International Journal of Advanced Science and Research
www.allsciencejournal.com
ISSN: 2455-4227
Received: 23-11-2025, Accepted: 22-12-2025, Published: 06-01-2026
Volume 11, Issue 1, 2026, Page No. 7-18

Development of an Xception-based convolutional autoencoder (Xnetcae) feature extraction technique for enhanced lung cancer detection

Olaleye O J[1], Olabiyisi S O[2], Ismaila W O[2], Achas M J[1], Ashade B T[1]
[1] Department of Computer Science and Information Technology, Bells University of Technology, Ota, Nigeria
[2] Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

Abstract
This study presents an enhanced deep learning feature extraction technique for early detection of lung cancer from low-dose computed tomography (LDCT) images. The Xception network was integrated into a standard convolutional autoencoder (CAE) architecture and the resultant Xception-based Convolutional Autoencoder (XnetCAE) was trained, validated and tested with a portion (20,000) of lung scan images from the Lung Image Database Consortium-Image Database Resource Initiative (LIDC-IDRI) dataset. The performance of the XnetCAE was evaluated and compared with that of a standard CAE using reconstruction loss, accuracy, precision, recall, F1-score and computation time metrics. With the XnetCAE technique, the reconstruction loss, accuracy, precision, recall, F1-score and computation time are 0.16, 71.97%, 0.79, 0.72, 0.75 in 36 seconds, while the standard CAE achieved 0.30, 65%, 0.65, 0.65, 0.65 in 48 seconds, respectively. These results show that the XnetCAE offers a superior performance over standard CAE.

Keywords: Deep learning, convolutional autoencoder, xception network, feature extraction, medical image analysis

Introduction

Nearly 2 million deaths are recorded annually from lung cancer alone making it the cancer with the highest mortality rate all over the world accounting for about 20% of all cancer-based deaths (Habbab *et al*., 2025; Balata *et al*., 2022; Sung *et al*., 2021) [1, 2, 3]. Detection of cancer early, particularly cancer of the lung gives room for timely and more successful treatment thereby reducing the rate of death in patients (Hu *et al*., 2023; Abbasian *et al*., 2022) [4, 5].

One notable technique used for the identification of cancer is by the critically assessing the image of the affected area. Some of the imaging techniques employed for the detection of lung cancer include chest x-ray (CXR), computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET) and Ultrasound (Grigory et. al., 2025; Tarnoki et. al., 2024). Amongst these imaging techniques, CT remains the most valuable for large-scale screening and produces faster results. Particularly, low-dose CT is considered very efficient for nodule detection due to its low radiation exposure (Tarnoki et. al., 2024; Prosper et. al., 2023;) [8]. However, LDCT images are often characterized by noise, artifacts, and lower contrast, which can reduce clarity and detection of small nodules that makes early detection difficult (Chen et. al., 2025; Tugwell-Allsup *et al*., 2021) [9, 10].

In the realm of enhancement, feature extraction, and image denoising, convolutional autoencoders (CAEs) have found great applicability in the medical imaging sector, too. They can be employed for the analysis of large-scale medical image datasets due to their excellent ability to learn representations of hidden features embedded in the input images (Pintelas *et al*., 2021) [11]. But standard CAE models have some limitations. They utilize standard convolutional layers, which make them inefficient, unable to capture complex spatial relationships (or fine details) in medical images, and prone to overfitting, especially in low-dose datasets where crucial diagnostic details are less (Shobayo and Saatchi, 2025; Berahmand *et al*., 2024; Iqbal *et al*., 2023; Lu et. al, 2022) [12, 13, 14, 15].

The Xception network utilizes depthwise separable convolutions and is able to capture features at a higher degree while utilizing fewer computational resources. Xception can remedy many of the shortcomings seen in traditional CAEs, and while being first in its class, Xception has yet to be used in the field of medical imaging pertaining lung cancer diagnostics (Chollet, 2017) [16]. This paper aims to develop an Xception-based Convolutional Autoencoder feature extraction technique for enhanced lung cancer detection. The objectives are to integrate Xception network with CAE to improve feature extraction capability, implement the XnetCAE with LDCT images in Python environment, determine the diagnostic potential of the XnetCAE tecnique and compare its effectiveness with standard CAE using reconstruction loss, accuracy, precision, recall, F1-score and computation time.

Literature Review

Structure of CAEs

CAEs are neural networks that combine convolutional layers and autoencoders. Medical imaging uses them for dimensionality reduction, feature extraction, and image denoising. CAEs have an encoder and a decoder, shown in Figure 1. The encoder reduces input image to the bottleneck layer (latent space). The decoder reconstructs compressed input data (Jayeprokash and Gonski, 2025; Berahmand *et al*., 2024) [13, 17]. Utilizing error functions such as MSE, Convolutional autoencoders (CAEs) attempt to lessen the differences between the input and its reconstruction output. Due to their high ability to represent the data, CAEs are good at feature extraction as well as reducing data dimensions (Vincent *et al*., 2008) [18].

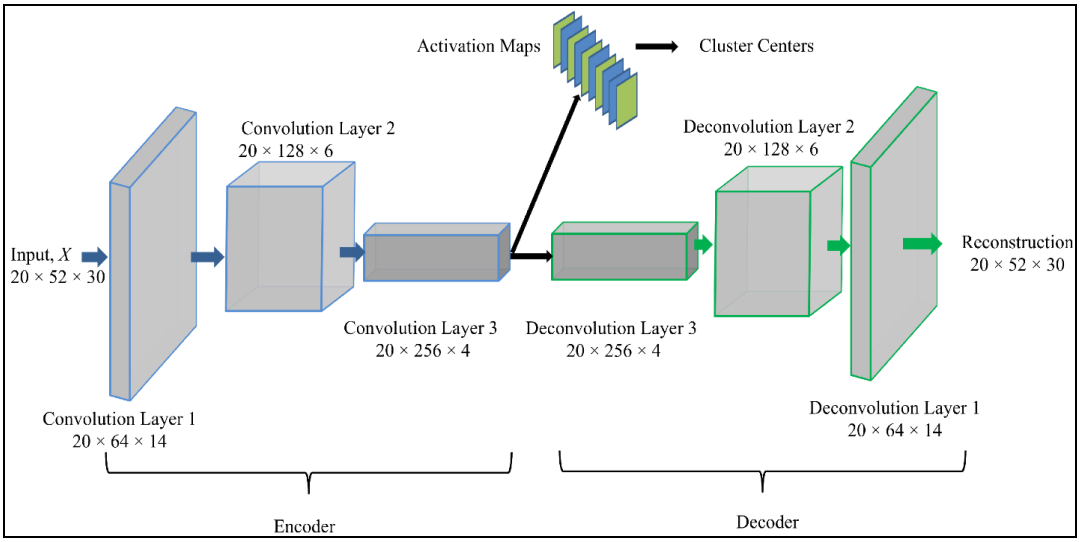

Fig 1: Architecture of CAE (Chadha *et al*., 2021) [19]

CAE Model for Feature Extraction and Image Reconstruction

CAEs reconstruct images using a convolutional encoding-decoding process to capture important features. For CAEs, image reconstruction involves the following processes (Xuan *et al*, 2025) [20]

1. **Encoder Function:** The encoder compresses the image into a lower representation.

$$Z = f_\theta(I) \quad (1)$$

where Z is the latent representation of the image, I is the input lung image, and f_θ is the encoder function with weights θ. The encoder is made up of many convolutional layers that are followed by functions of activation, which are commonly ReLU (Rectified Linear Unit)

$$Z = \sigma(W * I + b) \quad (2)$$

Where σ is ReLU function of activation, bias term is depicted by b, ∗ is convolution operation, and W is weight matrix of convolutional filters

$$\sigma(x) = \max(0, x) \quad (3)$$

The encoder uses max-pooling to minimize the input image's spatial sizes

$$Z = \text{MaxPool}(f_\theta(I)) \quad (4)$$

2. **Bottleneck Illustration:** The compressed feature representation Z at the bottleneck layer captures the lung nodules' most relevant spatial and structural properties. For image reconstruction (decoder), this format is helpful.

