Journal of Harbin Engineering University Vol 44 No. 7
ISSN: 1006-7043 July 2023

Effective Machine Learning Techniques for Cancer Prediction – An Exploratory Study

G. Sridevi
ResearchScholar, Department of Computer Science, School of Computing Sciences Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, India.
srimani1984@gmail.com

R. Shalini
ResearchSupervisor and AssistantProfessor in Department of Information Technology, School of Computing Science, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Pallavaram, Chennai 600117, India. rshalini.scs@velsuniv.ac.in

Abstract-One of the most serious health issues in the human body is cancer which causes expansion of atypical cell. To enable the successive medical organization of patients, it is now essential in cancer research to detect and prognosticate the kind of cancer as quickly as possible. Therefore, Machine learning (ML) approaches is extensively used in cancer research for the construction of prediction models, resulting in efficient and precise decision-making for effectively diagnose cancer. It is imperative for ML tools can identify features in challenging datasets. In addition to ML methods, new tools for image processing have been created to identify cancer. Being able to clearly classify a biopsy image is a difficult undertaking since the pathologist needs to be familiar with the specific characteristics of both normal and damaged cells. This research work discussed utility of these various ML techniques to identify various cancer types has been examined in some prior research articles. Also, this study concludes the findings and a discussion of the analysis with evaluation methods used in chosen publications. The goal of this study was to identify challenges in the field of cancer prediction and determine the most likely future directions for research aimed at filling those gaps.

Keywords: Prediction, Machine Learning, Cancer, Hyper Parameter Tuning

1. INTRODUCTION

When the term "cancer" first entered the medical lexicon in the 1600s, it was used to describe a group of cells that were developing irregularly had a potential to penetrate or migrate to other areas of the body [1]. Cancer metastasis is the process by which cells grow out of control and spread to different regions after beginning at one location in the human body [2, 3]. Malignant and benign cancer cells are two distinct characteristics in the disease. Family history, radiation therapy, hormones related to obesity and other variables are among the causes of cancer. Its extensive and costly therapeutic dosage is caused by high death and recurrence rates. Hence, accurate early diagnosis is necessary to increase the survival rate of cancer patients. Genetic mutations that affect our cells' activity, particularly how the cells evolve, and divide caused by triggering of genetic disease. Additional modifications will take place when the tumor cells expand. In essence, cancer cells contain more genetic alterations, like DNA mutations than healthy cells [4], [5]. In general, abnormal cells are typically excreted from the body by the immune system whereas few of cancer cells can hide it. Some potentially prescient elements can be discovered by looking more closely at the substantial correlation between positive instances and historical data that is clear from medical records. Therefore, it becomes essential to review data-driven and learning-based methodologies.

While lung and gastric cancers have also been covered, [6] particularly focuses a little more on breast cancer it affects most women globally because it is the most common kind of cancer. As demonstrated in Figure 1 and as reported in [6], the breast, lung, and stomach cancers have been emphasised as being among the most common cancers in humans. This study focuses mainly on breast cancer (as a specific form), which is totally treatable if caught early. Globally, lung cancer becoming more prevalent, and smoking has a strong correlation with the disease. Alcohol and smoking both significantly contribute to its etiogenesis. Similarly, it refers to gastric cancer, which is likewise treatable but only if caught early and prevented from spreading [7].

Even though the precise causes for each of the aforementioned malignancies are not yet understood with a high degree of certainty, it is accepted that number of risk factors may be connected to the disease. Any triggering factor that raises the likelihood of developing cancer might be considered a risk factor. These variables offer insightful data that can be organized as crucial prediction elements for a mathematical framework. As a result, it is crucial to talk about cancer risk variables since they can serve as indicators for a model of prediction [8] if they are correctly recognized and processed. Figure 2 depicts a potential framework for building a cancer predictive model and how understanding risk variables is also crucial. Obviously, various malignancies have unique risk factors attached to them. Certain risk factors may be connected to an individual's routines or behaviors (such as smoking, drinking, etc.) and food habits, and these can be changed. There are some risky variables that cannot be modified, such as a person's age, race, or family history. It's critical to recognize the many stages during which the cancer grows and spreads if we are to comprehend a number of risk factors specifically related with breast cancer. When discussing the risk factors, understanding of this issue becomes crucial.

