ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4

Journal of Coastal Life Medicine

A Statistical Modelling to Detect Carcinoma Cancer in Its Incipient Stages in Healthcare

Received: 24 October 2022, **Revised:** 18 November 2022, **Accepted:** 25 December 2022

Anthati Himavarnika
PG Scholar(M.Tech), Department of C.S.E., Annamacharya Institute of Technology and Sciences, Kadapa, India.
Email: lakshmihima5111@gmail.com

Patturi Prasanthi
Assistant Professor, Department of C.S.E, Annamacharya Institute of Technology and Sciences, Kadapa, India.
Email: patturiprasanthi1@gmail.com

Keywords
healthcare, cancer diagnosis, parametric analysis, carcinoma, deep learning

Abstract
Cancer is a public health problem on a global scale due to its high fatality rate and general complexity. The advancement of cancer prediction based on gene expression has been hastened by the rapid development of modern high-throughput sequencing methods and different machine-learning algorithms, offering insights into effective and precise treatment decision-making. Therefore, it is crucial to create Machine Learning (ML) algorithms that can tell cancer patients apart from healthy individuals. No one classification method has emerged as particularly successful, despite the widespread use of classification methods for cancer prediction. Using a multi-machine learning model optimization strategy, this research demonstrates how Deep Learning (DL) can be utilized to increase the accuracy of the models. We have chosen potential informative genes using statistical analysis, and we have been training five different classification models with these genes. The data from the five distinct classifiers is then "ensembled" using a deep learning technique. The great majority of cases are lung, stomach, and breast adenocarcinomas. Due to this, we applied deep learning-based methods to test the suggested inter-ensembles model using data from the cancer field. According to the research findings, using more than one set of classifiers or the conventional consensus approach improves the accuracy of cancer prognosis. The suggested deep learning-based inter-ensemble technique has been demonstrated to be accurate and effective for cancer diagnosis employing a wide range of classifiers.

1. Introduction

Cancer is distinguished by uncontrollable cell growth and metastasis. According to the GLOBOCAN study, 18.6% of cancer deaths will occur in 2021, totaling 18.5 million (excluding skin cancers other than melanoma). Early detection and diagnosis are critical for effective treatment because cancer is the leading cause of death and suffering. Cancer research has grown steadily in recent decades. Cancer is predicted using gene expression levels. Gene expression data analysis improves cancer detection and treatment. Doctors need better cancer prognosis methods [1]. Because of the rise in computer-aided procedures, ML methods have been applied to cancer detection, with researchers constantly investigating novel prediction algorithms. Researchers used Egypt's National Cancer Registry Program data to compare SVMs, kNNs, and Naive Bayes for feature selection and classification. Polynomial kernel SVMs outperformed kNN and N.B.s in classification accuracy. SVMs and random forests were investigated for cancer detection. SVMs beat random forests (R.F.s) in nine data sets and were equal in three others. The entire gene set was responsible for these results [2].

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

The results of gene selection were comparable. According to the extensive cancer prediction literature, all machine learning techniques have limitations and may fail during categorization. When solving the over-fitting problem of decision trees, SVMs have difficulty selecting a kernel function, and R.F.s may favor the group with more samples. A strategy that takes advantage of each machine learning approach's strengths and weaknesses should improve performance. Combining models has been studied to improve forecast accuracy. Bagging averages the outcomes of decision trees generated from randomly selected sections of the training samples to conclude. After each training cycle, boosting allows weighted votes to combine classification outputs based on the relevance of each training sample. Bagging and Boosting would use linear regression to connect neural network outputs and classify cancer using microarray data. They combined four classifier results from three conventional cancer data sets using the majority vote method. In Stacking and majority voting, various ML methods are used. Although it is too simple to reveal detailed information, majority voting is the most commonly used strategy in classification issues for combining classifiers [3].

Because stacking incorporates learning into the combining process, it is a more successful ensemble technique. Despite a lack of biological research, deep learning has evolved into a formidable learning technology with numerous advantages. DL can "learn" the complex structures of large data sets, including nonlinear systems, without interacting with humans, in contrast to popular voting, which only evaluates classifier concatenation [4]. To explain these strange correlations, we use deep learning in the stacking-based evolutionary algorithms of many classifiers. Deep neural networks combine five cancer prediction models: kNN, SVM, DT, R.F., GBDT, and tumor states. To avoid overfitting, we select informative genes using gene expression differential analysis. The selected genes are then fed into the five classifiers.

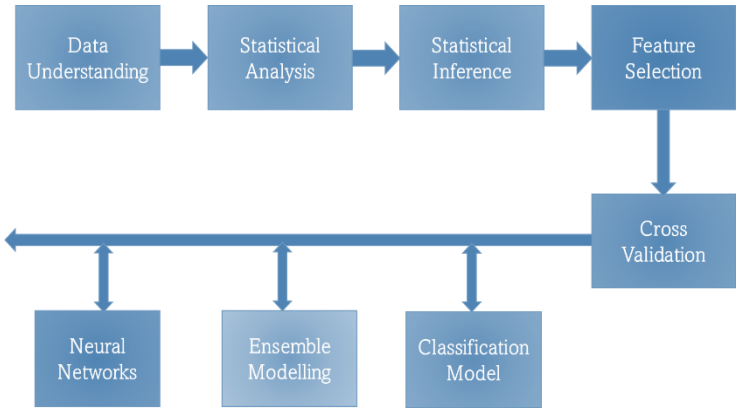

Figure 1. Functioning Research Process Structure

A deep neural network uses the outputs of five categorization models to make predictions. We tested the proposed method using publicly available breast, stomach, and lung data. The findings show that the deep learning-based multi-model clustering approach outperforms classification models and the majority voting method and better uses the limited clinical data. Figure 1 depicts the ensemble technique based on statistics [5] and deep learning. The most informative features are significant differences in gene expression, which are chosen and provided in the classification stage following differential expression analysis. The raw data is divided into S sets for training and testing using S-fold cross-validation. The data classification model is then built using several classifiers trained using training sets containing S 1 of the S groups. The remaining S group was included in the matching test set. A deep neural network classifier aggregates the first-stage predictions to reduce generalization errors and improve accuracy [6].

