RESEARCH ARTICLE

Efficient classification of different medical image multimodalities based on simple CNN architecture and augmentation algorithms

Walid El-Shafai[1,2] · **Amira A. Mahmoud**[2] · **Anas M. Ali**[2,3] · **El-Sayed M. El-Rabaie**[2] · **Taha E. Taha**[2] · **Adel S. El-Fishawy**[2] · **Osama Zahran**[2] · **Fathi E. Abd El-Samie**[2]

Abstract Convolutional neural networks (CNN) are the best deep learning architecture to perform tumors classification for different imaging modalities: Us, X-ray, CT, and MRI. The scarcity of medical images and scarcity of resources are the contemporary problem for achieving successful classification. Therefore, it is preferable to use a simple network that does not require training and implementation resources than to use complex or pre-trained CNN models. Simple networks make it easy to use in clinical diagnosis and on mobile platforms. In this paper, a proposed CNN architecture for medical image multimodalities classification is presented. This proposed network is simple and is directly trained by medical images which is better than using pre-trained deep learning networks. Firstly, the data augmentation process is applied to avoid data shortage, and then, the proposed CNN is trained using the resulted augmented data. Simulation results demonstrate the efficiency of the proposed CNN architecture for efficient classification. The proposed model is trained on medical Us, X-ray, CT, and MRI datasets from scratch, and it can achieve 92.7%, 91.1%, 100%, 100% accuracies for these datasets, respectively.

Introduction

Medical scans are great tools that help the specialists to identify the different abnormalities in the body organs. These scans can detect, diagnose, and treat different diseases. The main used medical scans are ultrasonic (Us), magnetic resonance imaging (MRI), computed tomography (CT), and X-ray [1, 2], which are formed of malignant and benign tumors which have become a major element of healthcare. Us imaging as a tool for medical diagnosis is excessively utilized in clinical practice, and in some situations is standard procedure because it is usually a painless scan, available, less expensive and uses non-ionizing radiation. X-ray is the most used diagnostic imaging test and is widely available. They use radiation to form the X-ray image of the body and bones which in some situations is harmful and precautions must be taken. CT scan combined X-ray scans with different angles to give a cross section image of the inside object under scan easily and the subject contrast

✉ Walid El-Shafai
walid.elshafai@el-eng.menofia.edu.eg

Amira A. Mahmoud
ameragreeda@yahoo.com

Anas M. Ali
anas5616242@gmail.com

El-Sayed M. El-Rabaie
elsayedelrabaie@gmail.com

Taha E. Taha
taha117@hotmail.com

Adel S. El-Fishawy
aelfishawy@hotmail.com

Osama Zahran
osama_zahran@hotmail.com

Fathi E. Abd El-Samie
fathi_sayed@yahoo.com

1 Security Engineering Lab, Computer Science Department, Prince Sultan University, Riyadh 11586, Saudi Arabia

2 Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

3 Alexandria Higher Institute of Engineering & Technology (AIET), Alexandria, Egypt

is clearly visible. MRI scan uses a powerful magnetic field, magnetic field gradients, and radio waves to form images of different organs. It can also create more visible image details compared to X-ray and CT scans. As well as, it uses non-damaging radiation.

The medical image diagnosis process is performed in two steps. At first, the most significant features are identified and extracted. Then, the most significant features are used in building the diagnostic model. The previous concept falls within medicine so that doctors use their experience to extract the most significant features and then determine the type of disease, which makes the diagnosis process a waste of time and there is a small percentage of human error. At the present time, CNN has proven a remarkable superiority in diagnostic tasks, as it can make a diagnosis that doctors are unable to do [3–10]. The authors of [11] proposed a CNN model based on k-mean to extract significant features and then applied a multi-class SVM model to diagnose the mammography dataset. In [12], the authors suggested an automatic computer-aided diagnosis model diagnose Us breast images, and the segmentation model was used to show the disease, and then, the classification models were implemented. The proposed approach also achieved 85.42% classification accuracy using CNN and between 80 and 77% classification accuracy in machine learning models. Yi Wang et al. [13] proposed a multi-view CNN diagnostic model on the Us breast images dataset divided into 135 malignant and 181 benign breast lesions. In [14], a CNN multi-organ CAD model is proposed to classify breast and thyroid in Us images.

