



Optimizing CNN based model for thyroid nodule classification using data augmentation, segmentation and boundary detection techniques

Rajshree Srivastava¹ · Pardeep Kumar¹

Received: 4 July 2022 / Revised: 22 September 2022 / Accepted: 2 March 2023 /

Published online: 17 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Thyroid nodule is an asymptomatic disorder which mostly occurs due to high production of thyroid hormones from the thyroid gland. The diagnosis is usually made by the radiologist and endocrinologists which heavily relies on their experience and expertise. Ultrasonography is one of the principal means for the initial assessment of nodules which is mainly performed when there is suspect of formation of nodules. In this research work, an optimized convolutional neural network model is proposed for the identification of thyroid nodules using various deep learning techniques like dense neural network, Alexnet, Resnet-50 and Visual geometry group-16. A total of 295 public and 654 collected thyroid ultrasonography datasets are considered in this work. The proposed model is evaluated on 1475 public and 3270 collected thyroid ultrasonography datasets with data augmentation technique. We experimentally determined the best optimized value for learning rate and drop out factor to enhance the performance of the models. The proposed model has achieved an accuracy of 93.75%, sensitivity of 94.62%, specificity of 92.53% and f-measure of 94.09% on the public dataset in experiment-I and an accuracy of 96.89%, sensitivity of 97.80%, specificity of 94.73% and f-measure of 97.26% on the collected dataset in experiment-II. The proposed model has shown an improvement of (4.57%, 7.84%), (5.06%, 8.24%), (4.43%, 6.63%) and (4.66%, 7.83%) in terms of accuracy, sensitivity, specificity and f-measure on (dataset –1, dataset-2) against other state of the art models.

Keywords Thyroid nodules · Deep learning · Classification · Detection · Ultrasonography · Boundary detection · Data augmentation · Segmentation · Convolutional neural network

✉ Pardeep Kumar
pardeepkumarkhokhar@gmail.com

Rajshree Srivastava
rajshree.srivastava27@gmail.com

¹ Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat, Dist. Solan, HP, India

1 Introduction

The traditional diagnostic techniques based on the expert knowledge of the doctors has some limitation as they depend on the doctor's expertise and experience. Thus, the performance of the diagnosis is limited and varies with the doctor's expertise. To overcome this problem, some hospitals advise a double screening scheme to confirm the disease, which results in extra time-consuming and high costs [33]. The development of imaging modalities in medical research like magnetic resonance imaging (MRI), ultrasonography (USG), tomography and X-ray led to an increase in demand for computer-aided diagnosis (CAD) systems [25]. X-Ray are widely used for diagnosis of chest problem and many other diseases [4]. These CAD systems work by capturing images of several parts of the human body and identifying the abnormal signs of the captured images in the presence of experienced doctors. This technique is helpful in many medical applications for identifying and detecting problems in the brain, kidney stones, lung, thyroid nodules, etc. [15]. In the case of thyroid nodules, many studies have designed systems to classify/detect benign and malignant nodules. These are an abnormal growth of cells that appear in the thyroid region. Some of the common symptoms found in patients are difficulty in swallowing food, pain in the neck area and lump in the neck [32]. When there is an excess of thyroxine hormone in the body, sudden weight gain/loss and rapid increase in a heartbeat are seen in the patient [3]. There is a set of scores based on thyroid nodules characteristics to differentiate benign and malignant thyroid nodules. These can be found in thyroid imaging reporting and data systems (TI-RADS) [27]. Benign nodules have thin/well-defined halo, regular margins, no lymph nodes, iso/hyperechoic pattern, hypovascular and coarse calcifications.

At the same time, malignant nodules have a thick or absent halo, irregular margins, taller or broader lesions, hypervascularity and micro-calcifications. In most cases, thyroid nodules are found to be benign when they undergo ultrasonography. Determining the benign/malignant nodules based on the symptoms is difficult for the doctors. Thus, it can be said that detection and classification of thyroid nodules plays a significant role in classifying thyroid nodules [39]. Figure 1 shows sample images of thyroid USG images.

Recent advancements in neural network (NN) architecture designs and training have given direction to experts and researchers to solve complex real-life problems using deep learning techniques. Deep learning (DL) is one of the branches of machine learning (ML), which when combined with convolution, is considered as an effective and efficient way to detect and classify disease. In general, models learn to perform classification tasks from various sounds, texts and pictures, which are fed as an input to the model. DL techniques use several neural network layers to extract specific characteristics from the given input data. These models sometimes outperform better than humans. The models are trained with a large amount of data. Its efficiency is increased when it is trained with a large amount of data [13]. Convolutional neural network (CNN) has some of the advantages due to which it is widely used, some of them are high computing power, excellent performance and low processing time. CNN comprises five layers, namely input layer, convolutional layer, max-pooling layer, fully connected layer and output layer [17]. The convolutional layer increases the invariance and reduce the map function that helps in cost reduction. The max-pooling layer computes the average and maximum function over the second layer. In a fully connected layer, all layers are clubbed and converted into a 1-D function layer. In the output layer, there is a classification result [43]. This research work focusses on detecting and classifying thyroid nodules using deep learning techniques.

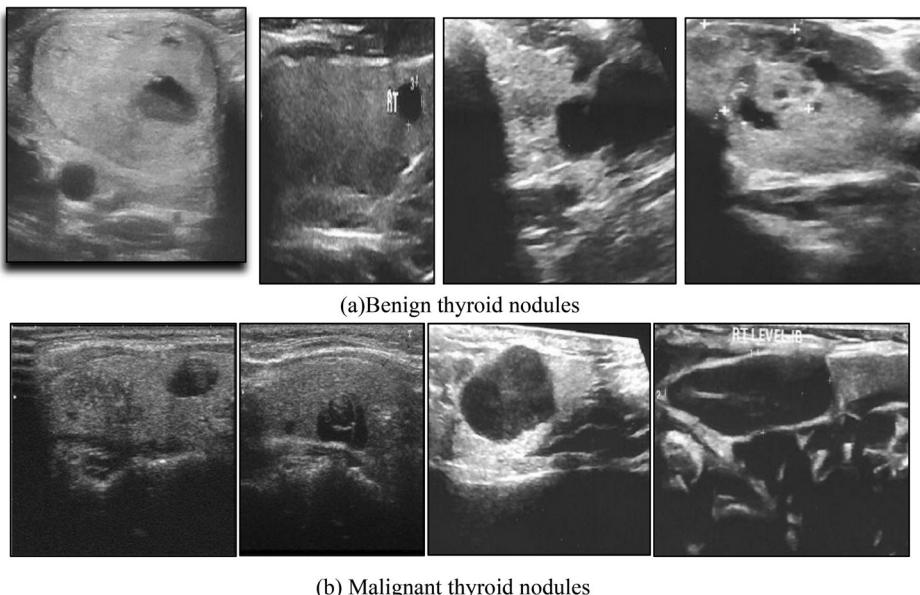


Fig. 1 a-b Sample images of thyroid USG images

1.1 Motivation and research contribution

In clinical examinations of any disease, the final interpretation is made by the senior radiologist. Therefore, they are not exempted from any diagnostic errors. For example, majority of thyroid nodules are heterogeneous and have different internal echogenicity (an ability to bounce an echo). As a result, specialists get confused about the variability of the echo patterns seen in a thyroid ultrasound machine. Some specialists advise a double screening scheme to confirm the disease, resulting in extra time-consuming and high costs. Therefore, in this research work an optimized CNN-based model is proposed for thyroid nodule classification using data augmentation, segmentation and boundary detection techniques to overcome this problem. This model will act as 2nd opinion for the clinicians. The paper is organized as follows: Section 2 provides the related work on detection and classification of thyroid nodules, Section 3 explains the proposed methodology, Section 4 throws lights on the experimental setup, Section 5 discusses the results and Section 6 focuses on the conclusion and future work.

1.2 Research contribution

- A CNN based model is proposed in this research work for the classification of thyroid nodules.
- The pre-processing steps like image resizing, grayscale conversion and noise removal are employed to improve the classification performance of the model. This helps to retain valuable information and have an improved high-level learning application.
- Further, data augmentation techniques are used to increase the size of the public Thyroid Ultrasound Image Database (TDID) and collected datasets from Kriti Scanning Center, Prayagraj, India.

- The major contribution of this work is optimizing the parameters like learning rate and drop-out factor parameters for the training of proposed model.
- The ultimate goal of diagnostic assessment of thyroid nodules is to accurately identify malignancy while avoiding overtreatment. Low-risk thyroid nodules can be safely monitored in many patients with minimal diagnostic intervention using this proposed model.

1.3 Background information

1.3.1 DNN

It is an artificial neural network (ANN) with multiple hidden layers between inputs and outputs layers having nodes in each layer [22]. The hidden layer performs mathematical computation as the data is fed into the input layer. The main advantage of using dense neural network (DNN) is integrating the feature and classification tasks to build good decision models. The various parameters, such as epoch size, learning rate, bias, initial weights, input and output layers, are initialized [35]. DNN is one of the successful classification models for the diagnosis of disease. Figure 2 shows architecture of DNN.

1.3.2 Alexnet

It is one of the CNN architectures designed by Krizhevsky, Sutskever [16] in 2012. It competed in the ImageNet Large Scale Visual Recognition Challenge on Sept. 30, 2012. It contained eight layers, having five convolutional (Conv) layers, followed by max-pooling layers and three fully connected (FC) layers. It uses the non-saturating ReLU activation function to improve the training performance of the models over tanh and sigmoid [12]. Figure 3 shows the architecture of Alexnet.

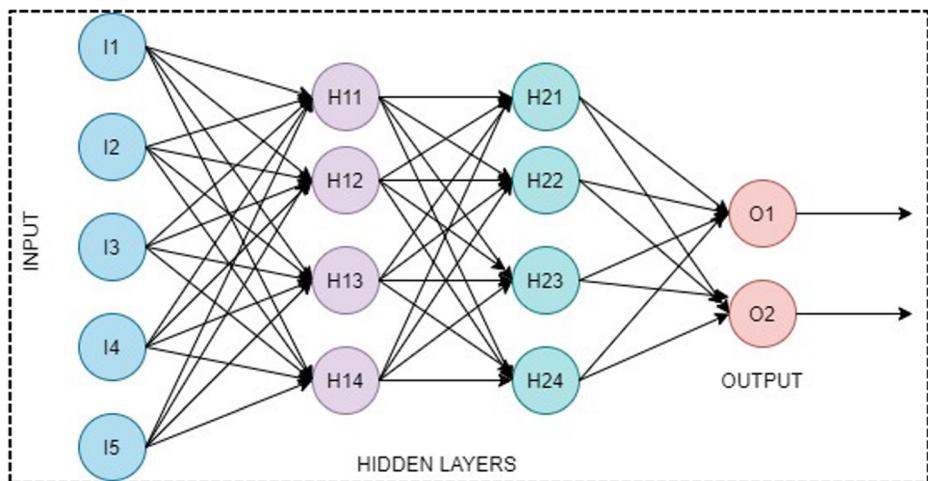


Fig. 2 Architecture of DNN

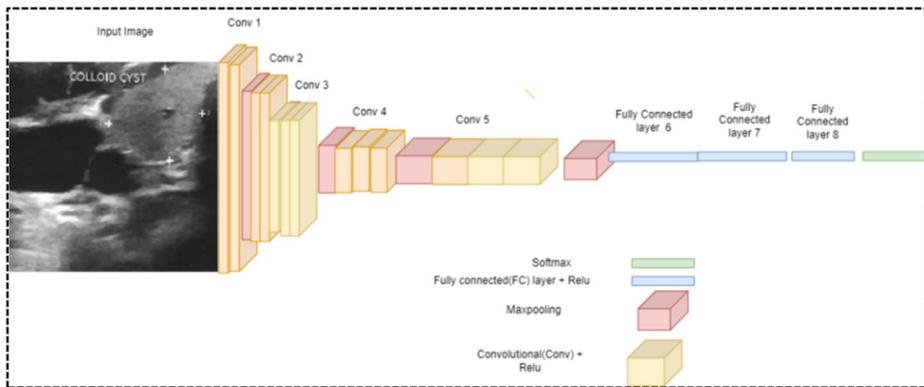


Fig. 3 Architecture of Alexnet

1.3.3 Resnet-50

He Zhang [7] from Microsoft research institute in Dec 2015 proposed Resnet. It has achieved the best result in ILSVRC 2015 competition by training 152- layers of neural networks. The Resnet-50 model consist of 5 stages having convolutional and identity block. Further, each convolutional block has three convolutional (Conv) layers followed by identity block with three convolutional layers [31]. Figure 4 shows an architecture of Resnet-50.