3. **Decoder Function:** Using the latent representation, the decoder recreates the initial input image

$$\hat{I} = g_\phi(Z) \quad (5)$$

where g_ϕ is the decoder function parameterised by weights ϕ, and I represents the reconstructed image. The decoder upscales the feature maps to the original picture dimensions using transposed convolutions, often known as deconvolution layers

$$\hat{I} = \sigma(W^T * Z + b) \quad (6)$$

Lastly, Mean Squared Error (MSE) is utilized to minimize reconstruction error

$$L_{recon} = \frac{1}{N} \sum_{i=1}^{N} (I_i - \hat{I}_i)^2 \quad (7)$$

N denotes the number of pixels in the image.

Xception Network

Extreme Inception, shortly called Xception, is a deep learning network developed by at Google. Chollet (2017) [16] explains the major advancements the Xception model has achieved due to the incorporation of depthwise convolutions, making it more efficient, and expanding the scope of the model to process more complex images and textured images with higher resolutions. Holding more layers than its predecessors, Xception applies separable convolutions and has added depth as a result streamlining the model with more efficient use of computational resources.

Depthwise Separable Convolutions

The single greatest advancement is the depthwise separations where standard convolution is partitioned into two parts. This design has shown to reduce the parameters needed as well as the complexity present within the computations, allowing the model to run more efficiently and faster.

Depthwise Convolution

In this stage, the convolution is performed on each of the input channels separately. Each of the channels is separated by reference to the kernel of a certain size in two dimensions of C by C and a single layer (the depth of the input). Each of the channels will also output some data in a separate depth of size K by K by Z. This separation has the purpose of increasing the efficiency of the computation since each of the channels can be processed as a separate convolution channel.

Pointwise Convolution

In this stage, outputs of the depthwise convolution are summed with a 1 × 1 × N kernel. Since all K channels are used, the pointwise convolution produces an output volume of K × K × N. This step maintains computing efficiency while allowing the model to capture cross-channel correlations. Jointly performing these two steps is what may allow the Xception model to achieve good performance with low parameter counts and fewer calculations.

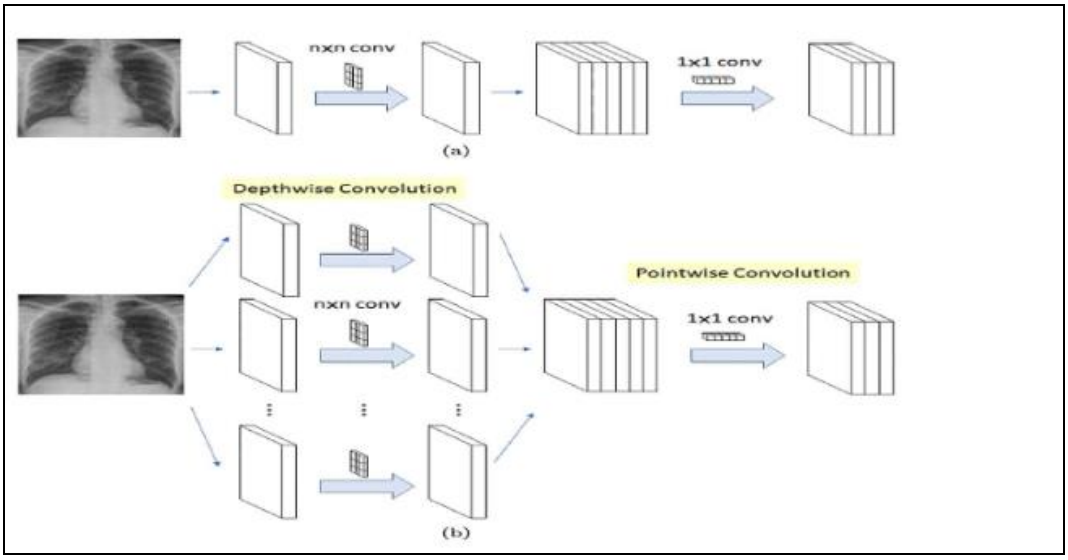

Fig 2: (a) Standard Convolution and (b) Depthwise Separable (Chollet 2017) [16]

Max Pooling for Efficiency

Max pooling layers boost Xception architecture efficiency. Figure 3 demonstrates how max pooling reduces feature map spatial dimensions by picking the maximum value from each pooling region. This method minimizes network's parameter count and computational cost while preserving its main features. Max pooling also reduces overfitting and increases model generalization by providing translation invariance. Max pooling optimizes the Xception architecture by focusing on critical input data, resulting in cleaner feature representations (Zhao and Zhang 2024; Zafar *et al*. 2022) [21, 26, 22].

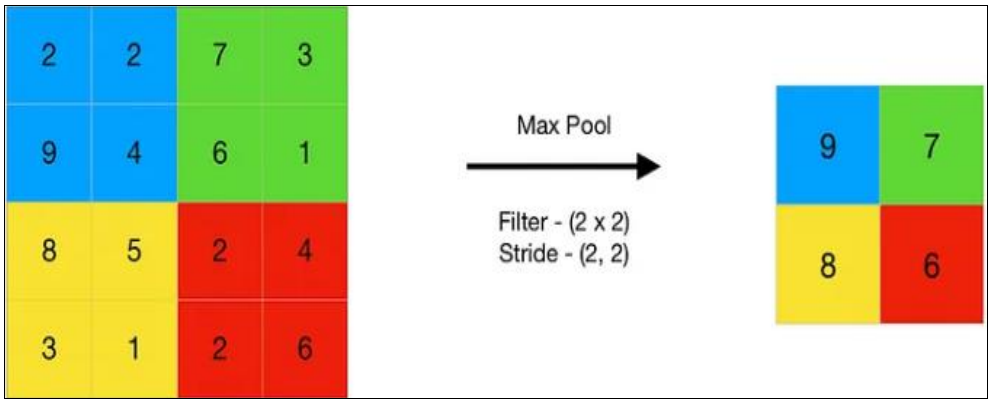

Fig 3: Max Pooling

Related Works

Medical image analysis integrates CAEs with typical convolutional layers for extraction of feature and learning of representation. These layers are computationally expensive, especially for high-dimensional and high-resolution medical pictures, which has been a concern in previous investigations. Research found that CAEs' local convolution procedures distorted reconstructions and prevented global context capture, making it difficult to identify small lung lesions (Berahmand *et al*. 2024; Zhuang *et al*. 2023) [13, 23]. Benchmarking revealed that deep CAEs use more GPU RAM and take longer to train, up to 2.3x longer than typical CNN classifiers. They found that deeper or higher-resolution CAEs increase computing cost without enhancing performance. For real-world medical applications, the research recommended depthwise separable convolutions and multi-objective optimisation to reduce redundancy and scale (Feng *et al*., 2024; Younesi *et al*., 2024) [24, 25]. Recent work created more efficient structures to overcome such restrictions.