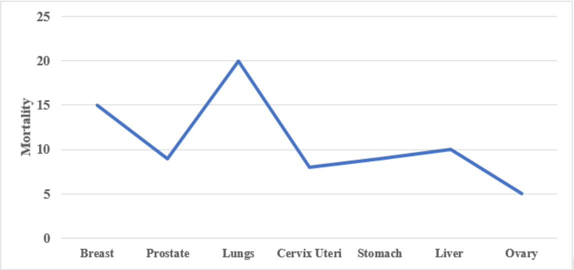

Figure 1. Mortality rate of most common kinds of cancer

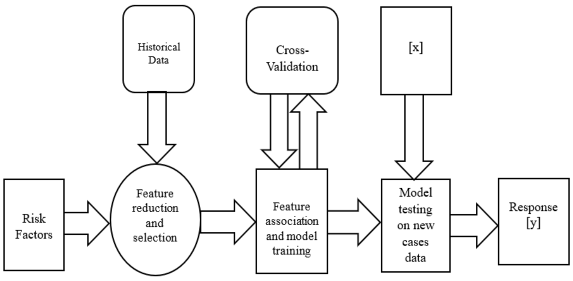

Figure 2. A General Scheme for Developing a Cancer Prediction Model

Numerous elements that impact how quickly cancer spreads are depicted in Figure 3. The top four global cancer risk factors are poor diet, tobacco, alcohol and a lack of physical activity whereas some persistent infections pose a significant risk of developing cancer and are very important in low- and middle-income nations.

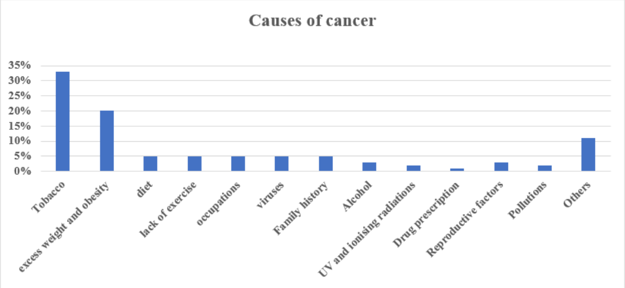

Figure 3. Reasons of cancers

The successful application of ML approaches has lately shown important advancements in image-based disease diagnosis and detection in cancer and oncology research. Utilising phenotypic and genomic data, ML models might also be used for cancer detection, classification, and treatment. This review paper focuses on ML-based strategies for cancer research.

1.1 MOTIVATION

The significant increase in cancer incidence and mortality cases around the world is what motivated this research [9]. There are numerous and complicated causes for cancer, but these are the result of demographic changes, including ageing and population increase, as well as shifts in the incidence and distribution. There are numerous studies in the field of cancer research. Rong et al. [10] examined mortality and survival rates according to gender identity. Dolatkhah et al. [11] work has made it possible to examine the pattern of breast cancer expansion in Iran and learn more about its persistence. Goodarzi et al. [12] introduced the assessment based on distinctive cross-sectional malignant growth investigations. Azamjah et al. [13] was to ascertain the death rate for breast cancer over a 25-year period in seven super regions as Identified by the Health Metrics and Evaluation (IHME). According to a study by Momenimovahed et al. [14], developed nations have a higher incidence of breast cancer than developing nations. With the help of the Cancer Prevalence in Five Continents, Zhang et al. [15] conducted an inquiry to acquire data on Circulating Tumor Cells (CTC) frequency. Incidence and country-specific socio-economic development were found to be positively correlated, according to Wong et al. [16]. According to Lee et al. [17] presents 800 individuals with a diagnosis of malignancy and symptomatic COVID-19 were examined. A modest COVID-19 illness history was present in 412 (52%) patients. The mortality rate was highly correlated with patient ageing, with 226 (28%) individuals dying. According to Shaikh and Rao [18], different ML algorithms are used to predict different types of diseases. These algorithms include Support Vector Machine (SVM), Adaboost, K-Nearest Neighbourhood (KNN), Transfer Learning, Logistic Regression (LR), Fuzzy Logic, Ensemble Learning, Transduction Learning, Naive Bayes (NB) and Neural Network (NN) which are mostly used in various contributions. Therefore, the purpose of the paper is to emphasize the progress provided by scholars in the domain of ML approaches for the timely diagnosis and identification of cancer.