It successfully addresses the degradation problem i.e., sudden rise and drop in the model's accuracy. This sudden rise and fall in the model's accuracy leads to overfitting. Resnet addresses this problem by adding several congruent mapping layers ($y = x$) once the model reaches the saturated accuracy, which helps by not increasing the model's error [37]. Figure 5 shows the residual convolutional block.

By using this network, the block performs an element-wise superposition of the input and output layers. It can be computed using Eq. 1:

$$H(x) = F(x) + x \quad (1)$$

where x : input to the neural network, $H(x)$: expected output.

This simple addition will not increase the additional parameters and computation time of the network. If the model's layers increase, this structure will adjust efficiently.

1.3.4 VGG-16

Visual Geometry Group (VGG-16) model was introduced in 2014 [37]. It was ranked 1st and 2nd respectively in the ImageNet competition held in 2014, where they have used this model for image localization and classification. This model has 16 layers and uses very few hyperparameters

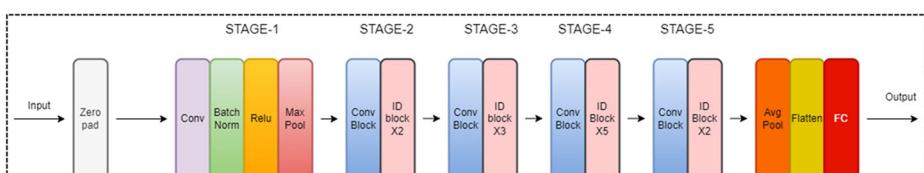
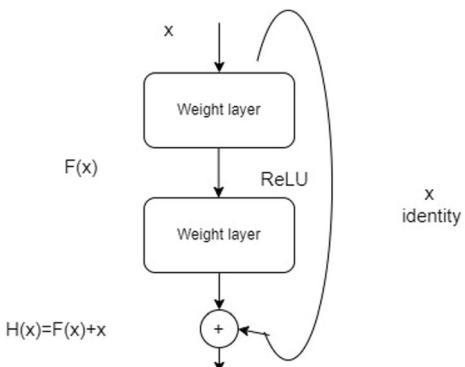


Fig. 4 An architecture of Resnet-50[45]

Fig. 5 The residual convolutional block



throughout the training process. Every CNN model has the same architecture, convolution stack, pooling layer, fully connected layer and softmax-layers [10]. The architecture uses 3×3 kernel shape throughout the single and double stride of the convolutional and pooling layers [29]. A complete model comprises 12 convolutional (Conv) 2-D layers, four fully connected (FC) layers with the softmax function at the end. Figure 6 shows the architecture of VGG-16.

2 Related work

In the recent years, many researchers have worked on machine learning (ML) and DL techniques for the diagnosis of the disease [18]. This section discusses the various studies done on thyroid nodules detection and classification. Song et al. [38] developed a multi-task cascade convolution neural network (MC-CNN) for automatic thyroid nodule detection. In their work, multi-scale single-shot detection is used for detecting nodules and MC-CNN algorithm for the classification. Different classifiers like MC-CNN, Naive Bayes, multi-layer perceptron (MLP), Alexnet, Googlenet and gradient boosting decision tree (GBDT) were used to evaluate the models. On the TDID dataset, the model has achieved an accuracy of 92.1%, sensitivity of 94.1% and specificity of 96.2%. Wang et al. [41] proposed a novel semi-supervised learning method for automatic thyroid nodule classification. In their work, there

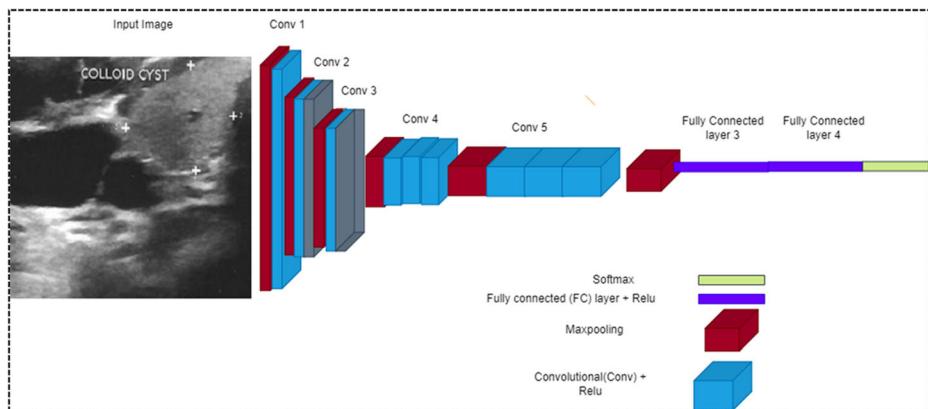


Fig. 6 Architecture of VGG-16

is a combination of expectation-maximization (EM) i.e., generating a bag of patient's images and instances as nodules of images and CNN to train the model and classify the thyroid nodules. Data augmentation techniques is used to increase the training size of the model. The model has achieved an accuracy of 80.91%, sensitivity of 80% and specificity of 80.91% with a pre-trained VGG-16 model on the TDID dataset. Ko et al. [14] proposed a deep convolutional neural network (DCNN) to diagnose thyroid malignancy on ultrasonography (USG) images. Using three trained CNN, model has achieved an accuracy of 86%, sensitivity of 84% and specificity of 90%. Nguyen et al. [23] analyzed the problem of frequency and spatial domains. A cascade classifier scheme is used to improve the model accuracy. Frequency and spatial domain features were extracted for detecting thyroid nodules and CNN is used to train the model. One of the limitations of their work is that their model is quite complicated rather than using a single method and did not consider the problem of the imbalanced dataset. The model has achieved an accuracy of 90.88% using TDID open database. Shi et al. [36] proposed knowledge-guided synthetic image adversarial augmentation (ACGAN) to classify thyroid nodules. The model has achieved an accuracy of 90.63%, sensitivity of 90.63% and specificity of 92.65% using 1937 USG thyroid images. Ajilisa et al. [1] used a different pre-trained CNN network (Alexnet, Xception, VGG-16, Inception v-3, Googlenet, VGG-19, Resnet-10 and Resnet-50) to classify thyroid USG images. K-means clustering technique is used to deal with imbalanced datasets. From their experiment, it is found that a deep neural network (DNN) outperforms from the rest of the networks with an accuracy of 89.93% and a sensitivity of 92.76% using TDID dataset. Some of the limitations of their model were the small dataset size, skewed dataset distribution and not an automatic detection model to classify thyroid nodules from USG images. Guo et al. [5] proposed an improved deep learning approach to diagnose thyroid nodules. The squeeze and excitation (SE) modules were used for adaptive feature selection and maximum inter-pixel relations (MPR) for inter-pixel relations problem. Authors have also used data augmentation techniques along with the classification. The model has achieved an accuracy of 90.17%, sensitivity of 86.99% and specificity of 92.35% using 407 thyroid USG images. Xie et al. [42] proposed a DNN based model for the identification and classification of thyroid nodules. They used the combination of local binary pattern (LBP) along with DL. The model has achieved an accuracy of 85%. Zhu et al. [48] proposed a deep convolutional neural network (DCNN) to classify thyroid USG images and achieved an accuracy of 86.5%, sensitivity of 86.7% and specificity of 87.7% using 719 thyroid USG images. Nguyen et al. [24] analyzed the problem of frequency and spatial domains. In their work, there is a combination of Resnet-50 and Inception-based neural networks to classify thyroid nodules. The model has achieved an accuracy of 92.05% on TDID open database using the weighted binary cross-entropy loss function for the handling of the imbalanced dataset. One of the limitations of their model is the small dataset size and not addressing the problem of noise removal. Yang et al. [44] proposed a multi-task cascade deep learning model (MCDLM) using integration of domain knowledge (DK) and multimodal ultrasound (USG) images for the diagnosis of thyroid nodules. U-net technique was used for the segmentation of thyroid nodules along with a pre-trained VGG-13 model. The model has achieved an accuracy of 90.01%, sensitivity of 87.47% and specificity of 92.15% using 3090 thyroid USG images. Lu et al. [20] proposed a generative adversarial network based deformable attention network for the identification of thyroid nodules. In this work, authors rather considering the segmented nodules, they have considered the whole ultrasound images. This helps them in incorporating rich semantic information. The TN-SCUI2020 challenge dataset is considered in this work. The model has achieved an accuracy od 82.81% and f-measure of

84.84%. Zhao et al. [46] proposed a local and global feature disentangled network (LoGo-Net) model for the classification of benign and malignant nodules. They design the self-attention mechanism that separates the tissue components and the anatomy components from images which results better diagnosis. The model has achieved an accuracy of 89.33%, 85.53% specificity and 91.36% sensitivity. Zhao et al. [47] proposed an automatic thyroid identification and classification model using feature fusion network. The model works in three phases (i) data collection and pre-processing, (ii) feature extraction using Gray level co-occurrence matrix (GLCM), Gray-level run-length matrix (GLRLM), Gray-level size zone matrix (GLSZM) and feature reduction using principal component analysis, (iii) classification using Resnet-18. The model has achieved 88.30% accuracy, specificity of 66.67%, sensitivity of 95.10% on 1874 images. Table 1 shows the various state-of-the-art of models for detecting and classifying thyroid nodules along with their findings and limitations.

2.1 Research gaps

This subsection throws light on the various research gaps identified in the existing studies. For evaluating and comparing the proposed model, we have considered those research papers that have used the public TDID dataset in their works. The various research gaps noted from the related work are: (i) Feature extraction techniques are not sufficient for the identification of thyroid nodules. Work can be extended to detection techniques. (ii) Medical ultrasound images are suspect to have noise. So, it is essential to address this problem. (iii) Deep learning techniques sometimes leads to the problem of overfitting because of less sample size. (iv) Optimizing the parameters of deep learning model plays a significant role to achieve the best performance of the model. To address all these gaps, the following solutions are incorporated in our work. The morphological operation, segmentation and boundary detection techniques were used to detect thyroid nodules from USG images. Along with the detection, various deep learning techniques like DNN, Alexnet, Resnet-50 and VGG-16 is considered for the classification of thyroid nodules. Since pre-processing plays an important role for the development of any model. The gaussian blur function is used to address the problem of noise present in medical USG images. To increase sample size various data augmentation techniques like RandRotation, RandomXReflection, RandomYReflection, RandomXTranslation, RandomYTranslation were applied on public TDID and collected datasets. Learning rate and drop-out factor is optimized to improve the model performance. The comparison of the proposed model is made with the reported literature. Various popular deep learning techniques like Resnet-50, DNN, Alexnet, VGG-16 is also explored for the evaluation of the model.

3 Proposed methodology

Optimizing CNN-based model for thyroid nodule classification is proposed using data augmentation, segmentation and boundary detection techniques. It works in the following four phases namely, (i) the data collection: public and local datasets, i.e., collected datasets, (ii) To enhance and maintain uniformity of the images various pre-processing steps were used like resizing, RGB to greyscale conversion, noise removal and augmentation techniques is imposed to increase the sample size, (iii) morphological operation (erosion and dilation), segmentation and boundary detection techniques have been used to detect the thyroid nodules from USG images and (iv) the parameters like drop out and learning rate is optimized for CNN for the

Table 1 The state-of-the-art of models for detecting and classifying thyroid nodules along with their findings and limitations

Ref. Ids	Proposed Models	Feature extraction techniques	Detection techniques	Classification	Datasets	Findings	Limitations
Song et al. [38] 2018	Proposed MC-CNN model	–	Muti-scale single detection	VGG-16, Naïve bayes, MLP, Alexnet, Googlenet, GBDT	TDID	VGG-16 performs better in comparison with the other classifiers. The introduction of the pretrained classifier can improve the performance of the models.	Failed to address the problem of noise removal and imbalanced dataset. Less sample size.
Wang et al. [41] 2018	Semi-supervised learning-based model	–	Proposal extraction method	CNN	TDID	Addressed the problem of imbalanced dataset.	Failed to address the problem of noise removal. Less sample size.
Ko et al. [14] 2019	DCNN model	–	–	Imagenet-vgg, imagenet--verydeep	Our Institution hospital	Model is competitive for identification of thyroid nodules.	Randomly selection of the training and testing dataset. The comparison between the other models is not present.
Nguyen et al. [23] 2019	An intelligent based model	–	–	Resnet-50, Resnet-34, Resnet-18	TDID	Addressed the problem of frequency and spatial domains. Cascaded classifier is used to improve the model performance.	Different scanners were used for capturing of images which caused different diagnostic performance of CNNs. A single image is used as a representative. Model is complicated. Less sample size. Failed to address the problem of noise and imbalanced dataset.