(Zhao *et al*. 2024) [21, 26] investigated CAEs on low-resolution CT images for lung cancer detection. CAEs were evaluated to denoise and reconstruct structural properties essential for appropriate categorization. CAE model training achieved high-quality image reconstruction from noisy, low-resolution CT inputs. Qualitative evaluations indicated that the model distorted lung nodule and tissue borders while maintaining PSNR and SSIM. Noise changed diagnostic characteristics in CAEs, investigators found. Traditional encoder–decoder convolutional layers lacked spatial

precision for clinical interpretation. Thus, although the rebuilt outputs were more appealing, they often lacked fine-grained structures for classification or segmentation, limiting their therapeutic potential.

Wang *et al*. (2023) [27] tested CAEs for lung cancer CT scan detection. CAEs trained on reconstruction loss (mean squared error) were compared to hybrid models optimised for reconstruction and feature representation categorisation. Their research indicated that CAEs maintain image structure and texture during reconstruction, but their learnt features lack downstream classification power. CAE bottleneck representations showed significant redundancy and overlap between benign and malignant nodules. They blamed the model's narrow optimisation objective, which minimises pixel-wise error rather than class semantics, for this repetition. They showed that CAE models struggled with latent space class separability even with deeper network configurations and data augmentation. Classification accuracy plateaued at 85.4%, lower than dual goal or integrated attention mechanism models. They found that CAEs taught with a single reconstruction aim overfit to superficial characteristics and fail to emphasise clinically significant differences, limiting their diagnostic efficacy.

(Da Cruz, 2022; Alberti *et al*., 2017) [28, 29] examined classic CAEs' architectural constraints in feature extraction and classification on thoracic imaging datasets such lung CT images. The research evaluated how single-objective training hinders CAE model generalization and discrimination. They trained multiple CAE versions on large synthetic and real-world CT datasets to test downstream classifiers that employed encoded features and picture quality. CAEs trained for image reconstruction provided visually consistent but task-inutility features. Latent vectors failed to detect small morphological differences between early-stage cancers and benign nodules, resulting in poor classification metrics. They showed how single-objective loss functions underutilize CAE model capabilities. The encoder compressed pixel intensity patterns across diagnostic spatial or contextual characteristics. Additionally, their ablation investigation showed that introducing an extra classification loss increased model performance by encouraging the encoder to acquire semantically relevant representations. Their conclusion stressed that CAEs are effective for noise reduction and unsupervised representation learning, but their single-objective nature slows classification tasks. They recommended multi-task learning frameworks or optimisation methods that account for task-specific restrictions and clinical priors to increase performance.

Tang *et al*. (2018) [30] analysed volumetric lung nodules using 3D CNNs. Their spatial context-based method identified benign and malignant nodules with high accuracy. However, 3D photo processing required more computer resources, making it challenging in therapy. A study according to Çiçek *et al*. (2016) [31] shows that computational load slows processing and limits experimentation with larger datasets and more complex designs. These results suggest that CAEs that use standard convolution layers may be less adaptive, especially in fast analysis and deployment.

A medical study on imaging-based lung cancer diagnosis has improved detection and classification. Comparable research has limits, performance, and contributions. Pathak *et al*, (2025) [32] stated that traditional convolutions' computational needs limit deep learning model scalability in medical imaging, especially for high-resolution input applications like radiology and histopathology image processing. Shan *et al*. (2018) [33] demonstrated that deep convolutional encoder–decoder models can enhance CT image quality, but models trained only to reduce reconstruction loss tend to prioritize pixel fidelity rather than diagnostic semantics, limiting generalization for downstream classification tasks.

After applying learnable filters to the input volume, conventional convolutional layers create several feature maps using element-wise multiplications and summations. This technique generates many multiply-accumulate (MAC) operations and parameters that expand exponentially with network depth and input size. Training and inference become computationally expensive and need high-performance hardware, which may not be viable in many medical contexts. Chollet (2017) [16] revealed that standard convolutional operations use more processing resources than depthwise separable convolutions in the proposed Xception architecture. He demonstrated that depthwise separable convolutions may minimize computation burden while maintaining accuracy, demonstrating the inadequacies of typical convolutional architectures for large-scale image processing.

Methodology

Data Acquisition

Due to its thorough annotations, the LIDR-IDRI datasets are widely used in lung cancer research. This work employed thoracic CT scan pictures from them. These datasets were downloaded from a public source (www.kaggle.com). The dataset comprises 2D cross-sectional slices from 3D thoracic volumes labelled for lung nodules. These annotations were used to train and test the XnetCAE for feature reconstruction and lung cancer diagnosis.

Dataset Preprocessing

Original CT scan images were DICOM. Preprocessing began by scanning each picture slice and converting pixel values into Hounsfield units (HU) utilizing image file metadata (Rescale Slope and Intercept). This stage verified that pixel values accurately represented air, lung, soft tissue, and bone. Next, HU values were trimmed between -1000 and 400. This removed air outside the body and thick regions like bones to concentrate on the lung area and potential nodules. This reduced noise and highlighted diagnostic features. After cutting, images were normalized to [0, 1]. This is essential for deep learning model deployment since it simplifies visual interpretation and learning. Image slices were then resized. The XnetCAE technique required comparable image sizes, and thus decreased the amount of data the model needed to analyze while preserving critical characteristics.

Formulation of the Xception-Based Convolutional Autoencoder (XnetCAE) Model for Detection of Cancer of the Lungs

In this research, an XnetCAE technique was developed by integrating Xception network into the standard CAE architecture to enable more effective feature extraction, reconstruction of LDCT images, and improved computation efficiency for the identification of cancer of the lungs. Figure 4 depicts the architecture of the developed XnetCAE technique. The primary layers of the XnetCAE model are the Xception Encoder, Decoder, and classification layer.

