Diagnosis of anomalies based on hybrid features extraction in thyroid images

Mahin Tasnimi 1 • Hamid Reza Ghaffari 1

Received: 17 April 2021 / Revised: 12 May 2022 / Accepted: 2 July 2022
Published online: 18 July 2022
 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Diagnosing benign and malignant glands in thyroid ultrasound images is considered a challenging issue. Recently, deep learning techniques have significantly resulted in extracting features from medical images and classifying them. Convolutional neural networks ignore the hierarchical structure of entities within images and do not pay attention to spatial information as well as the need for a large number of training samples. Capsule networks consist of different hierarchical capsules equivalent to the same layers in the convolutional neural networks. We propose a feature extraction method for ultrasound images based on the capsule network. Then, we combine those deep features with conventional features such as Histogram of Oriented Gradients and Local Binary Pattern together to form a hybrid feature space. We increase the accuracy percentage of a support vector machine (SVM) by balancing and reducing the data dimensions of samples. Since the SVM provides different training kernels according to the sample distribution method, the extracted textural features were categorized using each of these kernels to obtain the result. The parameters of classification evaluation using the researcher-made model have outperformed the other methods in this field. Experimental results showed that the combination of HOG, LBP, and CapsNet methods outperformed the others, with 83.95% accuracy in the SVM with a linear kernel.

Keywords Local binary pattern algorithm • Histogram of oriented gradients algorithm • Thyroid diagnosis • Deep learning • Capsule network

1 Introduction
Thyroid cancer is one of the diseases associated with the thyroid nodule [36]. This nodule is located in the human neck and this nodule preserves the body’s metabolism. Medical images
of this nodule help doctors in better diagnosing benign or malignant. Computer-aided diagnostic (CAD) systems plays important role in the anomaly diagnosis in the thyroid nodule images. The basis of this system is the extraction of key features of the images. In the study [51], the features are extracted and compared with each other. In [59, 60], the CAD systems were evaluated for breast cancer detection. The texture features along with other statistical features were used in breast cancer detection [27]. In [38], the accuracy of different computer-aided diagnostic systems for thyroid nodules classifications was determined. In recent years, thyroid cancer classification using ultrasound images was considered. The challenge of such classification problems commonly lies in how to select detectable features, thus much attention has been focused on the feature design of various types, such as morphometric features and traditional texture features [1, 58]. However, experimental results reveal that the internal simplicity and locality of low-level features determine their limitation on thyroid ultrasound images, for there exist built in disadvantages such as speckle noises and low contrast, along with the variations in shape, size, and stage of different glands. Thus, high-level features with semantic meaning should be induced to obtain better classification quality. The combination of those features may lead to a more wide representation of pathological characteristics of thyroid nodules. Among radiographic images of body structure, ultrasound is known as a valuable diagnostic method for the thyroid. In this case, the tissues of the target limb are examined to extract spatial and frequency characteristics (frequency range) therefrom. Next, the aggregation of textural features together forms the X vector and is assigned to classification models. The main challenge of diagnostic methods using ultrasound images is noise sensitivity and low accuracy due to the extraction of unnecessary features [29]. With the development of CAD systems and using deep learning as a branch of artificial intelligence, the accuracy of cancer diagnosis in ultrasound images was greatly improved [55]. In fact, the artificial intelligence is still in research state and it is growing day by day and increasing its scope in all possible fields [7]. The main challenge of diagnostic methods using ultrasound images is noise sensitivity and low accuracy due to the extraction of unnecessary features [29]. The effect of deep learning on the management of the thyroid glands in ultrasound images was also discussed in detail in the article [6]. Therefore, different models, such as convolutional neural networks (CNNs), receive images and extract their features with high precision by applying low- and high-pass filters [32]. These features can be a good description of key points in the image, but the important issue is the location of damaged tissues, which is less addressed in CNNs. Capsule networks have been introduced and designed to compensate for this feature. These networks can use gradient matrices to reverse the location and amount of variations in pixels, in addition to their frequency characteristics [43]. Additionally, a comprehensive review of the capsule network architecture and its use in different fields were discussed and then compared to CNN networks in the article [30]. The present study, therefore, uses a capsule network along with other traditional methods of extracting textural features. This combination can reverse the spatial variations of pixels as local features and the extent of their variations in various directions as global features. These points are a good descriptor for thyroid tissues and significantly increase the accuracy of classification models. To get better classification results, we propose a hybrid approach combining traditional features with deep features.

The main contribution of this study can be summarized in the following two aspects:

1- A capsule network (CAPSNET) was designed to remove the challenges of using CNN network in extracting deep features from images, and the extracted features were combined with the features of the Local Binary Pattern (LBP) and histogram of oriented
gradients (HOG) algorithms. Thus, finding these useful features was consequently led to better diagnosis.

2- The Principal Component Analysis (PCA) method was used to reduce the obtained features and the SVM classifier was also used to diagnose cancer and to use the benefits of deep learning and machine learning in combination.

The remained the article were organized as follows: Section 2 presents the previous related works, Section 3 describes the proposed method, Section 4 discusses the experimental results, and Section 5 reports the results.