Table 1 (continued)

Ref. Ids	Proposed Models	Feature extraction techniques	Detection techniques	Classification	Datasets	Findings	Limitations
Shi et al. [36] 2020	ACGAN model	—	—	ACGAN	Local dataset	Authors incorporated domain knowledge and deep learning together to improve model performance.	Work can be extended to semi-supervised learning and large dataset. Failed to address the problem of noise.
Ajilisa et al. [1] 2020	CAD System	—	—	Alexnet, Xception, VGG-16, Inception v-3, Googlenet, VGG-19, Resnet-10, Resnet-50, DNN CNN	TDID	K-means clustering method is used to handle imbalanced dataset problem. DNN performs best among other classifiers.	The sample size of dataset is small. Failed to address the problem of noise. Less sample size.
Guo et al. [5] 2020	An improved DL model	Squeeze and excitation module	—	Cooperated hospitals	A combination of SE and MPR is used for nodule diagnosis.	Failed to address the problem of noise.	
Xie et al. [42] 2020	Hybrid model	LBP	—	Resnet	The Shanghai Pudong People's Hospital.	Less sample size. Different other fusion of feature extraction models can be proposed.	
Zhu et al. [48] 2021	DCNN model	—	—	VGG-19	The Shanghai Pudong People's Hospital.	Data augmentation can be used to improve the performance of the model.	
Nguyen et al. [24] 2021	CAD System	—	—	Resnet-50, Inception	TDID	Model is competitive for the identification of thyroid nodules.	Less sample size. Failed to address the problem of noise. Different features extraction or detection technique can be explored.
						The weighted binary cross-entropy loss function is used to address the problem of unbalanced dataset and for training DCNN.	Model is complicated. Less sample size. Failed to address the problem of noise. Different other DL techniques can be explored for the

Table 1 (continued)

Ref. Ids	Proposed Models	Feature extraction techniques	Detection techniques	Classification	Datasets	Findings	Limitations
Yang et al. [44] 2021	MCDLM model	Supervised support vector machine	U-net	VGG-13	—	The combination of Resnet-50 and Inception models is used for the classification. The model has performed well in less sample size. It is an automatic detection model.	identification of thyroid nodules. Less sample size. Failed to address the problem of noise and imbalanced dataset. Different other types of feature extraction techniques can be used to improve the model performance.
Lu et al. [20] 2022	Generative adversarial based computer aided diagnostic system	Resnet	—	Resnet, Imagenet	TN-SCUI 2020 challenge dataset	Model is competitive for the identification of thyroid nodules.	Less sample size. Failed to address the problem of noise. Different features extraction or detection technique can be explored.
Zhao et al. [46] 2022	LoGo-Net model	—	—	DL model	West China hospital	The model can be used as a 2nd opinion by doctors. The model identifies nodules excellent with a small significant morphological difference.	Model can be further improved by incorporating segmentation or other deep learning model fusion for the identification of thyroid nodules. The problem of data imbalance can be better solved by using data augmentation and many other techniques.
Zhao et al. [47] 2022	Feature-fusion network DL model	GLCM, GLSZM, GLRLM	—	Resnet-18	Department of ultrasound of Chinese PLA general hospital	The model has achieved good result on less sample size. The work incorporated different feature extraction techniques along with DL model.	Work can be extended to segmentation and other DL models. Work needs to be evaluated on a large dataset.
						It can be used by the practitioners for their study.	

classification. Figure 7 shows the framework of the proposed methodology. The detailed description of the proposed model is discussed below:

3.1 Phase I: Data collection

In this work, two datasets were used:

3.1.1 Public TDID dataset

The dataset is collected from IDIME (Instituto de Diagnostico Medico) ultrasound department, one of the largest diagnostic imaging center in Bogota, Colombia. The dataset has 295 thyroid USG images, out of which 107 were benign, and 188 were malignant thyroid USG images by two experts [27]. The images were captured with Toshiba Nemio 30, Toshiba Nemio MX Ultrasound devices with 12 MHz convex and linear transducers. The images were classified by using Thyroid Imaging Reporting and Data system (TI-RADS) score.

3.1.2 Local dataset

The dataset is collected from Kriti Scanning Center, Prayagraj, Uttar Pradesh, India duly approved by National Accreditation Board for Hospitals and Healthcare Providers (NABH) [28]. The local/collected has total number of 654 images, out of which 428 were benign and 226 were malignant thyroid USG images. The duration of dataset collection was from July 2020 to March 2021. Thyroid ultrasound video sequences with Voluson E-10, Mindray-Resona 7, SIEMENS Healthineers, and Voluson E-8 Ultrasound devices with 3 to 11 MHz convex and linear transducers captured the thyroid images. In both cases, TI-RADS scores have been used to pre-classify the thyroid USG images [30]. Table 2 shows the TI-RADS score.

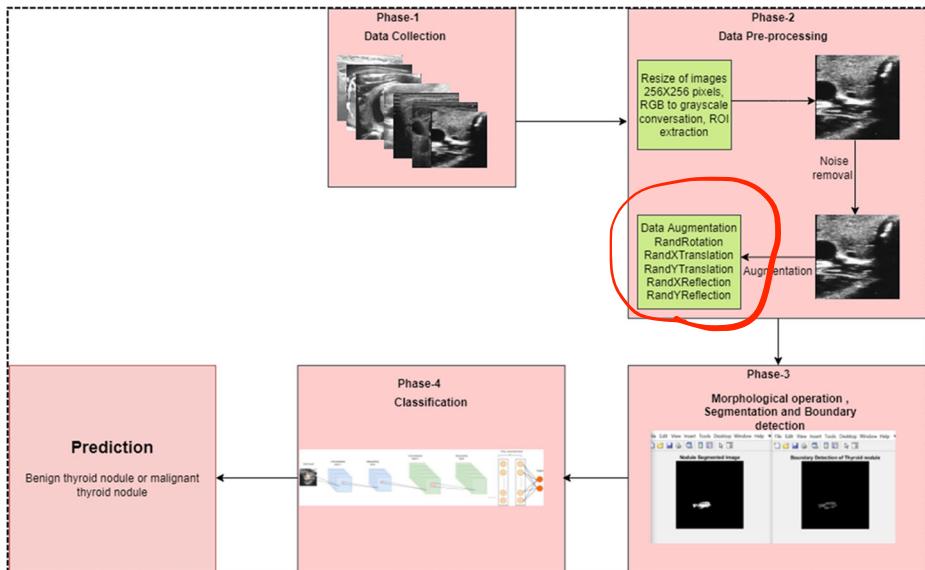


Fig. 7 Framework of proposed methodology

The criteria for the inclusion and exclusion of thyroid USG images are as follows:

Inclusion-

- (1) Images having clear boundary of the nodules
- (2) Images having one or multiple nodules
- (3) Patients having age greater than 16 years
- (4) Patients with the confirmed thyroid nodules problem

Exclusion –

- a. Excluded the images which were zoomed in or zoomed out
- b. Patients having age less than 16 years
- c. patients without confirmed thyroid nodules problem.

3.2 Phase II: Data pre-processing and data augmentation techniques

Data pre-processing is essential for the development of any model. Initially, the images were of size 560×360 pixels; it has been resized with 256×256 pixels using Eq. 2:

$$B = imresize(A, scale) \quad (2)$$

To reduce the complexity of the images, images are converted to gray scale from RGB scale using Eq. 3:

$$I = rgb2gray(RGB) \quad (3)$$

Two different noise removal techniques like **median blur** and **gaussian blur** were considered for this work. After the analyzing, it is found that **gaussian blur performs better**. It can be computed using Eq. 4:

$$h = fspecial(type) \quad (4)$$

where **fspecial**: function returns the kernel correlation, **h**: creates a 2D filter.

Later on, sharpening and enhancement techniques were applied on the images. To reduce complexity, images were converted to grayscale. Sometimes network gets overfitted, due to the availability of small datasets. Data augmentation techniques are applied to increase the sample size and it perform a geometric transformation to the image datasets to reduce the overfitting problem. It is applied using Eq. 5:

Table 2 TI-RADS score

TI-RADS Score	Type	Category
1	Benign	BENIGN
2	Not Suspicious	
3	Mildly Suspicious	
4	Moderately suspicious	MALIGNANT
5	Highly suspicious	

$$aug = imageDataAugmenter(Name, Value) \quad (5)$$

where Name: name of the techniques, Value: values assigned to techniques.

Some of them used in our work are RandRotation, RandomXReflection, RandomYReflection, RandomXTTranslation, RandomYTranslation. Algorithm 1 in appendix section shows the complete process of data pre-processing and data augmentation techniques.

3.3 Phase III. Morphological operations, segmentation and boundary detection techniques

These are a collection of non-linear operations related to the shape or morphology features of an image. It is applied on the binary image to find out the exact region of thyroid nodule within the image using some basic operations [9]. The most commonly used operations are dilation and erosion. Morphological operation is a broad set of image processing operations that process images based on shapes. It applies a structuring element to an input image, creating an output image of the same size. The morphological operations used in this work are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. However, in this research work we have tried to use fusion of traditional method (morphological operation and segmentation) for detection of nodules and deep learning techniques (like Alexnet, Resnet-50, VGG-16, DNN) for classification of thyroid nodules. The structure used in our proposed model is “disk.” Figure 8 shows the morphological operation.

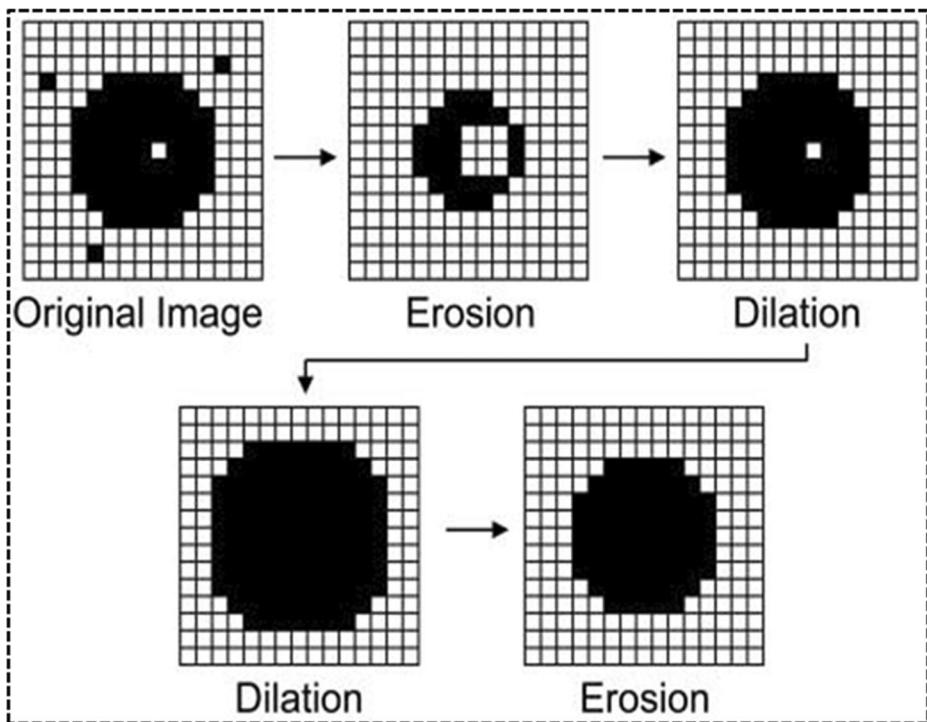


Fig. 8 The morphological operation [21]

The dilation function can be performed using Eq. 6:

$$I(\text{dilate})S = \{I + S; \text{for all pixels in } I \in S\} \quad (6)$$

where S: structuring element.