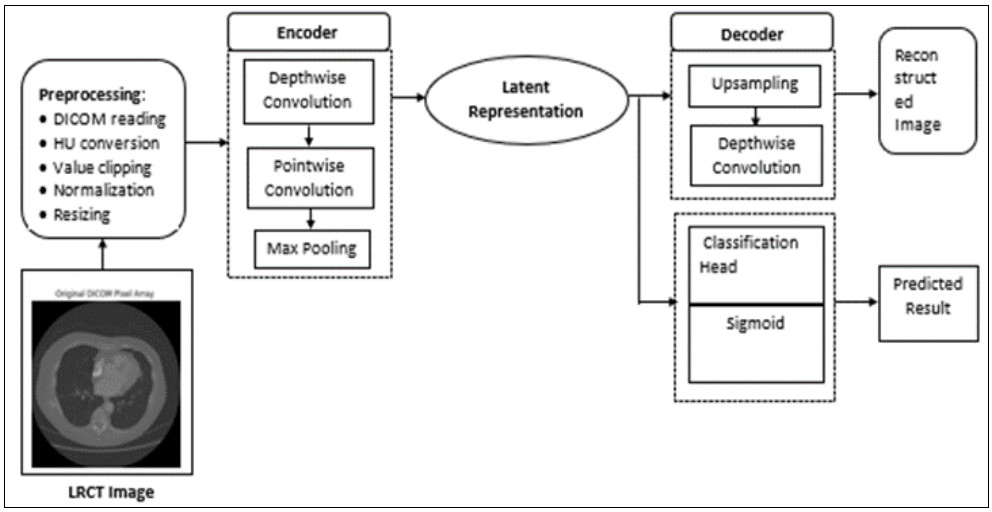

Fig 4: Architecture of the Developed XnetCAE Technique

1. Xception Encoder

Instead of convolutional layers, the encoder uses Xception to convert the LDCT image (I) to latent representations (Z) with lower dimension. Xception blocks perform depthwise and pointwise convolutions. First, channel-wise depthwise convolution filters spatial information

$$Z_{depth} = W_{depth} * I \quad (8)$$

Where Z_{depth} is depthwise convolution output, the depthwise convolutional filter weights is represented by W_{depth}, $*$ is a convolution operator. Pointwise convolution comes next by combining the outputs of the depthwise convolution using 1x1 filters

$$Z_{point} = W_{point} * Z_{depth} + b \quad (9)$$

Where b and W_{point} signify the pointwise filter biases and weights, respectively. To introduce nonlinearity to the output of each Xception block, the activation function σ(x), serving as the ReLU is utilized as indicated in Equation (3)

$$Z = \sigma(W_{point} * Z_{depth} + b) \quad (10)$$

After convolution, max pooling is used to decrease spatial resolution and form latent representation

$$Z = \text{MaxPool}\,(\sigma(W_{point} * (W_{depth} * I) + b)) \quad (11)$$

The latent representation Z represents lung nodule spatial information and is utilised for image reconstruction and classification.

2. Decoder

The original picture is rebuilt from latent characteristics Z. It restores spatial dimensions using upsampling layers and depthwise convolution

$$\hat{I} = \sigma(W^{I}_{depth} * Up(Z) + b) \quad (12)$$

where $\hat{I}$ denotes the reconstructed image(output of the decoder), W^{I}_{depth} are the learned transposed depthwise convolution filters employed to learn and reconstruct spatial features in the upsampled representation. Up(Z) denotes the upsampling function used in the latent space. Equation (7) specifies the MSE metric, used to evaluate the quality of reconstruction.

3. Classification Layer

A latent representation Z is part of a classification block that uses totally connected layers and a sigmoid activation function to classify lung cancer.

The categorization result ($\hat{y}$) is given by

$$\hat{y} = \text{sigmoid}\,(W_c Z + b_c) \quad (13)$$

The classification layer's bias and weights are denoted as b_c and W_c, respectively. To transform the result into probability scores, the function of sigmoid activation was used as given below

$$\text{sigmoid}\,(x) = \frac{1}{1+e^{-x}} \quad (14)$$

The categorization loss is calculated using Categorical Cross-Entropy (CCE)

$$L_{class} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (15)$$

C denotes the number of benign and malignant classes.

Results and Discussion

Conventional Autoencoder (CAE) Technique Implementation Results

An initial phase of this research employed CAE to serve as a benchmark for feature extraction and image reconstruction. CAE achieves image reconstruction by compressing LDCT scans and retaining a lower-dimensional latent representation. This research attempts to determine whether standard autoencoders are able to retain significant diagnostic characteristics during the compression of low-resolution computed tomographic images. This model was trained, validated, and tested on 20,000 datasets, divided and allocated according to a 70%-10%-20% proportion. Figure 5 depicts the CAE technique as having been trained for thirty epochs, wherein training was monitored via a loss measure that computed the variance between validation and training losses within examined epochs.

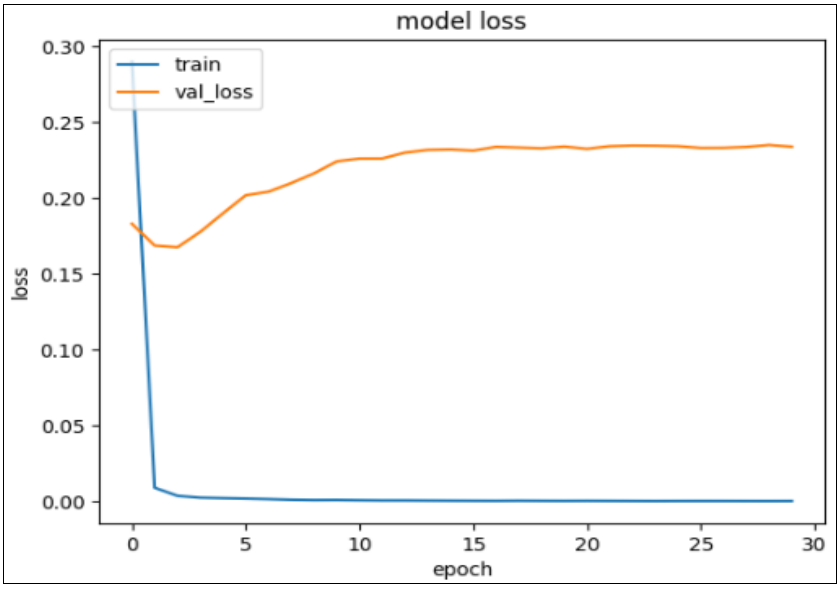

Fig 5: Training History of CAE technique

This clearly show that losses on training data (blue line) dropped almost completely to zero during the initial epochs and stabilized, meaning that this technique was able to fit the input data quite good. On the other hand, losses on the validation data (orange line) started out with a higher value than that of the training data, and during the initial epochs of training, drove validation losses down a bit. After that, however, the validation loss increased, and sustained a gap that signified that the model was trained on data it had never seen before. These circumstances are generally interpreted as overfitting and/or lack of model's latent characteristics to generalization of new data. This indicates that the model, after obtaining the training data, failed to generalize new information.

Content presented on Figure 6 depicts visuals concerning the performance of reconstructed images generated by the CAE on low dose Lung CT scans post training. The visuals have been arranged in three horizontal segments whereby each segment corresponds to the original image, compressed image, and decoded image. Topmost segment contains the original CT scan images through which the CAE was trained. Also shown in the images are different cross-sectional views of thorax, and images of lung regions which are either normal or abnormal. The middle segment includes the outputs of the encoder image as compressed encoded representations that are latent. The encoded compressed latent representations are less distinct in the images. Thus, the CAE was able to perform dimensionality reduction and sustain only the bare minimum content which resides in the image. The last segment of the CAE grids contains the outputs from the decoder which are images of the input features that had been encoded and are still noted to be reconstructed. The decoder did not perform pixel perfect performance as encoded images were not exact images of the original set. The CAE was able to perform the task successfully in the training set and the outputs have captured important features, while the irrelevant or excess information was removed.

