Punithavathy K

Research Scholar School of Electrical Sciences Hindustan Institute of Technology and Science Chennai, India

Sumathi Poobal

Professor Department of Electronics and Communication Engineering KCG College of Technology Chennai, India

M. M. Ramya

Professor Centre for Automation & Robotics Hindustan Institute of Technology and Science Chennai, India

Performance Evaluation of Machine Learning Techniques in Lung Cancer Classification from PET/CT Images

Lung cancer detection is highly challenging as it is asymptomatic till advanced stage. Early lung cancer detection helps to increase the patient's survival. Computer Aided Diagnosis (CAD) systems have been developed using Machine Learning (ML) and Artificial Intelligence (AI) techniques in detecting malicious regions from medical images. This study is intended to compare the classical ML techniques for lung cancer classification from Positron Emission Tomography/Computed Tomography (PET/CT) images. Significant texture and fractal descriptors extracted from PET/CT images generate non-linear data and were fed as inputs to the classifier. Various hyper-parameters and model parameters for the ML techniques have been tuned to fix optimal parameters for better performance. 10-fold cross validation was used to analyze the performance of the classifiers. Experimental study showed that Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel of width, $\sigma = 1$ outperformed and achieved highest accuracy of 98.10%.

Keywords: Computer aided diagnosis; lung cancer; PET/CT; texture features; fractal analysis; machine learning

1. INTRODUCTION

PET/CT is a supreme imaging technique in lung cancer diagnosis, staging and treatment planning as it provides complementary morphological and anatomical information [1]. Major drawback in PET/CT is the False Positives (FPs) due to inflammations or infection caused by other lung diseases [2]. Also, in PET/CT, the CT scans performed at minimal energy settings to reduce the radiation effects result in poor image quality; which affects interpretations and diagnostic accuracy [3].

Huge volume of images burdens the medical experts in interpretation which may result in error prone diagnosis. Therefore, CAD systems are highly essential to assist radiologists in lung cancer diagnosis with increased speed, accuracy, minimum diagnostic errors and less time [4]. CAD plays a major complementary role in medical diagnosis [5]. CAD systems developed using image processing techniques and AI for lung image interpretations using various image modalities. Current state-of- art techniques in lung cancer CAD systems, challenges faced in implementation steps, strengths and drawbacks of the existing approaches were reviewed and presented [6].

Texture Analysis (TA) plays a key role in discriminating between normal and abnormal regions in medical images. Texture features are found beneficial in image classification [7]. Critical comprehensions into the latest development of TA for quantifying the heterogeneity in PET/CT images were provided and offered recommendations for the use of TA in clinical research [8].

Received: August 2018, Accepted: October 2018 Correspondence to: Punithavathy K, School of Electrical Sciences, Hindustan Institute of Technology and Science, Chennai, 603103, India E-mail: punithavm@hindustanuniv.ac.in doi: 10.5937/fmet1903418P © Faculty of Mechanical Engineering, Belgrade. Allrights reserved First and second order statistical texture features to detect the lung cancer from CT images were analyzed [9]. However, texture features might be insufficient to identify and classify smaller lymph nodes. Medical images are characterized by irregular complex tissue structures which cannot be quantified by traditional Euclidean geometry. Hence, Fractal geometry has been used widely to analyze medical images and found significant to find the aggressiveness of the lung cancer [10].

Classification algorithms play an important role in CAD systems in detecting the malicious regions from medical images. Supervised ML techniques such as Artificial Neural Networks (ANN) and SVM have been recognized as robust models in classifying multi-dimensional, complex data. ANN based lung cancer detection system has been established and utilized first order texture features extracted from CT images [11]. Significant texture features yielded a better accuracy of 98.4% using ANN compared to 93.2 % by SVM [12]. Classification using Hop field neural network achieved 98% accuracy in detecting lung cancer [13]. Recent studies have demonstrated that SVM has been found as a popular ML tool in classification of medical images. A new computerized system for lung cancer classification was developed by several researchers using SVM classifier to improve the diagnostic accuracy and tested on texture features of PET/CT images [14 - 18]. Based on these deliberations, this study has been attempted to compare several ML methods for lung cancer classification using texture and fractal features from lung PET/CT images.

