







Detection of Pleuro Pulmonary Blastoma using Machine Learning Models

Raswitha Bandi* and T. Santhisri



Department of CSE, KoneruLakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, Andhra Pradesh-522502, India

E-mail/Orcid Id:

RB,  raswitha.29reddy@gmail.com,  <https://orcid.org/0000-0002-3916-4374>; TS,  santhisri@kluniversity.in,  <https://orcid.org/0000-0003-4763-0088>

Article History:

Received: 18th Feb., 2024**Accepted:** 19th June, 2024**Published:** 30th June, 2024

Keywords:

CT Image Processing,
Machine Learning
Algorithms, Medical Images,
Pleura PPB Cancer,
Pulmonary Blastoma

How to cite this Article:

Raswitha Bandi and T. Santhisri (2024).
Detection of Pleuro Pulmonary Blastoma
using Machine Learning Models.
*International Journal of Experimental
Research and Review*, 40(spl.), 151-163.
DOI:

<https://doi.org/10.52756/ijerr.2024.v40spl.012>

Abstract: Pleura Pulmonary Blastoma (PPB) is a type of lung cancer seen in children. PPB needs to be detected earlier when treating children. The mortality rate of PPB is higher if left untreated. It can be detected from CT images through various machine learning and classification algorithms. The earlier detection of PPB can save children's lives, for which several research works have proposed several machine learning models. Several researchers adopt traditional classification algorithms like random forest and decision tree algorithms for detecting PPB. However, these techniques provided lesser accuracy and were difficult for earlier detections. This paper considered several machine learning algorithms like SVM, LR and MP and experimented with CT images and DICER-1 data to understand their betterness and overcome such issues. The architecture of the following algorithms is discussed in detail, and the results are compared. Through this, the ideal machine learning algorithm for detecting PPB is found. All the algorithms are implemented with the Python software, and the performance metrics of the respective algorithms are recorded. The results show that the SVM algorithm provides better accuracy (96%) for the DICER-1 dataset, which is higher than CT images (95.60%).

Introduction

Pleuropulmonary Blastoma is a type of cancer, rare but aggressive, that occurs mainly in young children. Also found in adults, this disease is regarded as malignant. It originates either from the lungs or the pleura. This rare type of childhood tumour presents non-specific symptoms and can be divided into four subtypes that are discussed in the next part. Because of the non-specific nature and other symptoms and reasons that can be attributed to the cause of this disease, it is difficult to tell immediately if a child has Pleuropulmonary Blastoma (PPB). There are two common risk factors associated with PPB. They are a change in the DICER1 gene or a family history of DICER1 syndrome. The prevailing standard methods for identifying the abnormalities in patients with PPB, including heuristic, investigative, and untested methods, are neither encouraging nor satisfactory in diagnosing it in terms of accuracy.

The four subtypes of PPB are type I, type II and type III PPB. The first three types are generally found in children under the age of 8, and the fourth type may be found in any age. Type II includes both cystic and solid parts, found in children above 2 years of age. Type III represents a solid high grade and is usually diagnosed in children who are less than 7 years of age.

The two common sets of symptoms that may be found in children with PPB are stress or difficulty in breathing, which may be mild or severe. The other symptoms can be general illnesses associated with cough, fatigue, fever, loss of energy and appetite, and chest pain. There are two common risk factors associated with PPB. They are a change in the DICER1 gene or a family history of DICER1 syndrome. The basic form of the PPB present in the children is shown in Figure 1. The prevailing standard methods for identifying the abnormalities in patients with PPB, including heuristic, investigative and untested methods, are neither encouraging nor satisfactory in



diagnosing it in terms of accuracy. Even Deep learning methods cannot be wholly relied upon to give precise or accurate results in the testing process. Hence, the use of Machine Learning is suggested.

The main objective of this paper is to analyze the mutations of the DICER1 to predict the PPB-related conditions present in the gene data. The changes in the gene incorporated with the DNA (DICER1) are responsible for the high risk of cancer, mainly PPB. The mutations of DICER1 also create other diseases related to PPB. This paper experiments with various machine learning models with the DICER1 dataset to understand and choose the best learning model. The results conclude that the Multi-Class-SVM model is suitable for classifying the high volume of mutations of DICER1 and predicting the PPB and related diseases.

The main contributions of the paper are listed below

To compare the efficiency of the machine learning algorithms in detecting PPB. Discuss the anatomy of the PPB and check for features that will allow efficient PPB detection.

In this paper, we have used three machine-learning algorithms and experimented with them on two datasets. Using the CT and DNA datasets, the efficiency of the machine learning algorithm in terms of features PPB prediction is easily justified. This helps to understand the challenges in PPB prediction on a multi-model dataset and choose the most efficient model for the dataset. This paper also aims to understand the issues and challenges presented earlier.

Literature Review

Numerous researchers propose several methods to detect and identify PPB (Khant et al., 2023; Reddy and Khanaa, 2023). Some significant works are consolidated and listed in the survey presented in this work. Brodowska-Kania et al. (2016) presented a review to demonstrate the severity of PPB. Bueno et al. (2017) proposed a method to determine the abnormalities and characteristics of DICER1 syndromes in patients. The input imaging dataset contains various data like cystic nephroma, ovarian sex-stromal tumor, DICER1 syndrome, and ovarian Sertoli-Leydig. For this experiment, the data are taken from 16 patients. The experiment showed that 68.8% of the patients were affected by malignant lesions.

Addanki et al. (2017) proposed a study to improve the accuracy of diagnosing PPB tumors. The data was collected from 12 children with PPB. Through the mutations in the DCER1, the severity and reason for PPB are identified. The result of this study indicates that the

DCER1 mutation in the patient is similar to the report produced from the USA, UK and Japanese patients. K. van Engelen et al. (2017) experimented to diagnose PPB from the DCER1 mutations. The data are observed and collected from 78 patients. From the analysis, it is clear that PPB is a common type of tumor. It is found and diagnosed among children around five years and two months. Dehner et al. (2015) discussed that PPB is the most common type of tumor among children aged 0-6. The mutation in the DICER1 will result in PPB in the lungs. Nearly 75%-80% of Children are affected by PPB. Knight et al. (2019) presented a comprehensive review to describe PPB diseases. PPB is analyzed from the changes in the DICER1 and classified into four types: I, Ir, II and III.

Grigoletto et al. (2020) have investigated the inequalities in diagnosing pediatric tumors to study the possible correlation between diagnostic performance and activities. The data on the number of PPB cases registered from 2000-2014 was collected from some European pediatric oncology centers. This was compared with the number of cases expected. Kim and Lee et al. (2020) have studied the response of radiotherapy (RT) to an unresectable tumor of a child. The child had PPB with a DICER1 mutation that recurred even after surgery and led to respiratory problems. The purpose of the study was to evaluate the response of radiotherapy as it was one of the options to minimize tumors and facilitate treatment. The final analysis showed that the child was partially cured and maintained the status for a year. Further investigation of Radiotherapy (RT) for unrespectable PPB is recommended.

The lesions in the lung region affected by pediatric problems are some of the reasons for pulmonary issues, which are benign but, if left untreated, can cause cancer. A study was conducted on 11 children who developed such problems in the USA. Over 344 lesions were prenatally diagnosed; 177 had malignant pathology. The DICER1 mutations were also associated with the malignancy (Kunisaki et al., 2021). Also, the systemic feeding vessel was absent in the malignant lesions.