Rajeshwari S. Patil et al. [15] proposed a hybrid CNN and recurrent NN to detect lesions in mammogram images. Their basic phases are pre-processing followed by segmentation, feature extraction, and detection. Hua Li et al. [16] proposed a classification of benign and malignant mammogram images based on an improved DenseNet model for effective and accurate classification. The model was based on three stages: The first one is preprocessing and normalization. The second is replacing the first convolutional layer of their model with the Inception structure. Finally, the datasets are applied to pre-trained models and the DenseNet model. Umar Albalawi et al. [17] proposed a classification mammogram model based on CNN. They used the wiener filter to remove the noise and used the K-means clustering technique to segment the image followed by the CNN classifier. Shen, L. et al. [18] proposed a DL model in order to classify mammogram lesions. They compared their model with the previous models, and it acquired an AUC value of 0.91 on the CBIS-DDSM database and 0.95 on the FFDM database. Yuezhong Zhang et al. [19] proposed a CNN classifier for CT images based on the CDBN model. They used SVM as the feature classifier to enhance feature transfer and reuse to enrich the features. Applying the Adam optimizer algorithm,

they get both good accuracy and speed. Huseyin Polat and Homay Danaei Mehr [20] proposed a hybrid CNN lung classifier. They used the SoftMax radial basis function-based SVM to study their model performance. They also compared their model with AlexNet and GoogleNet. They acquired 91.81%, 88.53%, and 91.91% for accuracy rate, sensitivity, and precision, respectively. Li et al. [21] proposed a CNN classifier based on augmentation for a hyperspectral image. They proposed an augmentation technique to make the training samples number increased. Their method benefits deep CNN and extracts PBP features. They also used the decision fusion classifier.

Agrawal et al. [22] proposed a CNN model to classify gastrointestinal system features using a few samples in the training stage and transfer learning models. They also developed a metric to study model performance. This metric carried a correlation of 87% in the validation stage. Keita Saito et al. [23] proposed a CNN classifier for heart diseases. It was trained by heart disease images from scratch. Samir S. Yadav et al. [24] proposed a CNN classifier to diagnose a disease from chest X-ray images. It was shown that using augmentation techniques as well as transfer learning is very effective and leads to improved performance. Feng-Ping An et al. [25] proposed a CNN classifier for breast mass and brain tumor tissue. Their method constructed different CNN models suitable for the medical images' features using the adaptive sliding window fusion mechanism. The biggest problem with classifying the medical images using neural networks is the used database size. In addition, pre-processing is required; however, it is known that the pre-processing and the feature extraction in CNNs do not have to be performed. Table 1 shows an overview of the recent work using deep learning techniques for medical image classification.

Knowledge transfer learning has been used in many computer vision tasks. But there is a difference between natural images and medical images, which presents a difficult problem in building an effective CNN model for medical image diagnosis that outperforms other intelligent systems. In this work, a CNN architecture for diagnosing benign and malignant tumors is proposed. Also, a simple network was built that does not require many resources to implement the proposed network on mobile platforms. For effective evaluation of the proposed network, four different data sets were used: Us, X-ray, CT, and MRI. Detailed comparisons were also made with the latest transfer learning models, including the different measures of accuracy with the confusion matrix.

This paper aims firstly to classify the two types of tumors from four different database modalities with a CNN architecture between benign and malignant. Two methods are tested through experiments, namely transfer learning on three CNN models: VGG16, VGG19, and AlexNet. Also, a proposed training network is built from scratch. The more complex models were compared with the simple ones in terms of

Table 1 Overview of the recent work using deep learning techniques for medical image classification

