ORIGINAL ARTICLE

A comprehensive exploration of deep learning approaches for pulmonary nodule classification and segmentation in chest CT images

Murat Canayaz[1] · Sanem Şehribanoğlu[2] · Mesut Özgökçe[3] · M. Bilal Akıncı[3]

Abstract
Accurately determining whether nodules on CT images of the lung are benign or malignant plays an important role in the early diagnosis and treatment of tumors. In this study, the classification and segmentation of benign and malignant nodules on CT images of the lung were performed using deep learning models. A new approach, C+EffxNet, is used for classification. With this approach, the features are extracted from CT images and then classified with different classifiers. In other phases of the study, a segmentation between benign and malignant was performed and, for the first time, a comparison of nodes was made during segmentation. The deep learning models InceptionV3, DenseNet121, and SeResNet101 were used as backbone models for feature extraction in the segmentation phase. In the classification phase, an accuracy of 0.9798, a precision of 0.9802, a recognition of 0.9798, an $F1$ score of 0.9798, and a kappa value of 0.9690 were achieved. During segmentation, the highest values of 0.8026 Jacard index and 0.8877 Dice coefficient were achieved.

Keywords Deep learning · Pulmoner nodules · Classification · Segmentation

1 Introduction

When the statistical data published by the WHO International Cancer Research Unit are examined, the most common cancers in all age groups in the world are breast, prostate,and lung cancer [1]. While 5-year survival rates in breast and prostate cancer are 90% and 100% [2–4], 5-year survival rate in lung cancer is 18–20%, and it is the cancer with the highest mortality [3, 5] [6]. While it was stated in the articles of 2007 that 5-year survival rates in lung cancers diagnosed at an early stage reached 60–70%, these rates vary between 20.2 [3] and 22% [2] when it comes to

today. In this respect, early diagnosis of lung cancer is more important than other cancers.

Computed tomography images are used most frequently in lung cancer screenings and lesions are classified according to their visual characteristics. It is not always possible to detect a lesion at an early stage and to differentiate malignant benign with high accuracy [7]. The development of computer-aided diagnosis systems is thought to be beneficial for doctors in terms of early and accurate diagnosis. Therefore, it is very important to find an effective method to classify and segment nodules. Traditional methods may be insufficient to analyze such images, and therefore studies on classification and segmentation of nodules using deep learning (DL) methods have been carried out in recent years. DL methods consist of neural networks with the ability to learn datasets, and due to these networks, they have performed well in classifying and segmenting nodules. In particular, convolutional neural network (CNN) methods have been effective in extracting and classifying the features of images, greatly improving the quality and reliability of treatment, especially the early diagnosis and screening process [8].

✉ Murat Canayaz
  mcanayaz@yyu.edu.tr

1 Department of Computer Engineering, Van Yuzuncu Yil University, 65100 Van, Turkey

2 Department of Econometrics, Van Yuzuncu Yil University, 65100 Van, Turkey

3 Department of Radiology, Faculty of Medicine, Van Yüzüncü Yıl University, Van, Turkey

In this study, DL methods were used for the classification and segmentation of nodules on lung CT images, and the effectiveness of these methods was investigated. Studies on this subject may help early diagnosis and treatment of nodules by showing good performance in terms of accuracy and sensitivity in the classification and segmentation of nodules. Zheng et al. conducted a study using the XGBoost classifier, which uses machine learning algorithms with CT images and clinical data to classify benign and malignant tumors [9]. Zhu et al. used neural networks in their study of lung nodules, taking into account features such as fuzzy boundaries, sparse distribution, and subtle differences in CT images [10]. Prosper et al. conducted a study that included advances in the characterization of pulmonary nodules and early cancers using radiological features and deep learning architectures in addition to traditional image analysis approaches in chest CT [11]. Gugulothu et al. demonstrated the performance of the Logarithmic Layer Xception Neural Network (LLXcepNN) classifier using image processing methods to extract lung regions from CT images [12]. In their study, Lima et al. used pre-trained VGG16, VGG19, Inception, Resnet50, and Xception to extract features from each 2D layer of 3D nodules. They then used principal component analysis to reduce the dimensionality of the feature vectors [13]. Saied et al. used machine learning algorithms and deep learning models together with PCA in their studies [14]. Qiu et al. introduced a new deep learning model for the segmentation of small lung nodules [15]. Kido et al. targeted nodule segmentation in their study by developing a nested 3D fully connected convolutional network and a new loss function [16]. Bhattacharjee et al. proposed a fine-tuned dual skip-connection-based segmentation system that integrates the pre-trained Residual Neural Network (ResNet) 152 with the U-Net architecture to achieve a fast and accurate segmentation algorithm with fewer stages [17]. Savic et al. proposed a segmentation algorithm based on the fast-marching method [18].

When we review the literature, the lack of a comprehensive study that performs both classification and segmentation for the diagnosis and identification of pulnomer nodules, as in our study, constitutes the motivation for our study.

The innovative aspect of our study is that this model was used for the first time in the classification of pulmonary nodules. Our study aims to facilitate computer-aided diagnosis of a critical disease such as pulmonary nodules, as segmentation and classification of lung diseases is a technical challenge. Another contribution is the attempt to show that the model used for classification is structured in such a way that it can be used for different diseases within the lung. Our study addresses several shortcomings identified during our literature searches, including:

1. Overcoming the limitations of CNN networks in feature extraction by using attention blocks in the model when CNN fails.
2. Reducing computational costs by extracting the most effective features using the PCA method instead of a large number of features.
3. Demonstrate that more efficient results can be achieved when deep learning and machine learning models are used together.
4. Shows that segmentation can achieve results that can be easily evaluated by experts when different segmentation models with different backbones are used.

For this purpose, in our study, first of all, we studied the C+EffxNet model, which has previously shown success, in classifying pulnomer nodules as benign and malignant [19]. In the first version of this model, versions B0 to B3 of the EfficientNet model were used. In this study, the EfficientNet B4 version was added to these models and deep features were extracted from it.

The study consists of two stages (classification and segmentation). Primarily, our original dataset obtained from Van Yüzüncü Yıl University Dursun Odabaşı Training and Research Hospital was marked as benign and malignant on CT by two expert radiologists from this hospital. In the segmentation phase, three segmentation algorithms were used and the results were given comparatively. To increase the performance of these algorithms, three different backbones such as InceptionV3, DenseNet121, and SeResNet101 were used for better feature extraction. Our study presents an ablation study that researchers can compare by using three different segmentation algorithms and three different spine models used with these algorithms. In addition, another innovative aspect of our study comes to the fore as it is a study that provides an in-depth analysis for the pulnomer nodule.

Our study was organized as follows: First, relevant studies for classification and segmentation in Sect. 2. The methodologies used in the Sect. 3, the proposed method in the Sect. 3.5, the experimental studies in the Sect. 4, and the discussion and conclusion section in the Sect. 5.

2 Related works

In this part of our study, first the pulnomer nodule classification and then the literature studies published using segmentation and their use in our study are included in addition to these studies.

2.1 Classification works

Al-shabi et al. [20], developed a model called ProCAN on the LIDC-IDRI dataset. In the study, 98.05% AUC and 95.28% accuracy (Acc.) values were obtained. Fu et al. [21] presented a CNN approach that includes attention on CT images. In these studies, a success of 0.94 was obtained regardless of the size of the nodule. Heuvelmans et al. [22] trained the lung cancer prediction convolutional neural network (LCP-CNN) to generate a malignancy score for each nodule using CT data. They reported that the network they developed in their study performed well in identifying benign nodules by excluding malignancy with a high degree of accuracy in one-fifth of patients with small to medium nodules. Apostolopoulos et al. [23] used deep-convolutional Generative Adversarial Networks to train CNNs to eliminate the lack of large-scale data. They also developed a CNN network called feature fusion VGG19 (FF-VGG19) to improve feature extraction. This study shows that if the data to be used in the diagnosis of the disease are not sufficient, generative networks can be used and CNN models may be insufficient in feature extraction. In their study, they obtained an accuracy value of 0.92. Bening and malignant datasets were presented in their studies with these nodules completely removed. In our study, these nodules were tried to be determined with attention blocks on all CTs. Gu et al. [24]. lung has shown that it can be more successful than human performance during automatic nodule detection, especially for smaller nodules, in review studies examining nodule detection, segmentation, and classification performances. It has also been stated that when compared to traditional methods, deep learning methods show lower false positives although they provide high sensitivity. He et al. [25] proposed an interpretation-model-guided classification method based on ISHAP (improved Shapley Additive ExPlanations) for the classification of benign and malignant pulmonary nodules. They obtained a sensitivity of 0.862, a specificity of 0.885, and an accuracy of 0.873 in the Lung Image Database Consortium (LIDC) dataset. They reported that estimates made with the features extracted from the study are sometimes not understood or interpreted by clinicians. Our study presents a hybrid model to overcome this problem. Astaraki et al. [26] conducted a study using supervised and unsupervised methods using convolutional features. In this study, after feature extraction, the disease was diagnosed by the machine learning method. In the study, it was stated that more training may be needed to generalize machine learning models. The approach we used in our study has previously provided good performance in diagnosing COVID-19 on CT images. Bening malignant discrimination performance is also shown in this study. Halder et al.

[27] proposed the 2-Pathway Morphology-Based Convolutional Neural Network (2PMorphCNN), which classifies lung nodules by capturing both textural and morphological features. In their studies, the LIDC-IDRI dataset had 96.85% accuracy. In addition, it was stated in their studies that the convolution process was not effective in feature extraction alone. In order to overcome this in our study, the CBAM model consisting of attention blocks is included in the hybrid model. Huang et al. [28]. developed a manifold-based deep learning model and preprocessed the CT images in their study of benign malignant discrimination and removed relevant nodules. They are then classified through deep features. In their studies, it was emphasized that the high number of features could cause problems in classification. In our study, the PCA method was used to reduce the number of features. In this way, an increase in classification accuracy rates has been observed. Jin et al. [29] have published a detailed review article on the use of machine learning algorithms for the diagnosis of lung nodules. In this article, it is emphasized that deep learning applications can give better results than machine learning. In our study, an approach has been used that yields successful results as a result of the use of both deep learning applications and machine learning applications for pulmonary nodules. Yang et al. [30] proposed an improved U-Net framework for the 3D U-Net model in their benign and malignant identification studies, where they used U-Net for low-level features and CapsNet for high-level features.