2 Related works

For many years, human societies have been struggling with various health problems, one of which is the overgrowth of cells, known as cancer, in different parts of the body. Cancer is a disease that can be caused, for example, by hereditary and environmental factors, but changes in one’s DNA can lead to the disease in all cancer types [2]. There are different types of cancers, one of which is the thyroid gland, located below the larynx and above the clavicle, which controls the body’s metabolism and energy by secreting T4 and T3 hormones [5]. Diagnosis of thyroid-related diseases is known as one of the most important research areas related to the classification issues in data mining and machine learning science [23]. In [54], to compare the effect of several different data mining models on several different types of disease. The models used were decision trees, neural networks, logistic regression, support vector machine, and Naïve Bayes. Additionally, the diseases investigated were diabetes, breast cancer, and hyperthyroidism. Research results have shown that the support vector machine works better compared to the others. In [10] paper, analytical analysis is performed to find out the impact of COVID-19 in terms of total cases, total recovery and total death which has been reported during this pandemic that have been utilized Various Machine learning algorithms to perform this analysis. A data set of 7706 samples and 16 features was categorized through logistic regression and artificial neural networks (ANNs) which showed the superiority of ANNs in the classification of discrete features [22]. In a study [44] on a data set of 7547 samples with 30 features from the University of California, 776 samples had thyroid disorder and the rest were normal. A hybrid model of decision trees and random forests were used as classifiers in this research. The main challenge of this hybrid model is the imbalance in the input data set, which reduces the accuracy percentage to some extent. The use of three models of the decision tree, multilayer neural network (MLN), and RBF network on the University of California dataset with 215 samples (30 hypothyroid, 35 hyperthyroid, and 150 normal samples) were associated with an accuracy of 98.15 for MLN [41]. In [49], a linear analysis model was performed on a dataset with 3773 samples (184 thyroid cases) with 99.62% accuracy. An accuracy percentage of 95.38 was reported by combining statistical models, such as the Naive Bayes method [9] with the search algorithm. Decision tree models, MLNs, and RBF, along with Naive Bayes, were implemented on the datasets of the California and Romania Universities [20]. In this study, the decision tree with 96.5% accuracy on the California data set and the decision tree and MLN models with 82.4% accuracy on the Romanian data set outperformed other models. A version of the decision tree classification model, called CART, was responsible for detecting thyroid samples with 756 samples [21]. Due to the change in the selection of the branch point of different levels of the tree, this method
does not have the problems of the classic decision tree and records an accuracy of 94%. A combination of the SVM model with the particle swarm optimization algorithm was used for the University of California data set [4]. Here, the lost samples are filled using the nearest neighbor K algorithm. In another study, a different approach was proposed to solve the thyroid diagnosis problem [11], which aimed to analyze the medical data of the fissure and tongue to diagnose hyperthyroidism or hypothyroidism. Tongue tests can indicate the inner state of people. A semi-regulatory clustering model has been used for this purpose. The use of several separate indicators to extract features is another innovation of this research. In [39], images of the thyroid gland were collected by cytopathology, and these images were used to develop a deep learning model to diagnose thyroid bulges and disorders. The model was developed based on the concept of multi-sample learning. It assigns specific weights to each point of the image and the model is taught based on the local features of the image to discover the defects with the highest precision. The use of image features specific to thyroid failure improves the model performance. The detection precision of this model is 87%, which is more efficient than reference machine learning models. Applying association rules analysis methods is another approach to extract the rules for classifying thyroid samples. In [19], positive and negative rules are created concerning the features used in the database using the MS-Apriori algorithm. These rules are based on several support criteria and operate according to fuzzy logic. The function of this method is to map the features to a better answer space to fit the model and to rank the initial features. The most important ones remain and the weaker features are removed after ranking the features. Finally, thyroid disease is predicted using reference models, including decision tree, SVM, logistic regression, and Bernoulli Naïve Bayes. The proposed model in AUC outperformed the other models. Another study proposed a novel fuzzy anomaly detection system based on the hybridization of PSO and K-means clustering algorithms over Content-Centric Networks [25]. In the study [24], it was presented a hybrid approach of diagnosing anomalous run-time behaviors in distributed services from execution logs. Another study presented a survey of existing CAD systems that have been developed for the detection of gastric abnormalities based on their feature extraction techniques [3]. In [31], it was used hybrid features for human face recognition. Other research work presented a hybrid anomaly detection approach for seeded bearing faults [17]. In [52], it was presented a hybrid model where an Unsupervised DBN is trained to extract generic underlying features, and a one-class SVM is trained from the features learned by the DBN. The use of deep neural networks is one of the positive strategies in the extraction and classification of textural features. In one research, ultrasound data analysis, which is performed based on ultrasonic waves, was used to diagnose thyroid deficiencies. Of note, the algorithm used was the CNN method, and the accuracy of the diagnostic model was 95%. Regarding the claim of this article that there are great improvements in the extraction of thyroid-specific features in images of thyroid deficiency, improving the methods of extracting local features from the images turned to be the main goal in the proposed model of this study [33]. In [40], ultrasound images were categorized using the ResNet-50 pre-trained network, and the model was found to have higher accuracy than the VGG-19 pre-trained network. With the spread of pre-trained networks, a special type of network, called Google-Net [8], with a random forest classifier recorded a good percentage of accuracy. Vgg-f and Vgg-verydeep16 pre-trained networks are discussed in [28]. The combination of layers between deep nets in this structure is such that a dropout layer is used to prevent over-fitting. Another type of pre-trained network, called Image-Net, receives key points extracted from the image by SIFT methods and gradient-based histogram, with 92% classification accuracy [37]. Problems with deep classic networks, such as no consideration of
object placements in the image, led to the presentation of capsule neural networks (CapsNet). In [50], CIFAR and MNIST datasets were entrusted to the CapsNet and the results showed its high accuracy due to the dynamic routing of this network. In another study [48], several VGG16, GoogleNet, Alex-Net, and Inception-V3 pre-trained models were used to diagnose benign or malignant gland images by the learning transfer method. An important point in this study is the use of conventional ultrasound images and elasticity together, which in turn leads to better performance. Article [18] classified thyroid ultrasound images in a large range based on deep learning. Photo data was collected over two years and the Inception-V3 pre-trained network was used. The CIFAR-10 pre-trained network was used to extract high-level image features to reduce FP by the learning transfer method [53].

3 Suggested method

The model proposed by the researcher consists of three main phases: preprocessing, feature extraction, and final classification. Figure 1 shows the outline chart of the proposed model.

3.1 Pre-processing phase

Preliminary images include transcripts that contain information about patients and their condition by imaging systems. This section was deleted using the cropping method. In this method, the image border size is initially obtained using its main borders, and segmentation is done before the final deletion. The centroid points of the image objects are calculated and then the area of each part is returned in segmentation (each object is considered an island-shaped area of pixels with high brightness). Next, the areas with more area are considered as the main part of the image and the rest as the background. This can greatly reduce the sensitivity of images to noise (Fig. 2).

