Assorted Features Used in Machine Learning Algorithms to Detect Cancer- A Holistic Approach

S. Lavanya [1], Dr.S.Manoj Kumar [2]

[1]*Assistant Professor Department of Information Technology Karpagam College of Engineering, Coimbatore, Tamilnadu, India.*

[2]*Professor Department of Computer Science and Engineering KPR Institute of Engineering and Technology, Coimbatore, Tamilnadu, India.*

***Abstract -** Cancer is one of the chronic diseases with higher mortality rate compared to all other diseases. Early prediction and detection of cancer may reduce the mortality rate among the patients. Many researchers have put their effort on predicting and detecting the cancer using many machine learning algorithms. The first step of detecting the occurrence of cancer is selecting the attributes for the machine learning model. There are many types of cancer that affect lung, prostrate, liver, stomach, breast, uterus which affect both men and women in equal ratio. The motivation of this paper is to accumulate all the attributes which is used by the holistic machine learning algorithms to detect and predict different types of cancer in human beings. So that it will be helpful for the future work to segregate the features which are needed to detect the particular type of cancer.*

***Keywords:** Cancer Types, Machine Learning Algorithms, Feature Selection, Detecting Cancer.*

I. INTRODUCTION

Cancer is the huge collection of diseases which can start from any part of the body irrelevant of age, sex, diet, work place, hereditary etc.., as an abnormal cell which grows uncontrollably and grow beyond the usual boundaries to affect the other parts of the body. According to a report published by WHO (World Health Organization) on March 2021, there are more than 10 million deaths recorded due to cancer worldwide in 2020. In which most of the deaths in are caused by cancer in “lung (1.80 million), colon and rectum (935000), liver (830000), stomach (769000) and breast (685000)” [1]. Altogether the anxiety of cancer occurrence and mortality have been rapidly increasing worldwide.

Cancer is a genetic disease which is caused by changes occurring in genes that lead to abnormal cell division and growth. Metastatic cancer refers to cancer that has spread to other sections of the body, either as a result of the disease's aggressiveness or as a result of invasive procedures like biopsies. The manner of endearing the cancer to other parts of body is termed metastasis [2].

A. Basic Types of Cancer

There are assorted types of cancer which may occur in the human body, in which a holistic aspect of cancer is listed below [2].

TABLE I TYPES OF CANCER

S.no	Type of Cancer	Body Part Affected
1	Adeno Carcinoma	Breast, Colon, Prostrate
2	Basal Cell Carcinoma	Epidermis's bottom layer
3	Squamous Cell carcinoma	Stomach, Kidney, Lungs, Intestine and Bladder.
4	Transitional Cell Carcinoma	Lining of Bladder, Ureters and kidney.
5	Sarcoma	Bones and Soft Tissues
6	Leukemia	Blood Cells
7	Lymphoma	Lymph nodes and Lymph Vessels
8	Non-Hodgkin lymphoma	Blood Cells
9	Kaposi sarcoma	Skin lesion
10	Oropharyngeal cancer	Tonsil
11	Multiple Myeloma	Plasma Cells
12	Melanoma	Skin and Eyes
13	Tumors	Brain and Spinal cord
14	Colorectal cancer	Bowel, Colon and Rectal

In addition to this, a person may be affected by numerous other distinct forms of cancer. The cancer kinds listed above are simply broad categories.

B. Importance of Cancer Detection

Cancer detection is the process of identifying the occurrence of cancer based on the symptoms. Mortality rate of cancer is higher worldwide due to late diagnosis of cancer. According to the researchers if the cancer is detected at early stage than the mortality rate can be reduced in large extent. So it is very essential to detect the cancer at the early stage itself. Lamentably, there does not exist effectual screening tests for early detection of many cancers. In lieu, many works have intently insisted that screening even though has many advantages but also have some negatives in it.

In some cases there may be chance of over diagnosis and over treatment in which both are harmful for the patients. When the patients are suggested with unnecessary test in case to diagnose the cancer will lead to unwanted physical damage by means of radiation by scanning process or some invasive methods and psychological stress associated with the diagnosis. In order to handle the risk, the detection can be done based on the machine learning algorithm, as it will not create any physical harm to the patients like invasive biopsy test that lead to metastases and also the psychological stress.