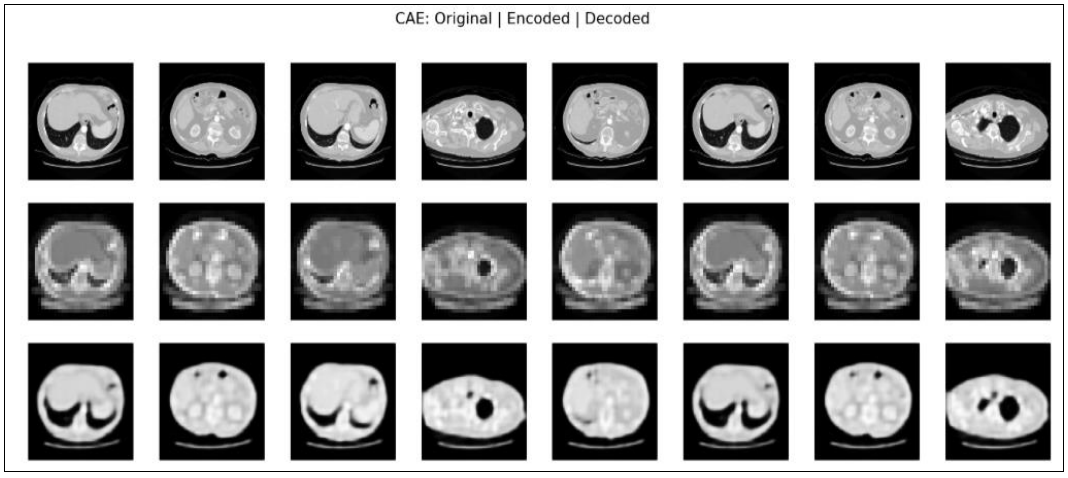

Fig 6: Visual Evidence of Image Reconstruction by CAE Technique

A classifier was trained on CAE features. Figure 7 and Table 1 illustrate how CAE predicted 1,303 benign (non-cancerous) samples as Class 0, True Negatives (TN). This means the model successfully captured and differentiated many healthy samples, thereby alleviating over-diagnosis in the clinical sense. However, the CAE technique incorrectly predicted 694 benign samples as malignant, (i.e., False Positives (FP)). This type of error presents overdiagnosis of the benign samples, leading to potential avoidable patient care in the form of stress and follow-up detailed procedure

plans to avoid patient psychological overburden. In contrast, CAE predicted 1,297 malignant (cancerous) samples correctly as Class 1, True Positives (TP). This is very important as the timely diagnosis of malignant cases is essential as a negative diagnosis in these samples can, in a sense, be leading. On the contrary, CAE did not detect 706 malignant samples of low class, misclassifying them as benign False Negatives (FN). This is the most concerning paradox, procrastinating on the diagnosis, the disease progresses more and more leading to a negative poor outcome of such a paradox. There is a clear need to reduce both FP and FN. This means a clear need to CAE in order to increase accuracy and avoid the dead cases.

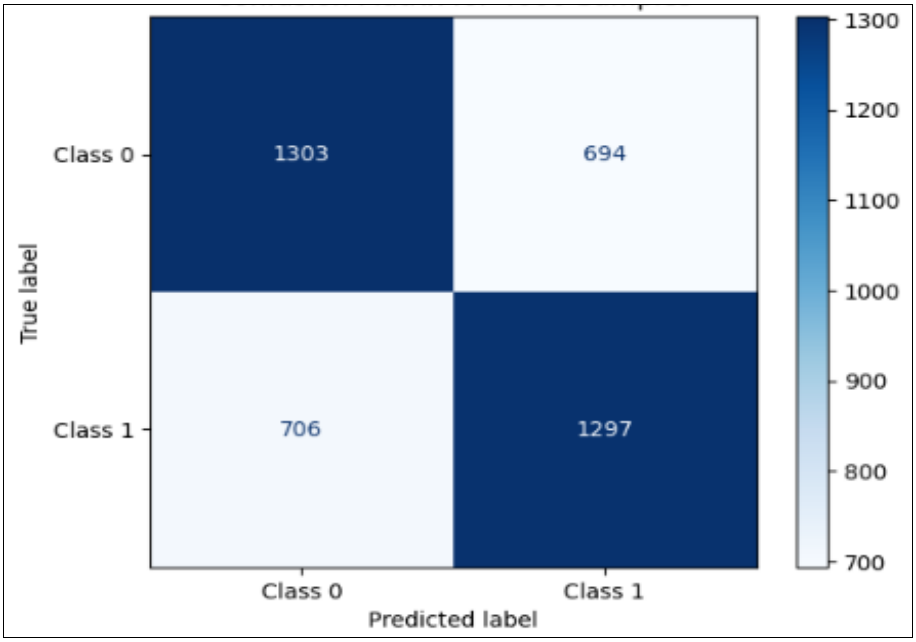

Fig 7: CAE Technique Test Set Confusion Matrix

Table 1: Results Obtained from CAE Technique Test Set Confusion Matrix

	Predicted Class 0	**Predicted Class 1**	**Total**
Actual Class 0	1303 (TN)	694 (FP)	1997
Actual Class 1	706 (FN)	1297 (TP)	2003
Total Predicted	2009	1991	4000

The CAE technique's classification performance on test data is summarized numerically in Table 2. The CAE technique achieved 65% accuracy, which means it showed a fair level of 65% correct recognition of the test dataset and 35% of incorrect recognition, suggesting a fair level of decision-making competence. An accuracy score of 0.65 would mean that this model could forecast the presence of a particular condition and would even have the ability to constrain the false positive rate to a fair level; however, it still requires refinement. A score of 0.65, which is a measure of recall, would mean that this model diagnosed positive cases, and to be precise, it missed 35 out of the 100 actual positive cases. The CAE technique does have a pretty close range of false positive results and false negative results, and the system does not lose performance in the process; however, the output shows that the CAE model performance is subpar, and an F1 score of 0.65 is a good indication of this. A loss of 0.30 is an indication of loss in reconstructing signals of fetching errors; indicating that the model is in a good place, but still needs an extra touch in terms of hyper-specialization of the neural network. With that computation time of 48 seconds, shows it is feasible to test the model in environments that require instant accuracy, but the model still needs work in terms of accuracy for it to be deployed into mission critical frameworks.

Table 2: CAE Technique Performance Evaluation Results

Metrics	**Result**
Accuracy	65%
Precision	0.65
Recall	0.65
F1-score	0.65
Loss	0.30
Computation Time (Sec)	48

Lastly, the information on Figure 8 shows the predicted classes using the feature from the CAE. Then, the Figure displays the samples of CT images with their true labels, predicted classes, and the associated probabilities. Along with the predictions, the CAE model shows this confidence for incorrect predictions and validates the need for improved methods for feature extraction.