3.2 Extracting key feature

Feature extraction refers to the process of converting the raw pixels of an image into useful and meaningful information. These features can be extracted both manually and automatically by neural networks. During this process, the low-level features in the images such as color, texture, and shape are firstly identified, and then the high-level features are discovered by low-level features. Thereafter, the features extracted from the images are shown in numerical values. The feature extraction operation in convolutional networks is mostly done in two parts.
as follows: convolutional layers and pooling layers. Thus, using convolutional layers, the filter is multiplied in the local area of the input, and with this operation, feature maps is automatically obtained in several steps. In fact, in the obtained feature maps, the main elements are identified among the pixels of the image. In the next step, using the pooling layers, the dimensions of the extracted feature maps are reduced in several steps.

In this phase, the spatial features of each image are extracted using traditional methods, such as local binary pattern (LBP), histogram of oriented gradients (HOG), and the capsule network. Any method has its capabilities and can ultimately increase the accuracy of the classification model. In LBP, the ratio of each pixel to the central pixel is presented in the form of a power function with base two, and its sum determines the status of each (local features). In the HOG method, the rate of change in color intensity of pixels is presented in the form of a matrix with predetermined angles (global features). In capsule networks, global features are restored using a color intensity gradient, in addition to extracting local features. The difference between capsule networks and traditional methods is their high accuracy in investigating low-pass frequencies and the location of their changes. In the following, a complete description is provided of the performed methods.

3.2.1 Feature extraction using local binary pattern

In this method, each pixel of the image is examined separately and shown with the symbol c. Next, the neighboring pixels $C (g_c)$ with radius $R (g_r)$ and the number $P$ are examined one by one (in a row) and the status of pixel $c$ is specified using Eq. (1) [42].

$$LBP_{P,R} = \sum_{i=0}^{P-1} S(g_i-g_c)2^i, \quad S(g_i-g_c) = \begin{cases} 1 & g_i \geq g_c \\ 0 & g_i < g_c \end{cases}$$

The main problem with the local binary pattern is the sensitivity to location, which makes the output different with each rotation (at any angle) of the image. This problem has been overcome using rotational patterns around the central pixel. These uniform patterns calculate the number of spatial transitions around $C$, and if the output is greater than $U$ (in this study, the
value of U is considered to be 2), it means that the image does not rotate and with the outcome is a different matrix. Equations 2 and 3 show the calculation of uniform patterns [46].

\[
LBP_{P,R}^{u \leq 2} = \begin{cases} 
\frac{\sum_{i=0}^{P-1} s(g_i - g_c)}{P+1} & \text{if } U_{(LBP,P,R)} \leq 2 \\
0 & \text{otherwise}
\end{cases}
\]  \tag{2}

\[
U_{(LBP,P,R)} = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{i=1}^{P-1} |s(g_i - g_c) - s(g_{i-1} - g_c)|
\]  \tag{3}

The calculation of the uniform local patterns results in a matrix the size of the input image with M rows and N columns (P matrix), which is used to extract statistical components as follows:

- **Contrast**: This property calculates the intensity of the color contrast between a pixel, its adjacent neighbors, and the P matrix (the contrast of a fixed image is zero).

\[
Contrast = \sum_{i=1}^{M} \sum_{j=1}^{N} |i-j|^2 \times p(i,j)
\]  \tag{4}

- **Correlation coefficients of variables**: This property describes the relationship between adjacent pixels and the P matrix. Numbers 1, 0, and \(-1\) mean complete relation, no relation, and inverse relation, respectively (\(\mu_x\) and \(\sigma_x\) show the mean and variance of the pixels, respectively, near the central pixel \(c\)).

\[
Corr = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (i,j)P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}
\]  \tag{5}

- **Dissimilarity**: This property returns the difference in the brightness of adjacent pixels and the P matrix.

\[
Dissimilarity = -\sum_{i=1}^{M} \sum_{j=1}^{N} |i-j|P(i,j)
\]  \tag{6}

- **Energy**: This property calculates the square of the elements that make up the P matrix (the energy property is known as the homogeneity component in the LBP matrix). When the distribution of gray surfaces in LBP is fixed or reproducible, the energy component gains more value and vice versa.

\[
Energy = \sum_{i=1}^{M} \sum_{j=1}^{N} p(i,j)^2
\]  \tag{7}

- **Entropy**: Entropy is the amount of anomaly in the P matrix (imbalance). When all the components of the P matrix are equal (no texture in the image), the entropy value will be zero.
Entropy $= \sum_{i=1}^{M} \sum_{j=1}^{N} p(i,j) \log(p(i,j))$ \hfill (8)

- Homogeneity: This property indicates the homogeneity or uniformity of the pixels. The more uniform it is, the extracted features are statistically more powerful because they retain the uniqueness of the patterns.

\[ \text{Homogeneity} = \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{1}{1 + (i,j)^2} p(i,j) \] \hfill (9)

### 3.2.2 Feature extraction using histogram of oriented gradients

In this method, the input image is divided into square blocks with a diameter of 8 units (pixels) and then the derivative of each pixel of the block (with coordinates r,c) is calculated in horizontal (IX) and vertical (IY) directions (Eq. 10). The gradient angle ($\theta$) with vertical to horizontal derivative inverse tangent is obtained by calculating the gradient of each pixel (Eq. 11). By determining the gradient angle as the slope of the pixel variations, the amount of variation (variation length or $\mu$) is also calculated using Eq. 12 [56].

\[ Ix(r,c) = Ix(r,c+1) - Ix(r,c-1) \] ,\[ Iy(r,c) = Iy(r-1,c) - Iy(r+1,c) \] \hfill (10)

\[ \theta = \tan^{-1} \frac{Iy}{Ix} \] \hfill (11)

\[ \mu = \sqrt{I_x^2 + I_y^2} \] \hfill (12)

Next, the blocks generated from the angle and gradient length matrices are examined one by one to determine the histogram vector. Here, the length of the vector is 9, which is set in the range of 10–170° with a step length of 20. In the last step, the histogram matrix is normalized using the zero (Euclidean) norm to reduce the effect of changes due to the contrast between images of an object.