Limitations and Motivation

PPB is commonly seen in patients with lung cancer, and there is a need for faster and more efficient detection of it. Due to the increase in the number of patients, several research works have focused on the quicker detection of PPB, which has led to lesser accuracy than other traditional methods. Various machine learning algorithms have been considered for their detection, and most of them provided better accuracy but consumed

more time. There was a need for an efficient solution for which classification algorithms were considered. This computational power and time can be reduced by making use of the features of PPB, which is achieved through regression algorithms like logistic regression and multilayer perceptrons. However, PPB's nature varies with the patient, making it difficult to train the models. To tackle this problem, SVM was widely considered. The SVM algorithm provided better accuracy and required less computational power to detect PPB.

Materials and Methods

Proposed Architecture

The CT images of the PPB are processed through machine learning algorithms, and their results are compared. The algorithms compared in this paper are logistic regression, multilayer perceptron, and support vector machine (Rao et al., 2023). The CT images are passed through all these algorithms, and their detection accuracy is tabulated. Through this, the best machine learning technique can be found.

Logistic regression

It is one of the supervised learning algorithms that label the input images by classifying them. The Logistic regression algorithm processes the input images from the training set and learns from the image set by Huang et al. (2019). Let the number of images be n ; each image is labelled as $D = (x_i, y_i)_i^n = 1$. Where D represents the dataset and x_i, y_i represents the data point. In the equation, the $y_i \in \gamma = 1, 2, \dots, K$ labels the images as ground truth in the set x_i , which may associated with a single or multiple classes (Rao et al., 2023). Let K be the number of possible classes. Most supervised learning algorithms adopt multiple classes and process them through probabilistic class distributions. Let p_k be the probability of the probabilistic distribution of the training images. Let the discrete elements in the probabilistic list be addressed as $\text{the } p_k = p(k \vee x), k \in Y$.

Estimating the probabilistic distributions

To obtain valid probabilistic values, the inputs are normalized using a logistic function (sigmoid) that nullifies the output values $z = \phi(x \vee \theta)$, and the intervals are maintained as $(0, 1)$ as:

$$p_k = p(k|x; \theta) = \sigma(z)_k = \frac{1}{1 + e^{-z_k}} = \frac{e^{z_k}}{e^{z_k} + 1}, k \in Y = 1, 2, \dots, \dots \dots \dots (1)$$

The model parameters are denoted as θ , and the x is given as the input for the model. It is done within a logit space z , which carries out the learned mapping of the features. It is based on the Bernoulli's distribution. The multiple classes help predict the positive probability in the $p(1|x)$ overall class k in that distribution (Yavanamandha et al., 2023). This is a multi-label

classification type in which each class is assigned multiple classes that combine to form various possible ways of features. This type of numerous combinations is widely seen in logistic regression models. Along with the multiple class assignment, each image is assigned a label that helps differentiate the images. The various classes assigned to the images help find similarities between the images. The probability distribution of the multiple classes can be extracted and forwarded to the vector p , which thus is $\sum_{k=1}^K p_k = 1$ where the $p_k \geq 0$. The softmax function is used to normalize the values obtained from the probability distribution and for a valid probability vector p ,

$$p_k = p(k|x; \theta) = h(z)_k = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}}, \text{ with } k \in Y = 1, 2, \dots, K \dots \dots \dots (2)$$

The mentioned soft max equation 2 represents the assignment of discrete variables to multiple classes collectively and interdependently. It is also termed as softmax regression. Logistic regression is a supervised learning model that uses the maximum probability of distribution, which is achieved using the cross-entropy model. However, a learning function is derived for the LR model to classify the classes. Two types of learning functions are framed, one for multiple classes and another for single classes. For multi-class classification, the objective function for learning is formulated in

$$L_{LR}(x, y) = -\sum_{k=1}^K (q_k \log(p_k) + (1 - q_k) \log(1 - p_k)) \dots \dots \dots (3)$$

The mentioned cross-entropy loss eq-3 uses the class-wise distribution functions of the Bernoulli in its opposite way of negative log-likelihood. For single-label classification, the cross entropy of the ground truth and predicted class is used as the learning objective function. The ground truth class $q = \delta_{k,y}$, and since it is a single label, it only provides two possibilities, either $k = y$ and $k \neq y$. The categorical cross entropy function is represented as:

$$L_{SR}(x, y) = -\sum_{k=1}^K q_k \log p_k = -\log p_y \dots \dots \dots (4)$$

After obtaining the learning parameters through both functions, the features are selected through two selection processes: the hard and soft selection methods. The hard selection comprises parameters that determine the PPB or lung cancer features. These features help in the exact selection of the PPB features. However, the hard selection methods leave out the least important or negligible features, which may also be important at certain stages of detection. The hard selection loss function is defined as:

$$L_{LR}^{hs}(x, y) = -\log(p_y) - \alpha \sum_{k \in f_{hs}(mvp, y)} (\log(1 - p_k)) \dots \dots \dots (5)$$

To overcome those issues, the soft selection method is used. It selects the parameters with the least and above the threshold probability level. This helps in covering the areas that are left during the soft selection. The α is used

to normalize the obtained value during the selection process. The m% of the predicted classes is returned from the better-predicted probabilities. The Specialized loss function is defined as:

$$L_{LR}^{SS}(x) = -\log(p_y) - \alpha \sum_{k=1, k=y}^K (p_k)^r \log(1 - p_k) \dots \dots \dots (6)$$

The results are converted into coordinates after both the hard and soft selection processes. The algorithm consists of a plotting model with two coordinates, x and y, which takes either of the binary values assigned to the data. Let the x be a set of data relating to a type of PPB that may point to the severity of the cancer, and let M be the variables depending on that. The probability of the protein that could be obtained from the CT images is $\pi = \Pr(y = 1/x)$. Let N be the number of images. The logistic regression of the DNA sequence is

$$\text{Logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta^T x \dots \dots \dots (7)$$

The intercepting parameter is α , and the varying coefficients are β^T . An observation is assigned to the response by the discriminant analysis. It is represented as $y(y \in 0,1)$ and looks for the most posterior probability. It is classified into several classes. If the class is 0, then the $p(0|x) > p(1|x)$; if the opposite is the class, it is 1. These posterior probabilities can be defined using Bayes' theorem and the following formula:

$$P(y \vee x) = \frac{p(x \vee y)p(y)}{p(x)} \dots \dots \dots (8)$$

Multivariate distributions like $p(x \vee y = 0)$ and $p(x \vee y = 1)$ are assumed to be the class-conditional distribution. These distributions have a mean vector μ_0 and μ_1 , $\sum_0^{-1}(x - \mu_1)^T, \sum_1^{-1}(x - \mu_1)$ and the classification rule can be classified as $y = 0$, which can be satisfied through the following Quadratic Discriminate analysis formula

$$(x - \mu_0)^T \sum_0^{-1} (x - \mu_1)^T \sum_1^{-1} (x - \mu_1) < 2 \left(\log(P(y = 0)) - \log(P(y = 1)) \right) + \log|\sum_0^{-1}(x - \mu_1)^T| - \log|\sum_1^{-1}(x - \mu_1)| \dots \dots \dots (9)$$

If the covariance matrices are equal, they can be simplified and obtained as the linear discriminant analysis. It can be used to cancel out the quadratic terms from the expression. The schematic representation of Logistic Regression is shown in Figure 1.

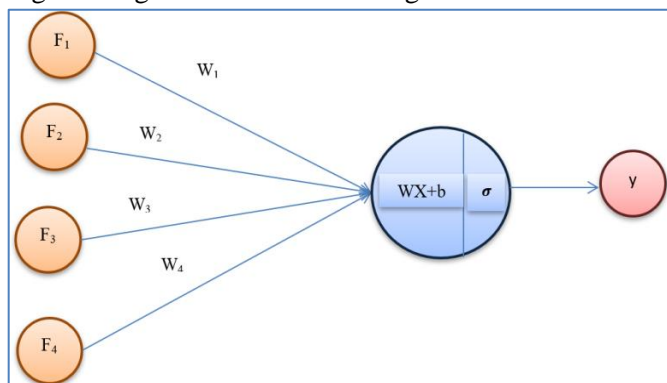


Figure 1. Schematic Representation of Logistic Regression.