2.2 Segmentation works

Dutande et al. [31] first performed pre-processing on CT images in their proposed study of the Deep Residual Separable Convolutional Neural Network for lung tumor segmentation. It has been emphasized that the standard U-Net model may be insufficient in feature extraction. In our study, backbones were used to overcome this inadequacy. Additionally, the performance of backbones in different segmentation models is also demonstrated. Tyagi et al. [32] obtained a Dice coefficient of 80.74% on the LUNA16 dataset in their study where they proposed a 3D conditional, generative, contentious network with simultaneous compression and excitation blocks for lung nodule segmentation. The use of patch-based processing, which affects the performance of GANs in the study, has also increased the transaction cost in proportion to accuracy. Liu et al. [33] proposed a method called the cascaded dual-pathway residual network (CDP-ResNet) to improve the segmentation of lung nodules in CT images. They obtained an 81.58% Dice coefficient on the LIDC dataset. In their study, they stated that even if it is not necessary to specify the location of the nodule in all layers, the ROI of a

particular layer where the nodule is located should be given. Even if the location of the nodule is not specified in our study, segmentation algorithms can detect the location of the nodule.

3 Materials and methods

3.1 Dataset

Our study is retrospective and covers the period from 2015 to 2021. The images were obtained by multislice tomography devices with 128 detectors (Siemens SOMATOM Definition AS+128, Forchheim, Germany) and 16 detectors (Somatom Emotion 16-slice; CT2012E- Siemens AG Berlin and München, Germany) at the Faculty of Medicine of Van Yüzüncü Yl University. Patients with nodular mass lesions on the lungs were detected in the films evaluated by specialist radiologists. Subsequently, patients with at least two years of follow-up or a definitive pathological diagnosis were included in the study, while patients with artifacts in their images and nodules smaller than 5 mm were excluded from the study.

Our study was in two stages and, in the first stage, it was aimed at evaluating the classification of lesions first. For this purpose, patients were evaluated histopathologically and clinicoradiologically by specialist radiologists, and two classes labeled as malignant and benign were formed. Then a total of 199 images were obtained from all axial sections, including the lesion from 29 (19e/10b) patients in the Malignant group, and a total of 202 images were obtained from all axial sections including the lesion of 68 (38e/30b) patients in the Benign group.

Since normal lung images without any lesions were needed for artificial intelligence training, 343 normal sections from 67 (38e/29b) patients were used as a third group, covering the upper, middle, and lower zones of the lung. The mean age of the patients included in the malignant group was 65 ± 10.4 (43–88 years), the mean age of the patients included in the benign group was 59 ± 12.2 (27–81 years), and the mean age of the patients included in the normal group was 56.9 ± 14.1 (26–81 years). For the second stage, the segmentation stage, there were mixed malignant and benign patients in the dataset. A total of 379 images were obtained from 80 (43e/37b) patients, 229 images from 57 (28 Male/29 Female) patients with benign diagnosis, and 150 from 23 (15 Male/8 Female) patients with malignant diagnosis. The mean age of this group was calculated as 61 ± 12 (27–88). Then, the location of the lesion in each image was determined by specialist doctors, and the lesions' borders were drawn manually using image processing programs. Thus, two separate copies of each image were obtained, marked and unmarked. The marked

images are shown in Fig. 1. Masks were removed from the marked benign and malignant patient data, as shown in Fig. 2.

3.2 Deep learning models

The C+EffxNet approach, utilized for classification, represents a novel hybrid deep learning method developed by Canayaz [19], specifically tailored for COVID-19 CT images. In the initial phase, we construct a CBAM-based model with an input layer of (256, 256, 3). This model incorporates channel attention, spatial attention, and residual blocks to extract crucial features from the images. Subsequently, the hypercolumn technique is employed to amalgamate these extracted features. Moving on to the second stage, EfficientNet models are integrated into the CBAM model, resulting in the creation of a third hybrid model. The final layer of the CBAM model is fused with the initial input layer of the EfficientNet models, adapting the shape of the first input layer to match the data output of the CBAM model. It is noteworthy that the models in this hybrid approach are not pre-trained; instead, hyperparameters are fine-tuned during the model training process. The training data for the hybrid model comprises images processed in these stages, yielding 1024 features for each model.

In summary, this approach involves a classification task employing two hybrid models. The first model consists of layers comprising CBAM blocks and feature maps of images utilizing the hypercolumn technique. In the second model of the hybrid approach, four versions of EfficientNet, a prominent deep learning architecture, are employed. In this study, the B4 version of EfficientNet is incorporated into this model, and a performance evaluation is conducted. This innovative approach is employed for feature extraction due to its demonstrated high performance in CT images.

The InceptionV3, DenseNet121 and SEResNet101 backbones were used for feature extraction from segmentation algorithms used for segmentation. If we briefly explain these models;

InceptionV3 consists of symmetric and asymmetric building blocks that include convolutions, mean pooling, maximum pooling, joins, dropouts, and fully connected layers [34]. In this network model architecture, there are factored convolutions to reduce the number of parameters. In addition, faster computation is provided by using small convolution windows. The fundamental features of Inception v3 are as follows: (a) Inception Module: The Inception v3 model employs Inception modules containing interconnected convolution filters of different sizes. These modules assist in learning various features of the network and contribute to a more effective representation of
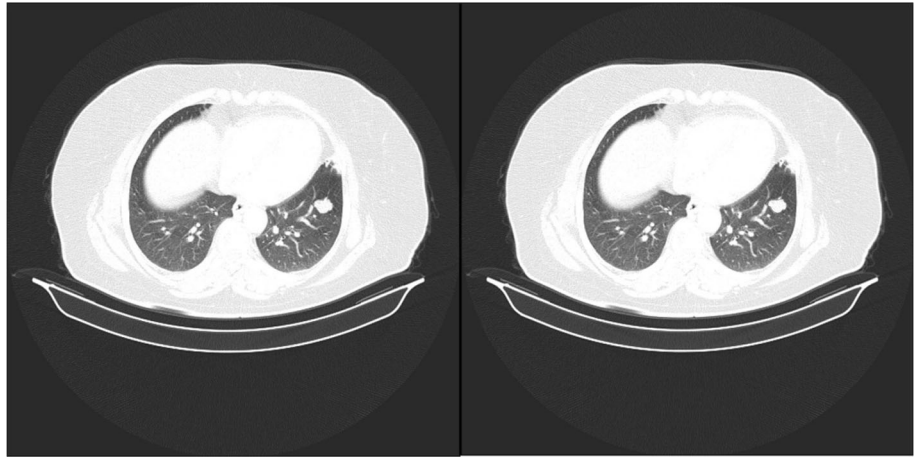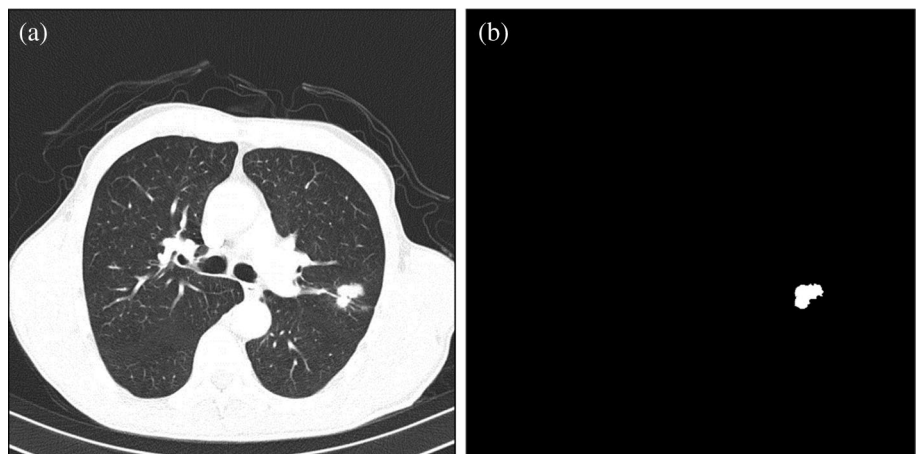
Fig. 1 Example of marked dataset

Fig. 2 Malignant patient image **a** original image **b** mask of image

information. (b) Auxiliary Classifiers: Inception v3 aims to facilitate the training of deeper networks by incorporating auxiliary classifiers during the training process. This approach can enhance learning by utilizing information from both earlier and deeper layers of the network. (c) Batch Normalization: Batch normalization is frequently utilized in Inception v3 as a technique that aids in training the network faster and improving generalization. This contributes to the network's more stable learning. (d) Factorization into Small Convolutions: In Inception v3, the computation cost is reduced by using factorized small convolution matrices instead of large convolution matrices. This helps achieve a lighter and more efficient model.

In the DenseNet121 architecture, each layer is directly linked to all other layers. In these networks, feature maps of previous layers are not collected in each layer, they are only combined and used as input [35]. In this architecture, there are 1 7×7 Convolution, 58 3×3 Convolutions, 61 1×1 Convolutions, 4 Average Pools, and 1 Fully Connected Layer. DenseNet architecture has denser connections compared to previous models, allowing the network to learn more effectively and facilitate information transfer.

The key feature of DenseNet architecture is the inclusion of direct connections from all layers to each other. In contrast to traditional CNN models, each layer is connected not only to the previous layer but also to all preceding layers. This enhances the flow of information, enabling the network to learn deeper and more effective features. DenseNet121 is a specific DenseNet model consisting of 121 layers. This model is commonly used in computer vision tasks such as object recognition, classification, and segmentation, particularly in tasks related to computer vision. Additionally, it is popular in transfer learning applications, as a pre-trained DenseNet model on extensive datasets tends to perform well in similar tasks.

SEResNet is a variation of ResNet that uses compression and excitation blocks. ResNet introduces a prominent architecture to facilitate the training of deep neural networks and enhance performance. SEResNet101 is an enhanced version of ResNet101. By incorporating Squeeze-and-Excitation (SE) blocks, the goal is to provide better learning capabilities to the model. These blocks aim to emphasize important features by focusing on learned feature maps. Consequently, the objective is to enable the

network to learn more effective features, ultimately improving overall performance. "Squeeze-and-Excitation" (SE) blocks are a mechanism used to assist deep neural networks in learning more effective features. These blocks aim to focus on crucial features in the feature maps. SE blocks typically consist of the following two main steps: In the Squeeze step, channel information in each feature map is compressed (squeezed) using a learned weight set. This involves measuring the importance level of each channel and performing a process that reduces the number of channels. In the Excitation step, weights are applied to the feature maps using the summary information obtained from the compression step. These weights express the importance level of each channel, creating a mechanism to determine which channels need more emphasis. This process determines the significance of each channel, and the network optimizes learning by placing greater emphasis on these important channels. As a result, the model learns more effective features, leading to an overall improvement in performance. Squeeze-and-Excitation blocks are designed to deliver better performance, especially in tasks that involve large and deep neural networks. It can better map the channel dependency with the squeeze and excitation block. In this way, it can better calibrate the filter outputs, leading to performance gains [36].

3.3 Classification

In this section, some of the classifiers used in the study will be briefly explained. Before moving on to classifiers, information about PCA, which is the feature reduction method we used in the study, will be given.

