlay, J., Shin, H., Bray, F., Forman, D., Mathers, C., and Parkin, D. (2011). Globocan 2008, cancer incidence and mortality worldwide: Iarc cancerbase no. 10. lyon, france: *International agency for research on cancer*; 20(1). 21-35..

[7] Friedman, J., Hastie, T., and Tibshirani, R. (2001). The elements of statistical learning, *Springer series in statistics New York, NY, USA.* 1(1). 111-125.

[8] Gholami, V., Chau, K., Fadaee, F., Torkaman, J., and Ghaffari, A. (2015). Modeling of groundwater level fluctuations using dendrochronology in alluvial aquifers. *Journal of hydrology*, 52(9). 1060–1069.

[9] Ho, T. K. (1995). Random decision forests. In Document analysis and recognition, 1995., *proceedings of the third international conference on IEEE.* . 1(1), 278–282.

[10] Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied logistic regression, *John Wiley & Sons*.398(1). 1-21.

[11] Hussain, M., Wajid, S. K., Elzaart, A., and Berbar, M. (2011). A comparison of svm kernel functions for breast cancer detection. *In 2011 Eighth International Conference Computer Graphics, Imaging and Visualization in IEEE*, 1(1). 145–150.

[12] Janghel, R., Shukla, A., Tiwari, R., and Kala, R. (2010). Breast cancer diagnosis using artificial neural network models. *In Information Sciences and Interaction Sciences (ICIS), 2010 3rd International Conference on IEEE.* 1(2). 89–94

[13] Kharya, S. and Soni, S. (2016). Weighted naive bayes classifier: A predictive model for breast cancer detection. *International Journal of Computer Applications*, 133(9). 32–73.

[14] Kohavi, R. (1998). Glossary of terms. *Special issue on applications of machine learning and the knowledge discovery process*, 30(271).127–132.

[15] Liu, L. (2018). Research on logistic regression algorithm of breast cancer diagnose data by machine learning. *In 2018 International Conference on Robots & Intelligent System (ICRIS) IEEE*. 1(1). 157–160.

[16] Lundin, M., Lundin, J., Burke, H., Toikkanen, S., Pylkkänen, L., and Joensuu, H. (1999). Artificial neural networks applied to survival prediction in breast cancer. *Oncology*, 57(4). 281–286.

[17] Marsilin, J. R. and Jiji, G. W. (2012). An efficient cbir approach for diagnosing the stages of breast cancer using knn classifier. *Bonfring International Journal of Advances in Image Processing*, 2(1). 01–05.

[18] Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, unforcedness, markedness and correlation. *International Journal of Computer Applications*, 133(9).2–17.

[19] Rennie, J. D., Shih, L., Teevan, J., and Karger, D. R. (2003). Tackling the poor assumptions of naive bayes text classifiers. *In Proceedings of the 20th international conference on machine learning (icml-03)*, 1(2). 616–623.

[20] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3). 210–229.

[21] Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2):87–108.

[22] Uddin, A. K., Khan, Z. J., and Johirul Islam, A. (2013). Cancer care scenario in bangladesh. *South Asian journal of cancer*, 2(2).102-120.

[23] van Gerven, M. and Bohte, S. (2018). Artificial neural networks as models of neural information processing. *Frontiers Media SA.* 3(2).12-32.

[24] Zhang, Z., Bast Jr, R. C., Yu, Y., Li, J., Sokoll, L. J., Rai, A. J., ... & Chan, D. W. (2004). Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer research*, 64(16), 5882-5890.

[25] Wolberg, W., Street, W. N., and Mangasarian, O. L. (1992). Breast cancer wisconsin (diagnostic) data set *[uci machine learning repository]*.1(1). 12-25.

[26] Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Philip, S. Y., et al. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37.