### 3.2.3 Feature extraction by the capsule network

CapsNets are a special type of CNNs that adapt local features by changing their structure and creating a layer, called a capsule, consisting of hundreds of neurons [50]. Three general layers are used in the CapsNet. There is a classic CNN in the first layer. This layer extracts local features (low-pass level) according to the color intensity of the pixels and provides them to
the capsule layer. The convolution layer has 256 channels and 9 × 9 size kernels with stride 1 (the activation function of this layer is RELU). The second layer consists of the CapsNet, which is referred to as the primary capsules. There are 32 channels in the second layer, each of which includes eight primary capsules with a convolution layer with a length of 256 channels and 9 × 9 kernels. Therefore, the output vector of each capsule (w) is 256 × 81 long, which can extract local features with high accuracy. Generally, the primary capsules in the second layer will have an output of 32 × 56 × 56 long (each capsule consists of eight dimensions) according to the input image size. In the final layer, there are numerical capsules (calculators) that produce the final features from the capsule layer (this layer indicates the possible presence of a sample from each class). Given that the capsule layer contains 32 channels with a size of 56 × 56, and since 10 capsules are considered in the third layer, the size of this layer is 56 × 56 × 32 × 10. Figure 3 shows the general structure of the CapsNet.

It should be noted that only feature extraction is performed in the capsule network used by the researcher. Therefore, it is critical to use a dynamic routing algorithm between the second and third layer capsules (because the output of the first layer is one-dimensional, routing does not take place between this layer and the primary capsule layer). Here, the output of each capsule (u) is sent to the third layer capsules (v0-v9) with the same probability vector (at the beginning, there is the same probability of receiving u vector, i.e. local features, by the numerical capsules). In other words, the \( i \)th capsule in the \( L \)-layer tries to detect the output of the \( j \)th capsule in the \( L + 1 \) layer [61]. Over several consecutive rounds, the output of capsule \( j \) (\( v_j \)) is finally sent to capsule \( i \) by a nonlinear function called squash (Eq. 13). Here, \( s_j \) is the length of the vector received by the numerical layer (calculator). Finally, the output of \( V_j \) for each calculator capsule will be the vector of the features extracted by the CapsNet (10 capsules each extracting 16 features according to the length of vector \( w \)) (Figs. 4, 5 and 6).

\[
V_j = \frac{||S_j||^2}{1 + ||S_j||^2} \times \frac{S_j}{||S_j||} \tag{13}
\]

### 3.3 Balancing in the texture features set

The thyroid cancer sample label has two general classes where the number of positive samples (with cancer) is minor compared to negative samples. This can cause an imbalance of the data

![Fig. 3 Schematic of the CapsNet](image)
set and directly affect the classification process of learning models. For this reason, this section uses the synthetic minority oversampling technique (SMOTE) algorithm, in which synthetic samples are generated from real data (textural feature vector in the minority class) and added to the model. The different steps of the SMOTE algorithm are as follows [15]:

1. Isolation of minority classes, including positively labeled specimens (with thyroid cancer).
2. For all minority samples \( x \), a sample \( y \) is randomly selected from the neighboring \( k \) (in this article \( k \) is equal to 5) directly to calculate the difference between \( x \) and \( y \).
3. Generation of a new sample by multiplying a normal random number by the output obtained from the second step and adding it to the original sample or \( x \).

By generating new samples, a balance is established in the textural feature set and conditions are provided for the selection of effective feature vectors.

### 3.4 Reducing the dimensions of textural features using the principal component analysis

Considering the basic logic of the PCA, each numerical matrix can be represented as a linear combination of part of its input values in the form of new variables. These variables describe the main features of the matrix and remove the content overlap from the data due to their orthogonal nature [16, 34, 35, 57]. In the first step, a variable (e.g. \( P_1 \)) is sought to be able to establish the condition of Eq. 14 in a matrix of size \( X_{n \times k} \).

![Fig. 4](image_url)  
**Fig. 4** Comparative diagram of the performance of different feature extraction and classification methods with a linear kernel
In statistics, the higher the variance at one point, the higher the accumulation of information in that area. Therefore, Eq. 15 can be expressed as maximizing the variance of variable \( P_1 \) (matrix \( V \)) [47].

\[
P_1 = X_{t1} \quad ||t1|| = 1
\]  

(14)

**Fig. 5** Comparative diagram of the performance of different feature extraction and classification methods with the Gaussian kernel

**Fig. 6** Comparative diagram of the performance of different feature extraction and classification methods with polynomial kernels
\[ \text{Var}(P_1) = \frac{1}{n} \|P_1\|^2 = \frac{1}{n} i'X'Xi_1 = i_1', \quad Vt_1V = \frac{1}{n} X'X \]  

(15)

Equation 15 can be maximized using the Lagrangian coefficient.

\[ L = i_1'Vt_1 - \lambda (i_1't_1 - 1) \]  

(16)

By calculating the partial derivative of \( L \) relative to \( t_1 \) and \( \lambda_1 \), \( t_1 \) is the normalized vector \( V \) and \( \lambda_1 \) is its corresponding value.

\[
\begin{align*}
\frac{\partial L}{\partial t_1} &= 2vt_1 - 2\lambda_1 t_1 = 0 \\
\frac{\partial L}{\partial \lambda_1} &= -(i_1't_1 - 1) = 0 \\
\end{align*}
\]

(17)

After the steps mentioned above, the unknown value of \( t_1 \) is the normalized vector corresponding to the maximum value of \( \lambda_1 \) or the principal diameter (eigenvector) of matrix \( V \). The structural variable \( P_1 = X_{t_1} \) is the principal component of the first entry in the \( V \) matrix (hence, \( m \) principal component of \( V \) can be obtained using \( P_m = X_{t_m} \)). The final output of the PCA algorithm (matrix \( V \)) is the same size as the input matrix and each input refers to a specific score or value. In the last step, the sum of the information in \( P \) is returned by Eq. 18 to reduce the data dimensions and select the effective features of matrix \( V \) to reach the threshold value, \( t_h \), as the final features (here, \( t_h \) value is considered 0.9).