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

2. Related Works

Several studies have examined breast cancer detection strategies using imaging and genetics. Furthermore, no studies have been conducted that use both approaches together.

The authors in [7] summarized the various histological image analysis approaches used to diagnose breast cancer. Convolutional neural network (CNN)designs are used in these methods. The authors classified their research by dataset. According to the findings of this study, HIA first used ANNs in 2012. The majority of the algorithms were ANNs and P.N.N.s. Textural and morphological features were the most commonly used in feature extraction. Deep Convolutional Neural Networks effectively detect and treat breast cancer early, improving treatment outcomes—forecasting noncommunicable diseases using a variety of algorithms.

Many classification strategies were analyzed and compared for effectiveness in [8]. Each of the eight separate N.C.D. datasets was run through a classification procedure using a 10-fold cross-validation method. The area under the curve was used to analyze these results for precision. The authors claim that the N.C.D. datasets are unreliable because they contain irrelevant attributes and noisy data. It was found that K.N.N., SVM, and N.N. were all robust in the face of this noise. They also mentioned that several pre-processing processes could help with the irrelevant attribute problem, leading to a higher percentage of correctness.

There have been many proposals, and implementations of natural inspiration computing (N.I.C.) approaches for diagnosing various human illnesses. In [9], the authors introduced five N.I.C. diagnostic algorithms that use insects and addressed their potential use in diagnosing diseases, including diabetes and cancer. The authors claim it successfully identified multiple tumors (breast, lung, prostate, and ovarian). Diagnostic accuracy for breast cancer was improved by integrating directed A.B.C. with neural networks. Furthermore, the authors developed a highly effective strategy for identifying diabetes and leukemia. They concluded that more accurate and encouraging results are produced when N.I.C.s are used in tandem with other categorization strategies. They stressed the importance of further research into diabetes and illness detection at different stages.

In [10], the authors provided data suggesting that N.N.s can help classify cancer diagnoses, especially in the early stages of the disease. Their findings show that certain N.N.s have shown promise in detecting cancerous cells. Unfortunately, the imaging method necessitates a large amount of computing power to pre-process the images.

A recent review study discussed several machine learning, deep learning, and data mining approaches related to breast cancer prediction [11]. This analysis of breast cancer research papers covered 27 machine learning publications, four articles addressing related issues, and eight convolutional neural network publications. The scientists found that while many papers made use of imaging, just a minority made use of genetics. Several algorithms were used, but the support vector machine (SVM), decision tree, and random forest were the most prevalent in analyzing breast cancer genetics. Contrarily, imaging methods have used many different kinds of algorithms, including convolutional neural networks and Naive Bayes.

Nevertheless, in contrast, the study authors [12] focused on gene mutation as a means of a breast cancer diagnosis. They indicated that gene annotation, gene discovery, and gene mutation detection would be performed as part of the reverse genetics classification phase to establish the presence or absence of malignancy. Several methods were identified as potential solutions, including regression, probabilistic models, SVMs, neural networks, and deep learning. They also discussed available methods for capturing the link between nucleotides and feature extraction. This is because genome sequencing generates a vast data set as a string. The authors in [13] examined deep learning breast cancer studies employing several imaging modalities. Datasets, architecture, applications, and evaluations guided these investigations. They focused on breast imaging deep learning frameworks utilizing three modality types (ultrasound, mammography, and M.R.I.). Their goal was to use DLR-based CAD systems to give current breast cancer imaging data. Their categorization used secret datasets and CNNs. After analyzing

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

these surveys, I will simultaneously explore genetic sequencing and imaging to predict breast cancer and help diagnose and treat it early. We will also advise researchers interested in this topic.

The authors of [14] propose an intuitive technique for classifying mammogram images as benign, malignant, or normal using machine learning methods. SVMs, CNNs, and R.F.s are compared. The experiment showed that CNN is the best classifier because its morphological and filtering operations intuitively categorize digital mammograms.

The authors in [15] utilize Dr. William H. Walberg's U.W. Hospital dataset. Logistic regression, k-nearest neighbors, SVM, naive Bayes, decision trees, random forest, and rotation forest were used to practice data visualization and machine learning. These machine-learning strategies and visualizations used R, Minitab, and Python. Comparing the procedures was done. The logistic regression model with all features had the highest classification accuracy (98.1%), while the proposed technique improved accuracy.

The authors in [16] compared SVM, Logistic Regression, Naive Bayes, and Random Forest. Wisconsin's breast cancer dataset corresponds. The Random Forest algorithm got the highest accuracy (99.76%) and lowest error rate. Anaconda Data Science Platforms was used to simulate all experiments.

Breast cancer subtypes can be identified using the authors' method [17]. Feature selection uses the Wisconsin Diagnosis and Analysis and Prognostic Breast Cancer datasets. It then classifies breast cancer using neural networks, focusing on M.L.P. and back-propagation neural RBF. This data set's nine features comprise the neural network's input layer. The neural network classifies input features into two cancer categories (benign and malignant). When tested on the database, RBF neural network classification had a 97% recurrence rate.