The erosion function can be performed using Eq. 7:

$$I(\text{eroded})S = \{I(\text{eroded})S\}(eroded)S \quad (7)$$

The morphological gradient function is performed for the enhancement of the edges. The subtraction function for this function can be performed using Eq. 8:

$$G = \left\{ I(\text{dilate})S - \{I(\text{eroded})S\} \right\} \quad (8)$$

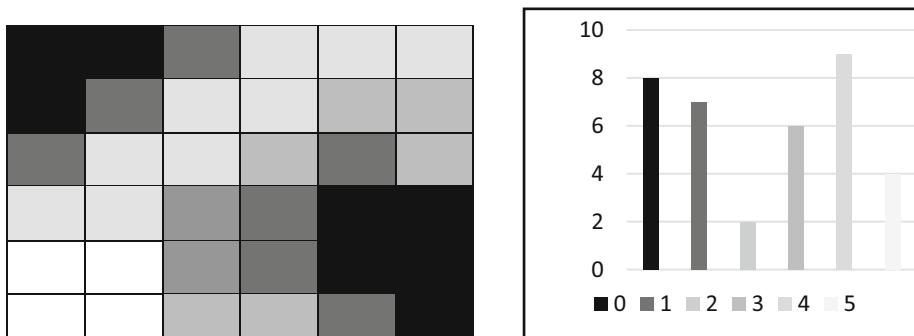
3.3.1 Otsu's thresholding for image segmentation

Thresholding is an important technique for image segmentation. It produces a uniform region based on the threshold criteria. Otsu's thresholding is an iterative process for all the possible threshold values thereby calculating and measuring the pixel levels for each side of the threshold [2]. It helps to have a clear picture of the tumor part in the thyroid USG images. Let suppose that there is threshold (T) in our problem, then the histogram is divided into C_0 and C_1 classes. In C_0 section and class, the intensity of the pixel will vary from 0 to T-1, while in C_1 section and class, it will vary from t to L-1. In the case of gray scale image, the value of L = 256. The threshold value which is obtained gives the minimal inter-class variance and maximal inter-class variance [8]. The “within-class variance” can be computed using Eq. 9:

$$\text{within class variance } \sigma_w^2 = w_b \sigma_b^2 + w_f \sigma_f^2 \quad (9)$$

where w_b = weight of the background, w_f = weight of foreground, σ_b = variance of background, σ_f = variance of foreground.

Illustrative example Let suppose there is an image of 6X6 matrix and its corresponding histogram as shown below in Fig. 9:



a Image of matrix 6x6

b Histogram of the image of matrix 6x6

Fig. 9 a-b Sample image matrix and its histogram

The calculation for background and foreground of an image will be as shown below:

$$\begin{aligned} \text{Weight } w_b &= \frac{8 + 7 + 2}{36} = 0.4722 \\ \text{Mean } \mu_b &= \frac{(0 * 8) + (1 * 7) + (2 * 2)}{17} = 0.6471 \\ \text{Variance } \sigma_b^2 &= \frac{\left((0-0.6471)^2 * 8\right) + \left((1-0.6471)^2 * 7\right) + \left((2-0.6471)^2 * 2\right)}{17} = 0.4637 \\ \text{Weight } w_f &= \frac{6 + 9 + 4}{36} = 0.5278 \\ \text{Mean } \mu_f &= \frac{(3 * 6) + (4 * 9) + (5 * 4)}{19} = 3.8947 \\ \text{Variance } \sigma_f^2 &= \frac{\left((3-3.8947)^2 * 6\right) + \left((4-3.8947)^2 * 9\right) + \left((5-3.8947)^2 * 4\right)}{19} = 0.5152 \end{aligned}$$

$$\text{within class variance } \sigma_w^2 = w_b \sigma_b^2 + w_f \sigma_f^2 = 0.4722 * 0.4632 + 0.5278 * 0.5152 = 0.4909$$

So, it can be concluded that the threshold value will be 3. The calculation will be same for all the threshold value (T) n to 5 as given in Fig. 9b. The one with the lowest within class variance will be finalized for the threshold value and is applied to the entire thyroid nodule images.

3.3.2 Boundary detection

Previously in medical imaging, identified tumors and their boundaries were drawn by experienced experts, contained some manual errors. Hence, to remove this error, the morphological operation ‘remove’ is used to remove all the interior pixels and leaving the boundary pixels [34]. Boundary detection is defined as joining a curve of all continuous points with the same colour or intensity. Edge and boundary detection techniques are successful techniques for shape analysis and object detection. Edge detection works by detecting boundaries of that region where there is a change in a single pixel [26]. Whereas in the case of boundary detection, it groups pixels of the same parameters to form a region. Figure 10 shows the sample images



Fig. 10 The sample images output where edge detection technique is applied [40]



Fig. 11 The sample images output where boundary detection technique is applied [40]

output where the edge detection technique is applied. Figure 11 shows the sample images output where the boundary detection technique is applied.

A contour line joins a point of equal height above a given level and is represented by different coloured boundaries. The contour-f function gives a better view of the system by each class with different colours. C-label adds height labels to a 2D contour plot, providing better insight into the image. Boundary detection of an image is computed using Eq. 10:

$$B(A) = A - (A \ominus B) \quad (10)$$

where A: Matrix of an image M x N, B: Structuring element, \ominus : Erosion function, B(A): Boundary detection.

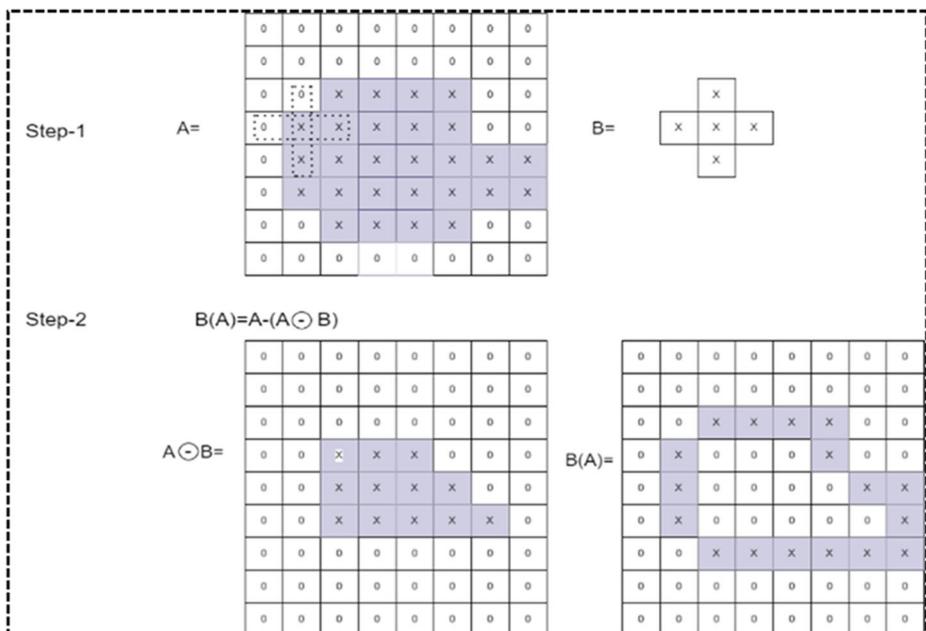


Fig. 12 Working of boundary detection technique

Figure 12 shows the working of the boundary detection technique. This figure considers a matrix A of size 8×8 and B as the structuring element. Here, the structuring element B is scanned with each element of matrix A; the one having the same pattern is marked with X. Thus, it can be said that the erosion function is performed. After that, the original matrix A is subtracted with $(A \ominus B)$ to have an image boundary $\mathbf{B}(\mathbf{A})$. Figure 13a-c shows the sample thyroid nodule images where segmentation and boundary detection techniques are applied. Algorithm 2 in appendix section shows the complete process of morphological operation, segmentation and boundary detection.

3.4 Phase IV: Classification

CNN is the most widely used deep learning (DL) model for visual imagery. To eliminate the overfitting problem, regularization method to eliminate overfitting. The classification takes place in the fully connected and softmax layers. In CNN, the most important layer is the convolution layer. The convolution function works by using Eq. 11:

$$x_j = I * W_j + b_j, \quad j = 1, 2 \dots f \quad (11)$$

where I: input, x_j : output, W_j : weights of the j th filter, b_j : j th bias.

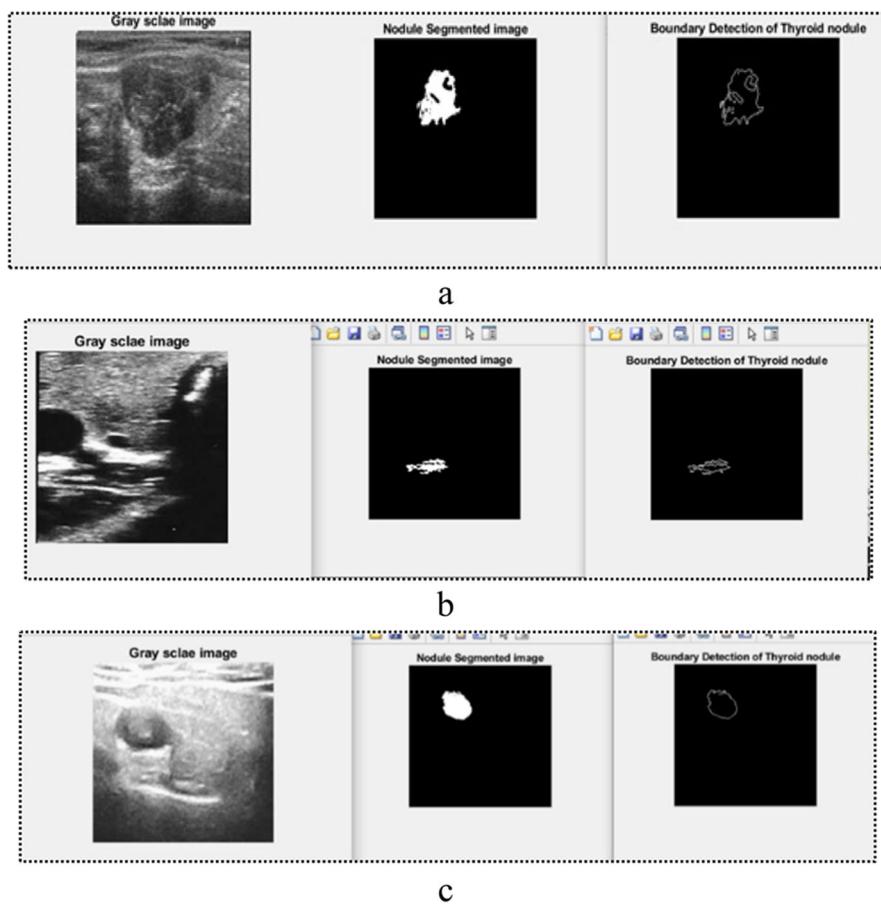


Fig. 13 a-c Sample thyroid nodules images with segmentation and boundary detection techniques

CNN architecture has multiple hidden layers, each of which has several 2D planes consisting of several neurons. Its input data is the 2D image and the feature extraction module is embedded in the CNN architecture. Algorithm 3 in appendix section shows the complete process for classification of thyroid nodules using CNN. Figure 14 shows the basic architecture of CNN.

CNN architecture is divided into five components, namely input layer, convolutional layer, max-pooling layer, fully connected layer and output layer. The detailed description of each component is explained below:

3.4.1 Input layer

It specifies a complete description of an image ($N_h \times N_w \times N_c$). The variables (h , w , c) correspond to the height, width and channel size of an image.

3.4.2 Convolutional layer

2D convolution operations are performed on the images using the kernel matrix, which gives a convoluted feature matrix. Different activation functions like ReLU, tanh, sigmoid functions, etc., are used. The obtained output becomes the input to the next layer. In our work, the ReLU activation function is used because it is faster than the other functions, replaces all negatives by zero and solves the gradient problem. It is used after each convolution layer. Figure 15 shows the operation performed by the convolutional layer. The ReLU function can be computed using Eq. 12:

$$f(x) = \begin{cases} x; & x \geq 0 \\ 0; & x < 0 \end{cases} \quad (12)$$

where: x is the input value to ReLU.