Multilayer Perceptron

It is also called the neural networks that represent the functioning of the human brain through mathematical models. It can find the non-linear relation between the input and target variables. The input data is processed, and the features extracted from them are transmitted to the subsequent layers. Finally, a single output is obtained (Dettori et al., 2018). The connection between the neurons of the subsequent layers is defined through their weights. The activation function is used to produce the final output that takes all the weighted inputs from the hidden layer and the bias term b_i :

$$h_i = f^{(1)}\left(b_i^{(1)} + \sum_{j=1}^M W_{ij} x_j\right) \dots \dots \dots (10)$$

The weight matrix is represented as W , and the input j and hidden neurons are connected with the weight W_{ij} . A logistic activation function in the output layer analyses the input data through a binary prediction. The response probability is obtained by $f2x = \frac{1}{1+e^{-x}}$, and the feed-forward neural network is represented as:

$$\pi = f^{(2)}\left(b^{(2)} + \sum_{j=1}^{n_h} v_j h_j\right) \dots \dots \dots (11)$$

Architecture of the Multilayer perceptron model

Initially, the proposed model is trained on input images. The input images are resized and maintained with 140x140 size. Before providing the input CT images, each pixel of the image is subtracted with a mean RGB value. This helps normalize the RGB value of the image and helps the proposed MP model process the pixels of the images. The multilayer perception model consists of several convolutional and pooling layers. The convolutional layer filters the images through 12x12, 10x10, and 9x9 receptive fields. Each consisted of a single-pixel stride or nil-pixel padding. The convolutional operation is provided through the following method,

$$y_j^r = f\left(b_j^r + \sum w_{i,j}^{r-1} * x_i^r\right) \dots \dots \dots (12)$$

Let r be the number of layers in the MP model, and let x_i and y_j be feature maps for the i th and j th input and output. Let $w_{i,j}$ be the convolutional kernel of the x_i and y_j . The bias between the layers is b_j . The convolutional operation is represented as "*". The max pooling is maintained at a 6x6 window. The following equation represents the functionality of the pooling layer.

$$y_{j,k}^i = \max_{0 \leq m, n < 5} (x_{j,5+m, k, 5+n}^i), \dots \dots \dots (13)$$

After the pooling, the y^i of the output feature maps are obtained through the overlapping layer 6x6 that passes input feature maps x^i through the pooling layer. In the end, a softmax layer is used to classify the features obtained from the medical image. Table 1 shows the architecture of the MP model with its layers.

Table 1. Layer architecture of the Multilayer perceptron.

Layer	Input size/stride	Output size
Convolution	12x12/1	129x129x32
Convolution	12x12/1	119x119x32
Max pool	6x6/2	60x60x32
Convolution	10x10/1	52x52x64
Max pool	6x6/2	24x24x64
Convolution	9x9/1	20x20x128
Convolution	10x10/1	10x10x256
Convolution	9x9/1	1x1x256
Rasterize		1x1x4
Softmax		1x1x4

Activation Function

Some of the most used activation functions, like $\tanh f(x) = \tanh(x)$ and sigmoid function $f(x) = (1 + e^{-x})^{-1}$, are considered in various research works. These functions map the larger values with the smaller values, in which the convergence rate is slow and the gradient diffusion problem arises. To overcome these issues, ReLU by Nair and Hinton et al. (2010) was used in those works. However, the softmax function provided a faster convergence rate and less gradient diffusion, which is employed in this paper.

Softmax

The softmax layer predicts the n number of classes that computes the probability of each category, and features obtained from them are rasterized into a single value as x, and the feature vector is represented as

$$P(y = j | x, \theta) = \frac{e^{\theta_j^T x}}{\sum_{j=1}^k e^{\theta_j^T x}} \dots \dots \dots (14)$$

The final target is obtained with k number of classes and with a weight vector θ_j^T .

$$p(y = j | x, \theta) = \frac{e^{\theta_j^T x}}{\sum_{j=1}^k e^{\theta_j^T x}} \dots \dots \dots (15)$$

Traditional Feature Extraction

In medical image classification, the widely extracted features from the CT images are the shape feature, histograms, color maps, and textures, according to Iyatomi et al. (2008), Barata et al. (2014) and Stanley et al. (2007). This also suits the feature extraction from the lung cancer datasets. Most of the traditional MP algorithms make use of global features to classify extracted features. The difference in color and texture helps identify the PPB and is also widely used in global features. The same method was also adopted by Celebi et al. (2007), Rao et al. (2022), and Rubegni et al. (2012), which provided better classification accuracy in their predictions. The same method is adopted in this paper, and we make use of the color changes and textural details to determine lung cancer or the PPB. Textural features

help represent the statistical features of the images, which show the state of the tissues and lesions of the lungs that help identify the exact state of the cancer. It requires the computation of multiple pixels instead of a single pixel, while the color features only involve individual pixels (Ramteke et al., 2012). The texture features in the CT images are extracted by processing the gray-level matrix G in image G. Through this process, the relation of the individual pixels with the neighboring pixels can also be evaluated. To analyze the textural features in the images, the angular second moment (ASM), Contrast (CON), correlation (COR), and entropy (ENT) are calculated. The grayscale matrix can represent the distance of separation and the angle of separation from 0 to 135 degrees.

$$ASM = \sum_i \sum_j G(i, j) \log(G(i, j)) \dots \dots \dots (16)$$

Every element in the matrix is squared and summed up to the angular second moment. The homogeneity and roughness of the image can be obtained from the ASM. If the ASM value is higher, then the roughness of the tissues will be higher.

$$ENT = - \sum_i \sum_j G(i, j) \log(G(i, j)) \dots \dots \dots (17)$$

The entropy function helps calculate the uncertainty in the images. Uncertain information needs to be found and excluded from the processing. This helps in removing the exorbitant values from the list that may affect the averaging process or the efficiency of the prediction.

$$CON = \sum_i \sum_j G(i - j)^2 G(i, j) \dots \dots \dots (18)$$

The Contrast helps measure the clarity of the image and the variation between each pixel in the image. With higher Contrast, images will have more details, and the elements in the images can be easily viewed.

$$COR = \frac{\sum_i \sum_j ij G(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \dots \dots \dots (19)$$

After calculating all these values, the mean and standard deviation of the parameters calculated are obtained, which thus provides a better feature vector for the texture of the image. The color features can be selected through the following Normalization equation, i.e., A_i ,

$$A_i = \frac{1}{N} \dots \dots \dots (20)$$

Feature Fusion

The high-level and traditional features extracted from the images are fused to provide better results. The fixed proportion λ represents the fusion of the R feature. The fusion feature is represented as NF, and the LF (Low frequency) and HF (High Frequency) represent the traditional and high-level features. The parameter %lambda helps to define the importance of LF and HF, and it makes predictions based on the local weights. And finally, it fuses the obtained features. However, this method is a linear method that does not take into account

the non-linearity in the image. It also requires repetition of the same process for different datasets.

$$NF = \lambda.LF + (1 - \lambda).HF \dots\dots\dots (21)$$

$$RF = \max(10) \dots\dots\dots (22)$$

Let $LF = l_1, l_2, \dots, l_i, \dots, l_n$ and $HF = h_1, h_2, \dots, h_j, \dots, h_m$ and the bias between LF and HF be b . RF represents the maximum variable or quantity. To overcome such issues, this paper proposes a new feature fusion method. It helps in training the MP model in a non-linear space. The Proposed MP model uses the fully connected and softmax layer to classify the features. It consists of a kernel function that maps the low-dimensional and high-dimensional features. This, in turn, provides better discriminative features that help identify the PPB from the CT images.