The authors [18] contrasted tree-augmented Naive Bayes and Markov blanket estimating networks to create an ensemble model for breast mass severity prediction. The algorithm helped doctors decide whether to biopsy a suspicious lesion based on mammography readings. The authors found that bayesian classifiers are a promising alternative to various medical applications.

In emergency care, authors [19] use Bayesian networks (B.N.) due to their powerful symbol, handling of ambiguity, and ability to consider multiple alternatives based on data. Bayesian networks work because of their symbolic representation.

3. Data Analysis using statistical methods

Many parameters can be found in the dataset. The statistical modeling software SAS JMP was utilized to get insight into the data and its meaning. The carcinoma dataset contains a wide variety of data and metric values. The distribution figures illustrate the ranges of a few crucial factors (Fig. 2 and Fig. 3).

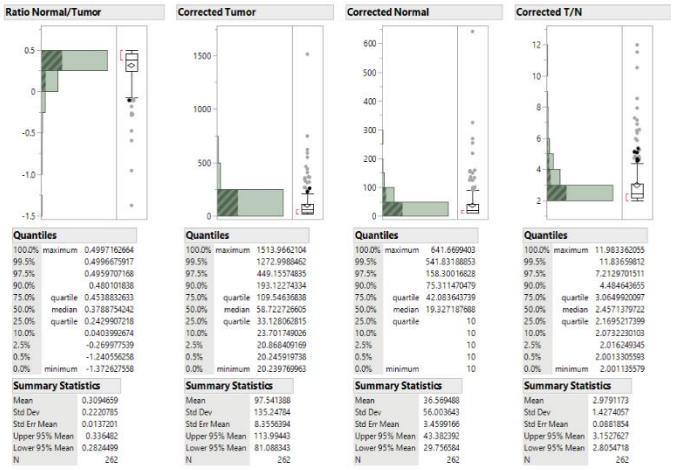

Figure 2: Tumor B2, C2 category – Statistics.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

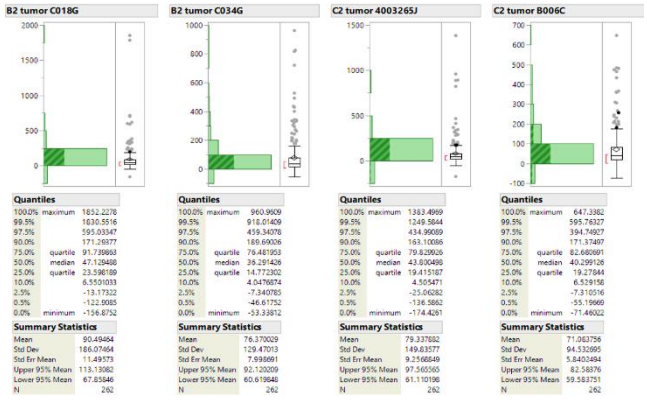

Figure 3. Tumor, Normally Distributed statistics

Creating classification models becomes increasingly difficult as more variables and data are employed in statistical research. Due to the limited quantity of cancer samples compared to the number of features, over-fitting and deterioration of classification capacity are more likely to occur in clinical practice. Feature selection is a valuable technique for dealing with such challenges. When training a classification model with limited data and many features, narrowing the subspace to a more manageable collection of features can help. Here, we employ the Distribution method to locate genes that may be useful in further classification. Researchers commonly use data distribution modeling to determine if a gene's reported change in read count warrants further investigation (i.e., more extraordinary than that which would be predicted owing to natural random fluctuation). Proteins that are significantly differently expressed can be eliminated from a differential expression analysis by using a BH-adjusted p-value and a fold change threshold.

Fundamental goals compared to pre-existing approaches classified under cancer detection models. The significant discovery demonstrates the dissimilarities between the current and suggested hybrid models, such as Ensemble techniques and Neural network-based Deep Learning approaches.

1) The Carcinoma Dataset was analyzed deeply to reveal hidden trends and insights.
2) Fundamental discoveries based on the statistical inference that illustrate the disparities between normal conditions and tumor situations
3) Using statistical, machine learning-based ensemble, deep learning/neural network-based modeling, it was possible to predict cancer from the Normal and Tumor states by analyzing the Tumor to Normal and Normal Tumor circumstances.

4. Data Modelling

The degree of difficulty required to use various classification methods can be affected by a wide range of variables. To achieve the maximum possible predictive accuracy on new data, it is necessary to identify the values of the complex variables that are most ideal for the specific application at hand. If data are abundant, a model can be chosen with little effort by separating the data into a training set, a validation set, and a test set. Training data is used to teach multiple models, which are then tested on a separate dataset to determine which performs best. The best complex model is selected among those trained, and it is proven effective by using the validation set.

Nevertheless, fewer data resources are available for training and testing in a real-world context, which raises the generalization error. To reduce the generalization error and prevent over-fitting, cross-validation is used. The data distribution for the S-fold cross-validation method with S = 4 employed in this paper is shown in Fig. 2 and Fig. 3. To conduct S-fold cross-validation, we split the entire dataset D into S equal-sized subsets, D1, D2, and D.S. We then utilized a random sampling method to divide the

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

remaining individuals into a test set and a set of S-1 groups for purposes of training.

The procedure was performed *S-1* times, and the average performance rating across all iterations was determined. We generate new data for the ensemble stage to avoid over-fitting by acting as a model selection for each classifier using S-fold cross-validation. As for the Carcinoma dataset, several features are portrayed in different capacities. The ratio between the tumor and normal tissues was the most prominent finding. Figure 4 compares cancer to a healthy state. Together with various additional criteria, they are laid out in great depth here. The average tumor and normal circumstances are displayed graphically in Fig. 4. Researchers found that several indicators indicate a high likelihood of cancer development [20]. This study's primary focus is on using tumor-specific probabilities to identify malignancy.