2. LITERATURE REVIEW

Makaju et al.'s [19] presented the evaluation of the various computer-aided procedures includes examining the most effective technique currently in use, identifying its shortcomings, and then suggesting a new model that incorporates those improvements. The strategy involved classifying and listing lung cancer detection methods based on accuracy of detecting the disease. Each stage of the procedures is examined, and overall shortcomings and limitations were highlighted. Some had low accuracy, while others had higher accuracy but none comes close to 100%. Subsequently, the precision of our research is to be increased to 100%.Win et al. [20] created a computer-aided decision method to identify cancer cells, using photographs of cytological pleural effusions. The picture quality was initially improved with the use of intensity correction and median filtering. Based on K-means clustering and basic linear iterative clustering, they employed a hybrid segmentation technique to isolate cell nuclei. A K-means clustering technique uses the Euclidean distance between each data point and the nearest centroid to calculate the error of every point of data.

According to Dhiliphan Rajkumar et al. [21], a tumor can be found on a CT liver picture. Liver is initially separated from CT scans of the abdomen. The normal and pathological liver structures are then categorized using SVM. The tumor will be separated from the liver's structural elements if it is aberrant. This method produces good results based on sensitivity, accuracy, and specificity in its computations. Significant progress is being made with this method. According to experimental findings, the proposed work obtains 98.2% total accuracy, 98.2% of sensitivity and 98.4% of specificity. According to Nanglia et al. [22] discussed for classification based on three-block mechanism, initially pre-processing the dataset, secondly feature extraction using Speeded Up Robust Features (SURF) technique, thirdly optimization using Genetic Algorithm (GA) and finally classification using Feed Forward Back Propagation Neural Network (FFBPNN). Kernel Attribute Selected Classifier is a hybrid classification algorithm with an average classification accuracy is 98.08%. This study mainly focusses to improve accuracy through the use of hybrid classification method.

Hsu et al. [23], proposes an efficient computer-aided diagnosis method assisted by intelligence learning approach. To enhance prediction performance, feature modelling based on ML is suggested. In order to carry out this experimental study, datasets on lung cancer, breast, Irvine repository and cervical were accessed from California University. The ideal features minimized by the suggested system are verified and developed using supervised learning algorithms. According to the study, accuracy was 99.62% for breast cancer, 96.88% for cervical cancer, and 98.21% for lung cancer based on training and performance using 10-Fold cross-validation method.Maleki et al. [24] developed a KNN technique for detecting the stage of a patient's disease and make the feature selection to be effective it uses the GA to minimize the dimensions of the dataset and speed up the classifier. The ideal value for k is chosen through an experimental process to increase the suggested algorithm's accuracy. When used on a lung cancer database, the suggested method produces results with 100% accuracy. This suggests that one might effectively help lung cancer staging detection using the algorithm to establish an association between the clinical data and data mining techniques.Ye et al. [25] presents feature of miRNAs used as biomarkers to diagnose Lung Squamous Cell Carcinoma (LSCC). To find variations between the microRNA expression of the para-carcinoma tissues and core tumors tissues from LUSC, this research work obtained clinical data and microRNA expression data from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. SVM and PCA techniques were used to enable the identification of 21 feature miRNAs between para-carcinoma tissues and lung SCC tumor samples. Due to considerable differences in expression between healthy and tumor tissues, the analysis revealed for particular collection of miRNAs was significantly associated with cancer patients' survival rate.