C. *The Need of Machine Learning Algorithm in Health Care*

In the fast growing world as technology blooms rapidly in similar way the health problems for the human also blooms enormously. In this respect health sectors generate tons of data each and every day all over the world. The data handled may include the details like patients personal details, their symptoms, their medications as well as their imaging and statistical details. The data which is generated in medical sector may be raw data which is un structured and distributed broadly through diverse sections like imaging, pharmacy and treatment. In order to analyze the data for some statistical purpose or the research purpose it needs to be organized and stored in the way it can be accessed easily and efficiently.

Handling the enormous amount of data manually may lead to may drawbacks like error in processing the data, false prediction due to manual error, mishandling the data, leakage of sensitive data and misinterpretation of data. In order to analyze the medical data which includes imaging data, clinical data, patient's personal experience or omics data it is difficult for the human to do it manually. As decisions which are made by human may contain errors due to handing the assorted features analyzed from different form of data. In order to handle these drawbacks health care data needs to be analyzed with machine learning algorithms. If machine learning algorithms are used for analyzing the data then the prediction of the diseases will be accurate and no misinterpretation will be occurred.

The results will be reliable when prediction and detection are carried out using machine learning algorithms, and early cancer detection may lower the patient mortality rate.

D. Data Analysis using Machine learning Algorithms for Cancer Prediction and Detection

Machine learning is one among the foremost common sorts of AI. It is used for decision making for the given test data set using the training data sets already available. Machine learning applications contain algorithms which have set of procedure to analyze the dataset and based on the features it will make the decision without any human intervention. Machine learning algorithm learns to make decision based on the training data set given to it.

Machine learning algorithms predict the results with higher accuracy, when proper training and test set is given as the input without the usage of programming codes. The three important components of the algorithm are representation, evaluation and optimization.

Representation of data is means the data which is to be analyzed by the machine learning algorithm needs to be classified and represented in the format which is understandable by the algorithm. Only if the representation of data is visualized in correct manner then only the next process of evaluation can be performed in effective manner. Evaluation means the data which is represented is classified according to the features and the algorithm. The third part of ML algorithms is optimization in which the algorithm selects the simplest model for deriving the most efficient and accurate outputs [3].

Extracting the necessary data from the mixed data collection and removing unwanted churns is the first and most important stage in data analysis. Data review, cleaning, transformation, and subsequent modeling are all steps in the process.

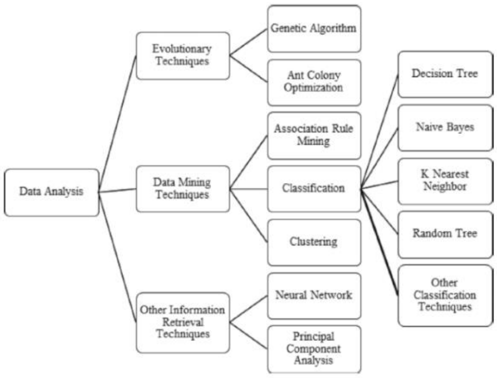

Fig 1. Types of Data Analysis using Machine Learning Algorithms

First and foremost step in data analysis feature extraction and selection. When the attributes we select for the prediction algorithm is appropriate, the result produced will be higher accuracy. In order to increase learning accuracy, save computation time, and produce a better model for more accurate result prediction, feature selection is the process of eliminating redundant and irrelevant data.

In order to forecast various forms of cancer, this work aims to perform a comparative analysis, compile a list of qualities, and provide an overview of the features utilized in various machine learning algorithms.

II. LITERATURE SURVEY

The new epicenter of authority is data. The exponential growth in the utilization of data analytics across all domains has rendered data analytics more attainable. There are many machine learning models proposed by many researchers which work on prediction and detection of different types of cancer. However, choosing the various attributes for the machine learning model is crucial if you want the best accuracy. The below given are some of the assorted attributes considered by the researchers for detecting the different types of cancer.

A.ASSORTED FEATURES USED IN DETECTION OF BREAST CANCER.