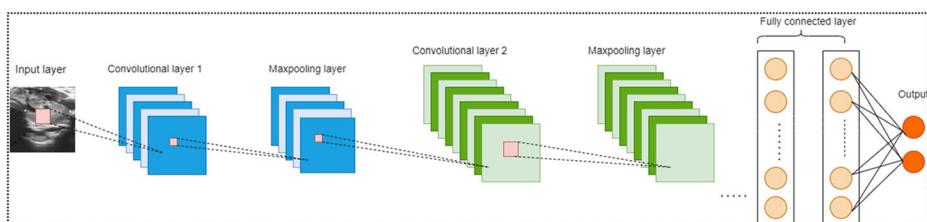


Fig. 14 Basic architecture of CNN

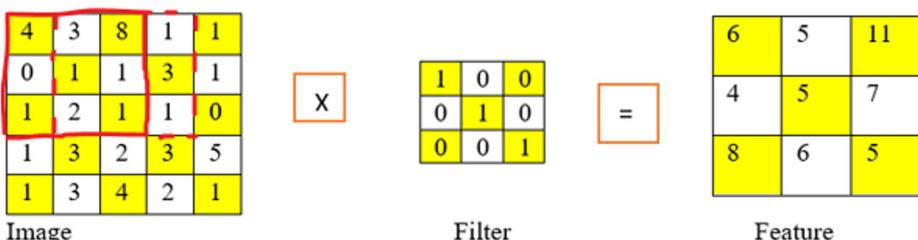


Fig. 15 The operation performed by convolutional layer



Fig. 16 The operation of max pooling layer

3.4.3 Max pooling layer

Convolutional layers (with activation function), sometimes followed by a down-sampling operation, reduces the spatial size of the feature map and remove the redundant spatial information. Down-sampling increases the number of filters in deeper convolutional layers without increasing the required computational time. Thus, it reduces the convoluted feature matrix, thereby improving the computation power [6]. Figure 16 shows the operation of the max-pooling layer using a 2×2 filter. Suppose y is pooling region, the activation set C included in Y ; then the max-pooling can be computed using Eq. 13:

$$f_M = \max(C_y) \quad (13)$$

3.4.4 Fully connected layer

This layer combines all the features learned by the previous layers to classify the images. Here flattening function is performed to convert a 2D matrix to the 1D array which helps in reducing

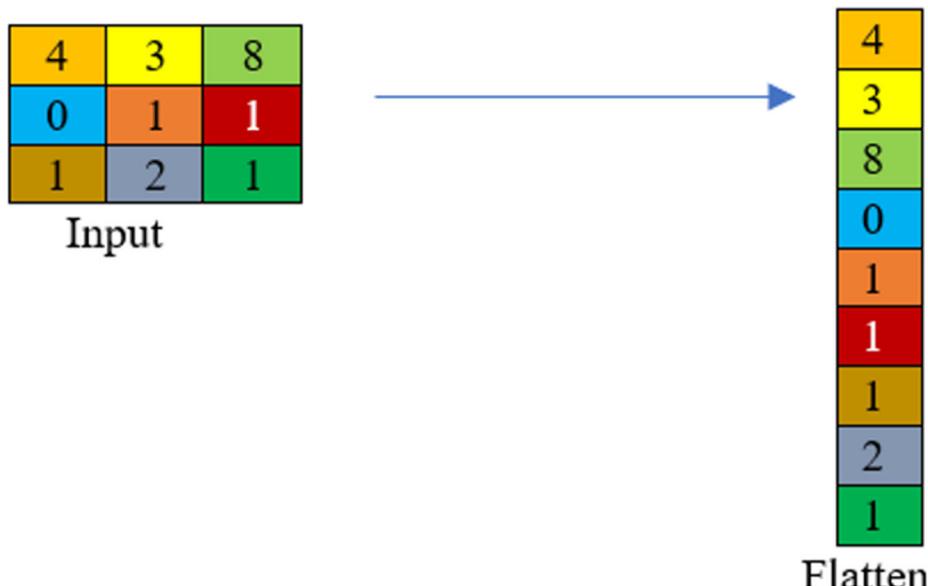


Fig. 17 The flattening operation performed in fully connected layer

dimension and computation complexity. Figure 17 shows the flattening operation performed in the fully connected layer. The working of the proposed model is shown in Fig. 18.

3.4.5 Output layer

Here, we have the classified output according to the given conditions. In our work, to normalize the output between zero and one softmax function is applied [11]. It can be computed using Eq. 14:

$$S(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (14)$$

where x: input; x_i : i-th element of the input vector; e^{x_i} : exponential; $\sum_{j=1}^n e^{x_j}$: normalized term.

4 Experimental setup

The complete experiment has been performed on MATLAB 2019b, 64bit operating system (OS), Intel (R) Core (TM) i5-8250U CPU, 1 TB solid-state drive (SSD), NVIDIA GeForce

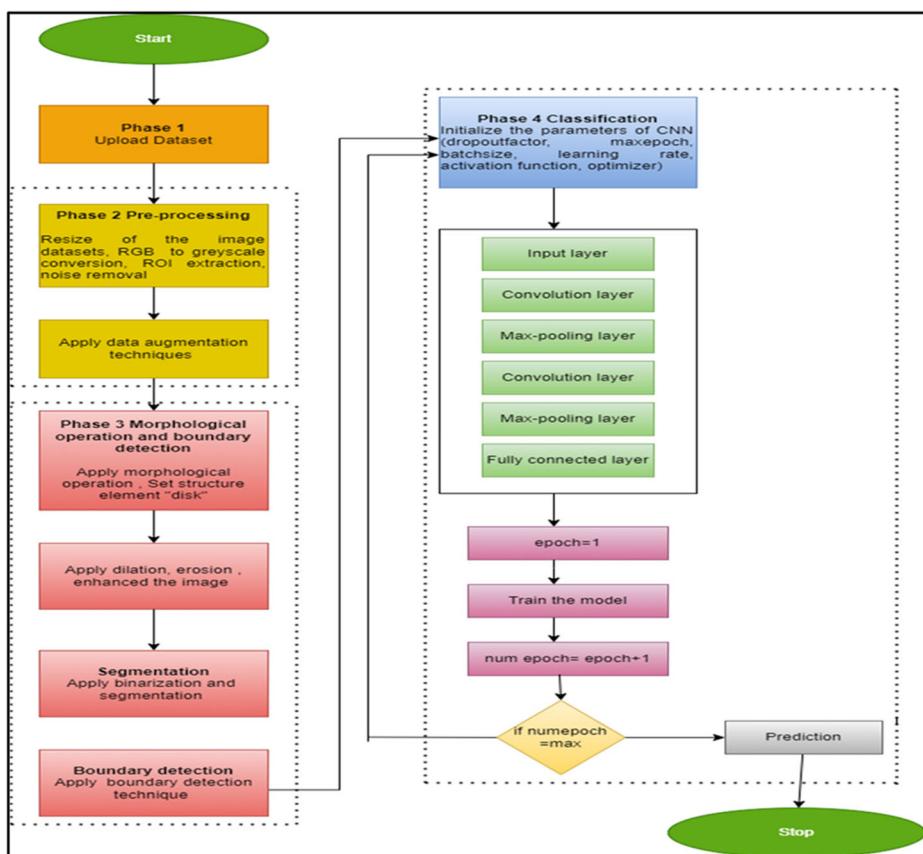


Fig. 18 Flowchart of the proposed model

and 16 GB RAM. Some parts of the experiment are also performed on python Go-laboratory with TensorFlow and keras as backend. Figure 19 shows the training and testing phase of CNN and Table 3 shows the parameter settings of CNN.

In order to represent the performance of the classification algorithm, a confusion matrix is defined having actual and predicated labels [19]. These are as follows:

- **True Positive (TP):** ROI is labelled as malignant thyroid nodule that is actually malignant nodule.
- **False Positive (FP):** ROI is labelled as malignant that is actually benign nodule.
- **True Negative (TN):** ROI is labelled as benign nodule that is actually benign nodule.
- **False Negative (FN):** ROI is labelled as benign nodule that is actually malignant nodule.

Based on all these four outcomes the following four parameters are computed in our work:

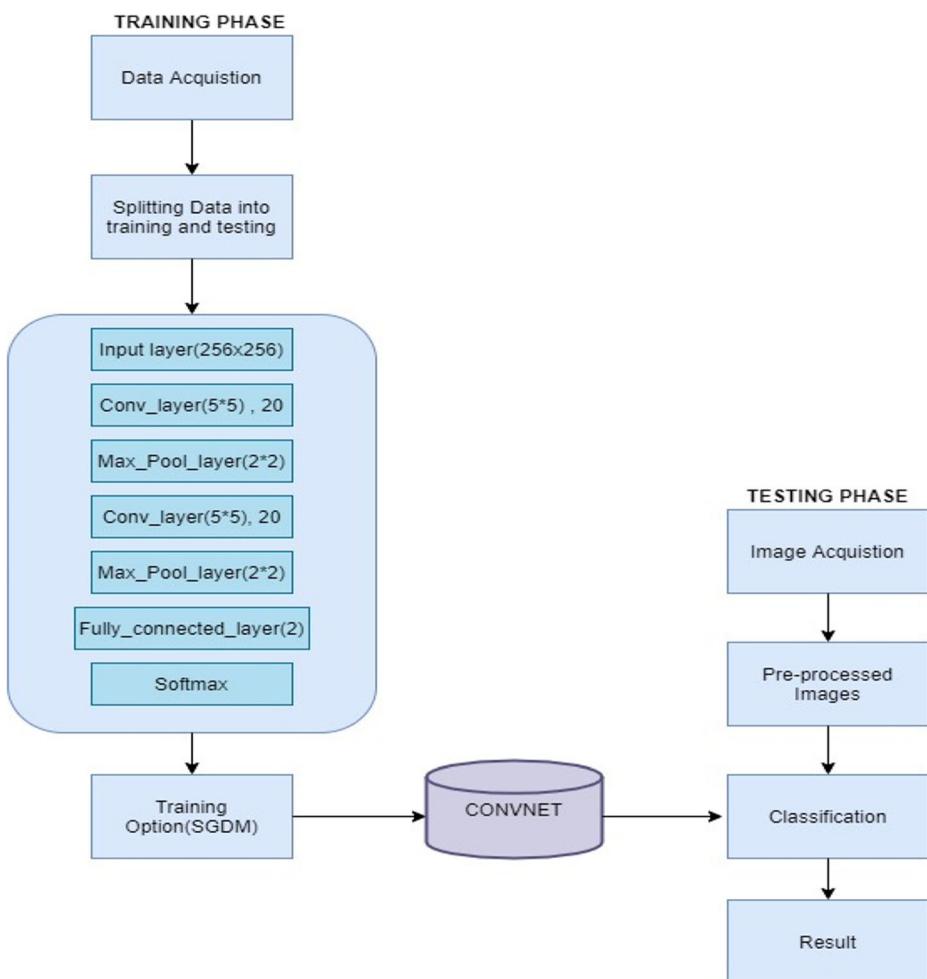


Fig. 19 Training and testing phase of CNN

Table 3 Parameter settings of CNN

Parameters	Values
Dropout Factor	0.2
Max Epoch	10
Batch size	15
Learning rate	0.01
Activation function	ReLU
Fully connected (Activation function)	SoftMax
Optimizer	SGDM
Max pooling	2X2

4.1 Sensitivity

It is defined as the ability to correctly identify the subjects/persons with the disease condition. It is computed using Eq. 15:

$$\text{Sensitivity} = \frac{TP}{TP + TN} \quad (15)$$

4.2 Accuracy

It is defined as the statistical measure to tell how well a classifier correctly identifies the condition. It is the percentage of true results in the given dataset. It is computed using Eq. 16:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

4.3 Specificity

The specificity of a diagnostic test quantifies its ability to correctly identify the subjects/persons with no disease condition. It is computed using Eq. 17:

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (17)$$

4.4 F-measure

It's the statistical analysis for binary classification. It is computed using Eq. 18:

$$F\text{-measure} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (18)$$

where precision = TP/TP + FP.

5 Results and discussions

This section discusses the simulation results of the proposed model. This work is divided into two experiments, experiment I and experiment II. The proposed model has been compared with the various state-of-the-art models and popular deep learning techniques like DNN,

Resnet-50, Alexnet and VGG-16. For better representation, public TDID dataset is renamed as dataset-1 and collected dataset as dataset-2. In most of the thyroid nodule detection and classification problems, authors have used local/private datasets. Therefore, we have decided to use public TDID and local datasets, i.e., the collected dataset in our work. The main reason for using the TDID dataset in our work is that authors in the previous studies of thyroid nodule problems [1, 23, 24, 38, 41] have used this dataset. This helped us to evaluate and compare the performance of the proposed model with the previous studies. The proposed model is also been compared with and without segmentation and boundary detection techniques.