[27] Xiao, Y., Wu, J., Lin, Z., and Zhao, X. (2018). Breast cancer diagnosis using an unsupervised feature extraction algorithm based on deep learning. *In 2018 37th Chinese Control Conference (CCC) IEEE*, 1(1). 9428–9433

[28] Zemouri, R., Omri, N., Devalland, C., Arnould, L., Morello, B., Zerhouni, N., and Fnaiech, F. (2018). Breast cancer diagnosis based on joint variable selection and constructive deep neural network. *In 2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME) IEEE*. 1(1). 159–164.

[29] Zhou, Z.-H., Jiang, Y., et al. (2003). Medical diagnosis with c4. 5 rule preceded by artificial neural network ensemble. *IEEE Transactions on information Technology in Biomedicine*, 7(1):37–42.

[30] Sharma, N., & Om, H. (2015). Usage of probabilistic and general regression neural network for early detection and prevention of oral cancer. *The Scientific World Journal*, *2015*.

[31] Sharma, S., Aggarwal, A., & Choudhury, T. (2018, December). Breast cancer detection using machine learning algorithms. In *2018 International conference on computational techniques, electronics and mechanical systems (CTEMS)IEEE*. 1(1). 114-118.

[32] Skates, S. J., Xu, F. J., Yu, Y. H., Sjövall, K., Einhorn, N., Chang, Y., ... & Knapp, R. C. (1995). Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. *Cancer*, *76*(S10), 2004-2010.

[33] Jahan, S., Islam, M. S., Islam, L., Rashme, T. Y., Prova, A. A., Paul, B. K., ... & Mosharof, M. K. (2021). Automated invasive cervical cancer disease detection at early stage through suitable machine learning model. *SN Applied Sciences*, *3(1)*, 1-17.

[34] Patel, D., Shah, Y., Thakkar, N., Shah, K., & Shah, M. (2020). Implementation of artificial intelligence techniques for cancer detection. *Augmented Human Research*, *5(1)*, 1-10.

[35] Al Mudawi, N., & Alazeb, A. (2022). A model for predicting cervical cancer using machine learning algorithms. *Sensors*, *22*(11), 4-132.

[36] Lilhore, U. K., Poongodi, M., Kaur, A., Simaiya, S., Algarni, A. D., Elmannai, H., ... & Hamdi, M. (2022). Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques. *Computational and Mathematical Methods in Medicine*, 1(1). 12-21.

[37] Książek, W., Gandor, M., & Pławiak, P. (2021). Comparison of various approaches to combine logistic regression with genetic algorithms in survival prediction of hepatocellular carcinoma. *Computers in Biology and Medicine*, *134(1)*, 104-431.

[38] Mehmood, M., Alshammari, N., Alanazi, S. A., & Ahmad, F. (2022). Systematic framework to predict early-stage liver carcinoma using hybrid of feature selection techniques and regression techniques. *Complexity*, *20(22)*, 1-11.

[39] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... & Xu, B. (2021). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *European radiology*, *3(1)*, 6096-6104.

[40] Jia, A. D., Li, B. Z., & Zhang, C. C. (2020). Detection of cervical cancer cells based on strong feature CNN-SVM network. *Neurocomputing*, *4(11)*, 112-127.

[41] Pan, Y., Zhang, L., Zhang, R., Han, J., Qin, W., Gu, Y., ... & Gu, J. (2021). Screening and diagnosis of colorectal cancer and advanced adenoma by Bionic Glycome method and machine learning. *American Journal of Cancer Research*, *11*(6), 3002-3019.

[42] Abunasser, B. S., AL-Hiealy, M. R. J., Zaqout, I. S., & Abu-Naser, S. S. (2022). Breast cancer detection and classification using deep learning Xception algorithm. *International Journal of Advanced Computer Science and Applications*, *13*(7). 51-68.

[43] Shukla, S. K., Kumar, B. M., Sinha, D., Nemade, V., Mussiraliyeva, S., Sugumar, R., & Jain, R. (2022). Apprehending the Effect of Internet of Things (IoT) Enables Big Data Processing through Multinetwork in Supporting High-Quality Food Products to Reduce Breast Cancer. *Journal of food quality*, *1(1). 21-45*