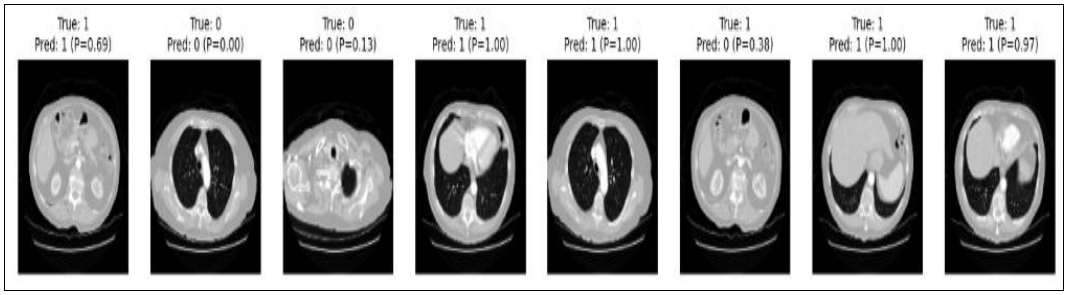

Fig 8: Visual evidence of classification predictions using the CAE technique extracted features.

Xception-based Convolutional Autoencoder (XnetCAE) Technique Implementation Results

Xception layers were added to the encoder network to improve the capacity of the CAE technique. This was done by separating the top classification layers from Xception network and using its convolutional base to obtain different and varied hierarchical spatial data from the input images. The pre-trained layers act as encoders by compressively transforming the original images to a latent representation. The original image is restored from the latent representation by the decoder, using convolutional layers that are transposed. This uses the efficient depthwise separable convolutions of Xception for extraction of feature. The preprocessed lung CT images with low resolution were used for training, validation and testing. Effectiveness of training of the XnetCAE model was determined by the validation and training loss for each epoch, shown in Figure 9.

In the early phases, the validation loss surpasses the training loss, suggesting that the technique's improvement is registered. Training and validation losses align and reach the same value by the 6th epoch. This strongly shows the technique is able to learn the training dataset and also does well on previously unused data. The technique, by design, shows no signs of overfitting on the data. Overfitting is the scenario where the model shows decreasing training loss and increasing validation loss. This is not the scenario, however. The two curves exhibit a small and consistent gap, which is a symptom of behaving well with unseen data.

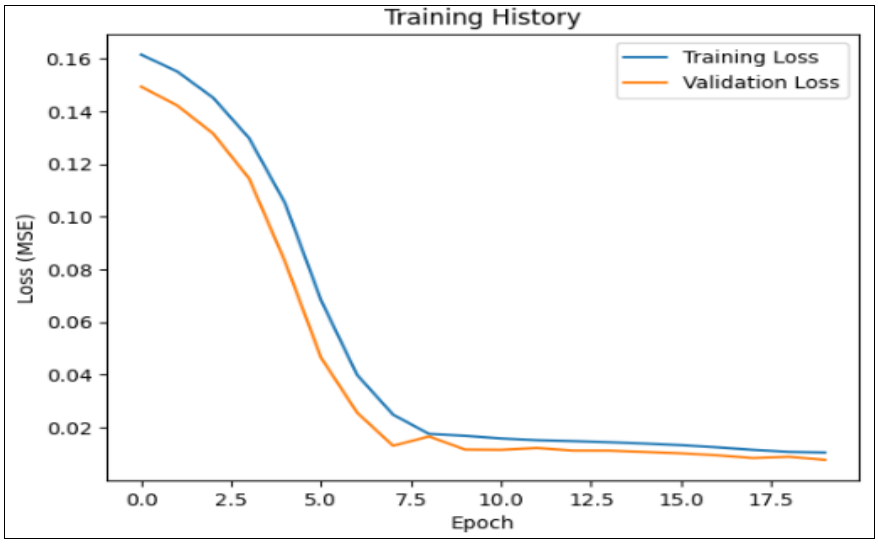

Fig 9: XnetCAE Technique Training History

Figure 10 shows the results of the XnetCAE technique applied on LDCT images with the images being reconstructed.

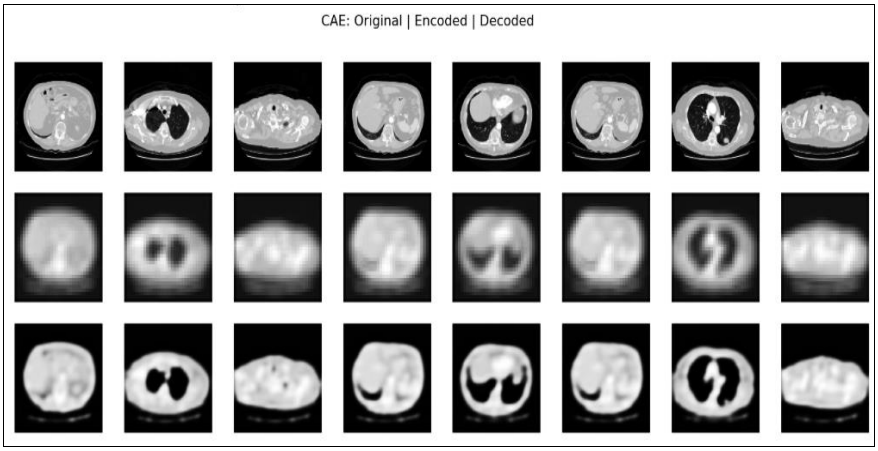

Fig 10: Visual Evidence of Image Reconstruction by XnetCAE Technique

First row depicts some original CT input images of the lung, the second row indicates the latent features encoded by the XnetCAE technique. The large amount of loss of detail and the blurriness of these encoded images is a show of the efficiency of the encoder in compressing the original input images by extracting important features for reconstruction. Third row shows images the decoder network generated using the compressed latent features. The pairs of reconstructed images and original examples are identical concerning structure and internal anatomical organization. These data indicate that the model is able to preserve and recover essential visual information through encoding and decoding processes. While the decoded images lose some detail due to the information compression that naturally occurs with dimensionality reduction, the images are still medically interpretable and could be classified. The images evidently show the XnetCAE captures and reconstructs relevant spatial structure in the LDCT images.

The latent representation obtained from the XnetCAE was entered into a classifier to test the strenght of the feature extracted by the XnetCAE technique. Figure 11 and Table 3 provides the classification outcomes, showing that the technique was able to correctly characterize 1,710 malignant and 1,169 benign cases. 670 of the malignant cases were incorrectly determined to be benign and there were 451 cases that were benign but were determined to be malignant.