5.1 Experiment-I

Initially the sample size for dataset-1 was 295, after applying five different data augmentation techniques the size of dataset-1 becomes 1475 i.e., 295×5 thyroid USG images. For the better evaluation of the models, 8:2 ratio is set for training and testing of the all cases of the deep learning techniques like DNN, Resnet-50, Alexnet and VGG-16 models. Figure 20 shows the initial and after applying data augmentation sample size of the dataset. Table 4 shows the comparison of the proposed model with and without segmentation, boundary detection and classification techniques on dataset-1. It is inferred from the table that the proposed model performs better with 1% to 2% improvement with segmentation, boundary detection and classification. Table 5 shows the performance comparison of the various state-of-the-art models for detecting and classifying thyroid nodules on dataset-1. It is found from the table that the proposed model performs better with an improvement of 2% to 4%. Table 6 shows a comparison of the proposed model with the popular deep learning techniques on dataset-1. From the table, it can be analysed that the proposed model has (achieved an accuracy of 93.75%, sensitivity of 94.62%, specificity of 92.53% and f-measure 94.09%) performs better followed by DNN, Resnet-50, VGG-16 and Alexnet models. Table 7 shows the results of different values set for learning rate and dropout factor on dataset-1. The model performs better with the learning rate 0.01 and drop out factor 0.2. Figure 21 shows the accuracy epoch graph for dataset-1. From the figure, it can be analyzed that the epoch and accuracy of the models are increasing simultaneously. The best accuracy of the models is achieved on the 10th epoch with

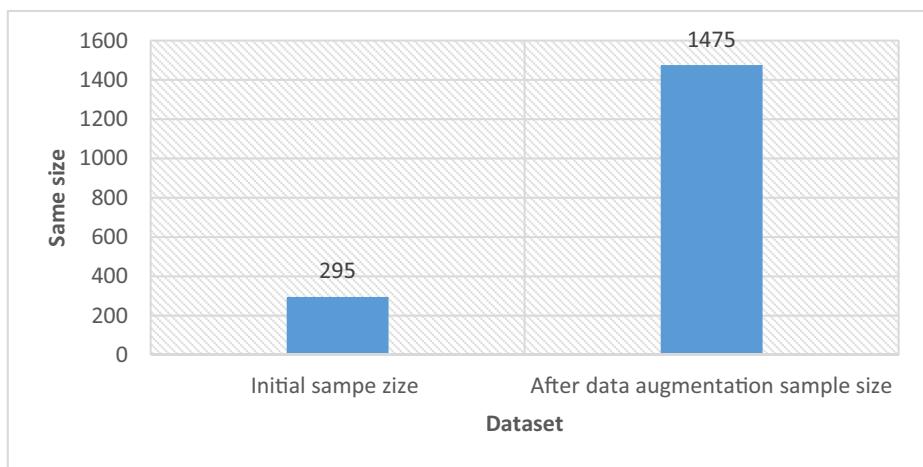


Fig. 20 The initial and after data augmentation sample size of the dataset

Table 4 Comparison of the proposed model with and without segmentation, boundary detection techniques and classification on dataset-1

CASES	Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
CASE 1	Without segmentation and boundary detection techniques	Dataset-1: 92.07	Dataset-1: 93.68	Dataset 1: 91.17	Dataset-1: 92.67
CASE 2	With segmentation and boundary detection techniques	Dataset-1: 93.75	Dataset-1: 94.62	Dataset-1: 92.53	Dataset-1: 94.09

Table 5 The performance comparison of the various state-of-the-art models for detecting and classifying thyroid nodules on dataset-1

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
Wang et al. (2018)	80.91	81.82	80	
Song et al. (2019)	92.1	94.1	96.2	
Nguyen et al. (2019)	90.88	—	—	
Ajilisa et al. (2020)	89.93	92.76	—	89.43
Nguyen et al. (2021)	92.05	—	—	
Proposed model on dataset-1	93.75	94.62	92.53	94.09

an accuracy of 93.75% on dataset-1. Figure 22 shows the comparative analysis of the of the various state-of-the-art models and proposed model accuracy on dataset-1. The model has achieved an accuracy of 93.75% on dataset-1. Figure 23 shows the comparative analysis of the various state-of-the-art models and proposed model sensitivity, specificity and f-measure on dataset-1. The results show the proposed model performs better than the other models having sensitivity of 94.62% and f-measure 94.09%. This proposed model has not shown a good improvement in specificity parameter in comparison with Song et al. having 96.2% specificity. Figure 24 shows the performance comparison of the popular deep learning techniques DNN, Alexnet, Resnet-50, VGG-16 and proposed model on dataset-1. An improvement of 2% to 5% is seen in comparison with other deep learning techniques followed by DNN, Resnet-50, VGG-16 and Alexnet on dataset-1.

5.2 Experiment-II

Initially the sample size for dataset-2 was 654, thyroid images after applying five different data augmentation techniques dataset-2 becomes 3270 i.e., 654×5 thyroid USG images. 8:2 ratio is

Table 6 Comparison of the proposed model with the popular deep learning techniques on dataset-1

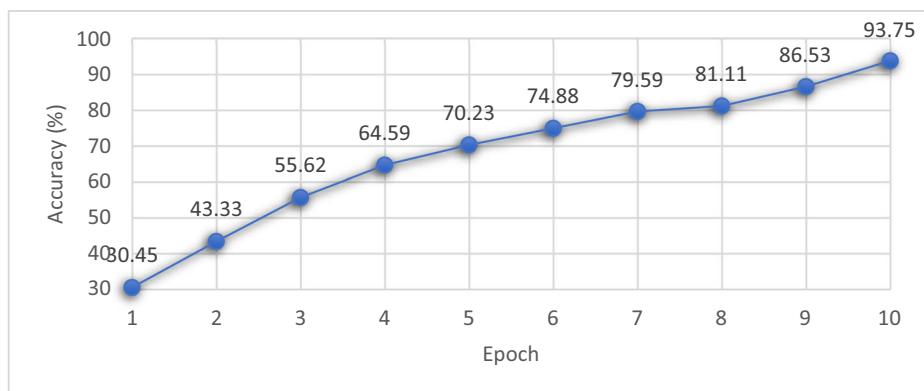
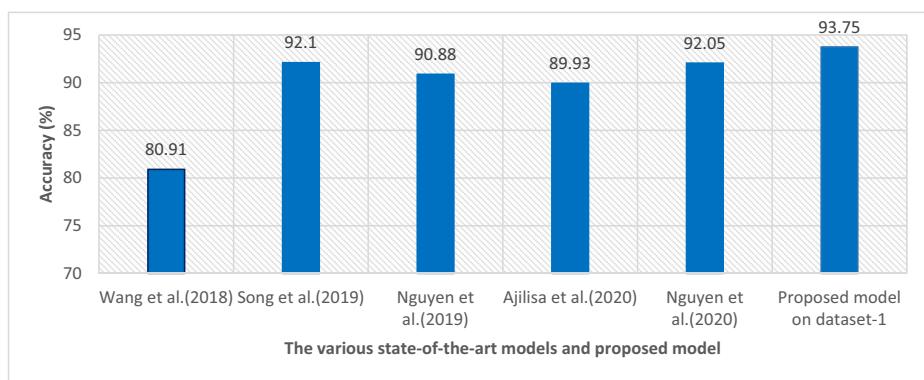
Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
DNN	89.59	90.81	88	90.34
Alexnet	82.84	83.17	81.63	83.17
Resnet-50	86.01	87.25	85.55	92.70
VGG-16	85.62	89.89	84.37	87.67
Proposed model	93.75	94.62	92.53	94.09

Table 7 Results of different values of parameters on dataset-1

Sr. No	Parameters		Accuracy
	Learning rate	Dropout factor	
1.	0.1	0.1	76.06
2.	0.04	0.02	80.90
3.	0.05	0.04	82.32
4.	0.07	0.06	85.16
5.	0.08	0.08	87.19
6	0.09	0.09	89
7	0.01	0.2	93.75

The best result is obtained with learning rate 0.01 and drop-out factor 0.2 on datatset-1

set for training and testing all the models as in experiment-1 for the better comparison. Figure 25 shows the initial and after data augmentation sample size. To further evaluate our proposed model, it has been compared with popular deep learning techniques like DNN, Alexnet, Resnet-50 and VGG-16. The reason for using these techniques is that these techniques have performed better in the previous studies [1, 23, 24, 38, 41] which helped us to

**Fig. 21** Accuracy epoch graph for dataset-1**Fig. 22** Comparative analysis of the of the various state-of-the-art models and proposed model accuracy on dataset-1

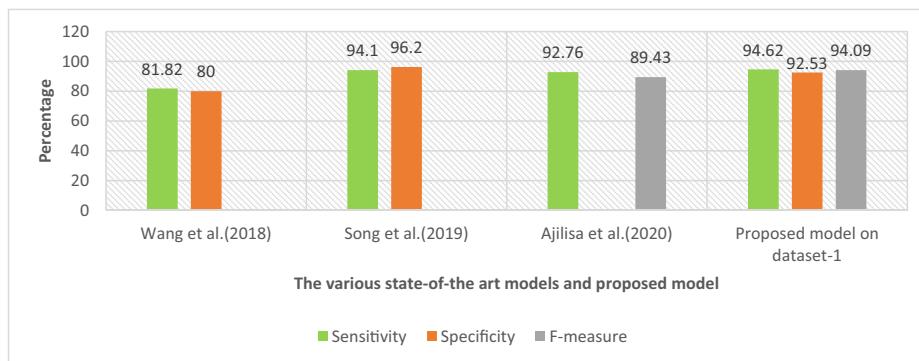


Fig. 23 Comparative analysis of the various state-of-the-art models and proposed model sensitivity, specificity and f-measure on dataset-1

compare with the proposed model. Table 8 shows the comparison of the proposed model with and without segmentation, boundary detection techniques and classification on dataset-2. In the first case, classification is performed without segmentation and boundary detection techniques, while in case 2, classification is performed with segmentation and boundary detection techniques. From the table, it is found the proposed model on dataset-2 performs **better with segmentation, boundary detection and classification** as an improvement of 1% to 2% in the performance of the model. Table 9 shows the performance comparison of the various state-of-the-art models for detecting and classifying thyroid nodules. The proposed model has achieved an accuracy of 96.89%, sensitivity of 97.80%, specificity of 94.73% and f-measure of 97.26%. The proposed model has performed better in comparison with the reported literature. Table 10 shows a comparison of the proposed model with the popular deep learning techniques on dataset-2. From the data collection to segmentation and boundary detection the steps are the same, only in the classification phase techniques (like DNN, Alexnet, Resnet-50, VGG-16) are changed to classify thyroid nodules. The proposed model on dataset-2 performs best followed by DNN, Resnet-50, VGG-16 and Alexnet models. Table 11 shows the various

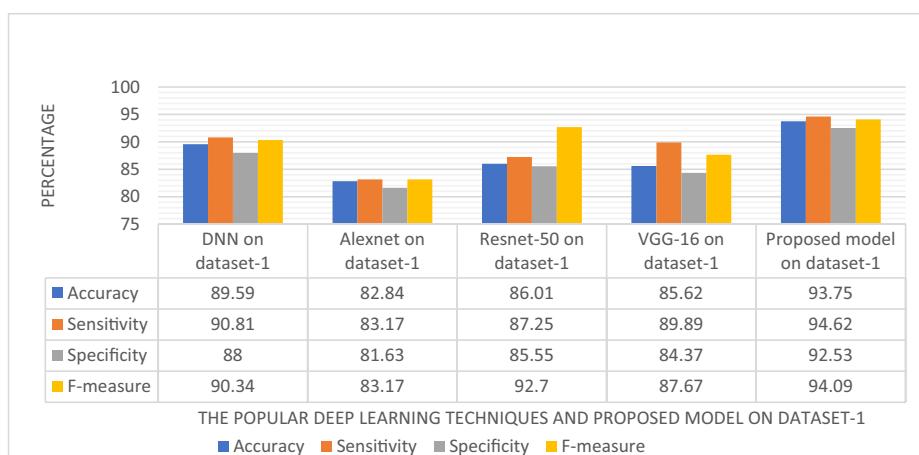


Fig. 24 Performance comparison of the popular deep learning techniques DNN, Alexnet, Resnet-50, VGG-16 and proposed model on dataset-1

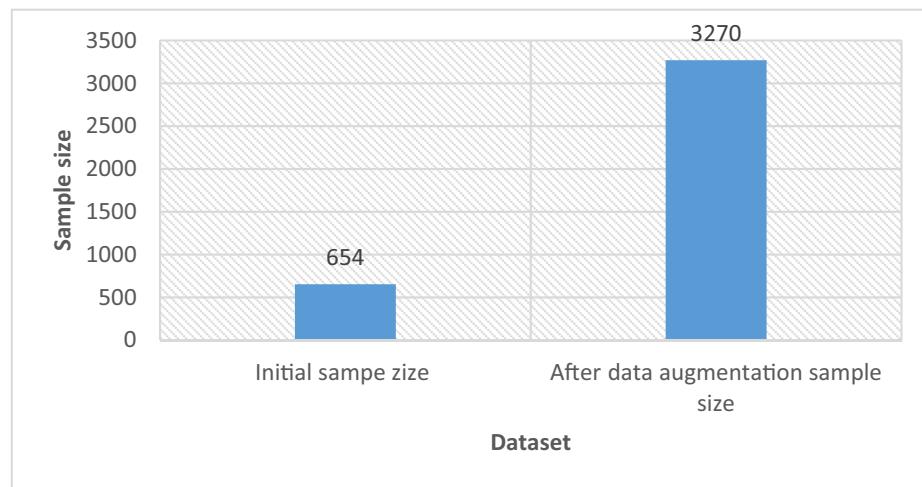


Fig. 25 The initial and after data augmentation sample size

Table 8 Comparison of the proposed model with and without segmentation, boundary detection techniques and classification on dataset-2

CASES	Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
CASE 1	Without segmentation, boundary detection and classification	Dataset-2: 95.38	Dataset-2: 96.73	Dataset-2: 92.10	Dataset-2: 96.20
CASE 2	With segmentation, boundary detection techniques and classification	Dataset-2: 96.89	Dataset-2: 97.80	Dataset-2: 94.73	Dataset-2: 97.26

parameters set for the learning rate and drop-out factor for the proposed model on dataset-2. From the table, it can be observed that the best result is obtained with drop-out factor 0.2 and learning rate 0.01 with an accuracy of 96.89% on dataset-2.