\[ \text{final indexes} = \sum_{i=1}^{m} \text{Var}(P_i) \leq th \]  

(18)

### 3.5 Classification using the SVM

In the last step of the proposed method, the principal components extracted from the features matrix are assigned to the SVM classifier. In this model, the choice of the location of the hyperplanes in the problem space and the maximization of the margins between them are directly related to the distribution style of samples. In general, the SVM has three main linear, polynomial, and Radial Basis Function kernels, each of which can select the location of the hyperplanes in the problem space with high accuracy. The basic challenge of SVM is low accuracy in dealing with low-repetition input samples. This feature has largely been eliminated using the K-Fold Cross Validation method. Here, the input dataset is assigned to the model in \( k \) steps, and each time \( \frac{1}{k} \) of data is considered as a test (main diameter of the data). In the current study, parameter \( k \) is set to five.

### 4 Experimental results

The dataset used in this study was thyroid ultrasound images data provided by the Society of Photo-Optical Instrumentation Engineers. Dataset is a publicly available thyroid ultrasound image database proposed by Pedraza et al. [45]. The thyroid dataset used in the present study
consists of 400 images with the .jpg extension and a size of 560 × 360 pixels, and each image has been resized to 128 × 128 pixels after applying the preprocessing phase. The numbers of positive and negative samples are 121 and 280 images, respectively. The number of positive samples has increased to 277 after producing artificial samples and creating a balance in the input database. The researcher’s proposed model is implemented in the Python programming language environment version 3.7 with a host system of 32 GB main memory, 2.9 GHz processor, 6 MB of cache, and 512 GB of SSD external memory. Success evaluation parameters of different feature extraction methods are model accuracy, sensitivity, specificity, and f1 score, as estimated in Eqs. (19)–(22).

\[
\begin{align*}
\text{Accuracy} &= \frac{TP + TN}{TP + FN + TN + FP} \\
\text{Sensitivity} &= \frac{TP}{TP + FN} \times 100 \\
\text{Specificity} &= \frac{TN}{TN + FP} \times 100 \\
F1\text{-score} &= \frac{TP}{TP + \frac{1}{2}(FP + FN)} \times 100
\end{align*}
\]

In the above equations, the TP parameter is the expected correctly classified samples in the current class, the TN is the other correctly classified samples from the other classes in the matrix. FP is the number of failed predictions of the current class in other classes, and the FN is the number of unsuccessfully predicted samples in the expected class. Tables 1, 2 and 3 show the results obtained from linear, Gaussian, and polynomial kernels (in all kernels, the gamma value is assumed to be 10).

The tests’ results of previous studies (which used pre-trained networks types) were compared with the results of the proposed method, based on several criteria. The tests’ results showed that the proposed method has better performance (Table 4).

According to Table 4, in papers [8, 40]and [53], which uses artificial neural networks; often the combination of texture features and reproduction of the input images are done by regarding the location of pixel variations (gradients), the substantial problems which can be mentioned as follows:

1. The main layout of different parts of the image is not properly examined and different parts of the target limb may be misidentified.
2. In the frequency domain; there is a possibility of interference in the high and low pass properties at the boundary among objects, which will ultimately reduce the accuracy of the model.
3. Frequency-based methods do not calculate minor changes such as curvature in distribution functions, which can be solved by the Hessian Matrix, but still high feature extraction time is a major problem with frequency domain methods.

As a result, it can be said that the researcher’s proposed methods in applying the local binary pattern matrix does not have the common problems of the frequency domain and the location of changes can be thoroughly observed.
In this regard, in the present study, a type of neural network called capsule has been employed that calculates the location of objects relative to each other and extracts the final features accordingly. Of course, capsule networks do also have challenges, the most important of which are:

1. High time in extracting tissue features
2. Long feature vector length that leads to fundamental complications in classification models
3. High computational complexity that requires powerful hardware

Examination of the outputs obtained from the model proposed by the researcher shows that the extracted patterns can isolate the adjacent tissues well and increase the accuracy by employing the local binary pattern matrix and the capsule neural network. Indeed, determining the maximum value of the threshold in the combination of different color channels perfectly states the rules governing medical images. These rules are based on the contrast distance of the pixels present in a particular color intensity, and when the tissues of the target limbs are considered to be the same spectrum as the healthy tissues due to the imaging conditions, the feature extraction is performed with high accuracy and all parts of the target limb are closely examined.

After this part, it is time to extract the spatial features in the category of different samples. To do so, first the polar coordinate calculation algorithm, which is a method of determining spatial points in image description, is implemented and it measures the amount of instantaneous variations in the borders and edges of the image.

The use of Gaussian filters with variance and prismatic scale as well as the use of capsule neural networks in determining the degree of curvature of the image function causes sharp and obviously instantaneous color intensity changes to be perfectly observed and key points to be extracted in predetermined neighborhoods. The advantages of using this method include the following:

| Feature extraction methods | Accuracy | Sensitivity | Specificity | F1_Score |
|---------------------------|----------|-------------|-------------|----------|
| HOG                       | 0.5954   | 0.5956      | 0.5976      | 0.5936   |
| LBP                       | 0.7569   | 0.7567      | 0.7562      | 0.7565   |
| CAPSNET                   | 0.4982   | 0.5017      | 0.3020      | 0.3511   |
| HOG + LBP                 | 0.7813   | 0.7812      | 0.7854      | 0.7804   |
| HOG + CAPSNET             | 0.5712   | 0.5712      | 0.5731      | 0.5677   |
| LBP + CAPSNET             | 0.7569   | 0.7569      | 0.7562      | 0.7565   |
| HOG + LBP + CAPSNET       | 0.8395   | 0.5397      | 0.7186      | 0.5239   |