Yuan et al. [26] discusses various ML techniques to examine the gene expression profiles of LSCCand lung Adeno Carcinoma (AC) samples obtained from Gene Expression Omnibus. An effective feature selection technique, termed as Monte Carlo technique, was used to first analyze the profiles. A feature list was created, ranking each feature in terms of relevance, and including some informative aspects. The best features were then extracted from the feature list using the incremental feature selection approach, then classify lung AC and lung SCC samples using SVM classifier with the highest level of accuracy. Shanthi and Rajkumar [27] introduced an improved Stochastic Diffusion Search (SDS) algorithm based on unique wrapper-based feature selection approach. Direct agent contact will help the SDS to determine the best feature subsets. To classify data, DT, NN and NB have all been employed. The experiment's findings demonstrate that the proposed method can perform at higher levels of quality than other feature selection approaches.

Iyswarya Lakshmi and Bharatha Devi [28] addresses the Novel Multi-Layer Perceptron (NMLP)in comparison to SVM using breast cancer dataset. With data from the github library, accuracy and loss calculations are made. The outcomes demonstrated that compared to the SVM's accuracy of 86.45% whereas the NMLP had a greater accuracy of 90.45%. The NMLP does not seem to match the SVM in terms of performance. It was discovered that there was a statistically significant difference of 0.194 ($p < 0.05$) between the SVM method and Novel Multi-Layer Perceptron. Mohammed Ebrahim et al. [29] offer a comparative analysis of ML methods to predict breast cancer. The 1.7 million data records in the dataset used in this investigation were received from the National Cancer Institute (NCH) in the USA. The impact of feature selection on accuracy was also studied. According to the data, ensemble procedures and DT performed better than other methods with accuracy of 98.7%. According to the study, Chhaya Gupta and Kirti Sharma [30] examined the Wisconsin Breast Cancer Dataset to categories breast cancer based on a various of ML methods. The primary goal is to compare the performance of various strategies using performance measures. According to experimental findings, LRachieves accuracy rates of 97.89% whereas sensitivity of 100%, F1 score of 97.35% and specificity of 90.69% respectively.

Shweta et al. [31] provides an overview of contemporary ML techniques used in cancer progression modelling. Different ML algorithms, including NB, SVM, KNN, Fuzzy logic, NN, Transduction learning, clustering, Random Forest, transfer learning, Adaboost, ensemble learning, and LR are mostly used in various contributions to predict the various types of diseases. According to the above discussion, having a technology that may help with early cancer detection and prevention would be excellent. This may improve the chances of survival for cancer patients. The hyperparameter tuning model increased accuracy from 94.15% to 98.83%, while the traditional approach improved from 93.56% to 97.08%.

Mohammed Mijanur Rahman et al. [32] analyze the efficacy of five widely recognized ML classifiers. The main strategy we used to counteract the influence of conventional methods was hyperparameter tuning using the grid search method, which enhanced the accuracy, recall, AUC & ROC curve, precision, and F1-score. ysis found that KNN performed better than any other classifier with an accuracy of 98.83%.Abhishek Gupta et al. [33] explains the image classification and ML methods were used on a lung cancer illness dataset to determine parameters like sensitivity, accuracy and so on. Applying Random Forest, K-NN and SVM algorithms is allowed for the analysis of early-stage of lung cancer using lung cancer dataset. SVM is the next most accurate, at 82.1%, followed by Random Forest, which delivers the best results. Table 1 illustrates the above-mentioned research work based on various ML techniques for cancer prediction has been discussed below.