Z. Huang et al. [4] proposed the Hierarchical Clustering Random Forest (HCRF) model, in which decision tree results are examined and clustered for improved prediction of results with higher accuracy. A machine learning model can only operate with more accuracy if the right features are chosen for it. The Variable Importance Measure approach was employed by the authors in this work to choose features from the dataset. The “Wisconsin Diagnosis Breast Cancer (WDBC) database and Wisconsin Breast Cancer (WBC) datasets” are used in this investigation. The suggested model operates with 97.05% accuracy when compared to the “Decision Tree, Adaboost, and Random Forest algorithms”.

Based on the study performed in different research papers, Twelve classification algorithms were used to detect the breast cancer, namely; “Ada BoostM1, Decision Table, J-Rip, J48, Lazy IBK, Lazy K-star, Logistics Regression, Multiclass Classifier, Multilayer–Perceptron, Naïve Bayes, Random Forest, and Random Tree”. In which the accuracy is fully depends upon the attributes selected by researchers for the prediction of the disease.

The researchers will choose which feature combinations to utilize based on the models and methodology they use. While some may anticipate using broad symptoms, others may use cell biology to diagnose cancer cells.

B. ASSORTED FEATURES USED IN DETECTION OF LUNG CANCER.

G. Zhang et.al [5] outlined attention embedded three dimensional convolution neural network to analyze the pulmonary nodules from CT imaging. The channel gives the feature selection a lot of consideration in the suggested strategy. For effective feature selection, the author employed “an adaptive feature fusion (AFF) network based on channel-space attention (CSA)”. Just as crucial as choosing the right feature for the model to increase accuracy is choosing the right model for cancer diagnosis. The researcher's suggested approach has a 97.7% sensitivity rate for lung cancer detection.

According to Y. Lu and et.al [6] utilizing both the imaging results and clinical information of the patients to make long-term mortality predictions, with minimum human input. The model use saliency maps for identifying the features form the CT images for the detection of both lung malignancy and other cardiac based diseases. Here, the clinical data of the patients is analyzed using SVM, GBM, and Random Forest, while the CT scans are analyzed using a 3D implementation of Res net. The “5-fold cross-validation technique” is used to evaluate the model. With an AUC value of 0.73, the current approach does reasonably well in forecasting long-term mortality in the NLST dataset.

C. ASSORTED FEATURES USED IN DETECTION OF THROAT CANCER.

According to the study performed in different research papers, the classification algorithms used for the diagnosis of throat cancer and its types are Naïve Bayes, Decision Tree.

A model that employed feature selection and machine learning to make predictions about oral cancer utilizing 31 oral cancer data from the “Malaysian Oral Cancer Database and Tissue Bank System (MOCDTBS)” was proposed by Siow-Wee Chang et al. [7] after taking clinicopathologic and genomic data into consideration. In the first step, five feature selection methods were examined using a dataset. The selected attributes from each feature group were combined with four classifiers—“Adaptive NeuroFuzzy Inference System (ANFIS), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Logistic Regression”—in the second phase of the study. For the processed data set, the suggested model had an accuracy of 74.76%.

D. ASSORTED FEATURES USED IN DETECTION OF COLORECTAL CANCER

Colorectal cancer is ranked third after breast and lung for its highest mortality rate. Many studies are available for the detection of colorectal cancer.

H. Abdul Rahman et.al[8] developed a ML algorithm for the detection of colorectal cancer using the global dietary data. The accuracy of the model has increased by employing suitable feature selection approaches, such as Boruta, Knockoff, and logistic regression (LR), in addition to the author's exclusive consideration of the patient's food intake for the forecast. The task is completed utilizing the proper classification techniques for both supervised and unstructured data sets. All methods worked extremely well in supervised classifiers when the values of kappa, sensitivity, specificity, and accuracy were greater than 0.90.