3.3.1 Reduced features with PCA

Principal component analysis (PCA) is an unsupervised linear transformation technique utilized for dimensionality reduction [37]. It allows the identification of new subspaces with the highest variance in high-dimensional data, effectively reducing its size [37, 38].While this reduction may result in the loss of certain properties, it primarily discards less informative traits about the population.

PCA brings together highly correlated variables, forming a reduced set of artificial variables known as 'principal components,' capturing the most significant variation in the data [38]. As an orthogonal statistical technique [39]. PCA aims to project high-dimensional data samples into a lower-dimensional space through linear transformation [40], preserving the original data features to the extent possible.

Through a linear transformation that minimizes redundant covariance information and maximizes variance information [41], PCA effectively combines highly correlated variables. The outcome is a concise set of dummy variables—'principal components'—that represent the primary sources of variation in the data [38].

Creating principal components in high-dimensional data begins with normalization of the data.

[p normalized random variables $Z' = (Z_1, Z_2, \ldots, Z_p)$]. Eigenvalues and eigenvectors of these standardized data are obtained from the variance–covariance Σ matrix. The eigenvalues of this matrix

$$\left|\Sigma - \lambda I_p\right| = 0 \tag{1}$$

are obtained from the roots of the Eq. 1. These obtained eigenvalues are ordered so that $\lambda_1 > \lambda_2 > , \ldots, > \lambda_p > 0$ effectively represents the decreasing variance in the data [42]. After establishing linear Eqs. 2–3 as:

$$Y = l^T Z, \tag{2}$$

l as the loading vector, the equation

$$\Lambda = P^T \Sigma P \tag{3}$$

is obtained, where Λ is the eigenvalue matrix and P is the eigenvalue matrix. The basic assumption of PCA is that the score and loading vector corresponding to the largest eigenvalues contain the most useful information, and the rest mainly contain noise. For this reason, these vectors are generally created in order of decreasing eigenvalues [43].

3.3.2 Classifiers

The purpose of SVM is to obtain the optimal separation hyperplane that will separate data belonging to different classes [44–46]. It uses Lagrangian multipliers to solve the optimisation problem to find the optimal separation hyperplane while solving the classification problem. In this way, the number of transactions is reduced [47]. SVMs are supervised learning models that analyze data for classification [45].

In SVM, the classification process for each sample x_i in the data set can be expressed as in Eq. 4 [46]:

$$f(x_i) = \text{sign}\left(\sum_{j=1}^{n} \alpha_j y_j K(x_i, x_j) + b\right) \tag{4}$$

where, $f(x_i)$: classification score of sample x_i, α_i: support vector weights, y_j: label (class) of sample x_j. $K(x_i, x_j)$: Kernel function, b: bias.

KNN is a non-parametric classification method. The input consists of the k closest training examples in a dataset [48]. It is determined by the k-value of the sample data point and the nearest neighbor [37]. The formula for KNN is given in Eq. 5.

$$\widehat{y}(x) = \text{mode}\{y_i | x_i \epsilon N_k(x)\} \tag{5}$$

here, $\hat{y}(x)$: estimated label of point x, $N_k(x)$: cluster containing the k closest neighbors of point x, y_i: label of point x_i.

The RidgeClassifier is a classifier that utilizes Ridge regression. Ridge regression [49] is suggested for data with multicollinearity issues to obtain predictors with smaller variances. In cases where classification is involved, it transforms the target variable by considering classes into $[-1, 1]$ and constructs the model using Ridge regression to solve the problem [50]. The loss function and formula for this classifier are given Eq. 6 [51].

$$L_0 = \text{MeanSquaredError} + L_2\text{penalty} \tag{6}$$

$$J(w) = \sum_{i=1}^{n} l\big(f(x_i), y_i\big) + \alpha \|w\|_2^2$$

where, $J(w)$: Loss function representing the total error, $l\big(f(x_i), y_i\big)$: Loss function measuring the error between the model's prediction and the actual label. α: regularization parameter, $\|w\|_2^2$: L_2 norm, represents the sum of squares of the weight vector.

The Ridge Classifier restricts the weights using this regularization term, enhancing generalization. As a result, the risk of overfitting decreases, and a more generalizable model is obtained.

XGBoost, a machine learning method based on decision-tree and gradient-boosting proposed by Chen and Guestrin in 2016 [52], has been defined as a fantastic blend of hardware and software optimization approaches that yield excellent results in a short period [53]. XGBoostClassifier is a gradient boosting classifier based on xgboost. XGBoost is an implementation of the widely recognized gradient boosting algorithm known for its efficiency and prediction accuracy [52]. This algorithm is equipped with methods such as Regularization, Missing values Imputation, Cross-Validation, Hyperparameter tuning, etc., to enhance the computation time and performance of the model [53]. The primary goal of XGBoost is to optimize the performance of the model through a target function. This function encompasses both loss and regularization terms. This classifier is expressed as in Eq. 7.

$$\text{Obj} = \sum_{i=1}^{n} l(y_i, \widehat{y_i}) + \sum_{k=1}^{K} \Omega(f_k) \tag{7}$$

Here, n: number of samples in the data set, $l(y_i, \widehat{y_i})$:: The loss function measures the error between the actual label y_i and the prediction $\widehat{y_i}$. K: Total number of trees, $\Omega(f_k)$: The regularization term controls the complexity of each tree.

3.4 Segmentation

3.4.1 U-Net

Fully Convolutional Network (FCN) architecture is a highly successful and frequently used basic architecture recommended for semantic segmentation [54]. The U-Net architecture is an architecture based on FCN architecture, which was proposed for the semantic segmentation of medical images [55]. The architecture of the network consists of two parts (contraction and expansion path) [55, 56]. Operations on the left side of the architecture, the shrinking or contraction path, are used to capture context information about the image (extract features from the image). These operations are exactly the same as the classical CNN architecture logic.

The operations on the right side of the architecture, the expansion path, are used to precisely position the parts that need to be segmented in the picture [55, 57]. For the skip connection between the shrinking path and the expanding path, a concatenation operator is applied instead of the sum. This enables spatial information to be applied directly to deeper layers and obtains a more accurate segmentation result [58]. Classical deep learning needs abundant examples and expensive computing resources. However, U-Net can adapt to minimum training sets [57, 59]. In particular, this network is suitable for medical image segmentation tasks [57, 60].

The main strategy that distinguishes U-Net from other segmentation architectures is to combine feature maps of the contraction phase and their symmetrical counterparts in the expansion phase. In this way, it is possible to disseminate context information into high-resolution feature maps [61]. The U-Net structure is shown in Fig. 3.

3.4.2 LinkNet

LinkNet has a lightweight deep neural network architecture that allows learning of semantic segmentation tasks without a significant increase in parameters. It is similar to U-Net and other segmentation networks. There is an encoder on the left and a decoder on the right [61]. The encoder functions by encoding the information in the source space, and the decoder maps this information in spatial categorization to perform the segmentation [61, 62]. The encoders used in segmentation processes in current neural network architectures perform more than one downsampling operation. This process causes some spatial information to be lost. Some information is lost when going through cascading convolutions in the encoder part [63]. This lost information is difficult to recover [62]. In Link-Net, the input of each encoder layer is also assigned to the

Fig. 3 U-Net structure [55]
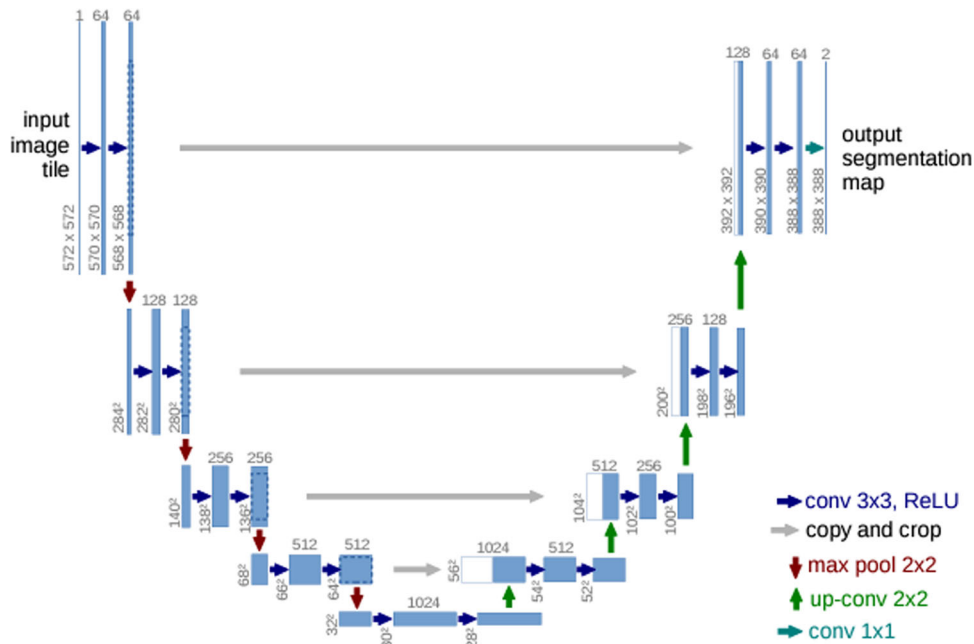
output of the corresponding decoder. The purpose of this process is to recover lost spatial information that can be used by the decoder and upsampling processes [62]. Fewer parameters are used because the information learned by the encoder is shared with the help of the decoder [62, 64]. The LinkNet structure is shown in Fig. 4.

Fig. 4 LinkNet structure [62]

3.4.3 FPN

FPN uses the pyramidal hierarchy of deep convolutional networks to construct feature pyramids at marginal extra cost [65]. Image pyramids are a data structure designed to support convolution through reduced imagery. It consists of a series of copies of the original image in which both sample density and resolution are reduced in regular steps [66]. Feature pyramids form the basis of a standard solution built on image pyramids [67]. It is believed that the pyramid can provide some conceptual unification to the problem of representing and manipulating low-level visual information. It offers a flexible, convenient multi-resolution format that matches multiple scales found in visual scenes and reflects multiple processing scales in the human visual system [66]. FPN is a feature extractor based on the pyramid concept with accuracy and speed [67]. ConvNets represent a high level of semantics and are resistant to variance, but pyramids are still needed to obtain the most accurate results. The main advantage of specifying each level of an image pyramid is that it produces a multi-scale feature representation where all levels, including high-resolution levels, are semantically strong. However, the disadvantage of an image pyramid is that it causes an increase in extraction time [67]. FPN combines features using bottom-up, top-down path and lateral connections by leveraging feature hierarchy. In this way, it creates a strong feature pyramid [68].