Authors	Methodology	Challenges	Outcomes
Makaju et al. [19]	Watershed algorithm with SVM	Cancer stages are not classified.	92% accuracy
Win. et al. [20]	Bagging Ensemble	The system generates false-negative results, due to its inability to recognise flaws in Pap smear images	Accuracy of 98.27%
Dhiliphan Rajkumar et al [21]	SVM	The manual interaction reduces the accuracy.	98.2% overall accuracy, 98.4% specificity, and 98.2% sensitivity
Nanglia et al [22]	SVM and NN	CT samples of lung cancer only done	Accuracy of 98.08%
Hsu et al [23]	MLP -NN	It does not supportcomplex and highdimensional, real-timedatasets.	98% ofaccuracy
Maleki et al [24]	KNN and GA	It was improved less by thegenetic-based classification	99% ofaccuracy
Ye et al. [25]	SVM andPCA	Only use SVM.	98% ofaccuracy
Yuan et al [26]	SVM	Only use the same features.	96% ofaccuracy
Shanthi et al [27]	NN-based stochastic diffusion search algorithm	Less improved accuracy	Accuracy of 89%
Iyswarya Lakshmi and Bharatha Devi [28]	NMLP and SVM	the shopping cart algorithm hashighest accuracy for a particular dataset.	accuracy of 90.45%
Mohamed Ebrahim et al. [29]	SVM, Linear Discriminant (LD), ensemble methods, LR, and traditional DT algorithms	Feature selection process need to be improved	98.4% accuracy
Chhaya Gupta and Kirti Sharma [30]	Ridge Classifier, SVM, Light Gradient Boosting Machine, KNN, Extra Trees Classifier, RF Classifier, NB, DT Classifier, Gradient Boosting Classifier, Ada Boost Classifier, andLR.	Only smaller dataset is used	accuracy of 97.89%
Shweta et al. [31]	ML	smaller number of training datasets are used	ovarian cancer prediction K- means accuracy of 97%, liver cancer prediction Fuzzy NN accuracy of 95.45%, colon cancer prediction Fuzzy Granular SVM

			accuracy of 99.71%,Breast cancer prediction RF accuracy of 99.24%.
Mohammed Mijanur Rahman et al. [32]	ML classifiers such as KNN, LR, NB with hyper parameter tuning, DT and RF algorithm	Smaller size dataset is used	The accuracy of KNN is 98.83%, LR is 97.08%, DT is 94.15%, NB is 95.91% and RF is 97.08%.
Abhishek Gupta et al. [33]	Algorithmic techniques SVM, K-NN, RF	Improved prototype software is not used	Accuracy of RF and SVM 84.2%, 82.1% .

3. RESEARCH GAP

Lung cancer ranks second globally in terms of prevalence. When it comes to cancer-related mortality in humans. Patients having lung cancer is an average 5-year survival rate of no more than 14%, on the other hand it is found in other organs such as cervix, bladder, breast, colon, and prostate which is much lower than the overall cancer survival rate [34]. Therefore, timely detection of lung cancer is crucial for choosing the right treatments and reducing mortality. Healthcare is one of the key suppliers of big data. The primary need for accurate healthcare data analysis is the early detection of lung cancer. To detect lung cancer using big data, numerous studies are now being designed. Therefore, a method of classification is required to increase the detection accuracy regarding time. Additionally, models for ML approaches have been created to improve detection accuracy. Lung cancer especially is not much recognized, so it is necessary to employ methodologies which will produce data with a high detection rate and data traits will be useful for the diagnosis process.

There are two well-known data mining techniques such as Artificial Neural Network (ANN) and DT, have been used with the aid of large datasets to introduce techniques for predicting breast cancer survivability. These approaches utilize a standard statistical methodology. The performance comparison was employed using 10-fold cross validation procedures to measure the evaluation of three detection models. The results showed that DT was the most accurate classifier to diagnose the disease, with a holdout sample of 93.6% accuracy rate whereas ANN was ranked second with a holdout sample accuracy of 91.2%. Like this, LR has 89.2% accuracy. In [35], research was carried out to create detection methods for determining the prognosis of prostate cancer, employing SVM in addition to the three previously described methods. Here, the results showed that, in addition to ANN and DT, the singled-out SVM also reached excellent accuracy. Additionally, ANNs, DTs, and LR techniques were used in [36] to assess prostate cancer survivorship. The SEER colon cancer dataset was used to compare several approaches for predicting survival rate in [37], and it was shown that NNs were the most accurate. Using three outperforming classifiers in ensemble voting from [38] produced the best prediction with the survival rate of AU-ROC curve for colon cancer. In certain studies, the SEER database was analysed using ML methods, such as unsupervised approaches, SVM, clustering based techniques and LR, to analyse the survival of lung cancer [39][40][41]. Data categorization techniques were evaluated in [42] for identifying individuals who had clear signs of lung cancer growth. In [43], the effectiveness of NB and DT classifiers was contrasted and applied to lung cancer data taken out of the SEER dataset. This achieved a patient survival detection accuracy of over 90%. The decision to use meta-classifiers and ensemble voting of five DTs from [44] was made for the purpose of better lung cancer survival rate prediction based on AU-ROC curve and precision.