Study/ Year	Image modality	Method	Classifier	Metric	Category
Hussein, S.et al. [3] 2019	CT lung and MRI and pancreas	Graph Regularized Sparse Multi-task learning	3D CNN and Transfer Learning	Acc 91.26% MSD 0.459	Categorization of lung nodules and pancreatic cysts (IPMN)
Shaikh, M. et al. [4] 2019	MRI brain/ Retina	(RAM) model	CNN	Acc 97% Acc 93.37%	Meningioma, Pituitary, and Glioma /Grading of diabetic macular edema
Afshar, P. et al. [10] 2019	MRI brain	Caps-Net Architecture	CNN	Acc 90.89%	Meningioma, Pituitary, and Glioma
Kaur, P. et al. [11] 2019	Mammography	multi-class SVM	CNN	Acc 96.9%,	M, B, N
Yi-Wei Chang et al. [12] 2020	Us breast	BI-RADS	SVM, RF, CNN	Acc 80.00% Acc 77.78% Acc 85.42%	M, B
Yi Wang et al. [13] 2020	Us breast	Inception-v3 architecture	CNN	AUC 94.68% Sen 88.60% Sp 87.60%	M, B
Xiaowen Liang et al. [14] 2020	Us breast and thyroid	Multiorgan CAD system	CNN	Sen 84.9%, Sp 69.0%, PPV 62.5%, NPV 88.2%, Acc 75.0%, AUC 0.769	Fibroadenoma, Invasive carcinoma/Nodular goiter Papillary carcinoma
Rajeshwari S. Patil et al. [15] 2020	Mammography	GLCM and GRLM	Hybrid CNN and RNN	Acc 90.5%, Sen 92.4%, Sp 89.8%, Pr 78.2%	M, B, N
Hua Li et al. [16] 2019	Mammography	DenseNet-II	CNN	Acc 94.55%	M, B
Umar Albalawi et al. [17] 2020	Mammography	Weiner filter k-means clustering	CNN	Acc97.143% Sen96.522% Sp 98.883%	M, B
Shen, L. et al. [18] 2019	Mammography	DL	CNN	AUC 0.91 Sen 86.1% Sp 80.1%	M, B
Yuezhong Zhang et al. [19] 2020	CT lung, liver, brain	SVM with RBF	CNN	Acc 93%	N, A
Huseyin Polat and Homay Danaei Mehr [20] 2019	CT lung	Hybrid 3D-CNN with SVM	CNN	Acc 91.81%, Sen 88.53% Pr 91.91%	N, A
W. Li et al. [21] 2018	Hyperspectral	PBP	CNN	Acc 93.14%	Assign a class label to each pixel
T. Agrawal et al. [22] 2019	GI	Transfer learning	CNN	87% correlation	GI landmarks
Keita Saito et al. [23] 2019	MRI heart	DL	CNN	Acc 96.17%	Heart diseases
Samir S. Yadav et al. [24] 2019	chest X-ray	(ORB and SVM) Transfer learning (VGG16-INV3) Caps Net	CNN	Acc 77.6% Acc 92.3% Acc 87.5% Acc 85.6%	Normal and Pneumonia
Feng-Ping An et al. [25] 2019	Mammography/ MRI brain/	Weight Initialization-Sliding Window Fusion	CNN	Acc 98.2% Acc 95.71%	M, B

C carcinoma, *N* normal, *B* benign, *M* malignant, *GLCM* gray-level co-occurrence, *MSD* mean score difference, *RAM* recurrent attention mechanism, *GAN* generative adversarial networks, *SVM* support vector machine, *RF* random forest, *CNN* convolution neural network, *Sen* sensitivity, *Sp* specificity, *Pr* precision, *PPV* positive predictive value, *NPV* negative predictive value, *Acc* accuracy, *AUC* area under the curve, *BI-RADS* Breast Imaging Reporting and Data System, *GI* gastrointestinal, *PBP* pixel-block pair

efficiency and training time. In addition to using a general model that works on various data, its implementation is realistic, easy, and guaranteed. The main contributions of the present study are as follows:

- Development of three AlexNet, VGG-16, and VGG-19 transfer learning models for classifying multitype medical images.
- Employing several pre-trained CNN models with fine-tuning and applying them to four different datasets, namely MRI, X-ray, Us, and CT with or without using data augmentation technique.
- Develop a proposed CNN architecture built from scratch that is characterized by low complexity and low training time.
- Apply the proposed CNN architecture consisting of 3×3 kernels and 1 stride to all convolutional layers, unlike other more complex models. The proposed model also achieved higher diagnostic accuracy compared to the state-of-the-art models.