The bottom-up path provides a top-down way to render higher-resolution layers than the rich layer. Each stage is defined as a pyramid level [66]. Among the reconstructed layers, the final layer output is selected as feature maps to
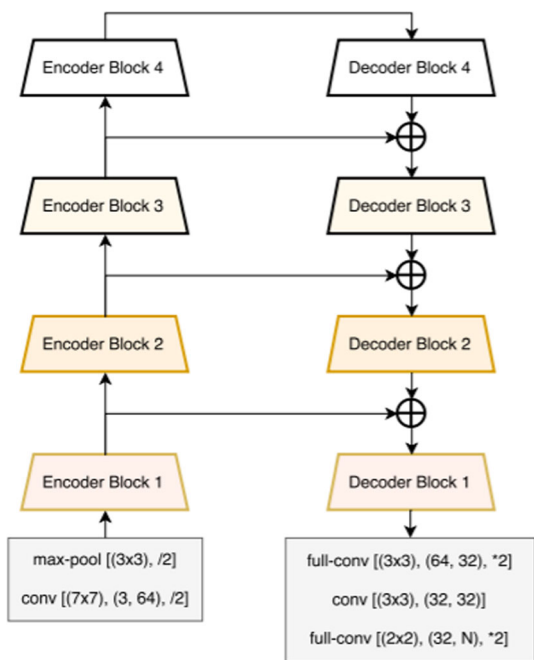
construct the pyramid [67]. The top-down path is enriched by adding lateral links between feature maps [64]. The top-down path provides a top-down way to create spatially coarser, but semantically stronger feature maps with higher resolution layers. These features are enhanced by features on the bottom-up path through lateral links. Each lateral link joins feature maps of the same spatial dimension from the bottom–up and top–down path [67]. The FPN structure is shown in Fig. 5.

3.5 Proposed method

Our benign-malignant analysis study, which is the subject of the study, consists of two stages. The first stage is the classification stage. At this stage, the classification performance of the dataset was analyzed. For this, the dataset given by the C+EffxNet method is trained. The models obtained as a result of this training were then used in the extraction of deep features. In the C+EffxNet method, versions of the EfficientNet deep learning model from $B0$ to $B4$ are used. The input value of this method is (256, 256, 3). From this new hybrid method, 1024 features were extracted for each model. We then subjected these features to feature selection using the PCA method and extracted 100 best features for each model. For our image dataset, this value was (744.100). In other words, 100 features of each image were selected. 20% of this feature dataset are reserved as test data. First of all, the classification process was carried out for this dataset using RidgeClassifier, SVM, KNN and XGBoostClassifier. The operations performed for the classification section are shown in Fig. 6.

Parameters used in training models; optimiser Adam, learning rate 0.0005, loss function categorical cross_entropy, batch_size 8, epochs set to 100.

The second stage is the segmentation stage of the disease. At this stage, results were obtained by using three powerful segmentation algorithms such as U-Net, LinkNet, and FPNet. InceptionV3, DenseNet121, and SeResnet are used as backbones for feature extraction in these algorithms. In the running of the algorithms, first of all, benign and malignant datasets were run separately and the results were obtained, and then these datasets were combined and the results were obtained. Parameters used in the segmentation phase; Adam optimizer was used as the optimizer, learning rate, Threshold value of 0.0001 was determined as 0.5, batch size was determined as 8. The segmentation process is shown in Fig. 7.

3.6 Metrics

Accuracy is one of the most common criteria in practice used to evaluate the generalization ability of classifiers [69]. and is the ratio of the number of correctly diagnosed nodules to the total number of nodules [37]. Precision is a measure of how well a model predicts only the positive outcome of a classification. Recall is the ratio of the number of correctly classified to the number of condition positive [37] and is used to measure the proportion of correctly classified positive patterns [69].

The $F1$ score is a measure that combines both precision and recall in a single measure. It is calculated by averaging the harmonic of precision and recall, and a higher $F1$ score performs better [69]. These metrics are Sensitivity (Se), Specificity (Sp), F-score (F-Scr), Precision (Pre), and Accuracy (Acc). The True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values are used to calculate the metrics. The equations for these metrics are given in Eqs. 8–12.

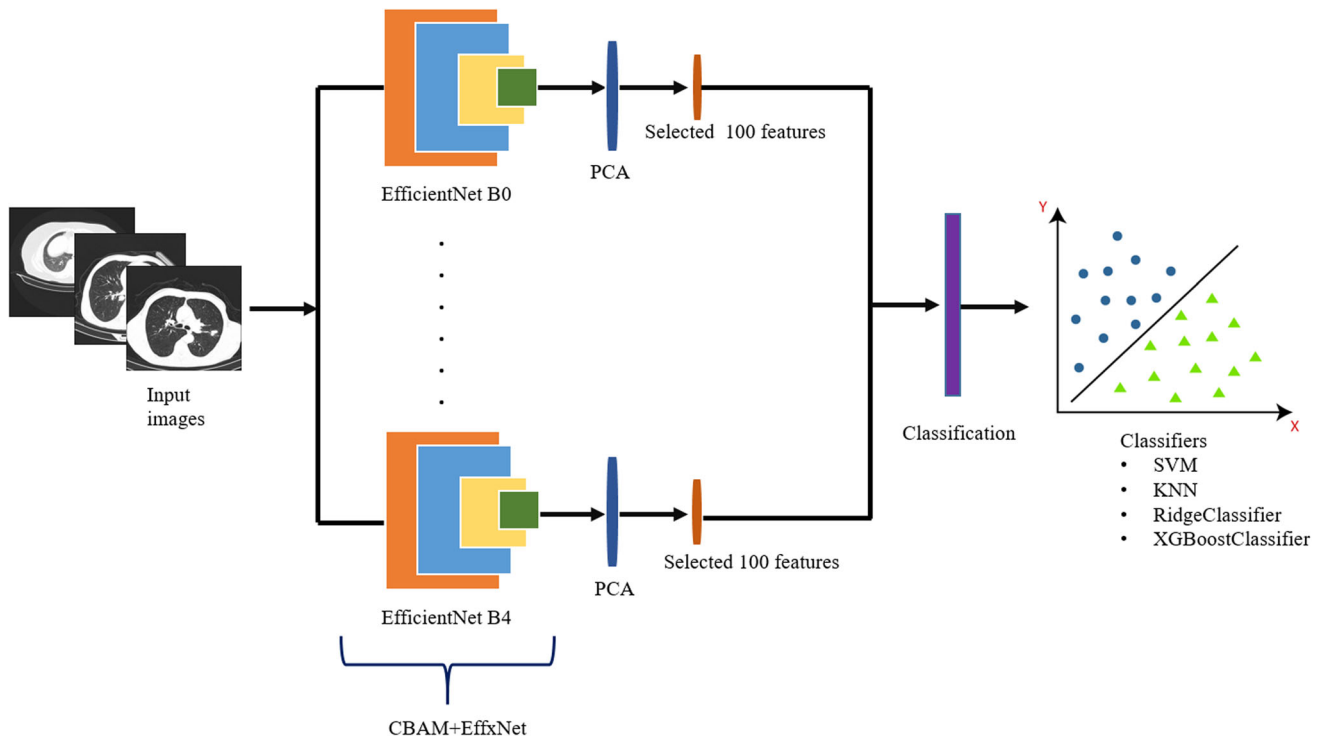$$Se = \frac{TP}{TP + FN} \tag{8}$$

Fig. 5 FPN structure [65]
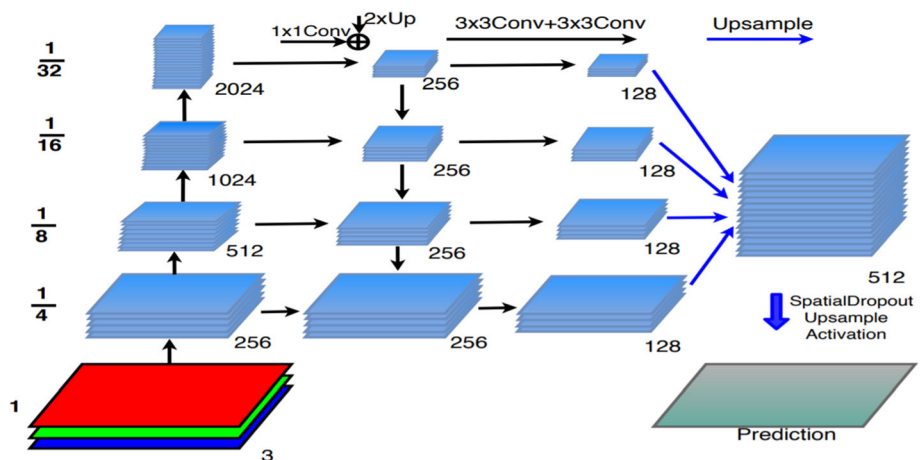
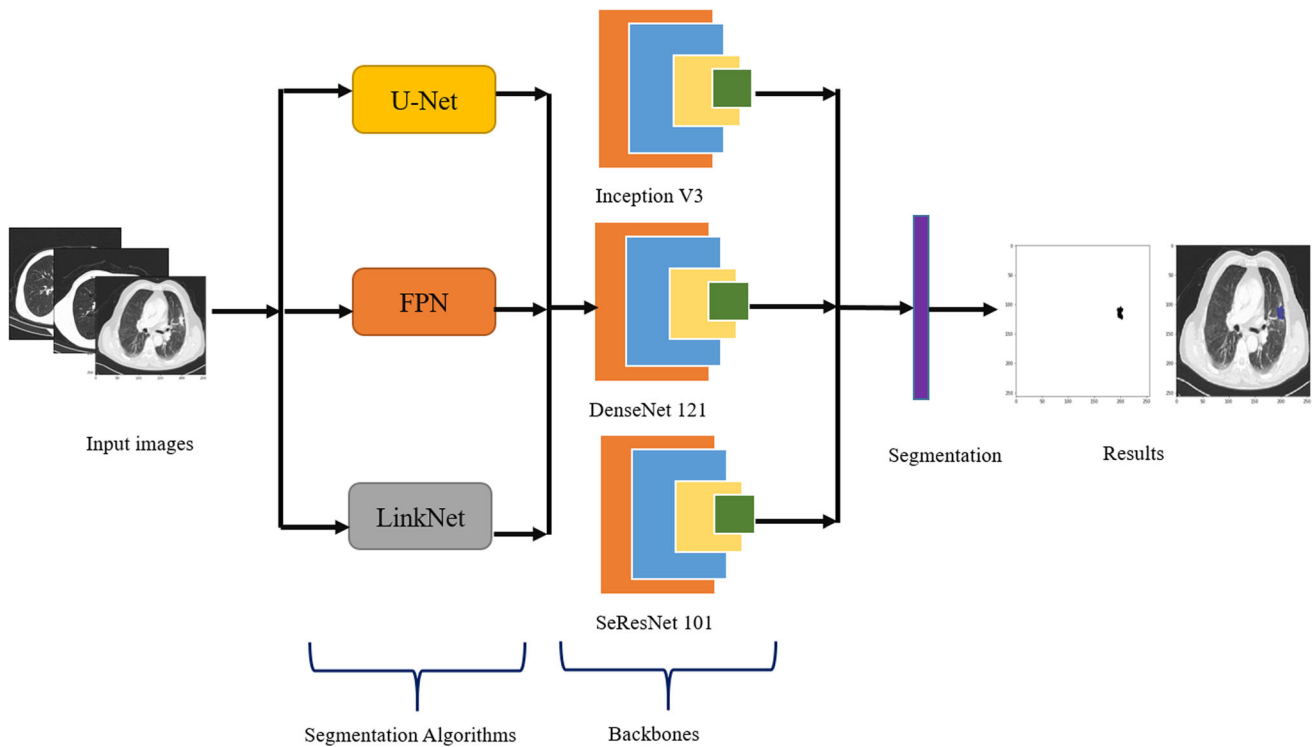**Fig. 6** Classification process