| Feature extraction methods | Accuracy | Sensitivity | Specificity | F1_Score |
|---------------------------|----------|-------------|-------------|----------|
| HOG                       | 0.9878   | 0.9877      | 0.9883      | 0.9878   |
| LBP                       | 0.9982   | 0.9981      | 0.9983      | 0.9981   |
| CAPSNET                   | 0.5128   | 0.5034      | 0.3491      | 0.3392   |
| HOG + LBP                 | 0.9982   | 0.9981      | 0.9983      | 0.9979   |
| HOG + CAPSNET             | 0.9878   | 0.9877      | 0.9983      | 0.9980   |
| LBP + CAPSNET             | 0.9975   | 0.9974      | 0.9983      | 0.9980   |
| HOG + LBP + CAPSNET       | 0.8681   | 0.6237      | 0.9312      | 0.6582   |
1. High speed in feature extraction by image descriptor.
2. Using a second-order Gaussian filter in the capsule network, which reduces algorithm’s sensitivity to noise samples (in these filters, which perform the softening process, the output will never be zero, which reduces the sensitivity to noise data).
3. The difference between Gaussian distributions, acting as a low-pass filter, examines all the details and reverses rapid changes.
4. The output of the local binary pattern matrix and the capsule neural network reveal boundaries that are appropriate points for extracting key points.
5. Investigation of the angle of changes of pixels by the capsule neural network: this process extracts global properties and eliminates the demand for the model to produce multiple matrices at different angles.

Therefore, the chief problem in models based on local patterns, which is the time required to extract image textures, is largely solved. In addition, the key points extracted from the input image are important parts that can identify differences in classification and determine the ultimate accuracy of the model. At this point, the local binary pattern algorithm only runs on the points that are the main features of the image and no longer need to examine all the pixels. The proposed model is not sensitive to image rotations and returns the same patterns. Moreover, the proposed model of the researcher can (through the use of mapping-reduction architecture) increase the percentage of classification accuracy and flexibility in encountering new coming samples by reducing noise sensitivity and the ability to function in large-scale real-time environments (by using hyperplanes in the input space and maximizing the margins between them).

A list of nomenclature and Acronyms is provided in Table 5.

| Table 3 | Classification results using support vector machine with polynomial kernel function |
|---------|---------------------------------|
| Feature extraction methods | Accuracy | Sensitivity | Specificity | F1_Score |
| HOG | 0.5955 | 0.5959 | 0.6553 | 0.5532 |
| LBP | 0.9513 | 0.9514 | 0.9529 | 0.9529 |
| CAPSNET | 0.8264 | 0.8267 | 0.8434 | 0.8234 |
| HOG + LBP | 0.6076 | 0.6079 | 0.6662 | 0.5714 |
| HOG + CAPSNET | 0.6042 | 0.6045 | 0.6670 | 0.5655 |
| LBP + CAPSNET | 0.9513 | 0.9514 | 0.9529 | 0.9513 |
| HOG + LBP + CAPSNET | 0.8252 | 0.5010 | 0.4126 | 0.4521 |

| Table 4 | Comparison of a different CAD system in classifying benign and malignant thyroid gland images |
|---------|---------------------------------|
| Specificity | Sensitivity | Accuracy | Classifiers | Method for Feature Extraction | Reference |
| 87.20 | 88.20 | 87.80 | SVM | Fine-tuned VGG-16 | Shi, Z et al. [53] |
| 99.30 | 82.80 | 96.34 | Cost-sensitive Random Forest classifier | Fine-tuned Google-Net | Chi, J et al. [8] |
| 64.17 | 80.69 | 97.33 | Artificial neural network | Fine-tuned ResNet-50 | Moussa O et al. [40] |
| 99.83 | 99.74 | 99.75 | SVM | CAPSNET + LBP + HOG | Proposed model |
5 Conclusion

In this paper, a hybrid feature extraction method for the thyroid nodules classification is proposed. Feature extraction refers to the process of converting the raw pixels of an image into useful and meaningful information. These features can be extracted both from traditional and deep methods. Deep learning in the field of medical images has occupied the heavy burden of feature extraction. CNN neural networks play a significant role in this area. Because CNNs cannot evaluate images from different angles, capsule networks have been introduced. Accordingly, a capsule neural network was shown that can be more successful in the field of feature extraction by considering the spatial connections of features in the image. In this article, the potential of implementing a capsule neural network to extract appropriate features from thyroid ultrasound images was investigated, and then the researcher combined those deep features with traditional methods features to diagnose the type of tissue cancer in the thyroid nodule.

In our database, the results indicate that the combination of HOG, LBP, and CapsNet methods outperformed the others, with 83.95% accuracy in the SVM with a linear kernel. Similarly, the LBP algorithm recorded a better performance in Gaussian and polynomial kernels, with 99.82% and 95.13% accuracies, respectively. According to the results, the HOG algorithm with 98.78% accuracy has higher accuracy than the other kernels, meaning the nonlinearity of this method and the LBP algorithm. In the linear kernel, the combination of all three methods shows more accuracy than the others, which can be attributed to the almost linear process of feature extraction by the CapsNet.

One of the problems with the proposed model could be the large size of the local binary pattern matrix, which needs to calculate the 180-degree suppression line at best. To overcome this problem, it is possible to use separate graphic memories called CUDA to promote processing power as a concurrency element. In addition, the possibility of real-time use of the proposed model in portable devices has become more achievable and it can be called a real-time method.