2. MATERIALS AND METHODS

2.1 Image Dataset

This method involved a retrospective study 18F-fluorodeoxyglucose (FDG) PET/CT images for lung

cancer detection and classification. Patients underwent PET/CT scan after administering an IV injection and images were acquired on a GE Discovery 16 slice PET/CT scanner. 1312 retrospective PET/CT images from 82 patients in the age group of 27 to 71 years, captured during 2015 to 2016 were utilized for evaluating the ML methods. Two mutually exclusive labels were considered in this study: (i) normal, i.e. images presenting no radiological abnormalities; (ii) abnormal, i.e. images reported as presenting nodules /lesions or masses. Images were collected from Anderson Diagnostics & Labs, Chennai, which include 34 normal studies and 48 abnormal studies which present Non-Small Cell Lung Cancer (NSCLC) in various stages. CT and PET images are of dimension 512×512 \times 178 and 128 \times 128 \times 178 respectively. PET/CT images are in Digital Imaging and Communications in Medicine (DICOM) format. DICOM PET/CT images were converted into JPEG images of dimension 256 \times 256×178 by the nuclear radiologists and provided for this study. Statistics of the data collected for this study is listed in Table 1.

Table 1: Patient Data

Parameter	Statistics
Patients count (male/female/total)	50/32/82
Patient age (min/max/median)	27/71/53
Benign / Malignant images count	34/48
Malignant (Stage II / III / IV)	12/21/15

Few normal and lung cancer PET/CT images in various stages used in this study are depicted (Figure 1a and Figure 1b).

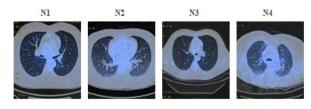


Figure 1a. Normal lung PET/CT images

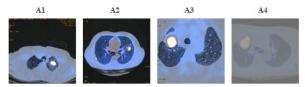


Figure 1b. Lung cancer PET/CT images in various stages

Abnormal image A1 has a lesion in left lung measuring 31×35 mm and in stage II. Image A2 is stage III lung cancer image having a lesion in left lung of size 35×28 mm. A3 is stage IV lung cancer image with a lesion in right lung measuring 37×43 mm. Image A4 is a poor contrast abnormal, stage IV lung cancer image with a lesion in right lung of 60×39 mm.

2.2 Machine learning techniques

Huge amount of data has been obtained due to the remarkable advancements in image acquisition devices and volume of images per single study. This may liable to human error and may have inter-observer variations in interpretations. ML techniques have made a remarkable progress in recent times and played a vital role in medical image analysis and assist medical experts for enable speedy, automatic diagnosis and with improved accuracy. Supervised learning algorithm develops a model based on training and makes predictions for new test data which are not in training set. These algorithms require a set of parameters tuned to optimal values that enable the model for best fitting of data and to perform in a best possible way. Various classical ML techniques were analyzed for comparative performance evaluation in lung cancer classification.

The classical ML methods covered Decision Trees (DT), Discriminant Analysis (DA), Naïve Bayes (NB), K-Nearest Neighbor (KNN), ANN and SVM.

Decision Tree is a popular ML algorithm and graphical representation which makes decisions based on certain conditions. Optimal decisions can be arrived by traversing through forward and backward calculation paths. This method analyzed 3 models such as simple decision tree with maximum 4 splits, medium tree with maximum 20 splits and complex tree with maximum 100 splits. Medium tree with 20 splits with Gini's diversity index as split criterion helped in making coarse distinction between normal and cancerous regions of lung PET/CT images.

Discriminant Analysis assigns objects to different classes based on the constructed discriminant equations. It predicts the class of a new set of inputs by maximizing the distance between the mean of each class and minimize the spreading within the class itself. Linear DA (LDA) is the preferred classification algorithm to classify linear data with more than two classes. Quadratic DA classifier is used to classify two or more classes by a quadratic surface.

K-Nearest Neighbors is the most widely used nonparametric technique in classification. It is a simple algorithm in which entire training dataset and all possible cases are stored. A new case is being assigned to a class and prediction is made by its 'K' nearest neighbors measured by a similarity. Euclidean distance metric was used to establish the nearest neighbors. Performance and behavior of the algorithm depends upon the value of 'K' and its selection depends upon the data. This study tried with various odd values of 'K' ranging from 1 to 13 (square root of the training samples/2) and analyzed the performance.