Fig. 7 Segmentation process

$$Sp = \frac{TN}{TN + FP} \qquad (9)$$

$$Pre = \frac{TP}{TP + FP} \qquad (10)$$

$$F\text{-score} = \frac{2TP}{2TP + FP + FN} \qquad (11)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (12)$$

The kappa value is a statistic that is used to measure the agreement between two raters or annotators who assign ratings or labels to a set of items. It is calculated by taking the ratio of the observed agreement between the raters to the agreement that would be expected by chance and subtracting the result from 1. This measure is useful because it takes into account the possibility of chance agreement between the raters, which can sometimes be high even if the raters are not actually in agreement [70]. A kappa value of 1 indicates perfect agreement, while a value less than 1 indicates less than perfect agreement. The Kappa coefficient Equation is given in Eq. 13.

$$k = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \qquad (13)$$

here, $Pr(a)$ is the ratio of the total number of matches observed for the two values, and $Pr(e)$ is the probability of this agreement occurring by chance.

The Jaccard index, also known as the Jaccard similarity coefficient, is a measure of the similarity between two datasets. Calculated by dividing the intersection size of two sets by the size of their union. This measure can be used to compare the similarity of two datasets, and a higher Jaccard index indicates more similarity [71]. The Jaccard similarity coefficient Equation is given in Eq. 14.

$$Jaccard\ index = \frac{TP}{TP + FP + FN} \qquad (14)$$

The Dice coefficient, also known as the Sørensen-Dice coefficient, is another measure of the similarity between two datasets. It is calculated by dividing twice the intersection size of two sets by the sum of the sizes of the two sets. The Dice coefficient can also be used to compare the similarity of two datasets, with a higher coefficient indicating greater similarity [72]. The equation for the Dice coefficient is given in Eq. 15.

$$Dice\ coefficient = \frac{2TP}{2TP + FP + FN} \qquad (15)$$

4 Experimental studies

4.1 Pre-processing

No preprocessing has been performed for the dataset used in the classification process. However, in the preprocessing stage of segmentation, experts examined the images in the dataset one by one and labeled the diseased parts in benign and malignant images. These tags are then separated into masks by image processing programs. These masks were used as segmentation tags in segmentation.

4.2 Bening malignant classification

At this stage, C+EffxNet models were run with the dataset. During the training phase, models saving the lowest validation loss function were retained. Initial performance evaluations on the validation dataset were conducted using these saved models, and the results are presented in Table 1. It's important to note that each sample in the dataset corresponds to an image, and the size of each image is (256, 256, 3).

When we examine this table, the best result on the validation dataset was achieved by EffNetB3 with a score of 0.88. Initially, our plan involved utilizing the C+Effx-Net model for feature extraction. Through training this model with the dataset, it learned from the previously unseen pulmonary nodule dataset and adjusted its weights accordingly. Consequently, the model became prepared for feature extraction from this dataset. The purpose of presenting Table 1 is to highlight the importance of deep feature extraction by showing the performance of this model before deep feature extraction. Following the extraction of 1024 features for each image in the dataset, PCA was employed to select the most influential 100 features from this set. Table 2 provides the classification algorithm performance with the selected features.

When we examine Table 2, we see that the best results are obtained with the features in which EffB0 is used. With the features obtained from EffB0, a performance of 0.97 was achieved in all classifiers. While the accuracy value obtained after training the hybrid model was 0.84, this

Table 1 Results obtained from the validation dataset

Model	Precision	Recall	$F1$-score	Accuracy
EffnetB0	0.89	0.85	0.87	0.8472
EffnetB1	0.82	0.81	0.81	0.8093
EffnetB2	0.86	0.85	0.85	0.8489
EffnetB3	0.86	0.88	0.87	0.8808
EffnetB4	0.81	0.80	0.80	0.8017

Table 2 Classification results of selected deep features

Model features	Selection	Classifier	Accuracy	Precision	Recall	F-Score	Kappa	MSE	RMSE	MAE
EffnetB0	PCA	SVM	0.9731	0.9742	0.9731	0.9731	0.9588	0.0469	0.2167	0.0335
		RidgeClassifier	0.9798	0.9802	0.9798	0.9798	0.9690	0.0402	0.2006	0.0268
		KNN	0.9731	0.9738	0.9731	0.9730	0.9587	0.0469	0.2167	0.03355
		XGBoostClassifier	0.9798	0.9802	0.9798	0.9798	0.9690	0.0402	0.2006	0.0268
EffnetB1	PCA	SVM	0.9597	0.9599	0.9597	0.9597	0.9383	0.0604	0.2457	0.0469
		RidgeClassifier	0.9530	0.9530	0.9530	0.9530	0.9279	0.0671	0.2590	0.0536
		KNN	0.9530	0.9537	0.9530	0.9531	0.9281	0.0671	0.2590	0.0536
		XGBoostClassifier	0.9530	0.9537	0.9530	0.9531	0.9281	0.0671	0.2590	0.0536
EffnetB2	PCA	SVM	0.9664	0.9663	0.9664	0.9663	0.9484	0.0335	0.1831	0.0335
		RidgeClassifier	0.9664	0.9663	0.9664	0.9663	0.9484	0.0335	0.1831	0.0335
		KNN	0.9597	0.9597	0.9597	0.9594	0.9380	0.0402	0.2006	0.0402
		XGBoostClassifier	0.9463	0.9464	0.9463	0.9459	0.9174	0.0738	0.2717	0.0604
EffnetB3	PCA	SVM	0.9530	0.9539	0.9530	0.9530	0.9283	0.0469	0.2167	0.0469
		RidgeClassifier	0.9597	0.9601	0.9597	0.9596	0.9384	0.0402	0.2006	0.0402
		KNN	0.9664	0.9666	0.9664	0.9663	0.9486	0.0335	0.1831	0.0335
		XGBoostClassifier	0.9597	0.9601	0.9597	0.9596	0.9384	0.0402	0.2006	0.0402
EffnetB4	PCA	SVM	0.9395	0.9418	0.9395	0.9395	0.9080	0.0805	0.2837	0.0671
		RidgeClassifier	0.9530	0.9540	0.9530	0.9529	0.9283	0.0671	0.2590	0.0536
		KNN	0.9395	0.9421	0.9395	0.9397	0.9080	0.0604	0.2457	0.0604
		XGBoostClassifier	0.9530	0.9540	0.9530	0.9529	0.9283	0.0671	0.2590	0.0536

value increased by 0.13 points with the use of deep features. The MSE, RMSE, and MAE values obtained with these features were 0.0402, 0.2006, 0.0268, respectively. The lowest value of these metrics was realized with the features obtained with EffnetB2. With these features, 0.0335 MSE values and 0.1831 RMSE values were obtained, and when we look at the table, it is seen that these values are the lowest values among other MSE and RMSE values. The confusion matrix of the EffnetB0 model, where the best result is obtained, is shown in Fig. 8.
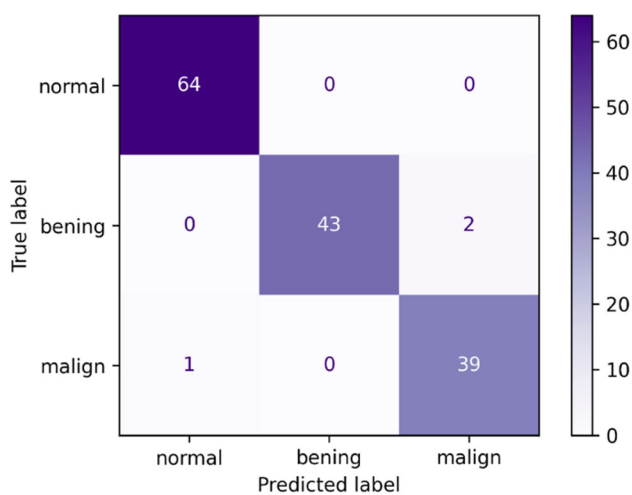
Fig. 8 EffB0 confussion matrix

To confirm the reliability of our results, classification results were obtained using Cross Validation (CV) and Leave One Out Cross Validation (LOOCV) with a K-fold 10 on the dataset. In these verification methods, in addition to the above classifiers, classification has been performed with many classifiers. The correlation between the results obtained when using these methods is also shown in Table 3.

When we look at Table 3, it is seen that the best results are obtained on the features obtained from EffB0 as in Table 2. In the classification made with these features, first, RidgeClassifier obtained the best result with 0.98 accuracy, while SVM became the second classifier with 0.977. The correlation between these characteristics and the results obtained is 0.99.

4.3 Bening malignant segmentation

In the segmentation phase, ablation studies were performed, and the results were obtained. First, only the dataset that contained malignant cells was trained. The results obtained for malignant are given in Table 4.