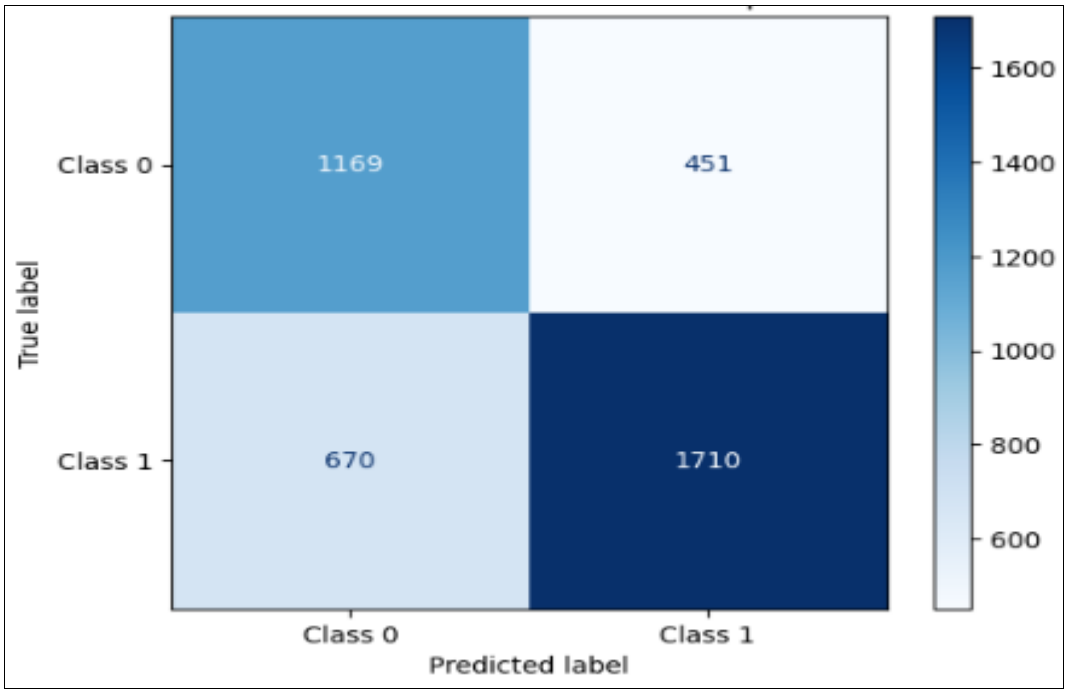

Fig 11: XnetCAE Technique Test Set Confusion Matrix

Table 3: Result Obtained from XnetCAE Technique Test Set Confusion Matrix

	Predicted Class 0	Predicted Class 1	Total
True Class 0	1169 (TN)	451 (FP)	1620
True Class 1	670 (FN)	1710 (TP)	2380
Total Predicted	1839	2161	4000

This results show that the XnetCAE can adequately tell the difference between benign and malignant instances of the lung. The number of false negatives (670) is troubling because of the medical nature of the task. It is possible that misclassifying malignant cases that need attention, could result in delays in diagnosis and treatment and lead to adverse outcomes for the patient. Besides having potential adverse outcomes, the 451 false positive (FP) cases could lead to dismissal of the model for having the potential to cause distress and unnecessary procedures for patients determined to have cancer. These findings imply that although the model has prospect, further optimization is necessary, particularly in enhancing recall to ensure fewer malignant cases are missed, which is important in medical applications where early diagnosis is critical. The quantitative test results of the XnetCAE technique's classification are presented in Table 4. The XnetCAE technique has 71.97% accuracy, identifying 72% of samples correctly and 28% incorrectly, exceeding all baseline CAE reliability parameters. A choice may be needed to boost performance. The model has fewer false positives than the baseline CAE model, as seen by its accuracy score of 0.79, which means 79% of positive predictions were right. In medicine, unrecognized positives may be harmful. The technique had a recall score of 0.72, meaning it correctly detected 72% of positive events and missed 28%. XnetCAE was more successful than CAE at balancing false positives and false negatives in medical settings, as shown by an F1-score of 0.75, although it might be better. The 0.16 reconstruction error decrease indicates that CAE preserved less information than its predecessors because the model learnt more correctly. A processing time of 36 seconds indicates the XnetCAE is capable of processing the test data for near real-time use in accuracy, processing time, and speed trade-offs.

Table 4: XnetCAE Technique Performance Evaluation Results

Metrics	Result
Accuracy	71.97%
Precision	0.79
Recall	0.72
F1-score	0.75
Loss	0.16
Computation Time (Sec)	36

The results of XnetCAE in classifying CT images are available in Figure 12. It contains several sampled CT images along with their true and predicted class labels and the models predicted probabilities for each class. Results of the model for these images indicate that the model is often overconfident in its predictions.

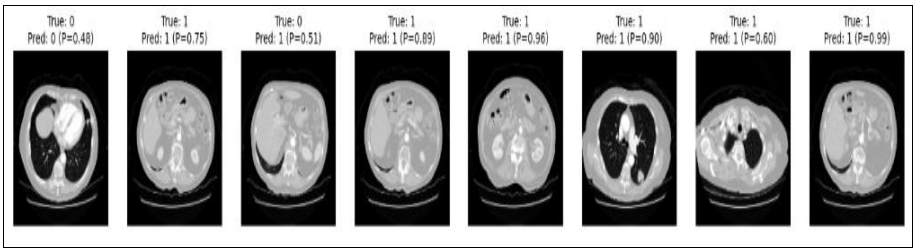

Fig 12: Visual evidence of classification predictions using the XnetCAE Technique extracted features.

Comparative Evaluation of CAE and XnetCAE Techniques

Table 5 shows the performance results the two techniques for detection of cancer of the lung from LDCT images. The baseline CAE technique had limited feature extraction; therefore it could only discriminate benign from malignant events with 65.0% accuracy while Xception's depthwise separable convolutions improved spatial feature representation and learning efficiency in the

Table 5: Experimental Results of CAE, and XnetCAE Techniques

Metrics	Accuracy (%)	Precision	Recall	F1-score	Loss (MSE)	Classification Time (Sec)
CAE	65	0.65	0.65	0.65	0.30	48
XnetCAE	71.97	0.79	0.72	0.75	0.16	36

XnetCAE technique to 71.97% demonstrating high diagnostic reliability and generalization. A steady increase is also observed in other performance metrics. The false positive (FP) precision was reduced from 0.65 with CAE technique to 0.79 with XnetCAE. Recall improves from 0.65 with CAE technique to 0.72 with XnetCAE, emphasizing its capacity to reduce FN which is vital for early cancer diagnosis. The F1-score grew from 0.65 with CAE technique to 0.75 in XnetCAE technique, demonstrating good and consistent categorization accuracy. Loss values fell from 0.30 to 0.16 in CAE technique and XnetCAE technique respectively. This shows that XnetCAE has superior convergence and forecast accuracy. The classification time for the testing dataset fell from 48 seconds with CAE technique to 36 seconds with XnetCAE technique, increasing computing efficiency. This time reduction and accuracy improvement suggest that architecture modifications in XnetCAE simplified feature extraction and decision-making. Thus, CAE gives a valid baseline, and XnetCAE makes beneficial architectural adjustments with high accuracies, sensitivity, reconstruction error, and classification time. Such qualities suggest that XnetCAE is best integrated into clinical processes when quick and accurate lung malignancy identification is crucial.

Conclusion

In this research, an XnetCAE technique that identifies lung cancer in low-dose CT scans was built to improve diagnostic accuracy, reconstruction error, and computational time. XnetCAE, was compared and contrasted with the standard CAE technique using a set of metrics of precision, accuracy, loss, recall, F1 measure, and computational time. The metrics showed that XnetCAE improved upon the CAE technique and was more generalizable with a better and faster classification of cancer of the lung. Results of the study demonstrated that the XnetCAE technique addressed some of the limitations of CAE, particularly, the excessive computational time and poor generalization. Specifically, computational inefficiency and feature extraction without loss of important but subtle anatomical features. This improvement makes the XnetCAE technique suitable for real-life usage where accuracy and speed are required.