5. Results and Discussion

After pre-processing the data sets, we assess the forecast performance of five popular classification methods for distinguishing between healthy and malignant tissue. K.N.N., D.T., SVM, R.F, and gradient-boosting decision trees are all used in the first stage of categorization (GBDTs). All five classification strategies below perform admirably in real-world applications, and their benefits are explained at length. kNN is helpful as a data categorization technique when only a little knowledge of the data's distribution is available. When using the k-nearest neighbor classifier, distances are determined by projecting data onto a metric space.

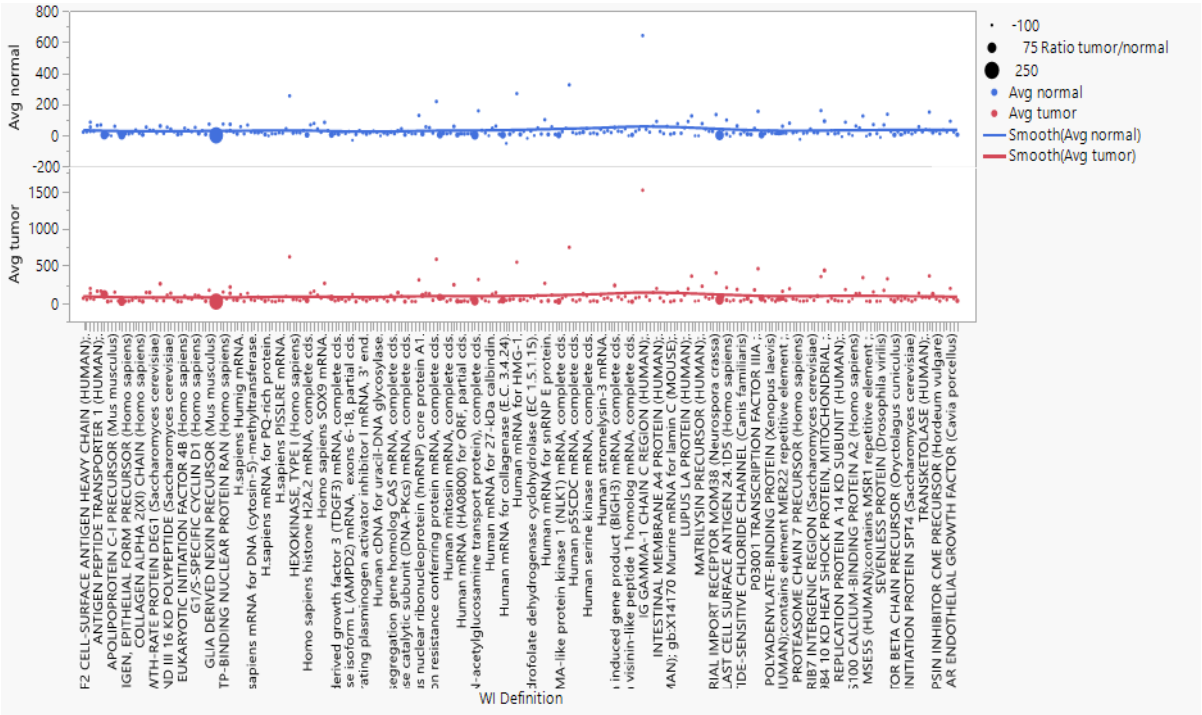

Figure 4. Tumor vs. Normal cancer

The purpose of k-nearest neighbors is to classify a test sample using the most common class in its k-nearest training samples. This is done by calculating the distance between a test sample and the training samples. The first step for SVMs is to transform the input sequence into a higher-dimensional feature space, which is then used to locate a hyperplane that splits the data into two classes. There is a vast chasm between the two communities. Then, fresh samples are projected into the same area, and their predicted class is based on which side of the divide they fall on with higher confidence. D.T.s, or "decision

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

trees," take the shape of trees, with the "nodes" reflecting the input parameters and the "leaves" representing the judgments that were reached. Because of the unique layout, we can reliably foretell the data type as we move down the tree to assign it to a category. It is only recently that researchers have begun using R.F.s to predict cancer.

R.Fs is an ensemble learning technique that combines tree predictors using the same distributional random vector. The class with the most votes from individual trees in the forest produces the best model. GBDTs, a machine-learning technique, combines several decision trees into a single, highly accurate model for prediction. Like traditional boosting techniques, GBDTs build the model incrementally. But, unlike those methods, they allow optimizing a fully-differentiable loss function. Three classic methods (kNN, SVMs, and D.Ts), plus two cutting-edge ones, are presented (R.F.s and GBDTs). Especially for data with unknown distribution, there is some evidence in the literature to suggest that kNN is one of the simplest categorization methods. However, a classifier's performance is susceptible to the value of k; kNN is vulnerable to redundant data, and adequate feature extraction is required before classification using kNN. We can confidently conclude that SVMs are the most widely used and effective method for classifying cancer types [10, 12]. Yet, SVMs have a challenging task: deciding which kernel is the most appropriate for a given situation.

The inability to guarantee the accuracy of the forecasts applies particularly to nonlinear cases for which there is no general solution. Despite their widespread use and popularity, D.T.s are notoriously unsuccessful at distinguishing between normal and malignant samples, despite being the most fundamental and widely utilized categorization strategy across many fields. Nonetheless, the classification result may be skewed towards the group with more data, even though the latter two methods, R.F.s, and GBDTs, are ensembles of D.Ts that evolve to tackle the over-fitting issue. Recognizing that each method has its limitations, we devise an ensemble technique to harness the strengths of multiple methods while sidestepping their weaknesses. To make our ensemble classifier more stable, we use a combination of evolutionary and traditional methods.