Naive Bayes Classifier is the most popular supervised learning method based on Bayes Theorem of Probability to predict each class. Naive Bayes classifier predicts with the assumption that all the features are unrelated to each other. NB algorithm is useful for a moderate to large training data set with several features.

Neural Networks is a promising area, successfully applied in medical field for cancer detection and to identify pathological conditions. ANNs contain input, hidden and output layers and each neuron in hidden layer is fully connected to every neuron in input and output layer. Initially, ANN is trained with known set of input and outputs using back propagation or gradient descent learning algorithms. Back propagation learning algorithms are most widely used in which the error is calculated and back propagated to the previous layers to adjust the weights and biases so that the mean square error (MSE) is minimum.

In constructing a model using ANN, the hyperparameters such as learning rate, momentum, number of hidden layers and number of hidden neurons have been tuned and optimal values were fixed based on the experimental results for MSE and accuracy. Learning rate of 0.3 and a momentum of 0.9 produced better performance with low MSE for most of the training functions. Performance goal of 0.0001 and 1000 epochs were set. The network was trained with 10, 15, 20, 25, 30 and 35 hidden neurons and observed the performance [19]. It is observed that 20 hidden neurons produced minimum MSE for many training functions. Hence, 20 hidden neurons were used.

SVMs now have become a popular tool in medical diagnosis [15, 16]. SVM is a supervised ML algorithm in which the dataset teaches SVM about the classes. New data is classified into various classes by finding a hyperplane using maximum margin rule.

The hyperparameters of SVM such as type of kernel and kernel width in RBF kernel have been investigated carefully to attain improved classification accuracy. Performance of SVM-RBF model is sensitive to the kernel width σ . Small width may cause over-fitting as the data points are close to hyperplane. Large width may end up with under-fitting, as the data points are very far from hyperplane. The optimal width is selected based on the accuracy. Performance of SVM with different kernels has been analyzed to identify a better classifier. Kernels are used in non-linear SVM to map the features into high dimensional space so that the linearly nonseparable data are linearly separable. SVM with RBF kernels is the most preferred classifier for complex, non-linear data as the training problem is convex. SVM hyperparameter tuning would help in achieving better classification accuracy.

2.3 Input features for ML methods

In this study, 14 second order texture features were calculated by using gray level co-occurrence matrix (GLCM). TA on lung PET/CT images resulted in identification of 3 significant texture features, which include auto correlation, sum average and sum variance [20]. 12 fractal features were calculated for PET/CT, which include Fractal dimension average (FD_{avg}) and lacunarity from fuzzy enhanced and modified images. Experimental study revealed that 10 significant fractal features were useful in lung cancer detection [21]. All these features were extracted from enhanced PET/CT images using Wiener filtering followed by fuzzy enhancement. This study identified 13 significant features which include 3 texture and 10 fractal features extracted from PET/CT images. A total of 1072 samples which contain samples from normal, at cancer boundaries and cancerous regions in a ratio of 50:25:25 are used for training, validation and testing the classifiers, in which normal and diseased samples hold the ratio 55:45.

The sample data generated with the texture and fractal features from lung PET/CT images was sparse,

diverse, multi-dimensional, linearly non-separable and slightly unbalanced

3. RESULTS

This study analyzed lung PET/CT images retrospectively to detect and classify the lung cancer using texture and fractal feature descriptors. MATLAB R2013a was used to implement this method and tested on lung PET/CT images of size 256×256 . Noise in PET/CT images due to the artefacts caused by metallic implants, respiratory motion and application of contrast degrade the image quality and the diagnostic accuracy. These artefacts lead to the possibility of FP results due to more uptake of radioactive tracer. Therefore, efficient de-noising techniques are needed for PET/CT images in lung cancer diagnosis.

Wiener filter was employed in de-noising the additive noise present in PET/CT images with preserving sharp edges and textures. Adaptive fuzzy image enhancement was designed and developed to enhance the contrast between different scale regions; especially for poor contrast images.

Extracted combined significant texture and fractal features were used in classification. These observations were arranged as a matrix and fed as input to the supervised classifier. 6 supervised classifiers including DT, DA, KNN, NB classifier, ANN and SVM with various kernels were tried and their performance was analyzed for finding the best classifier in classifying the lung cancer from PET/CT images. Ratio between training and testing dataset is 70:30. Performance measures such as sensitivity, specificity and accuracy obtained for all 6 classifiers were listed (Table 2).