Support vector machines

It is one of the most used supervised learning techniques, and it uses classification and regression to produce accurate results. This algorithm extracts the important features from the input data, as shown in Figure 2. The difference between the obtained features is calculated, and the margin of variation between the features is measured (Wang et al., 2010). In terms of detecting the PPB, we employ both the genetic and CT image datasets to detect and confirm PPB, which is a novel feature of the proposed model. Initially, the SVM algorithm compares the mutated genes with the original DICER1 mutations. Several harmless mutations may not cause PPB, So we detect genes with DICER1 mutations and support them through the CT image analysis, both of which are processed through the SVM algorithm. However, the processing of DNA sequences involves the adoption and plotting of the protein sequences and chromosomes. The variations in these sequences and chromosomes can lead to mutation. The DICER1 mutations are plotted by bisecting them into various protein sequences. Each protein sequence is considered as a feature. The features are not transformed directly but through a kernel substitution method that converts them into a general model. It generates a kernel with substitutes and transforms the input features continuously. This is called the least square support vector machine, and it is an adaptation of the SVM formulation proposed by Vapnik. The w is the original DICER1 gene, the b is the mutated gene, and e is the variation between them. The magnitude of the variation is calculated through the summation of e . The optimization problem of the LS-SVM is done through the following formulas,

$$\min_{w,b,e} J(w, b, e) = \frac{1}{2}w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 \dots\dots\dots (23)$$

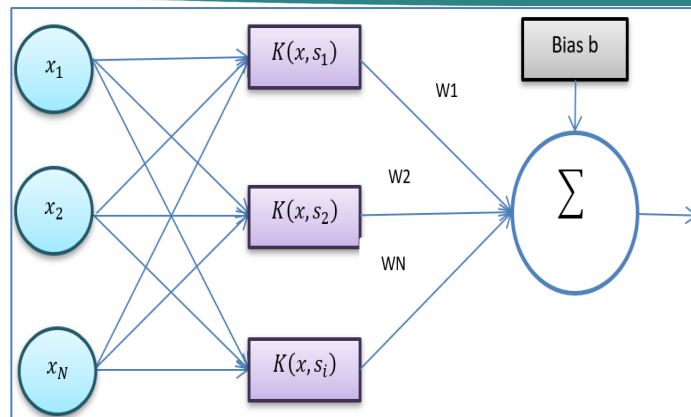


Figure 2. Support Vector Machine.

Based on the equality constraints, the subjects can be written as,

$$y_i[w^T \varphi(x_i) + b] = 1 - e_i, i = 1, \dots, N, \dots\dots\dots (24)$$

The weight of the vectors is W , the regularization parameter is represented as γ , and y_i is the desired DNA pattern of the DICER1 mutation. The required results can be obtained through lagrangian construction. It helps effectively compare the normal genes with the mutated ones. The kernel function is represented as (x, x_i) , and the inner products are computed and transformed. Through this, the classifier obtained is

$$y(x) = \text{sign}[\sum_{i=1}^N \alpha_i y_i K(x, s_i) + b], \dots\dots\dots (25)$$

The Mercer theorem is satisfied by the positive definite kernel, which can be represented as $K(x, x_i) = \varphi(x)^T \varphi(x_i)$. A 10-fold cross-validation process is used to tune the hyper-parameter γ of the LS-SVM classification technique and is shown in Figure 4. Support vector machines are one of the best classification algorithms for segmenting DNA patterns. The progressive tumor characteristics can be found in the miR-214 mutations, which is a kind of DICER1 mutation D. Dettori et al. (2018). These kinds of mutations can be seen in genes of MIRLET7C, MIRLET7B, MIRLET7A1, MIR9-1, MIR708 MIR214 characters. However, these values will be compared and plotted against various genes like MIRLET7A1 MIR9-1 MIR214 MIR140 MIR126 MIR125A, and MIR9-1 MIR483 MIR214 MIR140 MIR126 MIR125A genes. It compares the protein sequences (MIR9-1 in MIR9-1 MIR483 MIR214 MIR140 MIR126 MIR125A) of each gene and plots them and compares them. The variations that cause the mutation and that which affects the lungs and causes cancer are grouped. These grouped protein sequences are considered clusters. It tries to detect the closeness between the clusters to make decisions. Every cluster has a center value that is used to detect the location of the clusters and also helps in comparing them with others. The segmentation values are extracted through a soft function that processes the clusters. Finally, the DICER1

mutations are labelled and kept for CT image analysis results validation.

The SVM algorithm segments the images based on the support provided by the features extracted from the images. This helps in the soft segmentation of the images. For image segmentation, clusters need to be chosen and compared. Unlike the DNA patterns, the image segmentation process involves the processing of pixels that are plotted and grouped as clusters. The i and j represent the lung and muscle regions of the clusters. The proposed algorithm chooses the cluster, which acts as the reference value and helps complete the automation of the system.

$$K_1 = \frac{\sum_i^n (\sum_{i=1}^m \max(I(i,:)))}{N} \dots\dots\dots (26)$$

$$K_2 = \frac{\sum_i^n (\sum_{j=1}^n \max(I(:,j)))}{N} \dots\dots\dots (27)$$

Where the K_1 and K_2 represent the CT images compared and based on the input, the clusters are chosen. Then, they are processed independently, and the results are compared again with the proceeding clusters. Let the rows and columns be defined as m and n , and the non-zero pixels be denoted as N . The clusters represent the pixels, and their difference is computed to form two distinct matrices.

$$kd_1 = [k_1 - I(i,j)]^2 \dots\dots\dots (28)$$

$$kd_2 = [k_2 - I(i,j)]^2 \dots\dots\dots (29)$$

The matrix values of kd_1 and kd_2 are stored in the variables F_1 and F_2 . These values help efficiently segment the images.

$$D = \frac{1}{kd_1} + \frac{1}{kd_2} \dots\dots\dots (30)$$

$$D_1 = (1/kd_1) * D \dots\dots\dots (31)$$

$$D_2 = (1/kd_2) * D \dots\dots\dots (32)$$

$$F_1 = 1/D_1 \dots\dots\dots (33)$$

$$F_2 = 1/D_2 \dots\dots\dots (34)$$

The functioning of the SVM depends on the values obtained in the variables F_1 and F_2 . Both variables contribute to functions that provide the segmentation values.

$$S_1(i,j) = I(i,j), \text{ if } F_2(i,j) < F_1(i,j) \dots\dots\dots (35)$$

$$S_2(i,j) = I(i,j), \text{ if } F_1(i,j) < F_2(i,j) \dots\dots\dots (36)$$

The S_1 represents the images with the cluster id of k_1 compared with the image S_2 with cluster id of k_2 . The values obtained from the S_1 and S_2 are updated in the clusters k_1 and k_2 .

$$k_{1\text{ new}} = \frac{\sum_{ij} (s_1(i,j) \neq 0)}{N_1} \dots\dots\dots (37)$$

$$k_{2\text{ new}} = \frac{\sum_{ij} (s_2(i,j) \neq 0)}{N_2} \dots\dots\dots (38)$$

After comparing both clusters, the images are segmented according to the obtained value. The segmented portions are signified with lines. The lines represent the irregular nature of the diseases. The irregularities in the lungs are the reason for such

irregularities. However, the other structures need to be removed from the images to get an appropriate segmentation of the diseases. The irregular pixels in the images are identified and removed from the images. However, based on the eccentricity and area of the pixels are used to validate them.