Manual training has many drawbacks that are connected to ML methods. The difficulty of accurately recognizing type for pre-processing data appropriately before subjecting them to ML algorithms is more essential. Both the time commitment and the level of expertise were required very high. The research revealed that ML techniques do not consistently outperform traditional prediction techniques in terms of detection accuracy. This assertion is supported by the existing literature. It has been established through numerous studies that the results are different for ML models and traditional statistical models.

The prediction models employing the ML algorithms published in the actual works are less effective for detecting lung cancer with IoT integration, even though several methodologies were used to forecast various sorts of diseases. More effective hyper parameter tuning models could be implemented in the future with the potential to achieve the best prediction results. Furthermore, the optimisation methods are capable of modifying the ML models. There are a few drawbacks, including the difficulty in choosing parameters and the inability to identify the ideal solution to the elevated problem. The advantage of Particle Swarm Optimization (PSO) has been able to resolve challenging of optimisation problems. The convergence idea, however, is not relevant whereas Sequential Minimal Optimization (SMO) has some advantages, including the ability to help with quadratic issues that arise during SVM training and the ability to use less memory. However, it needs to be enhanced by adding a new variety. The success of ML in prediction research for particularly lung cancer which enabled by novel met-heuristic algorithms can be attributed to its capacity to tackle composite tasks based on dynamic environment and knowledge.

4. DISCUSSION

This research study discussed about various ML technique based on predicting cancer. In the past ten years, there have been numerous ML methods developed that produce reliable results when predicting cancer accurately. For distinguishing between clinical findings, it is also crucial to spot potential problems with study design, collect adequate data samples and validate classification outcomes. Furthermore, only few ML classification approaches have reached clinical practise, despite promises that they have helped with appropriate and effective decision-making. The use of gene expression signatures in hospitals won't begin until validation results are reliable, despite recent breakthroughs in omits technology that have helped us comprehend a wide spectrum of diseases even better. This study discussed about a significant limitation in terms of data samples. A fundamental criterion for using classification systems to represent a disease based on size of the required training data sets must be sufficient. To reasonably verify the calculators, a sizable dataset must be separated into test and training sets.

Depending on the training sample size and estimator, misclassifications may occur in contrast to the data's dimension or methodologies might be unstable and imperfect. It's obvious that a more affluent set of patients might anticipate their survival, which may increase the diagnosing method capability. However, a variety of factors, including the types of data collected, sample size, the time constraints, and the nature of the prediction findings, must be taken into consideration while choosing the appropriate method. To overcome the constraints mentioned above, novel techniques should be investigated for cancer modelling in future. Therefore, the goal of this research is to execute optimizations technique with respect to its hyperparameters, these focusses bridging the gap between ensemble learning and hyperparameter optimisation. Bayesian hyperparameter optimization may be used directly optimise a group of classifiers. This methods for expensive black-box functions, for which any new assessment necessitates with lot of computer resources, are the main application for Bayesian optimization. A branch of optimization that takes its inspiration from the Bayesian inference methodology. The Sequential Model-Based Optimization (SMBO) algorithm class that Bayesian optimization belongs to enables one to use the outcomes of a prior iteration to enhance the sampling strategy for a subsequent trial. The function of Bayesian optimization is to enhance the subsequent experiment based on the previous iteration results which belongs to Sequential Model-Based Optimization (SMBO) algorithm. Therefore, this research work concludes the efficient function evaluation is required to achieve the optimization using hyperparameter tuning based ensemble model.