E. ASSORTED FEATURES USED IN DETECTION OF ORAL CANCER

A multiparametric decision support system was developed by K. P. Exarchos et al. [9] to detect oral cancer recurrence. In this instance, oral squamous cell carcinoma is detected by the utilization of clinical, imaging, and genomic data. The selection of features is a crucial and initial stage in building a machine learning model. Here, the "wrapper algorithm and correlation-based feature subset selection (CFS)" are used to display the features with higher and lower correlation, respectively. The dataset will be shielded from overfitting by the wrapper technique. The dataset's distribution of impacted and non-affected patients will be balanced using the "Synthetic Minority Oversampling Technique". Following appropriate feature selection, the dataset is subjected to various models, including "Random Forests (RFs), Decision Trees (DTs), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Bayesian Networks (BNs)". The results, such as accuracy, specificity, and sensitivity, are compared to see whether algorithm the researcher proposed performs better along with the feature selection algorithm.

F. ASSORTED FEATURES USED IN DETECTION OF LEUKEMIA

Leukemia is a potentially lethal subtype of cancer that affects people of all ages, including adults and children, and is a major global cause of mortality. In past the detection of the abnormal blood cells are done manually and then deep learning algorithm like convolution neural network was used.

As per M. Bukhari et al.'s[10]construction, a novel Deep Learning algorithm variant focuses on selecting suitable features from the data set through a squeeze-and-excitation procedure. The author's model made use of the Acute Lymphoblastic Leukemia Image Database. Data augmentation is used to lessen overfitting because the author only used a small portion of the dataset. The suggested model is assessed and contrasted with various current models, such as "Naive Bayes, decision trees, K-nearest neighbors, and support vector machines", using the confusion matrix and the ROC Curve to show that the author's model is implemented with a 99.50% accuracy rate.

G. ASSORTED FEATURES USED IN DETECTION OF SKIN CANCER

Melanoma is a type of skin cancer that appears on the epidermis of the skin. Early detection of any type of cancer can be cured easily. S. T. Sukanya et.al [11] proposed the algorithm namely "Particle-Assisted Moth Search Algorithm (PA-MSA) that combines the concept of Moth Search Algorithm (MSA) and Particle Swarm Optimization (PSO)" [11] respectively. For the classification process,

- "Optimally chosen features (Fopt)" are fed as input and
- "Deep Convolution Neural Network (DCNN)" is used.

Finally, a performance-based comparison of the suggested PA-MSA and the existing models is done in terms of a variety of measures.

H. ASSORTED FEATURES USED IN DETECTION OF OVARIAN CANCER

Ovarian cancer (OC) is a tumor type that affects women's ovaries and is one of the leading causes of cancer-related fatalities. It is challenging to diagnose in its early stages.

Ahamad, M.M et al.[12] proposed a machine learning model for early detection of ovarian cancer using the clinical data. This study's main goal is to use machine learning models and statistical methodologies to predict outcomes for early diagnosis based on clinical data gathered from 349 distinct patients. Dataset consisted of 49 features for the diagnosis which is processed by means of Statistical Package for the Social Sciences (SPSS) for selecting the significant features. Author uses five-fold cross-validation in addition to “grid search to fine-tune hyperparameters” in our machine learning project. The proposed algorithm performed better than other traditional algorithms.

I. ASSORTED FEATURES USED IN DETECTION OF BRAIN CANCER

Adults are at significant risk for serious health complications because to the fast growth of abnormal brain cells that identify brain tumors, which can result in severe organ dysfunction or even death. MRI image processing is the only way to detect the brain cancer. The only technique to detect brain cancer is to process MRI images. However, manually recognizing brain tumors is a tough and time-consuming task that may result in errors. Image enhancement technologies that add various filters to the original images improve the visual representation of the MRI scans.

Abdusalomov, A. B et al [13] investigated more than 10,000 MRI imaging dataset for analyzing the brain cancer. The authors proposed a refined You “Only Look Once version 7 (YOLOv7) model” [13] for the accurate diagnosis of meningioma, glioma, and pituitary gland tumors in a brain tumor detection system. The authors included the “Convolutional Block Attention Module (CBAM) attention mechanism into YOLOv7” to improve its feature extraction capabilities, allowing for greater emphasis on prominent regions associated with brain cancers. Proposed model YOLOv7 performed better with 99% accuracy compared to other formal ML and DL algorithm.

J. ASSORTED FEATURES USED IN DETECTION OF CERVICAL CANCER

Women are primarily killed by cervical cancer, and effective treatment depends on early detection.