Another challenge for the proposed model is to select the correct value of the support vector machine parameters. As a solution in the future, algorithms can be used in this field that take advantage of the possibility of extensive scanning of the problem space, in order to improve this feature and minimize the time required to combine the candidate parameters. In this case,

| Acronyms | Definition |
|----------|------------|
| ANN      | Artificial Neural Network |
| CAD      | Computer-aided Diagnosis |
| CAPS-NET | Capsule Network |
| CNNs     | convolutional neural networks systems |
| DBN      | Deep Belief Network |
| FN       | False Negative |
| FP       | False Positive |
| HOG      | Histogram of Oriented Gradients |
| LBP      | Local Binary Patterns |
| PCA      | Principal Component Analysis |
| RBF      | Radial Basis Function |
| SIFT     | Scale Invariant Features Transform |
| SMOTE    | Synthetic Minority Oversampling Technique |
| SVM      | Support Vector Machine |
| TP       | True Positive |
| TN       | True Negative |
the hope for the correct selection of key parameters is increased and we will probably witness an increase in the accuracy of the support vector machine. It is also possible to use the optimization algorithm introduced in studies [12–14, 26] with SVM to achieve better results.

The proposed methods could be applied for the detection of many other potential lesions, such as mass and polyp in other cancers. Also, the feature extraction method could be used to recognize faces, signatures, and fingerprints of individuals.