$$\text{Eccentricity} = \frac{\max\{\sum_{i=1}^m I(i,:)\}}{\max\{\sum_{j=1}^n I(:,j)\}} \dots\dots\dots (39)$$

$$\text{Area} = \sum_{i=1}^m \sum_{j=1}^n I_b(i,j) \dots\dots\dots (40)$$

The pixels having abnormally high or low eccentricity or area will be removed, and the important areas in the images will be highlighted. After removing all the unwanted pixels, the pixels need to be processed to detect the cancer tissues. The centroid of the 2d image in the x and y direction is:

$$[C_x, C_y] = \left[\frac{\sum_{i=1}^m \sum_{j=1}^n i * I_b(i,j)}{\sum_{i=1}^m \sum_{j=1}^n I_b(i,j)}, \frac{\sum_{i=1}^m \sum_{j=1}^n j * I_b(i,j)}{\sum_{i=1}^m \sum_{j=1}^n I_b(i,j)} \right] \dots\dots\dots (41)$$

After all these processes, the images' Contrast, homogeneity, and auto-correlation are compared to find the area affected by the cancer. In the lung and bone marrow region, there will be calcium deposits in the parenchyma region. This process of formation of calcium deposits is called calcification. It does not affect the lungs or cause PPB in the lungs. However, it may affect the accuracy of the classification process. It needs to be removed from the CT images. The textural features of the images are analyzed to remove the textural features that correspond to the calcification process. It removes the erroneous nodules that are present in the CT images of the 2-dimensional data.

$$\text{Contrast} = \sum_{i=1}^m \sum_{j=1}^n (i-j)^2 * p(i,j) \dots\dots\dots (42)$$

$$\text{Homogeneity} = \frac{\sum_{i=1}^m \sum_{j=1}^n p(i,j)}{1+|(i-j)|} \dots\dots\dots (43)$$

$$\text{Auto - correlation} = \sum_{i=1}^m \sum_{j=1}^n n \frac{(i*j)*p(i,j) - (\mu_x*\mu_y)}{(\sigma_x*\sigma_y)} \dots\dots\dots (44)$$

The Contrast of the images helps to vary the features and detect cancer efficiently. The Contrast of the images also helps differentiate the features and solve the problems caused by PPB. The tissues affected by PPB can be easily viewed by improving the Contrast. Figure 3 shows the contrast value of the images in a scattered plot. The homogeneity of the features helps segment the images very much. The segmentation process needs to be carried out with the features extracted from the images that help in selecting an area in the image to be considered as malignant or benign tissue. The homogeneity equation provides the similarity between the features and helps detect an area that is tumor-affected or PPB. Figure 4 shows the homogeneity of the features considered. The obtained features need to be correlated to assign the overall area as PPB affected area. The SVM algorithm enables auto-correlation of the features and helps solve issues concerning the boundaries to which the

PPB has been affected. Figure 5 shows the Auto-correlation of the images.

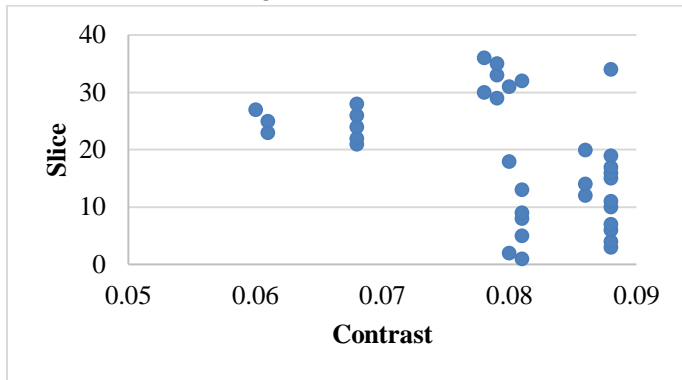


Figure 3. Contrast of the images.

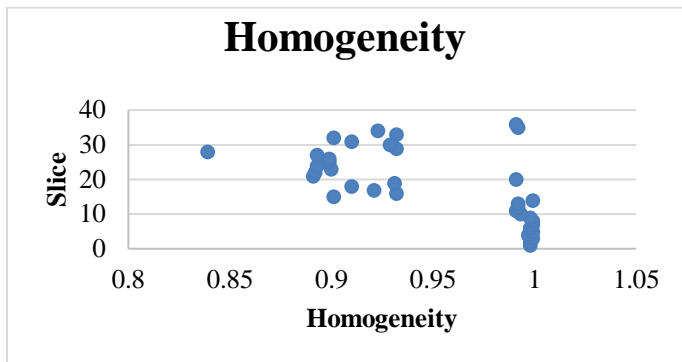


Figure 4. Homogeneity of the images.

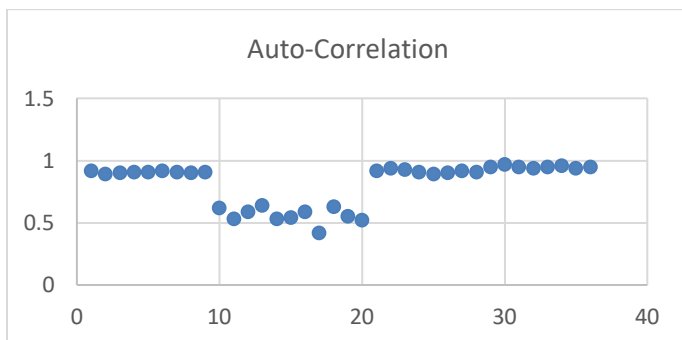


Figure 5. Auto-Correlation of the images.

Based on these parameters, the tumour-affected area is selected and arrived as PPB, and the results obtained are compared with other methods, which can be seen in other sections. The types of PPB cancer detected are discussed in the following sections. It can be seen that various types of PPB differ based on the mutation.

Implementation

Generally, PPB cancer is categorized as benign, aggressive, and malignant and occurs in children. It is also sub-categorized into T1- Pure cystic, T2- cystic and solid, and T3- solid. The lesions of T1 progress into T2 and T3. The central nervous system and bone metastasis are mainly affected by T1 and T3. These cancers are part of the PPB and dysplasia syndrome. The symptoms of the PPB disease can be identified at the age of 4 weeks to 12 months baby. The root cause of the PPB is the passage of

more affected gene copies from their parent to the children. The inheritance of the autosomal chromosome from their parents affects the children. The DICER1 dataset is taken from NCBI to analyze and predict PPB cancer, where more DICER1 data is given for malignant and other subtypes of PPB cancer diseases. It also has several mutations that represent the diseases under PPB. For example, MIRLET7C, MIRLET7B, MIRLET7A1, MIR9-1, MIR708, MIR214 are the genes representing PPB. If it is 31.8% of the total gene, it represents PPB cancer. The 29.9% of MIRLET7A1, MIR9-1, MIR214, MIR140, MIR126, MIR125A gene is called lung disease. The 29.5% of the gene MIR9-1, MIR483, MIR214, MIR140, MIR126, MIR125A is called lung cancer susceptibility. The 33% of the gene ARBP2, MIRLET7C, MIR125B1, DROSHA is called DICER1 syndrome. There are 20 diseases like this and mutations in genes cause these diseases to be related to PPB.

The existing and proposed machine learning algorithms are used to obtain the un-usual pattern of the DNA from DICER1 data. The internal core functions of the learning algorithms are programmed like pattern recognition to obtain the Oddness from the DNA data.

Results and Discussion

The PPB detection was performed using three machine-learning algorithms. An overall comparison was made between the three machine learning algorithms proposed in this work, and an efficient algorithm was derived. The overall experiment was carried out using Python software. It is implemented on a personal computer that has an i7-2630QM Core operating on a Windows 10 operating system at 2 GHz and 4GB of RAM space.