In order to help women diagnose cervical cancer early, Al Mudawi et al. [14] suggested a four-phase study technique that included data pre-processing, pseudo-coding, prediction model selection (PMS), and research dataset collection. In one instance, researchers employed real-time data on the activities, demographics, and medical histories of 858 individuals. The author employed predictive model selection techniques, which compare the outcomes of several machine learning models that are performed against the processed dataset. Four methodologies are used to analyze the results:

- “Empirical Consequence Report (ECP)”,
- “Exploratory Cervical Data Analysis (ECDA)”,
- “Computational Complexity Analysis (CCA)”, and
- “Comparative Analysis, And Survey Data Analysis (SDA)”.

Here, a variety of metrics, including time and space complexity, correlation analysis, accuracy, and sampling methods, are used to compare each algorithm. This inquiry additionally noted that the Microsoft Azure machine learning (ML) approach was used with the DT and RF algorithms to produce an appropriate data mining solution for cervical cancer prediction.

III. COMPARITIVE ANALYSIS OF ASSORTED FEATURES USED IN ML ALGORITHM

TABLE II LIST OF FEATURES ANALYSED BY ML ALGORITHM

S.no	Type of Cancer	ML Algorithm used to detect cancer	Features Used by ML Algorithm.
1	Breast Cancer	• “Ada BoostM1, • Decision Table, • J-Rip, J48, • Lazy IBK, • Lazy K-star, • Logistics Regression, • Multiclass Classifier, • Multilayer–Perceptron, • Naïve Bayes, • Random Forest, and • Random Tree. • Hierarchal clustering random forest Algorithm”[4]	Features selected for the ML for imaging data. • “Thickness of clump • Evenness of cell size • Evenness of cell shape • Margins of the affected cell. • Single Epithelial Cell Size • Features of Nuclei • Bland chromatin • Normal nucleoli • Based on the Radius • Texture • Perimeter

			• Area • Smoothness • Compactness • Concavity • Concave points • Symmetry • Fractural dimension” Features selected for the ML model for clinical data • “Age • Status of menopause • Size of the tumor • Invasive-nodes and lymps • Node-caps • Deg-malig • Left or right breast • Breast-quad • Irradiate”[4]
2	Lung Cancer	• “Convolutional neural network • 2D Faster R-CNN combined with the VGG-16 model • 3D Faster R-CNN with DPNs and encoder-decoder structure • 3D U-Net deep learning model • Support Vector Machine (SVM) classifier • Gradient Boosting Machine (GBM) classifier and • Random Forest (RF) Classifier”	According to G. Zhang et.al[5] the following attributes are considered from the pulmonary nodes of CT imaging. • “Dissection • Position of the tumor • Density • Intensity • Calcification • Transformation in surrounding tissues”. Y. Lu and et.al[6] utilizes the following features from the NLST dataset • “Age • Gender • Exposure to pollution in air • Usage of alcohol • Dust Allergy • Exposure to radiation in their occupation • Genetic Risk, Chronic Lung • Other disease • Person’s diet plan • Obesity • Smoking exposure (Active or passive) • Chest discomfort and ache • Coughing of blood • Weakness • Unusual loss of weight • Shortness of breath • Wheezing • Difficulty in swallowing the food • Clubbing of finger nails • Frequent cold • Dry cough