TABLE 1- K.N.N. Algorithm Analysis

K	Count	R-Square	RASE	SSE
1	262	-0.9650	0.082	1.789
2	262	-0.42	0.071	1.350
3	262	-0.350	0.069	1.229
4	262	-0.139	0.063	1.037
5	262	-0.139	0.061	0.946
6	262	-0.005	0.060	0.916
7	262	0.002	0.059	0.910
8	262	-0.001	0.059	0.912
9	262	-0.004	0.060	0.915
10	262	0.004	0.059	0.908

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

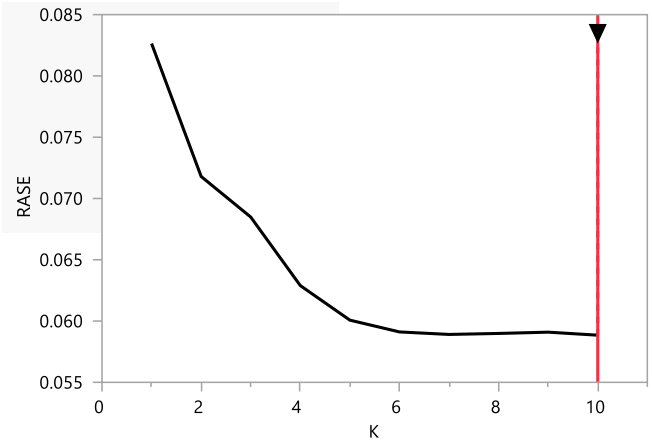

Figure 5: K N.N. Algorithm Analysis K – Fitting

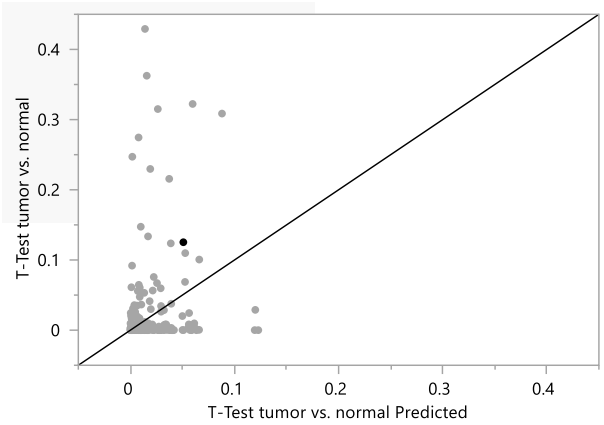

Figure 6: K N.N. Algorithm Analysis T- Test Tumor Vs. Normal condition

Values for K-Nearest Neighbors algorithms fitting models are shown in Table 1. K-N-N values for R-squared, RASE, and S.N.E. represent the accuracy with which the method fits the function. The training and testing environments for the kNN algorithm are depicted in Figs. 5, 6, and 7. Similarly, the fitting analysis for cancer prediction from the level of tumor and normal circumstances is illustrated in Fig. 8, and Fig. 9 was evaluated using the Support Vector Machines method. Fig. 10 shows the t-statistics concerning the SVM algorithm.

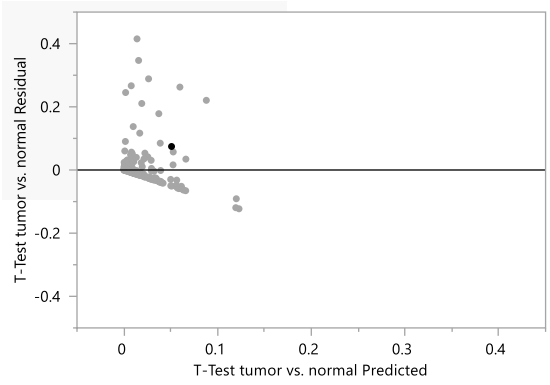

Figure 7: K N.N. Algorithm Analysis T- Test Tumor Vs. Normal condition

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

TABLE 2- SVM Algorithm Analysis

Response	T-Test tumor vs. normal
Validation Method	**KFold**
Kernel Function	Radial Basis Function

Measure	Training	Validation
Number of rows	210	52
Sum of Frequencies	210	52
RASE	0.0580823	0.0466788
R-Square	0.0973591	0.0950323
Number of Support Vectors	89	89