 Table 2. Comparison of performance measures of various classifiers analyzed in this method

Classifier Name	Sensitivity (%)	Specificity (%)	Accuracy (%)
Decision Trees	95.61	94.28	96.19
Linear DA	83.67	60.33	73.10
Quadratic DA	73.50	80.54	75.90
KNN	94.61	96.19	95.81
(K = 11)			
Naïve Bayes	98.61	91.42	96.26
ANN	92.60	92.90	92.90
SVM (RBF,	98.80	97.90	98.10
$\sigma = 1.0)$			

Figure 2 shows the plot of sensitivity, specificity and accuracy for the supervised classifiers analyzed in this study. Performance of the LDA and QDA classifiers are unsatisfactory with the nonlinearly separable data. ANN resulted in an accuracy of 92.9%. DT, KNN and NB classifiers produced a closer accuracy of around 96%. Out of all 6 classifiers analyzed in this study, it is demonstrated from the results that SVM with RBF kernel, $\sigma = 1.0$ outperformed all classifiers.

RBF kernel width σ has an impact on the accuracy of SVM classification significantly, particularly in the interval [0.5, 1.5]. SVM reasonably classified the data points whose distance from support vectors were equal to the width σ . Lower kernel widths such as 0.1 and 0.5 resulted in over fitting. In this case, the distance between

data points and support vectors were greater than σ , thus produced lower accuracy. Higher kernel widths like 1.5 and 2.0 caused under fitting, as distance between data points and support vectors were smaller than σ . In this SVM-RBF model, $\sigma = 1.0$ produced maximum accuracy. RBF kernel of width $\sigma = 1$ lead to smooth, sharp decision boundaries and generalize strongly without much under and over fitting. ROC curve of SVM-RBF ($\sigma = 1$) classifier is shown (Figure 3). SVM Classifier produced a sensitivity of 98.8% and specificity of 97.9%.

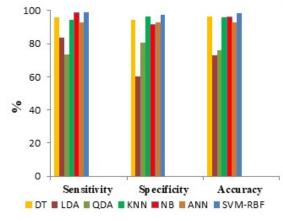


Figure 2. Performance comparison of various classifiers

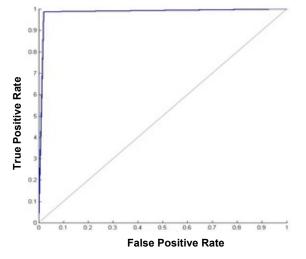


Figure 3. ROC curve of SVM with combined features

ROC curve is almost close to the gold standard of upper left corner, thus providing the highest accuracy. Overall accuracy attained is 98.10%, which is superior to the performance of various classifiers analyzed in this study. Increase in number of training samples unaltered the accuracy in SVM as SVM is least sensitive to sample size as it uses only support vectors to build the separating hyperplane.

4. DISCUSSION

Existing CAD systems in classifying lung cancer from PET/CT images utilized texture features and classical supervised classifiers ANN and SVM. These methods utilized thresholding for lung segmentation and extracted the first and second order texture features. All these methods used 10-fold cross validation for evaluating the performance of classifiers. The diagnostic accuracy of existing CAD systems established for lung

cancer classification from PET/CT images were compared with the results of this study utilizing texture and fractal features from PET/CT images. The observations are listed in Table 3.

 Table 3. Comparison of existing and developed CAD

 system in lung cancer classification using PET/CT images

Contributors	Features	Classifier	Accuracy (%)
Gao et al (2015)	Statistical texture features	SVM with RBF Kernel	86
Guo et al (2015)	Multiple image features, SUV, heterogeneity, CT textures	SVM	Augmented accuracy
Gutte et al (2007)	Intensity, Diagnostic features (SUV)	ANN	92
Wang et al (2017)	Diagnostic features	CNN	86
Zhao et al (2015)	First order texture features, SUV	SVM with RBF	Sensitivity of 95.6
Developed CAD system by this method	Second order texture fea- tures and fractal features	SVM with RBF	98.10