			• Snoring”
3	Throat Cancer	• “Adaptive Neuro-Fuzzy Inference System (ANFIS), • Artificial Neural Network (ANN) • Support Vector Machine (SVM) and • Logistic Regression”	Siow-Wee Chang et al. [7] used following demographic data for the diagnosis • “Risk factors, • Ethnicity, • Age, • Occupation, • Marital status” Following Clinical data is used along with above demographic features • “Type of lesion, • Size of lesion, • primary site, • Clinical neck node and etc..,” Along with the above attributes pathological data are also used • “Pathological TNM, • Neck node metastasis, • Bone invasion, • Tumor thickness • Bleeding from throat • Lack of hunger (Anorexia) • Loss of weight • Bumps in affected areas • Blockage of nose • Mouth breathing • Hyponasal speech (poor speech) • Halithosis (mouth odour) • Facial Asymmetry, Fatigue, Hoarseness, Dyspnoea (difficult breathing) ,Snoring etc..,”
4	Colorectal Cancer	Algorithm used for Feature selection, • “Logistic regression (LR), • Boruta, and • Knockoff selection” Algorithm used for classification, • “Neural network (Neuralnet), • K-nearest neighbors (knn), • Generalized linear model (GLM), and • Recursive partitioning (Rpart) • Uniform manifold approximation and projection (UMAP), • Apriori association rules, • Principal component analysis (PCA), • Factor analysis (FA)”[8]	Features derived as important contributor of CRC are • “Fiber, • Total fat, • Cholesterol, • Age, • Vitamin E, • Saturated fats, • Monounsaturated fats, • Carbohydrates, and • Vitamin b12”.[8]
5	Oral cancer	Algorithm used for Feature selection, • “Wrapper algorithm and	Clinical Features used in multiparametric decision support system • “Weight

		• Correlation-based Feature subset Selection (CFS)” Algorithm used for class imbalance, • “Synthetic Minority Oversampling Technique”. Algorithm used for classification, • “Bayesian Networks (BNs), • Artificial Neural Networks (ANNs), • Support Vector Machines (SVMs), • Decision Trees (DTs) and • Random Forests (RFs)” [9].	• Height • Diabetes • Allergies • Cholesterol • Family history of malignance • Smoking habit • Drinking habit • Mobile prosthesis • Infection • Eating Habit • BMI • Substance Exposition • Tumor details • P53_Stain • EGFR stain • Staging Details” [9] Imaging Features used in multiparametric decision support system • “Contrast Take Up Rate • Minor Axis Bigger than 10mm • Shape Deviation • Necrosis-details • Lymph node details” [9]
6	Leukemia	• “Convolutional Neural Networks • Squeeze and excitation learning (Feature Selection) • Naive bayes, • Decision tree, • K-nearest neighbor, and • Support vector machines”	• “Shape or size of blood cell • Anatomical characteristics such as, • Shape and edge features. • Texture • Color • GLCM” [10]
7	Skin Cancer	• “Fuzzy C-Means Clustering (FCM) • Particle-Assisted Moth Search Algorithm (PA-MSA) • Particle Swarm Optimization (PSO)”.	• “Local Vector Pattern (LVP), • Local Binary Pattern (LBP), • Local Directional Pattern (LDP) and • Local Tetra Pattern (LTrP)” [11]
8	Ovarian Cancer	• “Random Forest (RF), • Support Vector Machine (SVM), • Decision Tree (DT), • Extreme Gradient Boosting Machine (XGBoost), • Logistic Regression (LR), • Gradient Boosting Machine (GBM) and • Light Gradient Boosting Machine (LGBM)”	Following features are selected from Blood Routine Test • “Neutrophil ratio • Platelet count • Hemoglobin • Mean platelet volume • Red blood cell count etc..,” Following features are selected from general chemistry test • “Albumin • Indirect bilirubin • Uric acid • Total protein • Creatinine

			• Glucose etc..," Following features are selected from Tumor Marker • "Carbohydrate antigen 72-4 • Menopause • Age • Carcinoembryonic antigen • Human epididymic protein 4 "[12] etc..,
9	Brain Cancer	ML algorithms such as • "Only Look Once version 7 (YOLOv7) model[13] • Support vector machines (SVMs), • K-Nearest Neighbor (K-NN), • Decision trees, and • Naive Bayes". DL algorithms, such as • "Convolutional Neural Networks (CNNs), • VGGNets, • GoogleNet, and • ResNets".	Imaging data features • "Tumor location, • Shape, • Size, • Significant pivot length, • Euler number, • Minor hub length, • Robustness, region, and • Circularity"
10	Cervical Cancer	Method for Data preprocessing • "Principal Component analysis (PCA)" Method for Classification • "Support Vector Machine (SVM), • Decision Tree Classifier (DTC), • Random Forest (RF), • Logistic Regression (LR), • Gradient Boosting (GB), • Xgboost, Adaptive Boosting (AB),and • K-Nearest Neighbor (KNN)".	• "Age • Number of sexual partners • First sexual intercourse • Number of pregnancies • Smokes(all details) • Hormonal contraceptives(all details) • IUD(all details) • STDs(all details) • Dx: cancer • Dx: CIN • Dx: HPV"[14]