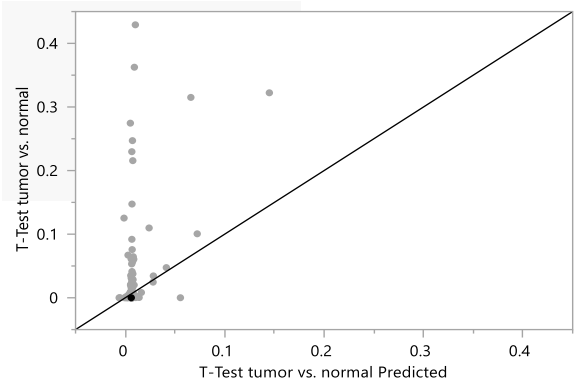

Figure 8: SVM Algorithm Analysis T- Test Tumor Vs. Normal training set

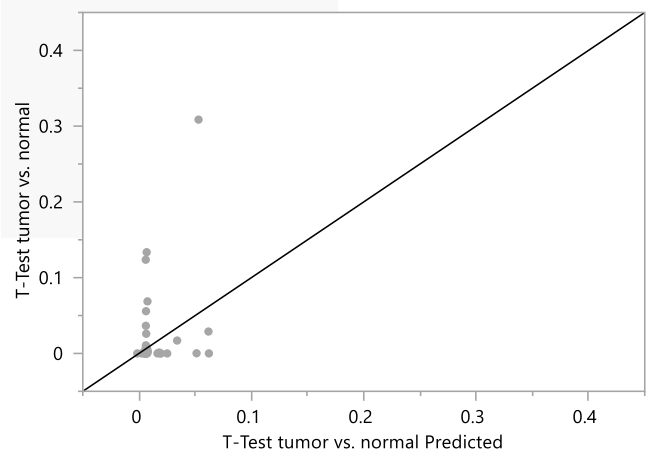

Figure 9: SVM Algorithm Analysis T- Test Tumor Vs. Normal testing set

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

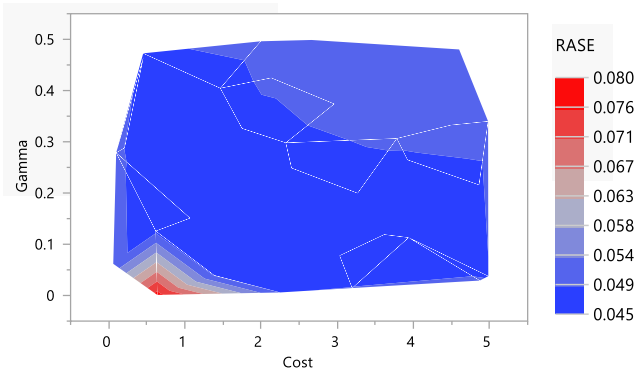

Figure 10: SVM Algorithm Analysis T- Test Tumor Vs. Normal modeling accuracy

6. Dl Based Analysis

Secondly, cancer prediction categorization methods are imperfect and may be erroneous. Categorization algorithms may work better together. A more advanced learning model uses a multi-model ensemble's predictions. The second-stage model training combines first-stage model predictions to make optimal predictions. This work employs deep learning as an ensemble model to aggregate many classifier outputs into a single exact estimate. Neural networks, which mimic the brain, are widely used. Neural networks can output from several inputs. Using one or more hidden nonlinear layers between the output and input layers, it can approximate nonlinear functions given a set of characteristics and an end goal. Deep neural networks with several hierarchical hidden units of nonlinear processing information discover complex patterns from high-dimensional raw input without supervision. Example neural network [21]. The leftmost layer, the input layer, contains input neurons. The rightmost layer has an output neuron. The middle layers are buried neurons. To classify samples accurately, we calculate the variance from actual scores to projected scores using an objective function. Then, the machine learns from the training samples and adjusts the input-output function variables internally, resulting in a small error [22].

The stochastic gradient descent (S.G.D.) algorithm is commonly used for this machine learning method. In a deep neural network, layer L1 is the input layer, layer Lnl is the output layer, where nl represents the number of levels and Ll represents each layer. Similarly, the total number of neurons in layer l will be denoted as sl. W = [W1, W2,..., Wnl] and b = [b1, b2,..., bnl] are the neural network parameters, and Wlij, j = 1, 2,..., sl1, I = 1, 2,..., sl, l = 2, 3,... Assume we have m samples of data labeled "(x 1, y1),(x 2, y2),...,(xm, ym)" to use as a training set and want to demonstrate how the S.G.D. can be used to train a neural network. For this discussion, let us refer to the cost function (objective function), where is a weight decay parameter that controls the relative importance of a mean-squared error term and a regulation term that limits the weight orders of magnitude to prevent over-fitting.

The most commonly used nonlinear function in this context in recent years [18] is the rectified linear unit (ReLU) f(z) = max0, z. The ReLU outperforms conventional nonlinear and logistic sigmoid functions in learning speed in inter-deep neural networks. The fraction of a given sample actively engaged in unit I in layer l is denoted by I, while the total represents z l I.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

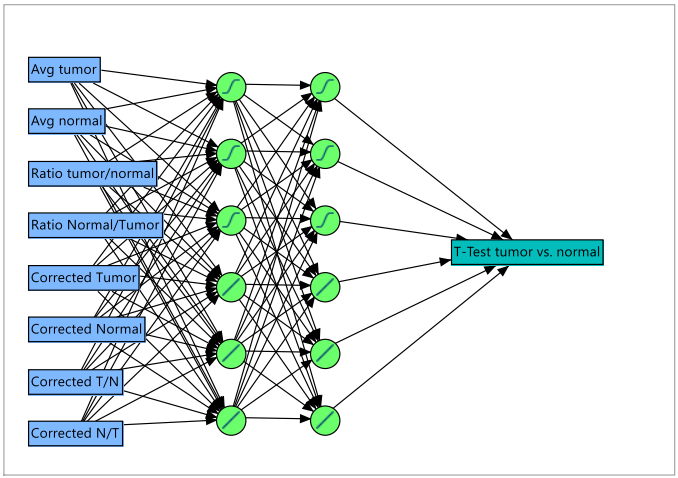

Figure 11: Neural Network Algorithm Analysis on T- Test Tumor Vs. Normal data

The weighted average and majority vote algorithms in regular ensemble strategy only evaluate linear correlations among classifiers and require operator participation, whereas the deep learning-based ensemble technique "learns" the relationships automatically. As the relationships between the various classifiers and sample labeling are uncertain, an essentially linear relation cannot predict with accuracy. Our second stage uses deep learning to automatically learn complex relationships, notably nonlinear ones, with minimum engineering. The deep learning-based inter-ensemble approach uses all data for reliable predictions. These data sets cover all cancer stages from various clinical situations, ages, sexes, and ethnicities. We examined tumor samples from non-chemotherapy or radiation-treated patients for this characteristic. Table 1 details datasets. Our method used normalized FPKM data and raw data counts. The formal recognition and ensemble approach used the normalized FPKM values of the significantly differentially expressed genes from the raw count data.