Figure 26 shows the accuracy epoch graph for dataset-2. We can analyze from the figure that the epoch and accuracy of the models are increasing simultaneously. The best accuracy of

Table 9 The performance comparison of the various state-of-the-art models for detecting and classifying thyroid nodules on dataset-2

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
Wang et al. (2018)	80.91	81.82	80	
Song et al. (2019)	92.1	94.1	96.2	
Nguyen et al. (2019)	90.88	—	—	
Ajilisa et al. (2020)	89.93	92.76	—	89.43
Nguyen et al. (2021)	92.05	—	—	
Proposed model on dataset-2	96.89	97.80	94.73	97.26

Table 10 Comparison of the proposed model with the popular deep learning techniques on dataset-2

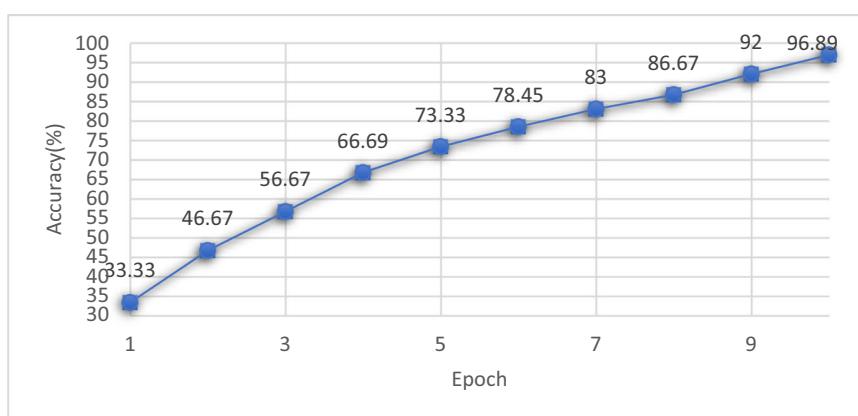
Techniques	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
DNN	92.63	93.68	91.17	93.18
Alexnet	83.08	84.76	82.97	85.98
Resnet-50	91.51	93.68	91.17	94.17
VGG-16	88.48	89	87.91	88.55
Proposed model	96.89	97.80	94.73	97.26

the models is achieved on the 10th epoch. The model has gained an accuracy of 96.89% on dataset-2. Figure 27 shows the comparative analysis of the various state-of-the-art models and proposed model accuracy on dataset-2. The proposed model has achieved an accuracy of 96.89% on dataset-2, which shows an improvement of 2% to 5% for detecting and classifying thyroid nodules from the previous studies. Figure 28 shows the comparative analysis of the various state-of-the-art models and proposed model results on sensitivity, specificity and f-measure on dataset-2. It shows that the proposed model has achieved sensitivity of 97.80%, f-measure of 97.26% on dataset-2. This proposed model has (achieved 94.73% specificity on dataset-2) not shown a good improvement in specificity parameter in comparison with Song

Table 11 Results of different values of parameters on dataset-2

Sr. No	Parameters		Accuracy
	Learning rate	Dropout factor	
1.	0.1	0.1	77.39
2.	0.04	0.02	84.76
3.	0.05	0.04	86.40
4.	0.07	0.06	88.86
5.	0.08	0.08	90.81
6	0.09	0.09	92.70
7	0.01	0.2	96.89

The best result is obtained with learning rate 0.01 and drop-out factor 0.2 on dataset-2

**Fig. 26** Accuracy epoch graph for dataset-2

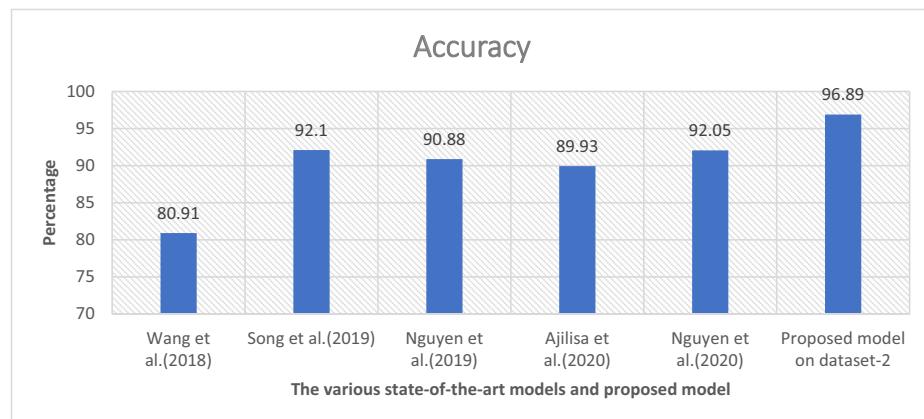


Fig. 27 Comparative analysis of the of the various state-of-the-art models and proposed model accuracy on dataset-2

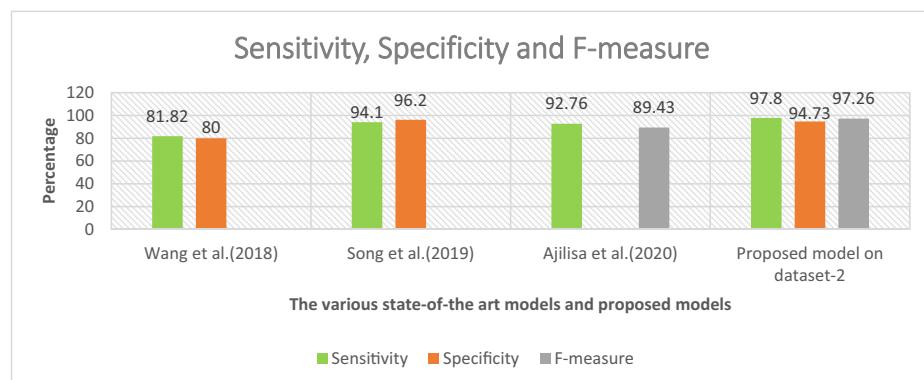


Fig. 28 Comparative analysis of the various state-of-the-art models and proposed model sensitivity, specificity and f-measure on dataset-2

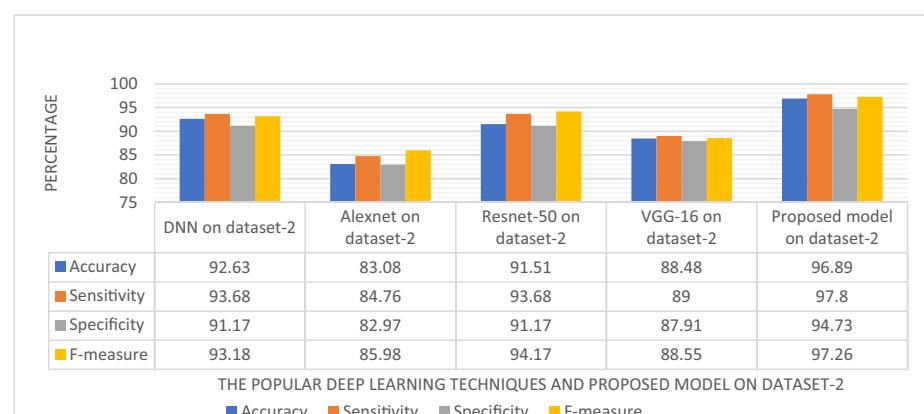


Fig. 29 Performance comparison of the popular deep learning techniques DNN, Alexnet, Resnet-50, VGG-16 and proposed model on dataset-2

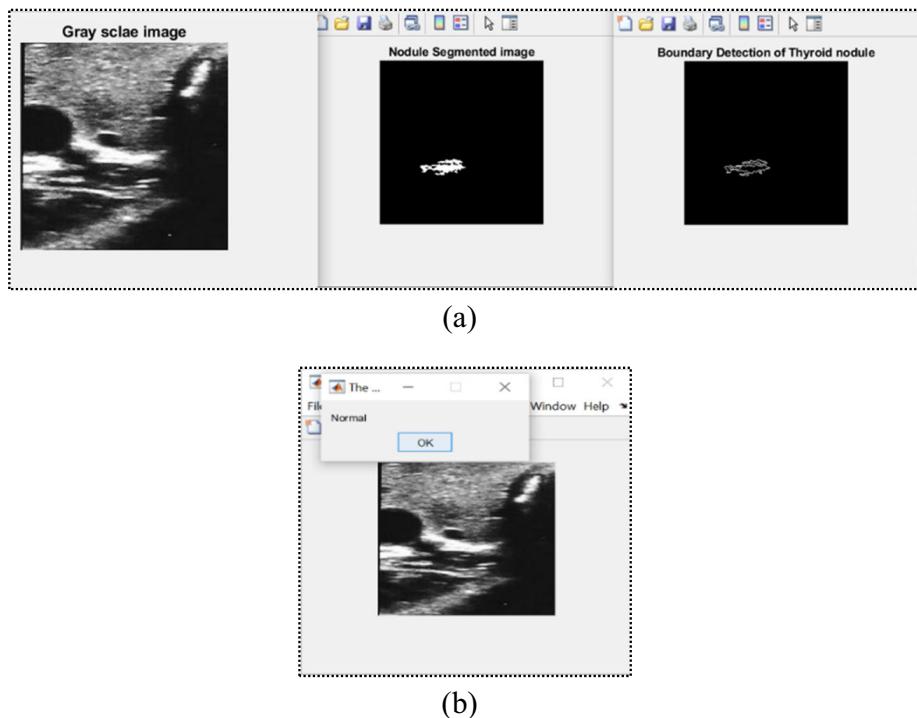


Fig. 30 a-b Simulation results

et al. having 96.2% specificity. Figure 29 shows the performance comparison of the popular deep learning techniques and proposed model on dataset-2. It is clear from the figure that the proposed model has performed better in comparison with other deep learning techniques followed by DNN, Resnet-50, VGG-16 and Alexnet on dataset-2. Figure 30a-b shows the simulation results run on MATLAB 2019b.

6 Conclusion and future work

In this work, optimizing CNN-based model for thyroid nodule classification is proposed using **data augmentation, segmentation and boundary detection techniques**. The model is evaluated on public (TDID) and collected datasets collected from Kriti Scanning Center, Prayagraj, India, duly approved by NABH. The various data augmentation techniques have been utilized to enhance the generalization capability of the proposed model. The proposed model has achieved an accuracy of 93.75%, sensitivity of 94.62%, specificity of 92.53% and f-measure of 94.09% on the public dataset using 1475 thyroid USG images in experiment-I and an accuracy of 96.89%, sensitivity of 97.80%, specificity of 94.73% and f-measure of 97.26% on the collected dataset using 3270 thyroid USG images in experiment-II. The major contribution of this work is also to optimize the learning rate and drop-out factor parameters for training the proposed model. The proposed model has shown an improvement of (4.57%, 7.84%), (5.06%, 8.24%), (4.43%, 6.63%) and (4.66%, 7.83%) in terms of accuracy, sensitivity, specificity and

f-measure on (dataset -1, dataset-2) against other state of the art models. It has shown an improvement of 2% to 5% in the performance evaluation compared with deep learning techniques like DNN, Resnet-50, Alexnet and VGG-16 for thyroid nodule classification with considering segmentation and boundary detection techniques. It works well even when there is a limited number of training images. The proposed model performs well in terms of accuracy, sensitivity and f-measure **but does not perform well in terms of specificity**. This study explores the use of deep learning techniques to assist doctors and practitioner's providing a better diagnosis model by improving the efficiency of large-scale disease screening and diagnosis capabilities. In future work, generative adversarial neural network (GAN) models & fusion of machine learning and deep learning techniques can be used to improve the performance of the model. A graphical user interface can also be designed for the clinicals study purpose. Currently, the number of the public dataset for thyroid nodule imaging is not enough. Therefore, the development of reliable and accessible datasets and the creation of uniform assessment metrics are issues that need to be covered in future research.