IV.CONCLUSION

Cancer detection refers to the process of evaluating whether cancer is present based on symptoms and other attributes from clinical and imaging data. This paper offers an overview of the holistic characteristics that various machine learning algorithms utilize to detect and forecast cancer early on. The objective was to conduct a comparative analysis, compile a list of features, and provide an overview of the attributes utilized in various machine learning algorithms for the prediction of various cancer kinds. The various qualities that are primarily taken into account by numerous researchers for the more accurate identification of cancer recurrence will be the focus of this paper's future study.

REFERENCES

[1] https://www.who.int/news-room/fact-sheets/detail/cancer

[2] National Cancer Institute. https://www.cancer.gov.

[3] https://healthinformatics.uic.edu/blog/machine-learning-in-healthcare/

[4] Z. Huang and D. Chen, "A Breast Cancer Diagnosis Method Based on VIM Feature Selection and Hierarchical Clustering

Random Forest Algorithm," in IEEE Access, vol. 10, pp. 3284-3293, 2022, doi: 10.1109/ACCESS.2021.3139595.

[5] G. Zhang, H. Zhang, Y. Yao and Q. Shen, "Attention-Guided Feature Extraction and Multiscale Feature Fusion 3D ResNet for Automated Pulmonary Nodule Detection," in IEEE Access, vol. 10, pp. 61530-61543, 2022, doi: 10.1109/ACCESS.2022.3182104.

[6] Y. Lu, S. Aslani, M. Emberton, D. C. Alexander and J. Jacob, "Deep Learning-Based Long Term Mortality Prediction in the National Lung Screening Trial," in IEEE Access, vol. 10, pp. 34369-34378, 2022, doi: 10.1109/ACCESS.2022.3161954.

[7] Chang SW, Abdul-Kareem S, Merican AF, Zain RB. “Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods.” BMC Bioinformatics. 2013 May 31;14:170. doi: 10.1186/1471-2105-14-170. PMID: 23725313; PMCID: PMC3673908.

[8] Abdul Rahman, H., Ottom, M.A. & Dinov, I.D. Machine learning-based colorectal cancer prediction using global dietary data. BMC Cancer 23, 144 (2023). https://doi.org/10.1186/s12885-023-10587-x

[9] K. P. Exarchos, Y. Goletsis and D. I. Fotiadis, "Multiparametric Decision Support System for the Prediction of Oral Cancer Reoccurrence," in IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 6, pp. 1127-1134, Nov. 2012, doi: 10.1109/TITB.2011.2165076.

[10] Maryam Bukhari, Sadaf Yasmin, Saima Sammad, Ahmed A. Abd El-Latif, "A Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Squeeze and Excitation Learning", Mathematical Problems in Engineering, vol. 2022, Article ID 2801227, 18 pages, 2022. https://doi.org/10.1155/2022/2801227

[11] S. T. Sukanya and S. Jerine,” Deep Learning-Based Melanoma Detection with Optimized Features via Hybrid Algorithm”, International Journal of Image and Graphics,vol. 2022/10/07, https://doi.org/10.1142/S0219467823500560

[12] Ahamad, M.M.; Aktar, S.; Uddin, M.J.; Rahman, T.; Alyami, S.A.; Al-Ashhab, S.; Akhdar, H.F.; Azad, A.; Moni, M.A. Early-Stage Detection of Ovarian Cancer Based on Clinical Data Using Machine Learning Approaches. J. Pers. Med. 2022, 12, 1211. https://doi.org/10.3390/jpm12081211

[13] Abdusalomov AB, Mukhiddinov M, Whangbo TK. Brain Tumor Detection Based on Deep Learning Approaches and Magnetic Resonance Imaging. Cancers (Basel). 2023;15(16):4172. Published 2023 Aug 18. doi:10.3390/cancers15164172

[14] Al Mudawi, N.; Alazeb, A. A Model for Predicting Cervical Cancer Using Machine Learning Algorithms. Sensors 2022, 22, 4132. https://doi.org/10.3390/s22114132