TABLE 3- Multilayer Neural Network Algorithm Analysis

Measures	Value
R-Square	0.021
RASE	0.059
Mean Abs Dev	0.020
-LogLikelihood	-393.00
SSE	0.595
Sum Freq	174

Five classification strategies (k-nearest neighbor, SVMs, DTs, R.F.s, and GBDFs) were utilized in the first stage, and their predictions were averaged using the 5-fold cross-validation method. The forecasts from the first step were then combined using a multi-model ensemble technique and a deep neural network. Deep neural networks can be employed to make more accurate predictions using the new data set's reduced dimensions and bigger sample size, made possible by 5-fold cross-validation.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

TABLE 4- Difference: Corrected T/N-T-Test tumor vs. normal

Corrected T/N	2.97	t-Ratio	33.99
T-Test tumor vs. normal	0.01	DF	261
Mean Difference	2.95	Prob > |t|	<.0001*
Std Error	0.087	Prob > t	<.0001*
Upper 95%	3.13	Prob < t	1.0000
Lower 95%	2.790	-	-
N	262	-	-
Correlation	0.326	-	-

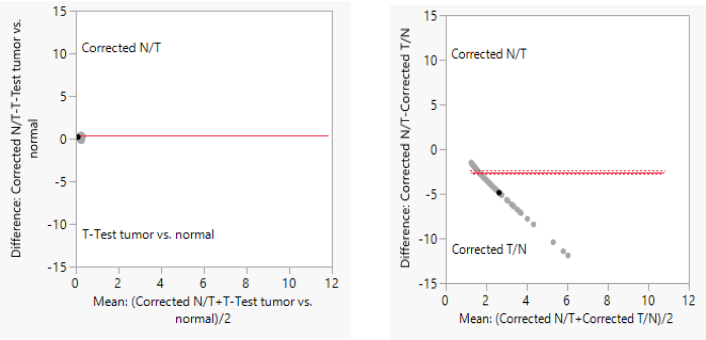

Figure 12. Finding the relation concerning Normal / Tumor – Corrected conditions

The figures show that combining multiple classifiers improves classification performance over a single classifier. Because it automatically learns and finds hidden structures, the deep learning-based ensemble solution outperforms the majority vote for all three datasets. Figure 12 depicts three related P.R. curves. The diagram shows that the ensemble technique outperforms individual classifiers and majority voting. The ensemble method is also practical, with skewed statistics reflecting clinical sample disparity. Fig. 13 displays the generalized regression for the ratio typical: tumor> average maximum likelihood with the validation column.

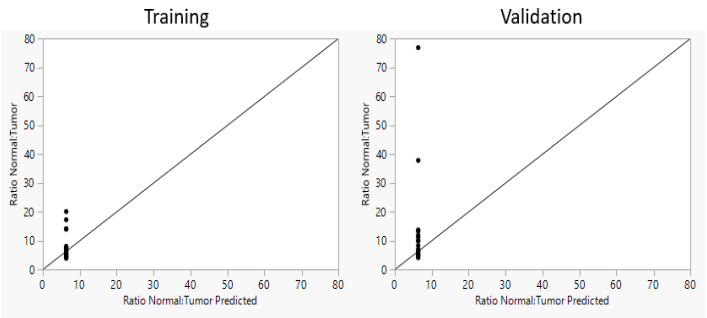

Figure 13. Training and Validation fitting with Ensembled Modeling With likelihood.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

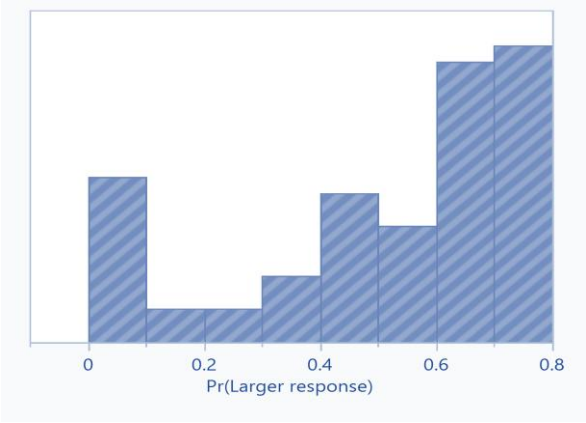

Figure 14. Training and Validation fitting with Ensembled Modeling With likelihood.

The Ratio Generalized Regression Analysis Tumor > Normal Maximum Likelihood with Validation Column > Diagnostic Bundle Presented in Fig. 14 Normal: Tumor > Normal Maximum Likelihood with Validation Column >

7. Conclusion

The effects of cancer on a worldwide scale are devastating. There is currently no gold standard for cancer prediction, despite the increasing popularity of machine learning approaches. In this research, we have brought a deep learning-based multi-model ensemble approach to cancer prediction. More specifically, we analyzed information about gene expression levels that was gathered from the lungs, the stomach, and the breasts. To avoid overfitting in categorization, we employed statistical analysis to identify genes whose expression levels differed considerably between normal and malignant phenotypes. The results indicate that differentially expressed analysis is critical for selecting the most pertinent data points and reducing data dimensionality, improving prediction accuracy, and reducing computational time. The next step in the multi-model ensemble method is to feed the predictions from numerous models into a deep neural network that has been trained to combine the inputs into a single, more precise forecast. The majority voting approach compares the results from different classifiers. Five classifiers were applied to the three cancer data sets, including a majority voting strategy and our proposed multimodal ensemble method. The proposed ensemble model outperforms state-of-the-art classifiers and majority voting on various evaluation metrics. The predictions from the first stage are used as features in the deep learning-based inter-ensemble model, leading to lower generation error and more data than when the model is trained independently.