When we examine Table 4, we see that the best results in train data are obtained by using the U-Net model and the InceptionV3 backbone. While the Jaccard index value obtained in the Train data was 0.9283, the Dice coefficient value was 0.9489. The test values obtained in this model and in the spine are 0.7129 and 0.8214, respectively. The

Table 3 CV and LOOCV results of selected traits

Model features	Fselection	Classifier	CV Acc	LOOCV Acc	Corr
EffnetB0	PCA	RidgeClassifier	0.979	0.980	0.99
		SGDClassifier	0.970	0.970	
		PassiveAggressiveClassifier	0.973	0.974	
		KNeighborsClassifier	0.973	0.973	
		DecisionTreeClassifier	0.957	0.953	
		ExtraTreeClassifier	0.806	0.815	
		LinearSVC	0.970	0.974	
		SVC	0.977	0.977	
		GaussianNB	0.965	0.965	
		RandomForestClassifier	0.976	0.976	
EffnetB1	PCA	RidgeClassifier	0.949	0.954	0.995
		SGDClassifier	0.935	0.937	
		PassiveAggressiveClassifier	0.945	0.944	
		KNeighborsClassifier	0.950	0.946	
		DecisionTreeClassifier	0.921	0.918	
		ExtraTreeClassifier	0.848	0.823	
		LinearSVC	0.941	0.946	
		SVC	0.956	0.956	
		GaussianNB	0.957	0.958	
		RandomForestClassifier	0.960	0.960	
EffnetB2	PCA	RidgeClassifier	0.966	0.966	0.99
		SGDClassifier	0.960	0.958	
		PassiveAggressiveClassifier	0.961	0.958	
		KNeighborsClassifier	0.958	0.960	
		DecisionTreeClassifier	0.935	0.952	
		ExtraTreeClassifier	0.835	0.835	
		LinearSVC	0.965	0.960	
		SVC	0.966	0.968	
		GaussianNB	0.946	0.948	
		RandomForestClassifier	0.962	0.962	
EffnetB3	PCA	RidgeClassifier	0.977	0.977	0.997
		SGDClassifier	0.972	0.970	
		PassiveAggressiveClassifier	0.969	0.972	
		KNeighborsClassifier	0.977	0.977	
		DecisionTreeClassifier	0.965	0.958	
		ExtraTreeClassifier	0.863	0.863	
		LinearSVC	0.970	0.972	
		SVC	0.975	0.977	
		GaussianNB	0.961	0.960	
		RandomForestClassifier	0.974	0.974	
EffnetB4	PCA	RidgeClassifier	0.940	0.945	0.996
		SGDClassifier	0.918	0.933	
		PassiveAggressiveClassifier	0.932	0.933	
		KNeighborsClassifier	0.934	0.935	
		DecisionTreeClassifier	0.909	0.910	
		ExtraTreeClassifier	0.766	0.751	
		LinearSVC	0.930	0.934	
		SVC	0.941	0.941	
		GaussianNB	0.923	0.923	
		RandomForestClassifier	0.946	0.946	

CV Cross-validation *K*-fold: 10, *LOOCV* LeaveOneOut Cross-Validation Accuracy

Table 4 Malignant results

Model	Backbone	Test	Train	Threshold
U-Net	DenseNet121	Jacard index: 0.6396	Jacard index: 0.9293	0.5
		Dice coefficient: 0.7317	Dice coefficient: 0.9487	
	InceptionV3	Jacard index: 0.7129	Jacard index: 0.9283	0.5
		Dice coefficient: 0.8214	Dice coefficient: 0.9489	
	SeResNet101	Jacard index: 0.6294	Jacard index: 0.9311	0.5
		Dice coefficient: 0.7254	Dice coefficient: 0.9515	
FPN	DenseNet121	Jacard index: 0.8026	Jacard index: 0.9248	0.5
		Dice coefficient: 0.8877	Dice coefficient: 0.9451	
	InceptionV3	Jacard index: 0.6782	Jacard index: 0.9193	0.5
		Dice coefficient: 0.7737	Dice coefficient: 0.9375	
	SeResNet101	Jacard index: 0.6076	Jacard index: 0.9109	0.5
		Dice coefficient: 0.6894	Dice coefficient: 0.9313	
LinkNet	DenseNet121	Jacard index: 0.7268	Jacard index: 0.9233	0.5
		Dice coefficient: 0.8360	Dice coefficient: 0.9426	
	InceptionV3	Jacard index: 0.6356	Jacard index: 0.9241	0.5
		Dice coefficient: 0.7277	Dice coefficient: 0.9449	
	SeResNet101	Jacard index: 0.6181	Jacard index: 0.9239	0.5

best values obtained in the test data were obtained with the FPN model and DenseNet121 backbone. The resulting Jaccard index value is 0.8026, while the Dice coefficient value is 0.8877. Taking into account the values obtained in the test data in the studies, we can clearly see that the best results for the Malignant dataset are obtained with the FPN model and DenseNet121 backbone. Then, model trainings were carried out on the dataset containing bening. The results obtained from this dataset are given in Table 5.

When Table 5 is examined, it is seen that the best results on train data are obtained with the LinkNet model and DenseNet121 backbone, with the Jaccard index 0.9140 and Dice coefficient 0.9284. A slight decrease in these values was observed compared with the train data in the malignant dataset. When we look at the test results for the Bening dataset, it is seen that the best performance is obtained with the U-Net model and the InceptionV3 backbone. The values obtained are 0.5127 for the Jaccard index and 0.5728 for Dice coefficient. When we compare it with malignant, this value is quite low. Finally, the training was carried out with the dataset containing both datasets. The results obtained as a result of this training are given in Table 6.

When we examine Table 6, we see that the model that provides the best performance in the training data is the U-Net model and the SeResNet101 backbone. The resulting Jaccard Index value is 0.9094, while the Dice coefficient value is 0.9301. In the test data, it was seen that the best results were obtained with the FPN model and the Dense-Net121 backbone. The Jaccard index value and Dice coefficient value obtained in this model and backbone are 0.3263 and 0.3890, respectively. These values were again observed to be quite low compared with the malignant results.

The application implemented for our study can be accessed at https://github.com/mcanayaz/PulnomaryNodules

5 Discussion and conclusions

Examination and classification of pulmonary nodules play an important role in the rapid diagnosis of lung cancer development. Our work consists of two stages. In the first stage, benign and malignant classification is made. At this stage, the performance of the C+EffxNet approach, which was previously recommended in the Covid-19 classification and had successful results, in the classification of benign and malignant was first measured. The published study of this approach used versions B0 through B3 of the EfficientNet model. In this study, new results were obtained from the B4 version. In the first stage, as a result of model training, the maximum success rate was 88%. However, the success rate increased to 97.98% as a result of the extraction of deep features and classification with feature selection from the extracted features. This clearly shows us the power of feature extraction and feature selection. The test size ratio in classification studies is 0.2. The best results were obtained with features extracted from the Ridge and XGBoost Classifier. In order to confirm the reliability of the results, CV and LOOCV cross-validation methods were applied to the obtained features. The correlation between the results obtained from these methods is 0.997.

The second phase of the study is the segmentation phase of benign and malignant nodules. At this stage, masks were obtained from the images marked by our radiologists. The

Table 5 Bening results

Model	Backbone	Test	Train	Threshold
U-Net	DenseNet121	Jacard index: 0.4521	Jacard index: 0.8879	0.5
		Dice coefficient: 0.5001	Dice coefficient: 0.89929	
	InceptionV3	Jacard index: 0.5127	Jacard index: 0.8961	0.5
		Dice coefficient: 0.5728	Dice coefficient: 0.9111	
	SeResNet101	Jacard index: 0.3018	Jacard index: 0.8892	0.5
		Dice coefficient: 0.3352	Dice coefficient: 0.9004	
FPN	DenseNet121	Jacard index: 0.4170	Jacard index: 0.8841	0.5
		Dice coefficient: 0.4759	Dice coefficient: 0.8953	
	InceptionV3	Jacard index: 0.3711	Jacard index: 0.8971	0.5
		Dice coefficient: 0.4443	Dice coefficient: 0.9110	
	SeResNet101	Jacard index: 0.3138	Jacard index: 0.8921	0.5
		Dice coefficient: 0.3642	Dice coefficient: 0.9051	
LinkNet	DenseNet121	Jacard index: 0.4058	Jacard index: 0.9140	0.5
		Dice coefficient: 0.4638	Dice coefficient: 0.9284	
	InceptionV3	Jacard index: 0.4801	Jacard index: 0.9024	0.5
		Dice coefficient: 0.5566	Dice coefficient: 0.9219	
	SeResNet101	Jacard index: 0.2699	Jacard index: 0.8818	0.5

Table 6 Bening and malignant results

Model	Backbone	Test	Train	Threshold
U-Net	DenseNet121	Jacard index: 0.3089	Jacard index: 0.8879	0.5
		Dice coefficient: 0.3559	Dice coefficient: 0.9037	
	InceptionV3	Jacard index: 0.1535	Jacard index: 0.6532	0.5
		Dice coefficient: 0.1832	Dice coefficient: 0.7239	
	SeResNet101	Jacard index: 0.2751	Jacard index: 0.9094	0.5
		Dice coefficient: 0.3208	Dice coefficient: 0.9301	
FPN	DenseNet121	Jacard index: 0.3263	Jacard index: 0.9052	0.5
		Dice coefficient: 0.3890	Dice coefficient: 0.9260	
	InceptionV3	Jacard index: 0.2688	Jacard index: 0.8907	0.5
		Dice coefficient: 0.3255	Dice coefficient: 0.9070	
	SeResNet101	Jacard index: 0.2756	Jacard index: 0.8056	0.5
		Dice coefficient: 0.3514	Dice coefficient: 0.8708	
LinkNet	DenseNet121	Jacard index: 0.2487	Jacard index: 0.8882	0.5
		Dice coefficient: 0.2934	Dice coefficient: 0.9066	
	Inceptionv3	Jacard index: 0.2686	Jacard index: 0.8824	0.5
		Dice coefficient: 0.3223	Dice coefficient: 0.9013	
	SeResNet101	Jacard index: 0.2926	Jacard index: 0.8934	0.5

results were obtained by running the dataset created from images and masks with 3 segmentation algorithms. In this section, the image dataset containing bening was first obtained, then the image dataset containing malignant, and finally, the segmentation results with separate backbones were obtained on the dataset obtained by combining the two. When we look at the results, the highest Dice coefficient of 0.88 was obtained from the image dataset containing malignant, while a jaccard index of 0.8026 was obtained. It should also be noted that the results obtained with this dataset are higher than those obtained with the Bening dataset and the combined dataset obtained by combining the two. Among the limitations of our study, it is thought that the number of images in the dataset should be increased to increase the success of performance metrics for segmentation. At the end of the study, segmentation results were obtained in patients that the models had never seen, and the results and radiologists' comments on the performance of the models are given in "Appendix". We will continue to work with new models for segmentation. In addition, it is planned to write an interface where both classification and segmentation can be used together.

Appendix

1.
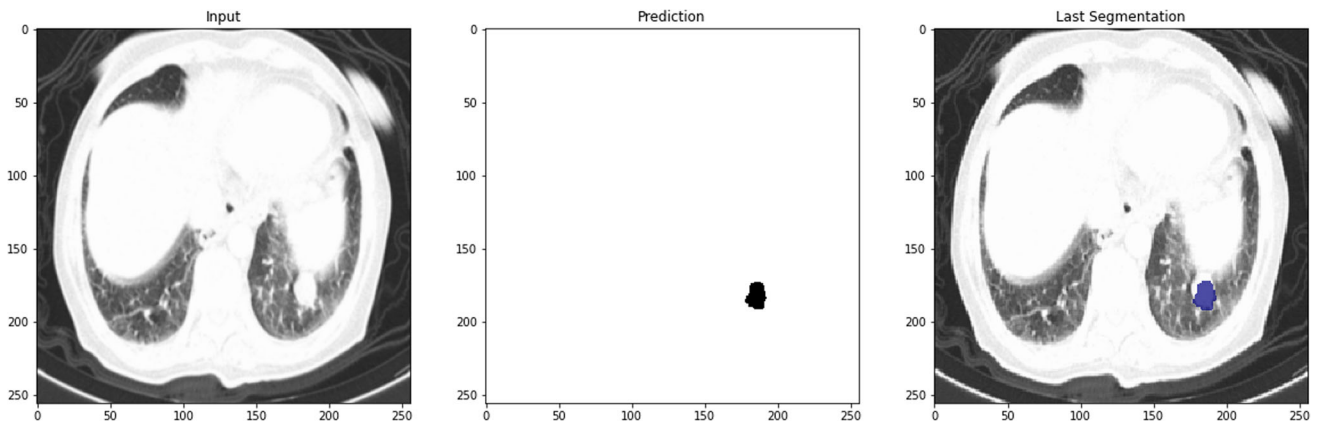
A nodular lesion with an anterior contour that contacts the pleura is observed in the lower lobe of the left lung. Our algorithm detected the posterior segments of our nodule with great accuracy. Additionally, no false inference was found in any focus other than the lesion. Here, the algorithms correctly detect the lesion borders adjacent to the lung parenchyma, and cannot detect adequately in the area adjacent to the lung membrane.