5. CONCLUSION

This paper describes and evaluates the cancer prediction based on the outcomes of different ML and in-depth learning. Several trends have been found in the types of ML techniques to be applied, the types of training data that must be utilized, the various malignancies being studied, the numerous endpoint projections to be created, overall effectiveness of cancer prediction and various outcome techniques are presented. It is evident that at least three different cancer kinds are predicted using a wider range of alternative learning techniques. It is also clear that ML approaches frequently increase the efficacy or expected accuracy of several prognostics, particularly when paired with conventional statistical or expert-based systems. Even if most of the research are well-planned and validated, more focus should be placed on how experiments are planned and carried out, particularly in terms of the quantity and quality of biological data. The results of this analysis reveal a combination of base learners' weighted predictions with the help of ML based ensemble hyperparameter model proposed that it is performing better than individual base learners, leading to superior ensemble predictions. While introducing an upgrade over the successful weighted average ensemble model, ensemble hyperparameter with ML model might achieve greater prediction accuracy among all when compared to the state-of-the-art ensemble methods. This supports the idea that improving prediction accuracy will be possible by modifying base learner using hyperparameter tuning with Bayesian optimization throughout the best ensemble creation approach. Since we applied the approaches to three other cancer-based datasets that address the same session and illustrate the adaptability of the created technique on numerous data sets, these findings highlight the ensemble hyperparameter with optimization model's applicability to real data sets.

REFERENCES

[1] Ramadan S (2020) Methods used in computer aided diagnosis for breast cancer detection using mammograms. J Healthc Eng 2020:1–21. https://doi.org/10.1155/2020/9162464

[2] Denil M, Bazzani L, Larochelle H, Freitas N (2012) Learning where to attend with deep architectures for image tracking. Neural Comput 2151–2184

[3] Win Y, Choomchuay S, Hamamoto K, Raveesunthornkiat M, Rangsirattanakul L, Poongsawat S (2018) Computer aided diagnosis system for detection of cancer cells on cytological pleural effusion images. Hindawi 2018:1–22. https://doi.org/10.1155/ 2018/6456724

[4] Milne A, Carneiro F, O'Morain C, Oferhaus G (2009) Nature meets nurture: molecular genetics of gastric cancer. Hum Genet 126:615–628

[5] Mehrotra R, Gupta D (2011) Exciting new advances in oral cancer diagnosis: avenues to early detection. Head Neck Oncol 3:1–9

[6] William R. Brown and Dennis J. Ahnen. The international health care burden of cancers of the gastrointestinal tract and liver. Cancer Research Frontiers, 1(1):1–9, February 2015.

[7] XiaohongR.Yang, Mark E.Sherman, and David L.Rimm. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prevention AACrJournals, 16(349), March 2007.

[8] Riccardo Bellazzi and Blaz Zupan. Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of Medical Informatics, 77(2):81–97, February 2008.

[9] Andriole G, Kramer B, Berg C (2009) Mortality results from a randomized prostate cancer screening trail. N Engl J Med 360:1310–1319

[10] Rong F, Gong W, Pan J, Wang W (2019) Analysis of mortality and survival rate of liver cancer in Zhejiang. Province in China: A general population-based study. Hindawi, 1–7

[11] Dolatkhah R et al (2020) Breast cancer survival and incidence: 10 Years cancer registry data in the northwest, Iran. Int J Breast Cancer, 1–6

[12] Goodarzi E, Moslem A, et al. (2021) Epidemiology, incidence and mortality of thyroid cancer and their relationship with the human development index in the world: an ecology study in 2018

[13] Azamjah N, Zadeh Y, Zayeri F (2018) Global trend of breast cancer mortality rate: A 25-year study. Global Trend Breast Cancer Mortal 20:1–6

[14] Momenimovahed Z, Salehiniya H (2019) Epidemiological characteristics of and risk factors for breast cancer in the world. Breast Cancer and Therapy, 151–164

[15] Zhang Y, Li M, Gao X, Chen Y, Liu T (2019) Nanotechnology in cancer diagnosis: progress, challenges and opportunities. J Hematol Oncol 12:1–13. https://doi.org/10.1186/s13045-019-0833-3

[16] Wong M, Goggins W, Fung F et al (2017) Incidence and mortality of kidney cancer: temporal patterns and global trends in 39 countries. Sci Rep 7:1–10

[17] Lee YM, Agelis V et al. (2020) COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: a prospective cohort study. Crossmark, 1–9

[18] F.J. Shaikh, D.S. Rao (2022) Prediction of Cancer Disease using Machine learning Approach,Materials Today: Proceedings vol-50, pp-40–47, https://doi.org/10.1016/j.matpr.2021.03.625

[19] Makaju S, Prasad PW, Alsadoon A, Singh AK, Elchouemi A. Lung cancer detection using CT scan images. Procedia Computer Science. 2018 Jan 1; 125:107- 14.