Dataset Description

The Midwest Pediatric Surgery Consortium, a cluster of 11 US tertiary clinics and children's medical centers in 7 adjacent states representing expected overall inhabitants of 58 million persons, authorized a central reliance alliance. These hospitals include Children's Mercy Hospital in Kansas City, Children's Wisconsin in Milwaukee, St. Louis Children's Hospital in St. Louis, and Norton Children's Hospital in Louisville, Kent (2017). Because there was no risk to patients, written permission was not required. With a surgical database of 521 primary lung lesions removed between January 1, 2009, and December 31, 2015, basic CT reports (designated as R0) and pathological reports were consulted (Kunisaki et al., 2021). The above-mentioned CT scans were done using intravenous Contrast contrast numbers of CT scanners having at least 16 slices.

Table 2. Result of SVM classifier.

Class	SVM classifier					
	TP rate	FP rate	Precision	Recall	F-score	Roc Area
1	0.989	0.110	0.953	0.987	0.966	0.942
2	0.883	0.010	0.941	0.883	0.910	0.985
3	0.668	0.001	1.000	0.670	0.804	0.969
4	0.859	0.007	0.948	0.854	0.905	0.947
5	1.004	0.002	1.003	1.001	1.000	1.000
Average	0.966	0.079	0.958	0.955	0.956	0.957

Table 3. Result of LR classifier.

Class	LR classifier					
	TP rate	FP rate	Precision	Recall	F-score	Roc Area
1	0.966	0.238	0.902	0.967	0.935	0.880
2	0.709	0.019	0.804	0.710	0.752	0.894
3	0.336	0.001	1.003	0.337	0.505	0.817
4	0.765	0.013	0.891	0.766	0.825	0.889
5	0.904	0.002	0.946	0.900	0.927	0.962
Average	0.910	0.170	0.899	0.900	0.892	0.885

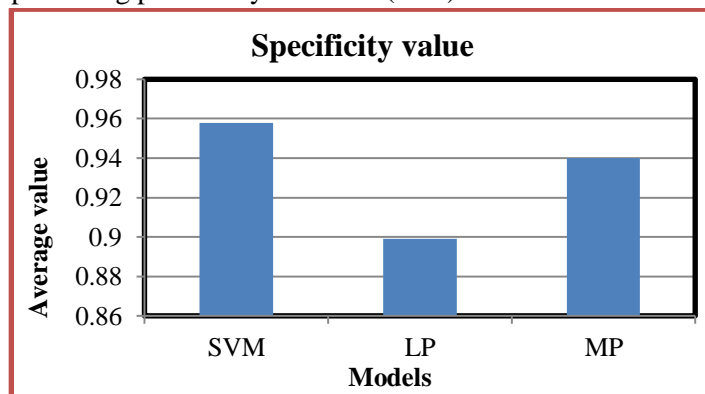
Table 4. Result of MP classifier.

Class	MP classifier					
	TP rate	FP rate	Precision	Recall	F-score	Roc Area
1	0.910	0.035	0.987	0.911	0.946	0.979
2	0.885	0.030	0.752	0.885	0.814	0.990
3	1.000	0.025	0.602	1.000	0.752	0.997
4	0.955	0.018	0.872	0.954	0.911	0.990
5	1.000	0.014	0.912	1.000	0.956	1.000
Average	0.927	0.032	0.940	0.927	0.929	0.986

The DNA dataset for the evaluation of the proposed method is taken from NCBI. The datasets are publicly available, and they provide the necessary genetic data to train the proposed model. The dataset in NCBI is the DICER1 gene that helps differentiate normal DICER1 mutations from cancerous ones. The dataset obtained from NCBI consists of RNA sequences of pleuropulmonaryblastoma that causes cancer. The dataset provides a detailed view of the mutations, protein sequences, and their size, range, and ID.

The machine learning algorithm is used to predict the level of pulmonary blastoma from the input image. The input datasets are classified using three different classifiers: SVM, LR, and MP. Then, using the 10-fold cross-validation process, the training and testing process is performed. For this, 70% of the data is taken for testing, and the remaining 30% of the data is taken for the testing process. After completing the testing process, the confusion matrix is calculated. In the confusion matrix,

true positive (TP), True Negative (TN), false positive (FP), and false negative (FN) values are calculated for the proposed machine learning classifiers. Through the confusion matrix values, the accuracy of each algorithm is evaluated. The main aim of the proposed method is to demonstrate the efficiency of the proposed SVM for predicting pulmonary blastoma (PPB) diseases.

**Figure 6. The average specificity value of the proposed models.**

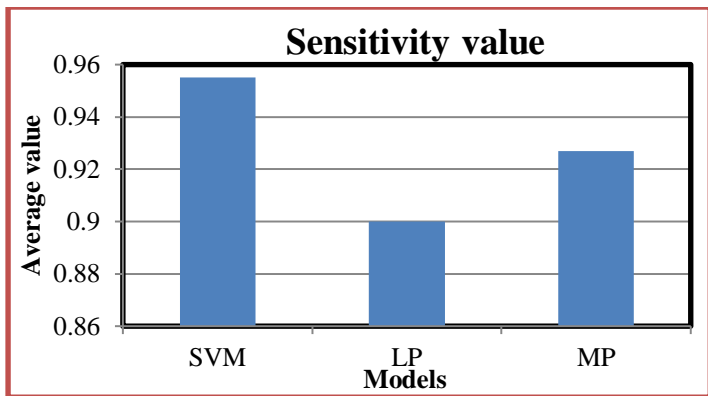


Figure 7. Average Sensitivity value of the proposed model.

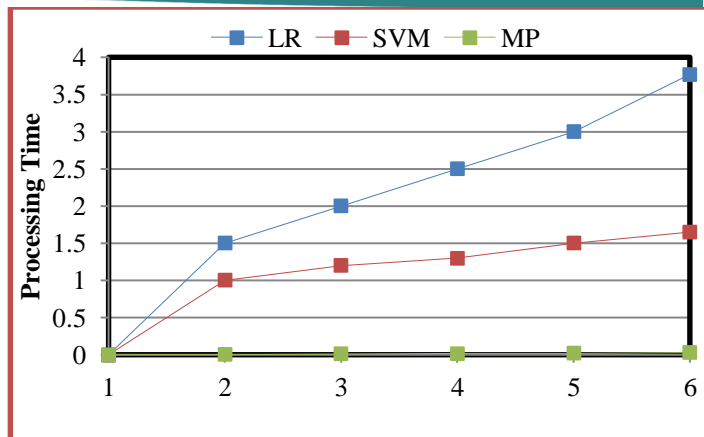


Figure 9. Processing Time of the Classifiers.

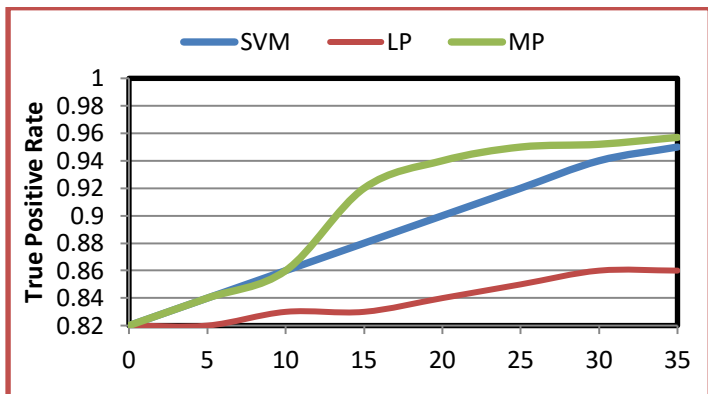
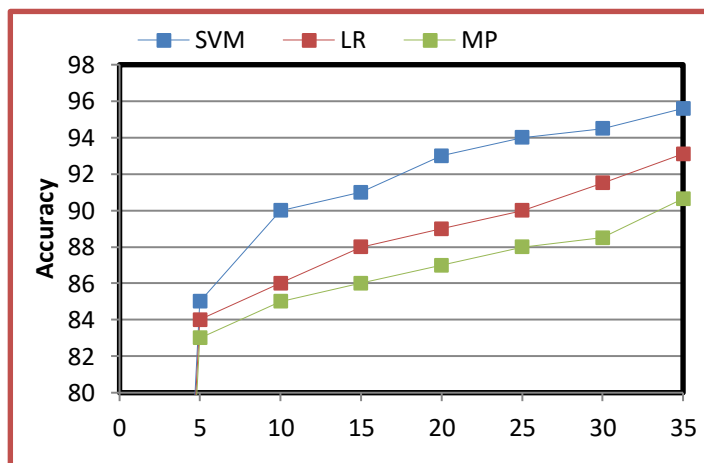


Figure 8. The AOC curve value of the proposed model.