Methods developed by (Gao et al, 2015; Wang et al, 2017) resulted in an accuracy of 86% whereas the methods by (Guo et al, 2015; Zhao et al, 2015) produced comparable accuracy of 92 and 95.6% respectively. Even though all these methods yielded a better accuracy, texture features might not be suitable for classifying small size lymph nodes (Gutte et al, 2007). Fuzzy enhancement designed for current study well enhanced the high textures with clear edges and facilitated to increase the diagnostic accuracy. Fractal features aided superior detection of cancerous regions and SVM is robust with strong generalization ability for moderate data using combined texture and fractal features for current study. It is clear from Table 3 that the developed CAD system by this method for lung cancer detection and classification from PET/CT images using SVM classifier with RBF kernel, $\sigma = 1$ achieved enriched accuracy of 98.10%.

5. CONCLUSION

Comparison of various ML techniques in lung cancer classification was envisioned in this study. CAD system was developed with texture and fractal descriptors from enhanced PET/CT images. Fuzzy image enhancement offered improved contrast for regions of high textures and found useful in lung cancer detection. Texture features based on statistical approach and fractal analysis were found highly useful for detecting and classifying distinct and subtle textures in dual-modality medical images. Significant features were selected from the extracted feature set based upon their relevance and contribution in lung cancer detection. This method identified 3 texture features and 10 fractal features as significant to detect and classify lung cancer from PET/CT images. The dataset obtained with these features were linearly non-separable. Several classical ML techniques such as DT, DA, NB, KNN, ANN and SVM have been considered for performance comparison. It is also observed that the classification accuracy produced by KNN and DT is comparable with the results by SVM. However, the specificity produced by SVM is higher than KNN and DT, thus yielding better classification accuracy. The RBF kernel used in SVM classifier mapped the linearly non-separable features well into linearly separable.

SVM classifier is robust with strong generalization ability for moderate data using combined texture and fractal features. SVM with RBF kernel, width $\sigma = 1$ yielded an accuracy of 98.10%; thus, concluding that SVM is the better classifier for lung cancer diagnosis compared to other classifiers analyzed in this method.

ACKNOWLEDGEMENTS

The authors express their sincere thanks to the nuclear radiologists, Anderson Diagnostics & Labs, Chennai for providing the image dataset, valuable inputs and critical validation. Authors also thank the management of Hindustan Institute of Technology & Science for their motivation and encouragement to carry out the research work.

REFERENCES

- [1] De Guevara Hernández, D.L.: The role of PET/CT imaging in lung cancer, Journal of Cancer Therapy, Vol. 6, No. 8, pp. 690, 2015.
- [2] Hochhegger, B., Alves, G. R. T., Irion, K. L., Fritscher, C. C., Fritscher, L. G., Concatto, N. H. and Marchiori, E.: PET/CT imaging in lung cancer: indications and findings, Jornal Brasileiro de Pneumologia, Vol. 41, No. 3, pp. 264-274, 2015.
- [3] http://eradiology.bidmc.harvard.edu.
- [4] Van Ginneken, B., Schaefer-Prokop, C.M., Prokop, M.: Computer-aided diagnosis: how to move from the laboratory to the clinic, Radiology, Vol. 261, No. 3, pp 719-732, 2011.
- [5] Doi, K.: Computer-aided diagnosis in medical imaging: historical review, current status and future potentia, Computerized Medical Imaging and Graphics, Vol. 31, No. 4-5, pp. 198-211, 2007.
- [6] El-Baz, A., Beache, G.M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., Soliman, A., Abdollahi, B.: Computer-aided diagnosis systems for lung cancer: challenges and methodologies. International Journal of Biomedical Imaging, 2013.
- [7] Haralick, R. M. and Shanmugam, K.: Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, Vol. 3, No. 6, pp. 610-621, 1973.
- [8] Hatt, M., Tixier, F., Pierce, L., Kinahan, P. E., Le Rest, C. C. and Visvikis, D.: Characterization of PET/CT images using texture analysis: the past, the present... any future?, European Journal of Nuclear Medicine and Molecular Imaging, Vol. 44, No. 1, pp. 151-165, 2017.