[20] Win KP, Kitjaidure Y, Hamamoto K, Aung TM (2020) Computer-assisted screening for cervical cancer using digital image processing of pap smear images. Appl Sci 10(5):1–22

[21] Dhiliphan Rajkumar, D. Deepa, J. Jeyaranjani (2019), Automatic Diagnosis of Liver Tumor in Ct Images, ISSN: 2249 – 8958, Volume-9 Issue-1S4, DOI:10.35940/ijeat. A1116.1291S419

[22] Nanglia, P., Kumar, S., Mahajan, A. N., Singh, P., &Rathee, D. (2021). A hybrid algorithm for lung cancer classification using SVM and Neural Networks. ICT Express, 7(3), 335-341. https://orcid.org/10.1016/j.icte.2020.06.007.

[23] Hsu, C. H., Chen, X., Lin, W., Jiang, C., Zhang, Y., Hao, Z., & Chung, Y. C. (2021). Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning. Measurement, 175, 109145. https://orcid.org/10.1016/j.measurement.2021.109145.

[24] Maleki, N., Zeinali, Y., &Niaki, S. T. A. (2021). A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications, 164, 113981. https://orcid.org/10.1016/j.eswa.2020.113981.

[25] Ye, Z., Sun, B., & Xiao, Z. (2020). Machine learning identifies 10 feature miRNAs for lung squamous cell carcinoma. Gene, 749, 144669. https://orcid.org/10.1016/j.gene.2020.144669.

[26] Yuan, F., Lu, L., & Zou, Q. (2020). Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1866(8), 165822. https://orcid.org/10.1016/j.bbadis.2020.165822.

[27] Shanthi, S., & Rajkumar, N. (2021). Lung cancer prediction using stochastic diffusion search (SDS) based feature selection and machine learning methods. Neural Processing Letters, 53, 2617-2630. 10.1007/s11063-020-10192-0.

[28] Iyswarya Lakshmi and Bharatha Devi (2023). Prediction Of Breast Cancer Using Novel Multi-Layer Perceptron In Comparison With Support Vector Machine To Improve Accuracy,Eur. Chem. Bull. 12 (S1), 4680 – 4685.

[29] Ebrahim, M.; Sedky, A.A.H.; Mesbah, S. Accuracy Assessment of Machine Learning Algorithms Used to Predict Breast Cancer. Data 2023, 8, 35. https://doi.org/10.3390/data8020035.

[30] Chhaya Gupta, Kirti Sharma (2022), Early Breast Cancer Detection using Various Machine Learning Techniques, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 11 Issue 06, June-2022, http://www.ijert.org.

[31] Shweta, Riya, Abhay Kumar (2022),Cancer Prediction Using Machine Learning Algorithm,International Journal of Science and Research (IJSR), Volume 11 Issue 5,ISSN: 2319-7064.

[32] Rahman, M.M., Rahman, A., Akter, S. and Pinky, S.A. (2023) Hyperparameter Tuning Based Machine Learning Classifier for Breast Cancer Prediction. Journal of Computer and Communications, 11, 149-165. https://doi.org/10.4236/jcc.2023.114007.

[33] Abhishek Gupta,ZuhaZuha,Israr Ahmad and Zeeshan Ansari (2022), A Study On Prediction Of Lung Cancer Using Machine Learning Algorithms,DOI: https://doi.org/10.21203/rs.3.rs-1912967/v1.

[34] Paweł Krawczyk, Tomasz Kucharczyk, Kamila Wojas-Krawczyk, Screening of Gene Mutations in Lung Cancer for Qualification to Molecularly Targeted Therapies, INTECH Open Access Publisher, 2012.

[35] E. Adetiba, O.O. Olugbara, Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features, Sci. World J. (2015).

[36] S.S. Alahmari, D. Cherezov, D.B. Goldgof, L.O. Hall, R.J. Gillies, M.B. Schabath, Delta radiomics improves pulmonary nodule malignancy prediction in