References

1. Zeng, Q., Klein, C., Caruso, S., Maille, P., Laleh, N. G., Sommacale, D., ... & Calderaro, J. (2022). Artificial intelligence predicts immune and inflammatory gene signatures directly from hepatocellular carcinoma histology. *Journal of Hepatology*, *77*(1), 116-127.
2. Zhang, G., Sun, J., & Zhang, X. (2022). A novel Cuproptosis-related LncRNA signature to predict prognosis in hepatocellular carcinoma. *Scientific reports*, *12*(1), 11325.
3. Yan, C., Niu, Y., Ma, L., Tian, L., & Ma, J. (2022). System analysis based on the cupro ptosis-related genes identifies LIPT1 as a novel therapy target for liver hepatocellular carcinoma. *Journal of Translational Medicine*, *20*(1), 1-18.
4. Allugunti, V. R. (2022). Breast cancer detection is based on thermographic images using machine learning and deep learning algorithms. *International Journal of Engineering in Computer Science*, *4*(1), 49-56.
5. Ramana, K., Kumar, M. R., Sreenivasulu, K., Gadekallu, T. R., Bhatia, S., Agarwal, P., & Idrees, S. M. (2022). Early prediction of lung cancers using deep saliency capsule and pre-trained deep learning frameworks. *Frontiers in Oncology*, *12*.

ISSN: 2309-5288 (Print)
ISSN: 2309-6152 (Online)
CODEN: JCLMC4
Journal of Coastal Life Medicine

6. Rudra Kumar, M., Pathak, R., & Gunjan, V. K. (2022). Diagnosis and Medicine Prediction for COVID-19 Using Machine Learning Approach. In *Computational Intelligence in Machine Learning: Select Proceedings of ICCIML 2021* (pp. 123-133). Singapore: Springer Nature Singapore.
7. Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., ... & Teng, Y. (2020). A comprehensive review for breast histopathology image analysis using classical and deep neural networks. *IEEE Access*, *8*, 90931-90956.
8. Saravana Kumar, K., & Ramasubramanian, S. A clinical decision support system for heart disease prediction with ensemble two-fold classification framework. *Journal of Intelligent & Fuzzy Systems*, (Preprint), 1-18.
9. Kumari, N., & Acharjya, D. P. (2022). Data classification using rough set and bioinspired computing in healthcare applications-an extensive review. *Multimedia Tools and Applications*, 1-27.
10. Alshayeji, M. H., Ellethy, H., & Gupta, R. (2022). Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. *Biomedical Signal Processing and Control*, *71*, 103141.
11. Fatima, N., Liu, L., Hong, S., & Ahmed, H. (2020). Prediction of breast cancer, comparative review of machine learning techniques, and their analysis. *IEEE Access*, *8*, 150360-150376.
12. Wisesty, U. N., Mengko, T. R., & Purwarianti, A. (2020, April). Gene mutation detection for breast cancer disease: A review. In *I.O.P. Conference Series: Materials Science and Engineering* (Vol. 830, No. 3, p. 032051). I.O.P. Publishing.
13. Pang, T., Wong, J. H. D., Ng, W. L., & Chan, C. S. (2020). Deep learning radiomics in breast cancer with different modalities: Overview and future. *Expert Systems with Applications*, *158*, 113501.
14. Vasundhara, S., Kiranmayee, B. V., & Suresh, C. (2019). Machine learning approach for breast cancer prediction. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(1).
15. Ak, M. F. (2020, April). A comparative analysis of breast cancer detection and diagnosis using data visualization and machine learning applications. In *Healthcare* (Vol. 8, No. 2, p. 111). MDPI.
16. Taghizadeh, E., Heydarheydari, S., Saberi, A., JafarpoorNesheli, S., & Rezaeijo, S. M. (2022). Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods. *Bmc Bioinformatics*, *23*(1), 1-9.
17. Raad, A., Kalakech, A., & Ayache, M. (2012). Breast cancer classification using neural network approach: M.L.P. and RBF. *Ali Mohsen Kabalan*, *7*(8), 105.
18. Elsayad, A. M. (2010). Predicting the severity of breast masses with ensemble of Bayesian classifiers. *Journal of computer science*, *6*(5), 576.
19. Sylviaa, M. S. M., & Sudha, N. (2022). Review On Diagnosis for Early Stage of Breast Cancer. *JOURNAL OF ALGEBRAIC STATISTICS*, *13*(3), 784-789.
20. Visser, I. J., Levink, I. J. M., Peppelenbosch, M. P., Fuhler, G. M., Bruno, M. J., & Cahen, D. L. (2022). Systematic review and meta-analysis: Diagnostic performance of D.N.A. alterations in pancreatic juice for the detection of pancreatic cancer. *Pancreatology*.
21. Haznedar, B., & Simsek, N. Y. (2022). A Comparative Study on Classification Methods for Renal Cell and Lung Cancers Using RNA-Seq Data. *IEEE Access*, *10*, 105412-105420.
22. Chalapathi, M. M., Kumar, M. R., Sharma, N., & Shitharth, S. (2022). Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal. *Security and Communication Networks*, *202*