References

1. Habbab FM, Bédard ELR, Joy AA, Alam Z, Abraham AG, Roa WHY. Early Detection of Lung Cancer: A Review of Innovative Milestones and Techniques. Journal of Clinical Medicine,2025:14(21):7812. https://doi.org/10.3390/jcm14217812
2. Balata H, Quaife SL, Craig C, Ryan DJ, Bradley P, Crosbie PAJ, *et al*. Early Diagnosis and Lung Cancer Screening. Clinical Oncology,2022:34(11):708–715. doi:10.1016/j.clon.2022.08.036
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians,2021:71(3):209–249. https://doi.org/10.3322/caac.21660

4. Hu S, Tao J, Peng M, Ye Z, Chen Z, Chen H, *et al*. Accurate detection of early-stage lung cancer using a panel of circulating cell-free DNA methylation biomarkers. Biomarker Research,2023:11:45. https://doi.org/10.1186/s40364-023-00486-5
5. Abbasian MH, Ardekani AM, Sobhani N, Roudi R. The Role of Genomics and Proteomics in Lung Cancer Early Detection and Treatment. Cancers,2022:14(20):5144. https://doi.org/10.3390/cancers14205144
6. Grigore B, Peters JL, Hamad W, Calanzani N, Asare L, Walter FM, *et al*. Evaluation of imaging techniques for early detection of intrathoracic cancers in symptomatic patients in primary care: a systematic review. BMJ Open,2025:15:091435. doi:10.1136/bmjopen-2024-091435
7. Tárnoki ÁD, Tárnoki DL, Dąbrowska M, Knetki-Wróblewska M, Frille A, Stubbs H, *et al*. New developments in the imaging of lung cancer. Breathe,2024:20(1):230176. https://doi.org/10.1183/20734735.0176-2023
8. Prosper AE, Kammer MN, Maldonado F, Aberle DR, Hsu W. Expanding Role of Advanced Image Analysis in CT-detected Indeterminate Pulmonary Nodules and Early Lung Cancer Characterization. Radiology,2023:309(1):222904. https://pubs.rsna.org/doi/abs/10.1148/radiol.222904
9. Chen L, Kao H, Wu P, Sheu M, Tu H, Huang L. *et al*. Hybrid iterative reconstruction in ultra-low-dose CT for accurate pulmonary nodule assessment: A Phantom study. Medicine, 2025, 104. https://doi.org/10.1097/md.0000000000041612
10. Tugwell-Allsup J, Owen B, England A. Low-dose chest CT and the impact on nodule visibility. Radiography, 2020. https://doi.org/10.1016/j.radi.2020.05.004
11. Pintelas E, Livieris IE, Pintelas PE. A Convolutional Autoencoder Topology for Classification in High-Dimensional Noisy Image Datasets. Sensors,2021:21(22):7731. https://doi.org/10.3390/s21227731
12. Shobayo O, Saatchi R. Developments in Deep Learning Artificial Neural Network Techniques for Medical Image Analysis and Interpretation. Diagnostics,2025:15(9):1072. https://doi.org/10.3390/diagnostics15091072
13. Berahmand K, Daneshfar F, Salehi ES, Li Y, Xu Y. Autoencoders and their applications in machine learning: a survey. Artificial Intelligence Review,2024:57(2):28. https://doi.org/10.1007/s10462-023-10662-6
14. Iqbal S, Qureshi AN, Li J, Mahmood T. On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks. Archives of Computational Methods in Engineering,2023:30(5):3173–3233. https://doi.org/10.1007/s11831-023-09899-9
15. Lu J, Verma N, Jha N. Convolutional Autoencoder-Based Transfer Learning for Multi-Task Image Inferences. IEEE Transactions on Emerging Topics in Computing,2022:10:1045–1057. https://doi.org/10.1109/tetc.2021.3068063
16. Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 1800–1807. https://doi.org/10.1109/CVPR.2017.195
17. Jayeprokash D, Gonski J. Convolutional Autoencoders for Data Compression and Anomaly Detection in Small Satellite Technologies. Information,2025:16(8):690. https://doi.org/10.3390/info16080690
18. Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. Proceedings of the International Conference on Machine Learning, 2008, 1096–1103. https://doi.org/10.1145/1390156.1390294
19. Chadha AS, Khoja A, Shukla S. Deep learning for medical image analysis: A comprehensive review. Journal of Medical Systems,2021:45(3):1–20. https://doi.org/10.1007/s10916-021-01720-1
20. Xuan W, Chen P, Li R, Wang F, Fu K, Wen Z. A targeted one dimensional fully convolutional autoencoder network for intelligent compression of magnetic flux leakage data. Scientific Reports, 2025, 15. https://doi.org/10.1038/s41598-025-96282-2
21. Zhao L, Zhang Z. A improved pooling method for convolutional neural networks. Scientific Reports, 2024, 14. https://doi.org/10.1038/s41598-024-51258-6
22. Zafar A, Aamir M, Nawi N, Arshad A, Riaz S, Alruban A, *et al*. A Comparison of Pooling Methods for Convolutional Neural Networks. Applied Sciences, 2022. https://doi.org/10.3390/app12178643
23. Zhuang JX, Luo L, Chen H. Advancing volumetric medical image segmentation via global-local masked autoencoder. arXiv, 2023. https://arxiv.org/abs/2306.08913
24. Feng Y, Zhang Y, Zhou Z, Huang P, Liu L, Liu X, *et al*. Memristor-based storage system with convolutional autoencoder-based image compression network. Nature Communications,2024:15(1):1132. https://doi.org/10.1038/s41467-024-45312-0
25. Younesi A, Ansari M, Fazli M, Ejlali A, Shafique M, Henkel J. *et al*. A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. IEEE Access,2024:12:41180–41218. https://doi.org/10.1109/access.2024.3376441
26. Zhao H, Qian L, Zhu Y, Tian D. Low Dose CT Image Denoising: A Comparative Study of Deep Learning Models and Training Strategies. AI Medicine,2024:1(1):7. https://doi.org/10.53941/aim.2024.100007
27. Wang C, Shao J, Song L, Ren P, Liu D, Li W. *et al*. Persistent increase and improved survival of stage I lung cancer based on a large-scale real-world sample of 26,226 cases. Chinese Medical Journal,2023:136(16):1937–1948. DOI:10.1097/CM9.0000000000002729
28. Da Cruz SD. Autoencoder and partially impossible reconstruction losses. Sensors,2022:22(13):4862. https://doi.org/10.3390/s22134862
29. Alberti M, Seuret M, Ingold R, Liwicki M. A pitfall of unsupervised pre-training. arXiv, 2017. https://arxiv.org/abs/1712.01655
30. Tang H, Kim DR, Xie X. Automated pulmonary nodule detection using 3D deep convolutional neural networks. IEEE International Symposium on Biomedical Imaging, 2018, 523–526. doi:10.1109/ISBI.2018.8363630

31. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Medical Image Computing and Computer-Assisted Intervention, 2016. https://doi.org/10.1007/978-3-319-46723-8_49
32. Pathak P, MS, MD, RP, Sheeba A, RA. Deep Learning Models for Image Classification Advances in Convolutional Neural Network Architectures. ITM Web of Conferences, 2025. https://doi.org/10.1051/itmconf/20257601004
33. Shan H, Zhang Y, Yang Q, Kruger U, Kalra M, Sun L. *et al*. 3-D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning From a 2-D Trained Network. IEEE Transactions on Medical Imaging,2018:37:1522–1534. https://doi.org/10.1109/tmi.2018.2832217