In the future, further investigations include evaluating it on more clinical data and promoting it in clinical practice with the help of radiologists and surgeons. As a solution in the future, noise removal methods based on low-pass filters can be used in the preprocessing phase to reduce the sensitivity of the model to noise. Next, key points of images can be isolated using extraction algorithms, such as Speed up Robust Features (SURF), and the number of texture changes can be calculated only in those areas. This not only reduces the time consumed but also increases both the accuracy of the model and its flexibility in the face of new samples. Also, in future works, we can use the combination of capsule network and CNN network to extract better features.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Agarwal D, Shriram KS, Subramanian N (2013) Automatic view classification of echocardiograms using histogram of oriented gradients. In: ISBI
2. Ahn HS, Kim HJ, Welch HG (2014) Korea’s thyroid-cancer “epidemic”—screening and overdiagnosis. N Engl J Med 371(19):1765–1767
3. Ali H, Sharif M, Yasmin M, Rehmani MH, Riaz F (2020) A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract. Artif Intell Rev 53.4: 2635–2707
4. Aswathi A, Antony A (2018) An intelligent system for thyroid disease classification and diagnosis. In: 2018 second international conference on inventive communication and computational technologies (ICICCT). 2018. IEEE
5. Azizi G, Keller JM, Mayo ML, Piper K, Puett D, Earp KM, Malchoff CD (2015) Thyroid nodules and shear wave elastography: a new tool in thyroid cancer detection. Ultrasound Med Biol 41(11):2855–2865
6. Buda M, Wildman-Tohbriner B, Hoang JK et al (2019) Management of Thyroid Nodules Seen on US images: deep learning may match performance of radiologists. Radiol 292(3):695–701
7. Chatterjee I (2021) Artificial intelligence and patentability: review and discussions. International Journal of Modern Research 1:15–21
8. Chi J, Walia E, Paul Babyn P et al (2017) Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. J Digit Imaging 30:477–486
9. Dash S, Das M, Mishra BK (2016) Implementation of an optimized classification model for prediction of hypothyroid disease risks. In: 2016 international conference on inventive computation technologies (ICICT), IEEE
10. Deepika M, Kalaiselvi K (2018) An empirical study on disease diagnosis using data mining techniques. In 2018 second international conference on inventive communication and computational technologies (ICICCT), IEEE
11. Devi GU, Anita EAM (2019) A novel semi-supervised learning algorithm for thyroid and ulcer classification in tongue image. Cluster Comput 22:11537–11549
12. Dhiman G, Kaur A (2019) STOA: a bio-inspired based optimization algorithm for industrial engineering problems. Eng Appl Artif Intell 82:148–174
13. Dhiman G, Kumar V (2017) Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications. Adv Eng Softw 114:48–70
14. Dhiman G, Kumar V (2018) Emperor penguin optimizer: a bio-inspired algorithm for engineering problems. Knowl-Based Syst 159:20–50
15. Douzas G, Baco F, Fonseca J, Khudinyan M (2017) Adaptive fuzzy sliding mode control algorithm simulation for 2-DOF articulated robot. Int J Wirel Mob Comput 13(4):306–313
16. Georgoulas G, Loutas T, Stylios CD, Kostopoulos V (2013) Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. Mech Syst Signal Process 41:1–2:510–525
17. Hao Y, Zuo W, Shi Z, Yue L, Xue S, He F (2018) Prognosis of Thyroid Disease Using MS-Apriori Improved Decision Tree. In: Liu W, Giunchiglia F, Yang B (eds) Knowledge Science, Engineering and Management. KSEM 2018. Lecture notes in computer science, vol 11061. Springer, Cham. https://doi.org/10.1007/978-3-319-99365-2_40
18. Ionita I, Ionita L (2016) Prediction of thyroid disease using data mining techniques. BRAIN 7(3):115–124
19. Ionita I, Ionita L (2016) Applying data mining techniques in healthcare. Stud Inform Control 25(3):385–394
20. Jiang L, Wang P, Wang J, et al. (2019) A deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck 41:6:1269–1279
21. Jiang L, Wang P, Wang J, et al. (2019) A deep convolutional neural network for the diagnosis of thyroid papillary cancer in ultrasound images. Sci Rep 8(1):1
22. Kaur S, Awasthi LK, Sangal AL, Dhiman G (2020) Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization. Eng Appl Artif Intell 90:103541
23. Kavitha S, Thyagarajan K K, “Features based mammogram image classification using weighted feature support vector machine,” Communications in Computer and Information Science Vol. 270, 2012, pp. 320-329. ISBN: 978-3-642-29215-6 (print) 978-3-642-29216-3 (online), ISSN: 1865-0929 (print) 1865-0937 (online), Publisher: Springer Berlin Heidelberg https://doi.org/10.1007/978-3-642-29216-3_35, 2012.
24. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, Jung I, Kim EK, Moon HJ, Park VY, Lee E, Kwak JY (2019) A deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck 2019:1–7. https://doi.org/10.1002/hed.25415
25. Kulkarni NN, Bairagi VK (2017) Extracting salient features for EEG-based diagnosis of Alzheimer’s disease using support vector machine classifier. IETE J Res 63(1):11–22
26. Kaur S, Awasthi LK, Sangal AL, Dhiman G (2020) Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization. Eng Appl Artif Intell 90:103541
27. Kavitha S, Thyagarajan K K, "Features based mammogram image classification using weighted feature support vector machine," Communications in Computer and Information Science Vol. 270, 2012, pp. 320-329. ISBN: 978-3-642-29215-6 (print) 978-3-642-29216-3 (online), ISSN: 1865-0929 (print) 1865-0937 (online), Publisher: Springer Berlin Heidelberg https://doi.org/10.1007/978-3-642-29216-3_35, 2012.
28. Ko SY, Lee JH, Yoon JH, Na H, Hong E, Han K, Jung I, Kim EK, Moon HJ, Park VY, Lee E, Kwak JY (2019) A deep convolutional neural network for the diagnosis of thyroid nodules on ultrasound. Head Neck 2019:1–7. https://doi.org/10.1002/hed.25415
29. Kavitha S, Thyagarajan K K, "Features based mammogram image classification using weighted feature support vector machine," Communications in Computer and Information Science Vol. 270, 2012, pp. 320-329. ISBN: 978-3-642-29215-6 (print) 978-3-642-29216-3 (online), ISSN: 1865-0929 (print) 1865-0937 (online), Publisher: Springer Berlin Heidelberg https://doi.org/10.1007/978-3-642-29216-3_35, 2012.
30. Lincango-Naranjo E, Solis-Pazmino P, El Kawkgi O, Salazar-Vega J, Garcia C, Ledesma T, … Brito JP (2021) Triggers of thyroid cancer diagnosis: a systematic review and meta-analysis. BMC Cancer 1–16
31. Liu T, Xie Sh, Yu J, Niu L, Sun W (2017) Classification of thyroid nodules in ultrasound images using deep model-based transfer learning and hybrid features. In: 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)
32. Liu R, Li H, Liang F, Yao L, Liu J, Li M, Cao L, Song B (2019) Diagnostic accuracy of different computer-aided diagnostic systems for malignant and benign thyroid nodules classification in ultrasound images. Medicine 98(e16227):29
33. Malignancy Prediction from Whole Slide Cytopathology Images (n.d.)
34. Moussa O, Khachnaoui H, Guetari R, Khilifa N (2019) Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network. Int J Imaging Syst Technol 2019:1–11. https://doi.org/10.1002/ima
41. Nallamuth R, Palanichamy J (2015) Optimized construction of various classification models for the diagnosis of thyroid problems in human beings. Kuwait J Sci 42(2):189–205
42. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 4(7):971–987
43. Pande S, Chetty M (2018) Analysis of capsule network (Capsnet) architectures and applications. J Adv Res 10(10):2765–2771
44. Pandey S, Kumar Gour D, Sharma V (2015) Comparative study on the classification of thyroid diseases. Int J Eng Technol 28(9):457–460
45. Pedraza L, Vargas C, Narváez F, Durán O, Muñoz E, Romero E (2015) An open access thyroid ultrasound image database
46. Peng SH, Kim DH, Lee SL, Lim MK (2010) Texture feature extraction based on a uniformity estimation method for local brightness and structure in chest CT images. Comput Biol Med 40(11–12):931–942
47. Qi J, Jiang G, Li G, Sun Y, Tao B (2020) Surface EMG hand gesture recognition system based on PCA and GRNN. Neural Comput Appl 32(10):6343–6351
48. Qin P, Wu K, Hu Y, Zeng J, Chai X (2019) Diagnosis of benign and malignant thyroid nodules using combined conventional ultrasound and ultrasound elasticity imaging. IEEE J Biomed Health Inform 24(4):1028–1036
49. Rasitha Banu G (2016) Predicting thyroid disease using linear discriminant analysis (LDA) data mining technique. Commun Appl Anal 4(1):4–6
50. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. CoRR, abs/1710.09829
51. Salimian M, Rezai A, Hamidpour SSF, Khajeh-Khaliili F (2019) "Effective features in thermal images for breast Cancer detection," 2nd National Conference on new Technologies in Electrical and Computer Engineering
52. Sarah M, Erfani SR, Karunasekera S, Leckie C (2016) High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. Pattern Recogn. https://doi.org/10.1016/j.patcog.2016.03.028
53. Shi Z, Hao H, Minghua Zhao M et al (2018) A deep CNN based transfer learning method for false-negative reduction. Multimed Tools Appl 78:1017–1033
54. Vaishnav PK, Sharma S, Sharma P (2021) Analytical Review Analysis for Screening COVID-19. International Journal of Modern Research, 1, 22–29. Spotted hyena optimizer: A novel bio-inspired based metaheuristic technique for engineering applications
55. Wang S, Liu J-B, Zhu Z, Eisenbrey J (2019) Advanced Ultrasound in Diagnosis and Therapy. Artif Intell Ultrasound Imaging: Current Res Appl 03:053–061
56. Xing H, Huang Y, Duan Q et al (2018) Abnormal event detection in crowded scenes using histogram of oriented contextual gradient descriptor. EURASIP J Adv Signal Process 2018:1–15
57. Yin Q, Li G, Zhu J (2017) Research on the method of step feature extraction for EOD robot based on 2D laser radar. Discrete Contin Dyn Syst Ser S (DCDS-S) 2018.1:1415–1421
58. Zakeri FS et al (2012) Classification of benign and malignant breast masses based on shape and texture features in sonography images. JMS 36(3):1621–1627
59. Zarei M, Rezai A, Hamidpour SSF (2021) "Breast cancer detection using RSFS-based feature selection algorithms in thermal images," Biomed Eng: Appl, Basis C
60. Zarei M, Rezai A, Hamidpour SSF (2021) Breast cancer segmentation based on modified Gaussian mean shift algorithm for infrared thermal images. Comput Methods Biomech Biomed Eng: Imaging Visual. https://doi.org/10.1080/21681163.2021.1897884
61. Zhao T, Liu Y, Huo G, Zhu X (2019) A deep learning iris recognition method based on capsule network architecture. IEEE Access 7:49691–49701

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