In this image, a nodular lesion is observed in the upper lobe of the anterior part of the left lung. Although our lesion is close to the pleura, contact is not visible. Our algorithm showed the lesion boundaries almost completely accurately. There is an inability to detect in the inner parts of the lesion. In addition, in the right lung, there are foci that are incorrectly detected in the vicinity of the vascular structures.

2.
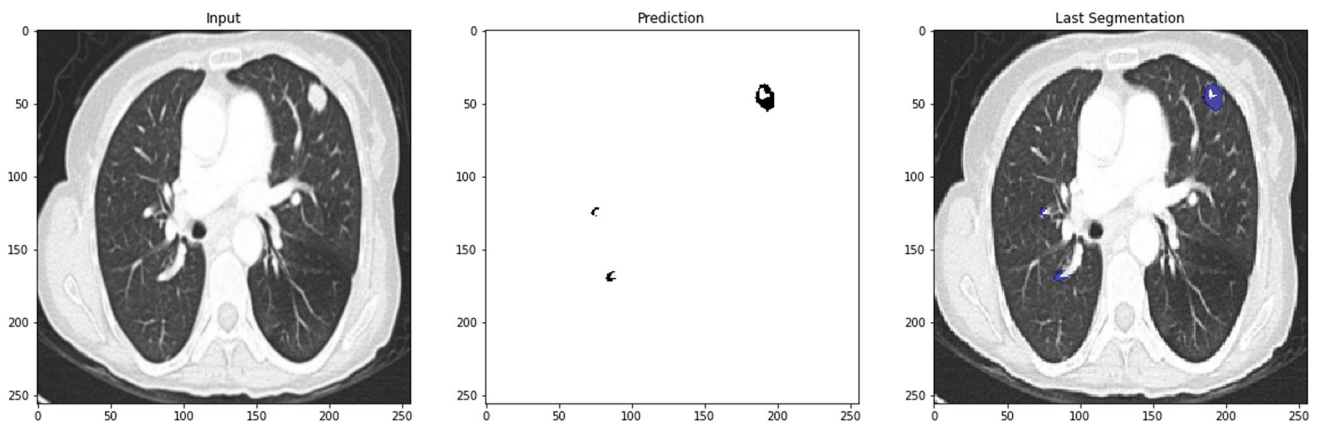
3.
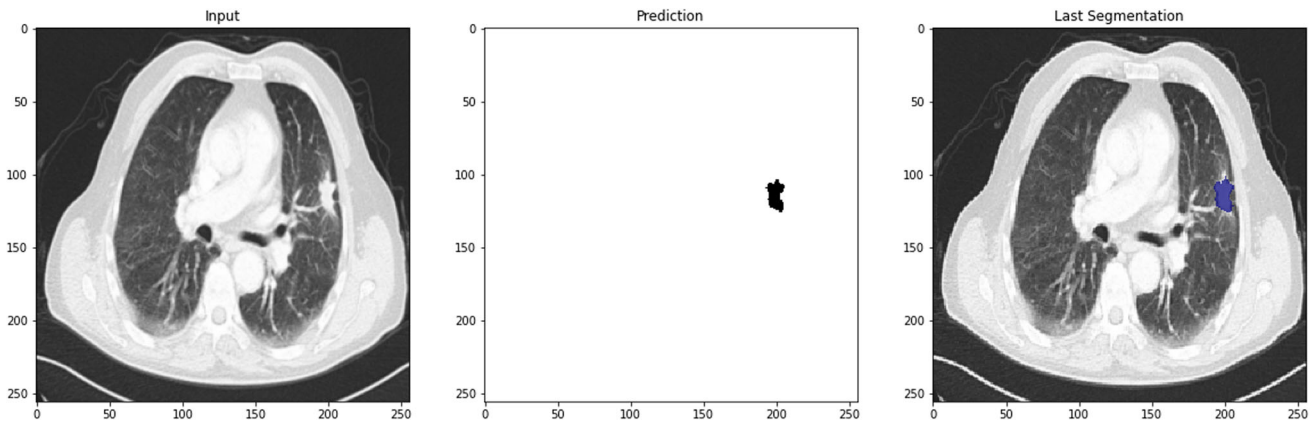
A nodular mass lesion with irregular borders and lobulated contours is observed in the upper lobe of the left lung. The lesion is close proximity to the pleura (lung membrane), but the lung parenchyma (black tissue) is selected in between. Our algorithm detected the lesion with high accuracy and did not detect any false focus.

4.
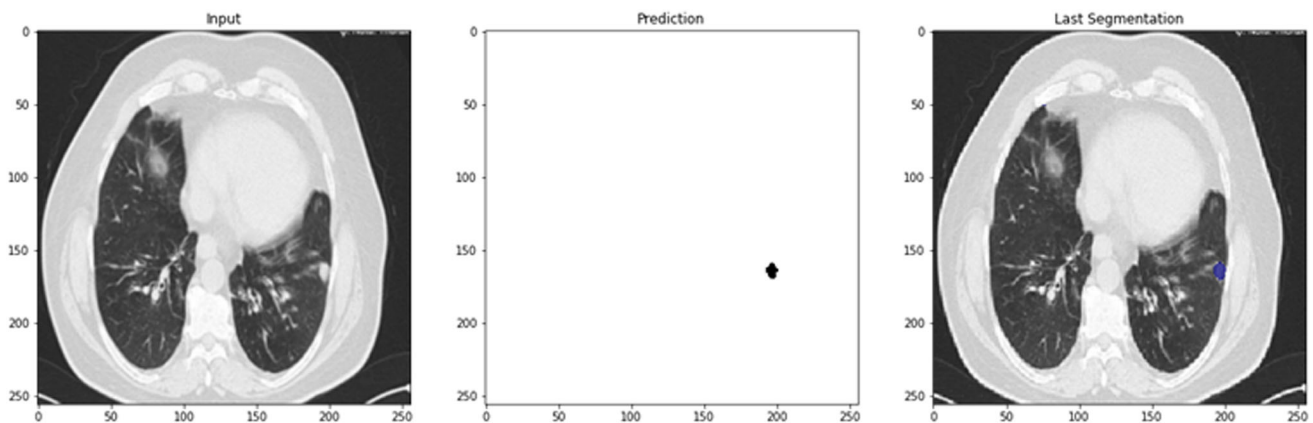
There is a nodule close to the pleura in the lower lobe of the left lung. The algorithm was successful in showing the lesion and did not extract an incorrect focus outside the lesion.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest There is no conflict of interest between the authors.

Ethics approval Ethical permissions were obtained from Van Yuzuncu Yil University Hospital for the study.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. WHO (2022) Cancer. In: WHO. https://www.who.int/news-room/fact-sheets/detail/cancer
2. Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics, 2022. CA Cancer J Clin 72:7–33. https://doi.org/10.3322/caac.21708
3. AIHW (2021) Cancer in Australia 2021. Australian Institute of Health and Welfare, Australia
4. Siegel R, Ma J, Zou Z, Jemal A (2014) Cancer statistics, 2014. CA Cancer J Clin 64:9–29. https://doi.org/10.3322/caac.21208
5. WHO International Agency for Research Cancer (2021) The Global Cancer Observatory
6. Debevec L, Debeljak A (2007) Multidisciplinary management of lung cancer. J Thorac Oncol 2:577. https://doi.org/10.1097/JTO.0b013e318060f16d
7. Swensen SJ, Jett JR, Hartman TE, et al (2005) Radiology CT screening for lung cancer: five-year prospective. Cancer 259–265
8. Agarwal A, Patni K, Rajeswari D (2021) Lung cancer detection and classification based on Alexnet CNN. In: 2021 6th International conference on communication and electronics systems (ICCES). IEEE, pp 1390–1397
9. Zheng Y, Dong J, Yang X et al (2023) Benign-malignant classification of pulmonary nodules by low-dose spiral computerized tomography and clinical data with machine learning in opportunistic screening. Cancer Med 12:12050–12064. https://doi.org/10.1002/cam4.5886
10. Zhu H, Liu W, Gao Z, Zhang H (2023) Explainable classification of benign-malignant pulmonary nodules with neural networks and information bottleneck. IEEE Trans Neural Netw Learn Syst pp (1–12). https://doi.org/10.1109/TNNLS.2023.3303395
11. Prosper AE, Kammer MN, Maldonado F et al (2023) Expanding role of advanced image analysis in CT-detected indeterminate pulmonary nodules and early lung cancer characterization. Radiology 309:e222904. https://doi.org/10.1148/radiol.222904
12. Gugulothu VK, Balaji S (2023) An automatic classification of pulmonary nodules for lung cancer diagnosis using novel LLXcepNN classifier. J Cancer Res Clin Oncol 149:6049–6057. https://doi.org/10.1007/s00432-022-04539-4
13. Lima T, Luz D, Oseas A et al (2023) Automatic classification of pulmonary nodules in computed tomography images using pre-trained networks and bag of features. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-14900-5
14. Saied M, Raafat M, Yehia S, Khalil MM (2023) Efficient pulmonary nodules classification using radiomics and different artificial intelligence strategies. Insights Imaging 14:91. https://doi.org/10.1186/s13244-023-01441-6
15. Qiu J, Li B, Liao R et al (2023) A dual-task region-boundary aware neural network for accurate pulmonary nodule segmentation. J Vis Commun Image Represent 96:103909. https://doi.org/10.1016/j.jvcir.2023.103909
16. Kido S, Kidera S, Hirano Y et al (2022) Segmentation of lung nodules on CT images using a nested three-dimensional fully connected convolutional network. Front Artif Intell 5:1–9. https://doi.org/10.3389/frai.2022.782225
17. Bhattacharjee A, Murugan R, Goel T et al (2023) Pulmonary nodule segmentation framework based on fine-tuned and pre-trained deep neural network using CT images. Front Artif Intell 7:1–9. https://doi.org/10.3389/frai.2022.782225
18. Savic M, Ma Y, Ramponi G et al (2021) Lung nodule segmentation with a region-based fast marching method. Sensors 21:1–32. https://doi.org/10.3390/s21051908
19. Canayaz M (2021) C+EffxNet: a novel hybrid approach for COVID-19 diagnosis on CT images based on CBAM and EfficientNet. Chaos Solitons Fractals 151:111310. https://doi.org/10.1016/j.chaos.2021.111310
20. Al-Shabi M, Shak K, Tan M (2022) ProCAN: progressive growing channel attentive non-local network for lung nodule classification. Pattern Recognit 122:108309. https://doi.org/10.1016/j.patcog.2021.108309
21. Fu X, Bi L, Kumar A et al (2022) An attention-enhanced cross-task network to analyze lung nodule attributes in CT images. Pattern Recognit 126:108576. https://doi.org/10.1016/j.patcog.2022.108576
22. Heuvelmans MA, van Ooijen PMA, Ather S et al (2021) Lung cancer prediction by deep learning to identify benign lung nodules. Lung Cancer 154:1–4. https://doi.org/10.1016/j.lungcan.2021.01.027
23. Apostolopoulos ID, Papathanasiou ND, Panayiotakis GS (2021) Classification of lung nodule malignancy in computed tomography imaging utilizing generative adversarial networks and semi-supervised transfer learning. Biocybern Biomed Eng 41:1243–1257. https://doi.org/10.1016/j.bbe.2021.08.006
24. Gu D, Liu G, Xue Z (2021) On the performance of lung nodule detection, segmentation and classification. Comput Med Imaging Graph 89:101886. https://doi.org/10.1016/j.compmedimag.2021.101886
25. He W, Li B, Liao R et al (2022) An ISHAP-based interpretation-model-guided classification method for malignant pulmonary nodule. Knowledge-Based Syst 237:107778. https://doi.org/10.1016/j.knosys.2021.107778
26. Astaraki M, Zakko Y, Toma Dasu I et al (2021) Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features. Phys Med 83:146–153. https://doi.org/10.1016/j.ejmp.2021.03.013
27. Halder A, Chatterjee S, Dey D (2022) Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification. Biomed Signal Process Control 72:103347. https://doi.org/10.1016/j.bspc.2021.103347