[15] Kumar V., Mishra B.K., Mazzara M., Thanh D.N.H., Verma A. (2020) Prediction of Malignant and Benign Breast Cancer: A Data Mining Approach in Healthcare Applications. In: Borah S., Emilia Balas V., Polkowski Z. (eds) Advances in Data Science and Management. Lecture Notes on Data Engineering and Communications Technologies, vol 37. Springer, Singapore. https://doi.org/10.1007/978-981-15-0978-0_43

[16] B. S. Rao et al., "A Novel Machine Learning Approach of Multi-omics Data Prediction," 2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICSES55317.2022.9914340.

[17] T. Shanmuga PriyaDr. T. Meyyappan(2021) Disease Prediction by Machine Learning Over Big Data Lung Cancer,International Journal of Scientific Research in Computer Science, Engineering and Information Technology,Volume 7, Issue 1 Page Number: 16-24 Publication Issue : January-February-2021

[18] Patra R. (2020) Prediction of Lung Cancer Using Machine Learning Classifier. In: Chaubey N., Parikh S., Amin K. (eds) Computing Science, Communication and Security. COMS2 2020. Communications in Computer and Information Science, vol 1235. Springer, Singapore. https://doi.org/10.1007/978-981-15-6648-6_11

[19] Prof O. Olabode1 et al Classification Of Head And Neck Cancer Types Using Machine Learning Algorithm EPRA International

Journal of Research and Development (IJRD) Volume: 5 | Issue: 4 | April 2020 http://gco.iarc.fr/tomorrow/home

[20] ZandHKK(2015)Acomparative survey on data mining techniques for breast cancer diagnosis and prediction. Indian J Fundam Appl Life Sci 5(2005):4330–4339

[21] Agrawal A, Misra S, Narayanan R, Polepeddi L, Choudhary A (2011) A lung cancer outcome calculator using ensemble data mining on SEER data. BIOKDD 2011, San Diego, CA, USA. ACM, New York, pp 1–9

[22] Khan MT, Qamar S, Massin LF (2012) A prototype of cancer/heart disease prediction model using data mining. Int J Appl Eng Res 7(11):1241–1249

[23] Suji RJ, Rajagopalan DS (2013) An automatic oral cancer classification using data mining techniques. Int J Adv Res Comput Commun Eng 2(10):3759–3765

[24] Abdelaal MMA, Sena HA, Farouq MW, Salem AM (2010) Using data mining for assessing diagnosis of breast cancer. In: Proceedings of the international multiconference on computer science and information technology, Wisla, pp 11–17.

[25] https://www.cancerresearchuk.org/about-cancer/cancersymptoms/why-is-early-diagnosis-important

[26] Kharya S (2012) Using data mining techniques for diagnosis and prognosis of cancer disease. Int J Comput Sci Eng Inf Technol 2(2):55–66

[27] Christopher T, Banu JJ (2016) Study of classification algorithm for lung cancer prediction. Int J Innov Sci Eng Technol 3(2):42–49

[28] Kumar GR, Ramachandra GA, Nagamani K (2013) An efficient prediction of breast cancer data using datamining techniques. IJIET 2(4):139–144

[29] Ada KR (2013) A study of detection of lung cancer using data mining classification techniques. Int J Adv Res Comput Sci Softw Eng 3(3):2277

[30] Thein HTT, Tun KMM (2015) An approach for breast cancer diagnosis classification using neural network. Adv Comput Int J 6(1):1–11

[31] Balachandran K, Anitha R (2010) Supervised learning processing techniques for pre-diagnosis of lung cancer disease. Int J Comput Appl 1(5):17–21

[32] Yeh WC, Chang WW, Chung YY (2009) A new hybrid approach formining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Syst Appl 36(4):8204–8211

[33] Arutchelvan K, Periasamy R (2015) Analysis of cancer detection system using data mining approach. Int J Innov Res Adv Eng 11(2):57–60

[34] Williams K, Idowu P,Balogun J, OluwarantiA(2015) Breast cancer risk prediction using data mining classification techniques. Trans Netw Commun 3(2):1–11