Table 5 shows the computational time of the three classifiers (SVM, LR, and MP) along with the prediction accuracy range. The table shows that compared to the other two methods, the KNN model performs at 1.10 sec time. Likewise, the SVM and CNN models computed the process with 1.65sec and 3.77sec, respectively. Figure 1 graphically represents the computation time of these classifiers.



10. Accuracy Analysis.

Finally, the accuracy percentage of the proposed is evaluated and graphically illustrated in the figure-10. Figure 10 shows that when compared with the other two methods, the SVM model outperforms the other methods. The SVM predicted the PPB cancerous cell from the input medical image with 95.60%, while LR predicted the cell with an accuracy of 93.11%, and the MP provided 90.65% accuracy.

In above Table 2, five different attributes are used to demonstrate the accuracy value of the SVM classifiers. The TP, FP Recall, F-score, ROC, and precision value of the SVM classifiers are evaluated and defined in Table 2. Figures 6, 7, and 8 graphically represent the proposed models' specificity, sensitivity, and AOC values. The proposed SVM classifier achieved a 95.82% prediction ratio compared to other methods. Similarly, the above table-3 and table-4 also define the analysis results of the LR and MP classifiers, respectively. The comparison results of three classifiers are given in Table 5.

Table 5. Comparison result of the models.

Classifier	Comparison Result of the Model By Sensitivity, Specificity, Accuracy		
	Specificity	Sensitivity	Accuracy
SVM	94.35	95.40	95.60
LR	91.45	92.89	93.11
MP	87.65	89.76	90.65

Table 6. Results obtained from Genetic datasets of DICER1 and PPB.

Classifier	Classes	TP rate	FP rate	Precision	Recall	F-score	Roc Area	Specificity	Sensitivity	Accuracy
SVM	1	0.989	0.11	0.953	0.987	0.966	0.942	95.6	94.8	94.2
	2	0.883	0.01	0.941	0.883	0.91	0.985	95.8	95	94.8
	3	0.668	0.001	1	0.67	0.804	0.969	95.9	95.7	95.1
	4	0.859	0.007	0.948	0.854	0.905	0.947	96.5	96.3	95.8
	5	1.004	0.002	1.003	1.001	1	1	96.6	96.3	95.4
LR	1	0.966	0.079	0.958	0.955	0.956	0.957	97.1	96.2	95.7
	2	0.929	0.07	0.923	0.947	0.906	0.922	93.11	92.51	91.81
	3	0.853	0.03	0.931	0.793	0.87	0.925	94.01	93.51	93.31
	4	0.578	0.079	0.97	0.6	0.784	0.889	94.31	93.71	92.81
	5	0.809	0.073	0.928	0.804	0.845	0.877	95.11	94.71	93.81
MP	1	0.914	0.088	0.913	0.921	0.95	0.91	95.31	95.01	94.41
	2	0.976	0.159	0.988	0.965	1.026	1.017	90.65	90.25	90.15
	3	0.989	0.09	1.013	1.007	0.916	0.972	91.45	90.55	89.65
	4	0.923	0.02	1.021	0.813	0.88	0.955	92.35	91.45	91.15
	5	0.658	0.069	1.03	0.64	0.824	0.959	93.25	92.65	92.15
		0.839	0.053	0.948	0.854	0.895	0.927	93.55	92.75	92.05

The machine learning algorithm's performance, as explained in this paper, is also verified by experimenting with the DICER1 dataset. Some of the performance factors, such as sensitivity, specificity, and accuracy, are calculated from the experiment. The obtained result is also compared with the results using CT images. Table 6 provides the performance metrics of the algorithms considered while evaluating the DNA datasets. These datasets provide accurate genetic variations in the cancer tissues, which can be seen from the accuracy of the proposed algorithms. It can also be seen that the proposed algorithm provides better and more efficient results with the genetic datasets. The average accuracy, specificity, and sensitivity are relatively higher compared to the CT images.

Conclusion

In this work, the process of PPB detection is performed with better accuracy. PPB is one of the

uncommon cancerous diseases that usually develop in the chest of children who are below the age of 6. Early detection and treatment of PPB are significant as PPB is capable of taking away the life of budding infants. There are several techniques involved in detecting PPB. Algorithms such as random forests and decision trees do not meet the accuracy level necessary for the timely diagnosis of PPB. Therefore, machine learning techniques such as SVM, LR, and MP are employed in this work to detect PPB. The methods are implemented in Python, and the computational duration is recorded as time detection is required. Also, a comparison between the three methods was made in terms of computational duration and accuracy. While considering these parameters, SVM performed better than the other two methods. The SVM obtained an overall accuracy of 95.60% in detecting PPB with CT images, which is less than the accuracy of DNA datasets (96%). In the future, this research work will be carried out using deep learning algorithms with DNA datasets.

Acknowledgement

I would like to express my gratitude to my supervisor Dr.T.Santhisree, for her immense expertise and knowledge to implement my research.

Conflict of Interest

The authors declare that they have no confidence in their interest.