28. Huang H, Li Y, Wu R et al (2022) Benign-malignant classification of pulmonary nodule with deep feature optimization framework. Biomed Signal Process Control 76:103701. https://doi.org/10.1016/j.bspc.2022.103701

29. Jin H, Yu C, Gong Z et al (2023) Machine learning techniques for pulmonary nodule computer-aided diagnosis using CT images: a systematic review. Biomed Signal Process Control 79:104104. https://doi.org/10.1016/j.bspc.2022.104104

30. Yang K, Liu J, Tang W et al (2020) Identification of benign and malignant pulmonary nodules on chest CT using improved 3D U-Net deep learning framework. Eur J Radiol 129:109013. https://doi.org/10.1016/j.ejrad.2020.109013

31. Dutande P, Baid U, Talbar S (2022) Deep residual separable convolutional neural network for lung tumor segmentation. Comput Biol Med 141:105161. https://doi.org/10.1016/j.compbiomed.2021.105161

32. Tyagi S, Talbar SN (2022) CSE-GAN: a 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation. Comput Biol Med 147:105781. https://doi.org/10.1016/j.compbiomed.2022.105781

33. Liu H, Cao H, Song E et al (2019) A cascaded dual-pathway residual network for lung nodule segmentation in CT images. Phys Med 63:112–121. https://doi.org/10.1016/j.ejmp.2019.06.003

34. Szegedy C, Vanhoucke V, Ioffe S, et al (2016) Rethinking the inception architecture for computer vision. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2016-Dec, pp 2818–2826. https://doi.org/10.1109/CVPR.2016.308

35. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proc 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017-Jan, pp 2261–2269. https://doi.org/10.1109/CVPR.2017.243

36. Hu J, Shen L, Albanie S et al (2020) Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 42:2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372

37. Gunaydin O, Gunay M, Sengel O (2019) Comparison of lung cancer detection algorithms. In: 2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT). IEEE, pp 1–4

38. Jolliffe IT, Cadima J (2015) Principal component analysis: a review and recent developments. Philos Trans A 374:1–16. https://doi.org/10.1098/rsta.2015.0202

39. Salihasan BM, Abdulazeez AM (2021) A review of principal component analysis algorithm for dimensionality reduction. J Soft Comput Data Min. https://doi.org/10.30880/jscdm.2021.02.01.003

40. Ma J, Yuan Y (2019) Dimension reduction of image deep feature using PCA. J Vis Commun Image Represent 63:102578. https://doi.org/10.1016/j.jvcir.2019.102578

41. Khalid S, Khalil T, Nasreen S (2014) A survey of feature selection and feature extraction techniques in machine learning. In: 2014 Science and information conference. IEEE, pp 372–378

42. Mwangi B, Tian TS, Soares JC (2014) A review of feature reduction techniques in neuroimaging. Neuroinformatics 12:229–244. https://doi.org/10.1007/s12021-013-9204-3

43. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2:37–52. https://doi.org/10.1016/0169-7439(87)80084-9

44. Raghavendra S, Chandra DP (2014) Support vector machine applications in the field of hydrology: a review. Appl Soft Comput 19:372–386. https://doi.org/10.1016/j.asoc.2014.02.002

45. Sathishkumar R, Kalaiarasan K, Prabhakaran A, Aravind M (2019) Detection of lung cancer using SVM classifier and KNN algorithm. In: 2019 IEEE international conference on system, computation, automation and networking (ICSCAN). IEEE, pp 1–7

46. Anil Kumar C, Harish S, Ravi P et al (2022) Lung cancer prediction from text datasets using machine learning. Biomed Res Int 2022:1–10. https://doi.org/10.1155/2022/6254177

47. Osowski S, Siwek K, Markiewicz T (2004) MLP and SVM networks: a comparative study. In: Proc 6th Nord Signal Process Symp 2004 NORSIG 2004, pp 37–40

48. Zhang Z (2016) Introduction to machine learning: K-nearest neighbors. Ann Transl Med 4:218. https://doi.org/10.21037/atm.2016.03.37

49. Hoerl AE, Kennard RW (2000) American society for quality ridge regression: biased estimation for. 42:80–86

50. Peng C, Cheng Q (2021) Discriminative ridge machine: a classifier for high-dimensional data or imbalanced data. IEEE Trans Neural Netw Learn Syst 32:2595–2609. https://doi.org/10.1109/TNNLS.2020.3006877

51. Ghosh UK, Al Abir F, Rifaat N et al (2022) Most dominant metabolomic biomarkers identification for lung cancer. Inform Med Unlocked 28:100824. https://doi.org/10.1016/j.imu.2021.100824

52. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proc ACM SIGKDD Int Conf Knowl Discov Data Min 13–17 Aug:785–794. https://doi.org/10.1145/2939672.2939785

53. Walia H, Jeevaraj S (2021) Early mortality risk prediction in Covid-19 patients using an ensemble of machine learning models. Int Conf Comput Perform Eval ComPE 2021:965–970. https://doi.org/10.1109/ComPE53109.2021.9751945

54. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: 2015 ieee conference on computer vision and pattern recognition (CVPR). IEEE

55. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical image computing and computer-assisted intervention—MICCAI 2015. Lecture notes in computer science, vol 9351. Springer. Munich, Germany, pp 234–241

56. Liu Y, Fang Q, Jiang A et al (2021) Texture analysis based on U-Net neural network for intracranial hemorrhage identification predicts early enlargement. Comput Methods Programs Biomed 206:106140. https://doi.org/10.1016/j.cmpb.2021.106140

57. Jiao L, Zhao J (2019) A survey on the new generation of deep learning in image processing. IEEE Access 7:172231–172263. https://doi.org/10.1109/ACCESS.2019.2956508

58. Li L, Wei M, Liu B et al (2021) Deep learning for hemorrhagic lesion detection and segmentation on brain CT images. IEEE J Biomed Heal Inform 25:1646–1659. https://doi.org/10.1109/JBHI.2020.3028243

59. Cao G, Wang Y, Zhu X, et al (2020) Segmentation of intracerebral hemorrhage based on improved U-Net. In: 2020 IEEE conference on telecommunications, optics and computer science (TOCS). IEEE, pp 183–185

60. Oghli MG, Shabanzadeh A, Moradi S et al (2021) Automatic fetal biometry prediction using a novel deep convolutional network architecture. Phys Med 88:127–137. https://doi.org/10.1016/j.ejmp.2021.06.020

61. Araújo RL, de Araújo FHD, Silva RRV (2021) Automatic segmentation of melanoma skin cancer using transfer learning and fine-tuning. Multimed Syst. https://doi.org/10.1007/s00530-021-00840-3

62. Chaurasia A, Culurciello E (2017) LinkNet: exploiting encoder representations for efficient semantic segmentation. https://doi.org/10.1109/VCIP.2017.8305148

63. Sobhaninia Z, Rezaei S, Karimi N, et al (2020) Brain tumor segmentation by cascaded deep neural networks using multiple image scales. In: 2020 28th Iranian conference on electrical engineering (ICEE). IEEE

64. Singh J, Tripathy A, Garg P, Kumar A (2020) Lung tuberculosis detection using anti-aliased convolutional networks. Proc Comput Sci 173:281–290. https://doi.org/10.1016/j.procs.2020.06.033

65. Seferbekov SS, Iglovikov VI, Buslaev A V., Shvets AA (2018) Feature pyramid network for multi-class land segmentation. https://doi.org/10.48550/arXiv.1806.03510

66. Adelson EH, Anderson CH, Bergen JR et al (1984) Pyramid methods in image processing. RCA Eng 29:33–41

67. Lin T-Y, Dollar P, Girshick R, et al (2017) Feature pyramid networks for object detection. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 936–944

68. Hu M, Li Y, Fang L, Wang S (2021) A2-FPN: attention aggregation based feature pyramid network for instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 15343–15352

69. Hossin M, Sulaiman M (2015) A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 5:01–11. https://doi.org/10.5121/ijdkp.2015.5201

70. Morris R, MacNeela P, Scott A et al (2008) Ambiguities and conflicting results: the limitations of the kappa statistic in establishing the interrater reliability of the Irish nursing minimum data set for mental health—a discussion paper. Int J Nurs Stud 45:645–647

71. Costa L da F. (2021) Further generalizations of the jaccard index. https://doi.org/10.48550/arXiv.2110.09619

72. Antorán J, Adel T, Weller A, Hernández-lobato JM (2021) Continuous dice coefficient: a method for evaluating probabilistic segmentations, pp 1–34

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.