Appendix 1 Algorithm for data pre-processing and data augmentation techniques

Algorithm 1 Data pre-processing and data augmentation techniques

Input: Imd → Image

Output: AugImd → Augmented Images

// Define the image browsing option available for the image dataset uploading //

Step 1 *Pathname =Browse (Folder_name, Title of the folder)*
 // Reading of images//

Step 2 *for i→ NumThy*
 Full-path= (Pathname with its filename)
 Images(i)= Read (Full-path)
 //Data pre-processing and data augmentation//

Step 3 *for i→row(Imd)*
 for j→column(Imd)

Step 4 *B= imresize (Imd, scale)*

Step 5 *B= rgb2gray (RGB)*

Step 6 *denoise← Gaussian_blur*

Step 7 *Data Augmentation← imageDataAugmenter (Name, Value)*

Step 8 *End for*
 End for

Step 9 **Return:** AugImd

Step 10 **End**

Appendix 2 Algorithm for morphological operation, segmentation and boundary detection

Algorithm 2 Morphological operation, segmentation and boundary detection

Input: Imd→ image

Output: BdImage→ *Boundary detected image*

// Performing morphological operation and segmentation//

Step 1 Set structure element se=strl('disk',3)

Step 2 $gi \rightarrow imdilate(Imd, se)$

Step 3 $ge \rightarrow imerode(Imd, se)$

Step 4 $Close_Imd \rightarrow gi-ge$

Step 5 *for* $i=i \rightarrow R$

for $j=j \rightarrow C$

Step 6 If $Imd(i,j) > avg(Imd)$

$Bimg(i,j)=1$

Step 7 else if $Imd(i,j) < avg(Imd)$

$Bimg(i,j)=0$

Step 8 $TH \leftarrow Ostu_Threshold$

Step 9 $Segment \leftarrow SegmentThyroidNodule$

$Mask \leftarrow morphological\ (Bimg)$

//Applying boundary detection (BD) technique//

Step 10 $Boundaries \leftarrow boundaries\ (Mask)$

for each I *do*

$BD\ (BdImage)$

end for

end for

end for

Return: Boundary detected Image

End

Appendix 3 Algorithm for classification of thyroid nodules using CNN

Algorithm 3 Classification using CNN

Input: BDImage

Output: Type of thyroid nodule

//Initialization of parameters of CNN//

Step 1 Initialize the parameters of CNN

- *dropoutfactor*
- *maxepoch*
- *batchsize*
- *learning rate*
- *activation function*
- *optimizer*

Step 2 Apply **Relu** activation function

Apply **maxpooling** layer of window size 2 x2.

Convert obtain matrix into 1D feature vector

Add **fully connected layer** and **softmax** as activation function for the classification of thyroid nodules.

Step 3 //Set training and testing ratio//

Convnet ← Train (Net, Training data)

Classification result ← Train (Net, Training data, test data)

//Classification//

Step 4 if (*Classification result* = true)

 Show classification result and predict the nature of the thyroid nodule

 else

 Update the parameters of CNN and go to step 1

Step 5 Prediction

Return: Type of thyroid nodule

End

Funding The authors have no relevant financial or non-financial interests to disclose.

Data Availability Not available due to ethics approval condition.

Declarations

Conflict of interest We declare no conflict of interest.

Research involving human participants and/or animals In this study there is no involvement of human participants and/or Animals.

References

1. Ajilisa, OA, Jagathyraj, VP, Sabu, MK (2020) Computer-Aided Diagnosis of Thyroid Nodule from Ultrasound Images Using Transfer Learning from Deep Convolutional Neural Network Models. In 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA) (pp 237–241). IEEE
2. Al-Rahlawee ATH, Rahebi J (2021) Multilevel thresholding of images with improved Otsu thresholding by black widow optimization algorithm. *Multimed Tools Appl* 80(18):28217–28243
3. Baldini E, Lauro A, Tripodi D, Pironi D, Amabile MI, Ferent IC, Ulisse S (2022) Thyroid Diseases and Breast Cancer. *J Personal Med* 12(2):156
4. Das NN, Kumar N, Kaur M, Kumar V, Singh D (2020) Automated deep transfer learning-based approach for detection of COVID-19 infection in chest X-rays. *Irbm*
5. Guo X, Zhao H, Tang Z (2020) An Improved Deep Learning Approach for Thyroid Nodule Diagnosis. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, pp 296–299. <https://doi.org/10.1109/ISBI45749.2020.9098637>
6. Gurunathan A, Krishnan B (2021) Detection and diagnosis of brain tumors using deep learning convolutional neural networks. *Int J Imaging Syst Technol* 31(3):1174–1184
7. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
8. Ikebe M, Asai T (2005) A digital vision chip for early feature extraction with rotated template-matching CA. *J Robot Mechatronics* 17(4):372
9. Ilhan U, Ilhan A (2017) Brain tumor segmentation based on a new threshold approach. *Procedia Comput Sci* 120:580–587. <https://doi.org/10.1016/j.procs.2017.11.282>
10. Kaur M, Kumar V, Yadav V, Singh D, Kumar N, Das NN (2021) Metaheuristic-based deep COVID-19 screening model from chest X-ray images. *J Healthcare Eng*. <https://doi.org/10.1155/2021/8829829>
11. Khan A, Sohail A, Zahoor U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. *Artif Intell Rev* 53(8):5455–5516
12. Khan ID, Khan MH, Farooq O, Khan YU (2021) A Comparative Analysis of Seizure Detection via Scalogram using GoogLeNet, AlexNet and SqueezeNet. In 2021 Smart Technologies, Communication and Robotics (STCR) (pp 1–5). IEEE
13. Khanna M, Agarwal A, Singh LK, Thawkar S, Khanna A, Gupta D (2021) Radiologist-Level Two Novel and Robust Automated Computer-Aided Prediction Models for Early Detection of COVID-19 Infection from Chest X-ray Images. *Arab J Sci Eng*, 1–33
14. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, Kwak JY (2019) Deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. *Head Neck* 41(4):885–891
15. Koundal D, Gupta S, Singh S (2018) Computer aided thyroid nodule detection system using medical ultrasound images. *Biomed Signal Process Control* 40:117–130. <https://doi.org/10.1016/j.bspc.2017.08.025>
16. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Proces Syst* 25:1097–1105. <https://dl.acm.org/doi/10.5555/2999134.2999257>
17. Kumar N, Gupta M, Gupta D, Tiwari S (2021) Novel deep transfer learning model for COVID-19 patient detection using X-ray chest images. *J Ambient Intell Humaniz Comput* 14(1):469–478. <https://doi.org/10.1007/s12652-021-03306-6>
18. Kumar N, Narayan Das N, Gupta D, Gupta K, Bindra J (2021) Efficient automated disease diagnosis using machine learning models *J Healthcare Eng*. <https://doi.org/10.1155/2021/9983652>
19. Li C, Tofghi MR, Schreurs D, Horng TSJ (2016) Principles and applications of RF/microwave in healthcare and biosensing. Elsevier, Amsterdam
20. Lu J, Ouyang X, Shen X, Liu T, Cui Z, Wang Q, Shen D (2022) GAN-Guided Deformable Attention Network for Identifying Thyroid Nodules in Ultrasound Images. *IEEE J Biomed Health Inf* 26(4):1582–1590. <https://doi.org/10.1109/JBHI.2022.3153559>
21. Merzban MH, Elbayoumi M (2019) Efficient solution of Otsu multilevel image thresholding: A comparative study. *Expert Syst Appl* 116:299–309. <https://doi.org/10.1016/j.eswa.2018.09.008>
22. Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM (2018) Classification using deep learning neural networks for brain tumors. *Future Comput Inf J* 3(1):68–71. <https://doi.org/10.1016/j.fcij.2017.12.001>
23. Nguyen DT, Pham TD, Batchuluun G, Yoon HS, Park KR (2019) Artificial intelligence-based thyroid nodule classification using information from spatial and frequency domains. *J Clin Med* 8(11):1976
24. Nguyen DT, Kang JK, Pham TD, Batchuluun G, Park KR (2021) Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence. *Sensors* 20(7):1822. <https://doi.org/10.3390/s20071822>
25. Owais M, Arsalan M, Choi J, Park KR (2019) Effective diagnosis and treatment through content-based medical image retrieval (CBMIR) by using artificial intelligence. *J Clin Med* 8(4):462
26. Park K, Chae M, Cho JH (2021) Image Pre-Processing Method of Machine Learning for Edge Detection with Image Signal Processor Enhancement. *Micromachines* 12(1):73

27. Pedraza L, Vargas C, Narváez F, Durán O, Muñoz E, Romero E (2015) An open access thyroid ultrasound image database. In 10th International Symposium on Medical Information Processing and Analysis (Vol. 9287, p. 92870 W). Int Soc Opt Photon
28. Quality Council of India (n.d.) [Online] Available: <https://www.nabh.co/frmViewCGHSRecommend.aspx?Type=Diagnostic%20Centre&cityID=94>. Accessed 07 Sept 2022
29. Rai HM, Chatterjee K (2021) 2D MRI image analysis and brain tumor detection using deep learning CNN model LeU-Net. *Multimed Tools Appl* 80:36111–36141. <https://doi.org/10.1007/s11042-021-11504-9>
30. Richman DM, Benson CB, Doubilet PM, Wassner AJ, Asch E, Cherella CE et al (2020) Assessment of American college of radiology thyroid imaging reporting and data system (TI-RADS) for pediatric thyroid nodules. *Radiology* 294(2):415–420. <https://doi.org/10.1148/radiol.2019191326>
31. Rohith, G, Kumar, LS (2022) Design of Deep Convolution Neural Networks for categorical signature classification of raw panchromatic satellite images. *Multimed Tools Appl*, 1–38
32. Ruggeri, RM, Giovanella, L, Campenni, A (2022) SARS-CoV-2 vaccine may trigger thyroid autoimmunity: real-life experience and review of the literature. *J Endocrinol Investig*, 1–7
33. Saba T (2021) Computer vision for microscopic skin cancer diagnosis using handcrafted and non-handcrafted features. *Microsc Res Tech* 84(6):1272–1283
34. Saba T, Khan MA, Rehman A, Marie-Sainte SL (2019) Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction. *J Med Syst* 43(9):1–19
35. Shaban WM, Rabie AH, Saleh AI, Abo-Elsoud MA (2021) Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network. *Appl Soft Comput* 99:106906. <https://doi.org/10.1016/j.asoc.2020.106906>
36. Shi G, Wang J, Qiang Y, Yang X, Zhao J, Hao R, Kazihise NGF (2020) Knowledge-guided synthetic medical image adversarial augmentation for ultrasonography thyroid nodule classification. *Comput Methods Prog Biomed* 196(105):611
37. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
38. Song W, Li S, Liu J, Qin H, Zhang B, Zhang S, Hao A (2018) Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition. *IEEE J Biomed Health Inf* 23(3):1215–1224
39. Srivastava R, Kuma P (2021) A hybrid model for the identification and classification of thyroid nodules in medical ultrasound images. *Int J Model, Identification and Control (IJMIC)*. [In Press]
40. Tejas R (2022) Contour Detection & Edge Detection, Accessed: Sep. 03, 2022[Online], Available: <https://medium.com/@tejas9723/contour-detection-edge-detection-with-opencv-96a74097ef6>
41. Wang, J, Li, S, Song, W, Qin, H, Zhang, B, Hao, A (2018) Learning from weakly-labelled clinical data for automatic thyroid nodule classification in ultrasound images. In 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, pp 3114–3118. <https://doi.org/10.1109/ICIP.2018.8451085>
42. Xie J, Guo L, Zhao C, Li X, Luo Y, Jianwei L (2020) A hybrid deep learning and handcrafted features based approach for thyroid nodule classification in ultrasound images. In *J Phys Conf Ser* (Vol. 1693, No. 1, p. 012160). IOP Publishing
43. Xu Z, Sheykhamad FR, Ghadimi N, Razmjooey N (2020) Computer-aided diagnosis of skin cancer based on soft computing techniques. *Open Med* 15(1):860–871
44. Yang W, Dong Y, Du Q, Qiang Y, Wu K, Zhao J, Zia MB (2021) Integrate domain knowledge in training multi-task cascade deep learning model for benign–malignant thyroid nodule classification on ultrasound images. *Eng Appl Artif Intell* 98(104):064
45. Yao X, Wang X, Wang SH, Zhang YD (2020) A comprehensive survey on convolutional neural network in medical image analysis. *Multimed Tools Appl*, 1–45
46. Zhao SX, Chen Y, Yang KF, Luo Y, Ma BY, Li YJ (2022) A Local and Global Feature Disentangled Network: Toward Classification of Benign-malignant Thyroid Nodules from Ultrasound Image. *IEEE Trans Med Imaging*
47. Zhao X, Shen X, Wan W, Lu Y, Hu S, Xiao R, Li J (2022) Automatic Thyroid Ultrasound Image Classification Using Feature Fusion Network. *IEEE Access* 10:27,917–27,924
48. Zhu YC, AlZoubi A, Jassim S, Jiang Q, Zhang Y, Wang YB, Hongbo DU (2021) A generic deep learning framework to classify thyroid and breast lesions in ultrasound images. *Ultrasonics* 110:106300. <https://doi.org/10.1016/j.ultras.2020.106300>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.