References

- Addanki, A., Chaitanya, K., Bartakke, S., & Sethuratnam, S. (2017). A case report of pleuropulmonary blastoma presenting as tension pneumothorax. *Indian Journal of Medical and Paediatric Oncology*, 38(1), 70. <https://doi.org/10.4103/0971-5851.203515>
- Barata, C., Ruela, M., Francisco, M., Mendonca, T., & Marques, J. S. (2014). Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems Journal*, 8(3), 965–979. <https://doi.org/10.1109/jsyst.2013.2271540>
- Brodowska-Kania, D., Kotwica, E., Paturej, A., Sośnicki, W., Patera, J., Giżewska, A., & Niemczyk, S. (2016). What do we know about pulmonary blastoma?: A review of literature and clinical case report. *Nagoya Journal of Medical Science*, 78(4), 507.
- Bueno, M. T., Martínez-Ríos, C., De La Puente Gregorio, A., Ahyad, R. A., Villani, A., Druker, H., Van Engelen, K., Gallinger, B., Aronoff, L., Grant, R., Malkin, D., & Greer, M. C. (2017). Pediatric imaging in DICER1 syndrome. *Pediatric Radiology*, 47(10), 1292–1301. <https://doi.org/10.1007/s00247-017-3875-0>
- Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(6), 362–373. <https://doi.org/10.1016/j.compmedimag.2007.01.003>
- Dehner, L. P., Messinger, Y. H., Schultz, K. a. P., Williams, G. M., Wikenheiser-Brokamp, K., & Hill, D. A. (2015). Pleuropulmonary Blastoma: Evolution of an Entity as an Entry into a Familial Tumor Predisposition Syndrome. *Pediatric and Developmental Pathology*, 18(6), 504–511. <https://doi.org/10.2350/15-10-1732-0a.1>
- Dettori, D., Orso, F., Penna, E., Baruffaldi, D., Brundu, S., Maione, F., Turco, E., Giraud, E., & Taverna, D. (2018). Therapeutic silencing of MIR-214 inhibits tumor progression in multiple mouse models. *Molecular Therapy*, 26(8), 2008–2018. <https://doi.org/10.1016/j.ymthe.2018.05.020>
- Grigoletto, V., Tagarelli, A., Sparber-Sauer, M., Koscielniak, E., Orbach, D., Duplan, M., Stachowicz-Stencel, T., Bien, E., López-Almaraz, R., Ben-Ami, T., Pourtsidis, A., Österlundh, G., Mazanek, P., Zsiros, J., Farinha, N., Ferrari, A., & Bisogno, G. (2020). Inequalities in diagnosis and registration of pediatric very rare tumors: a European study on pleuropulmonary blastoma. *European Journal of Pediatrics*, 179(5), 749–756. <https://doi.org/10.1007/s00431-019-03566-7>
- Hirschl, R. B., Minneci, P., Gadepalli, S., Saito, J. M., Askegard-Giesmann, J., Deans, K., Downard, C., Fallat, M., Finnell, M., Helmraath, M., Kabre, R., Lal, D., Leys, C., Mak, G., Ostlie, D., Rescorla, F., Sato, T., St Peter, S., Von Allmen, D., & Warner, B. (2017a). Development of a multi-institutional clinical research consortium for pediatric surgery. *Journal of Pediatric Surgery*, 52(7), 1084–1088. <https://doi.org/10.1016/j.jpedsurg.2016.11.042>
- Hirschl, R. B., Minneci, P., Gadepalli, S., Saito, J. M., Askegard-Giesmann, J., Deans, K., Downard, C., Fallat, M., Finnell, M., Helmraath, M., Kabre, R., Lal, D., Leys, C., Mak, G., Ostlie, D., Rescorla, F., Sato, T., St Peter, S., Von Allmen, D., & Warner, B. (2017b). Development of a multi-institutional clinical research consortium for pediatric surgery. *Journal of Pediatric Surgery*, 52(7), 1084–1088. <https://doi.org/10.1016/j.jpedsurg.2016.11.042>
- Huang, F. L. (2019). Alternatives to logistic regression models in experimental studies. *Journal of Experimental Education*, 90(1), 213–228. <https://doi.org/10.1080/00220973.2019.1699769>
- Iyatomi, H., Oka, H., Celebi, M., Hashimoto, M., Hagiwara, M., Tanaka, M., & Ogawa, K. (2008). An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. *Computerized Medical Imaging and Graphics*, 32(7), 566–579. <https://doi.org/10.1016/j.compmedimag.2008.06.005>
- Khant, A., Dave, Y., Tare, H., Udugade, B., Udugade, S., Choudante, S., & Pulate, C. (2023). Effectiveness of respiratory muscle training on pulmonary function and quality of life in cotton industry workers. *Int. J. Exp. Res. Rev.*, 32, 160-165. <https://doi.org/10.52756/ijerr.2023.v32.013>
- Kim, J. S., & Lee, J. H. (2020). Exceptional response to radiotherapy in unresectable pleuropulmonary blastoma of a child. *Radiation Oncology Journal*,

- 38(2),148–150.
<https://doi.org/10.3857/roj.2020.00290>
- Knight, S., Knight, T., Khan, A., & Murphy, A. J. (2019). Current Management of Pleuropulmonary blastoma: a Surgical perspective. *Children*, 6(8), 86. <https://doi.org/10.3390/children6080086>
- Kunisaki, S. M., Lal, D. R., Saito, J. M., Fallat, M. E., St Peter, S. D., Fox, Z. D., Heider, A., Chan, S. S., Boyd, K. P., Burns, R. C., Deans, K. J., Gadepalli, S. K., Hirschl, R. B., Kabre, R., Landman, M. P., Leys, C. M., Mak, G. Z., Minneci, P. C., Wright, T. N., & Helmuth, M. A. (2021). Pleuropulmonary blastoma in pediatric lung lesions. *Pediatrics*, 147(4). <https://doi.org/10.1542/peds.2020-028357>
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International conference on machine learning (ICML-10), Haifa, Israel, 2010, pp. 807-814.
- National Center for Biotechnology Information (NCBI). (n.d.). Official full name: Dicer 1, ribonuclease III (Homo sapiens (human)). Retrieved from <https://www.ncbi.nlm.nih.gov/gene/23405>.
- National Center for Biotechnology Information (NCBI). (n.d.). Gene expression omnibus: GSE163678. Retrieved from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163678>.
- Ramteke, R. J., & Monali, K. Y. (2012). Automatic medical image classification and abnormality detection using k-nearest neighbour. *International Journal of Advanced Computer Research*, 2(4), 190.
- Rao, M. S., Kumar, S. P., & Rao, K. S. (2023). Effective medical leaf identification using hybridization of GMM-CNN. *International Journal of Experimental Research and Review*, 32, 115–123. <https://doi.org/10.52756/ijerr.2023.v32.009>
- Rao, M. S., UmaMaheswaran, S., Sattaru, N. C., Abdullah, K. H., Pandey, U. K., & Biban, L. (2022). A critical understanding of integrated artificial intelligence techniques for the healthcare prediction system. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). <https://doi.org/10.1109/icacite53722.2022.9823678>
- Reddy, N. S., & Khanaa, V. (2023). Diagnosing and categorizing of pulmonary diseases using Deep learning conventional Neural network. *Int. J. Exp. Res. Rev.*, 31(Spl Volume), 12-22. <https://doi.org/10.52756/10.52756/ijerr.2023.v31sp1.002>
- Rubegni, P., Cevenini, G., Nami, N., Argenziano, G., Saida, T., Burrioni, M., Bono, R., Quaglino, P., Barbini, P., Miracco, C., Lamberti, A., & Fimiani, M. (2012). Dermoscopy and digital dermoscopy analysis of palmoplantar ‘Equivocal’ pigmented skin lesions in Caucasians. *Dermatology*, 225(3), 248–255. <https://doi.org/10.1159/000343928>
- Stanley, R. J., Stoecker, W. V., & Moss, R. H. (2007). A relative color approach to color discrimination for malignant melanoma detection in dermoscopy images. *Skin Research and Technology*, 13(1), 62–72. <https://doi.org/10.1111/j.1600-0846.2007.00192.x>
- Van Engelen, K., Villani, A., Wasserman, J. D., Aronoff, L., Greer, M. C., Bueno, M. T., Gallinger, B., Kim, R. H., Grant, R., Meyn, M. S., Malkin, D., & Druker, H. (2017). DICER1 syndrome: Approach to testing and management at a large pediatric tertiary care center. *Pediatric Blood & Cancer*, 65(1). <https://doi.org/10.1002/pbc.26720>
- Wang, H., & Huang, G. (2010). Application of support vector machine in cancer diagnosis. *Medical Oncology*, 28(S1), 613–618. <https://doi.org/10.1007/s12032-010-9663-4>
- Yavanamandha, P., Keerthana, B., Jahnavi, P., Rao, K. V., & Kumar, C. R. (2023). Machine Learning-Based Gesture Recognition for Communication with the Deaf and Dumb. *International Journal of Experimental Research and Review*, 34 (Special Vol.),26–35. <https://doi.org/10.52756/ijerr.2023.v34spl.004>

How to cite this Article:

Raswitha Bandi and T. Santhisri (2024). Detection of Pleuro Pulmonary Blastoma using Machine Learning Models. *International Journal of Experimental Research and Review*, 40(spl.), 151-163.

DOI: <https://doi.org/10.52756/ijerr.2024.v40spl.012>



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.