- [9] Zayed, N. and Elnemr, H.A.: Statistical analysis of haralick texture features to discriminate lung abnormalities, Journal of Biomedical Imaging, Vol. 2015, pp 12, 2015.
- [10] Al-Kadi, O.S. and Watson, D.: Texture analysis of aggressive and nonaggressive lung tumor CE CT images, IEEE Transactions on Biomedical Engineering, Vol. 55, No. 7, pp. 1822-1830, 2008.
- [11] Kuruvilla, J., Gunavathi, K.: Lung cancer classification using neural networks for CT images, Computer Methods and Programs in Biomedicine, Vol. 113, No. 1, pp. 202-209, 2014.
- [12] Kohad, R. and Ahire, V.: Application of machine learning techniques for the diagnosis of lung cancer with ANT colony optimization, International Journal of Computer Applications, Vol. 113, No. 18, pp. 34-41, 2015.
- [13] Taher, F., Sammouda, R.: Lung cancer detection by using artificial neural network and fuzzy clustering methods,: in: *Proceedings of GCC Conference and Exhibition* (GCC), 2011 February, IEEE, pp. 295-298.
- [14] Gutte, H., Jakobsson, D., Olofsson, F., Ohlsson, M., Valind, S., Loft, A., Edenbrandt, L. and Kjær, A. : Automated interpretation of PET/CT images in patients with lung cancer. Nuclear Medicine Communications, Vol. 28, No. 2, pp. 79-84, 2007.
- [15] Gao, X., Chu, C., Li, Y., Lu, P., Wang, W., Liu, W. and Yu, L.: The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer. European Journal of Radiology., Vol. 84, No. 2, pp. 312-317, 2015.
- [16] Guo, N., Yen, R. F., El Fakhri, G. and Li, Q.: SVM based lung cancer diagnosis using multiple image features in PET/CT, in: *Proceedings of the Nuclear Science Symposium and Medical Imaging Conference* (NSS/MIC), October 2015, IEEE, pp.1-4.
- [17] Zhao, J., Ji, G., Qiang, Y., Han, X., Pei, B., Shi, Z.: A new method of detecting pulmonary nodules with PET/CT based on an improved watershed algorithm, Plos One, Vol. 10, No. 4, pp e0123694, 2015.
- [18] Wang, H., Zhou, Z., Li, Y., Chen, Z., Lu, P., Wang, W., Liu, W. and Yu, L.:Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images, EJNMMI Research, Vol. 7, No. 1, pp. 11, 2017.
- [19] Punithavathy, K., Sumathi Poobal, M. M. Ramy: Artificial neural network based lung cancer detection for PET/CT Images, Indian Journal of Science and Technology, Vol. 10, No. 42, pp. 1-13, 2018.
- [20] Punithavathy, K., M. M. Ramya and Sumathi Poobal: Analysis of statistical texture features for automatic lung cancer detection in PET/CT images,: in: Proceedings of International Conference on Robotics, Automation, Control and Embedded Systems (RACE), February 2015, pp 1-5.

[21] Punithavathy, K., Sumathi Poobal, M. M. Ramya: Fractal based lung cancer detection from fuzzy enhanced PET/CT Images, International Journal of Pure and Applied Mathematics, Vol. 117, No. 7, pp 427–444, 2017.

ЕВАЛУАЦИЈА ПЕРФОРМАНСИ ТЕХНИКА МАШИНСКОГ УЧЕЊА КОД КЛАСИФИКАЦИЈЕ КАРЦИНОМА ПЛУЋА НА ОСНОВУ СНИМАКА РЕТ/СЕТ СКЕНЕРА

Пунитавати К, С. Побал, М.М. Рамја

Откривање карцинома плућа представља велики изазов јер је болест асимптоматска до узнапредовале фазе. Рано откривање карцинома повећава шансе пацијента за преживљавање. САD системи су развијени уз помоћ ML и AI техника за откривање малигних регија на основу снимака скенера.

Циљ овог рада је да се изврши поређење између класификације карцинома плућа применом класичних ML техника и слика добијених помоћу PET/CT снимака. Сигнификантни дескриптори текстуре и сегмената добијени PET/CT скенером дају нелинеарне податке који су употребљени као инпут за класификатор.

Различити хиперпараметри и параметри модела за технике ML су подешени да буду оптимални параметри бољих перформанси. Анализа перформанси класификатора је обављена десет пута унакрсном евалуацијом. Експериментална проучавања су показала да је SVM са језгреном RBF ширине $\sigma = 1$ била знатно боља и постигла тачност од 98,1%.