The rest of this paper is organized as follows. Sect. "Convolutional neural networks (CNN)" gives short notes about CNN. Sect. "Material and methods" describes the material and methods used in our work. Sect. "Deep features extraction and classification via transfer learning" gives short notes on the transfer learning methodology. Sect. "Proposed CNN Model" illustrates the proposed model architecture. Sect. "Experimental results and discussion" shows the experimental results and discussion. Sect. "Conclusions and future work" provides the conclusions followed by references.

Convolutional neural networks (CNN)

In the past few years, he showed that interest in the medical field is a priority for human beings. Therefore, a lot of research has been developed in the medical field. Most of the new research focused on the use of artificial intelligence in many medical branches because of its superiority over traditional techniques. The CNN architecture consists of an input layer, a convolution layer, a classification layer, and an output layer [4, 5, 26–29]. The input layer consists of the dimensions of the input images. The convolution layer is the main layer that performs two operations of feature extraction and feature selection. It also depends on the trainable filters, and each filter consists of a number of weights that adapt to the images entered during training. The convolution layer also contains padding, which is adding rows and columns of zeros to the borders of the entered images so that the image dimensions do not change. In addition, the number of convolutional layers reflects the complexity of the network. At the end of each

convolution layer is a sub-layer called an activation layer. The activation layer is responsible for selecting the best values or weights for the filters used in the convolutional layers. The choice of values varies by choosing the type of activation layer, and the most used is the ReLU layer, which chooses the values of weights between zero and infinity. The ReLU layer carries out a threshold process for each one of the inputs, and the values that are smaller than zero are replaced by zero.

$$f(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases} \tag{1}$$

As mentioned earlier, convolutional layers select the best features, yet features are reduced relatively slowly. Therefore, pooling layers such as max-pooling, mean-pooling, and average-pooling are usually used. Pooling layers reduce the number of features without training, so they do not count in memory. Classification layers are neural networks (NNs) and are called fully connected layers (FC). FC layers combine all the features you learned in the previous layers to identify patterns and then classify the images. The output layer is based on the SoftMax activation function. In addition, the output layer calculates the cross-entropy loss.

$$y_r(x) = \frac{exp(a_r(x))j}{\sum_{j=1}^{k} exp(a_j(x))} \tag{2}$$

where $0 \leq y_r \leq 1$ and $\sum_{j=1}^{k} y_j$.
The SoftMax function is:

$$P(c_r|x,\theta) = \frac{P(x, \theta|c_r)P(c_r)}{\sum_{j=1}^{k} P(x, \theta|c_j)P(c_j)} = \frac{exp(a_r(x,\theta))}{\sum_{j=1}^{k} \exp(a_j(x,\theta))} \tag{3}$$

where $0 \leq P(c_r|x,\theta) \leq 1$ and $\sum_{j=1}^{k} P(c_j|x,\theta) = 1$ d.

Moreover, $a_r = \ln(P(x, \theta|c_r)P(c_r))$, $P(x, \theta|c_r)$ is the sample conditional probability of the given class r, and $P(c_r)$ is the probability of the class. The SoftMax function output values are taken and assigned to one class of the two classes using the cross-entropy function [19]:

$$loss = -\sum_{i=1}^{N} \sum_{j=1}^{K} t_{ij} \ln y_{ij} \tag{4}$$

where N is the number of samples, t_{ij} is the indicator that the ith sample belongs to the jth class, and y_{ij} is the output for sample i for class j, which, in this case, is the value from the SoftMax function. That is, it is the probability that the network associates the ith input with class j.

Material and methods

In this paper, a simple CNN structure is proposed to classify tumors in multimedia medical images. In this section, all the proposed data sets in this research are described. Table 1 also summarizes the details of the databases.

Dataset

Simulation results are conducted on four different examples of medical images (Us breast images [30], X-ray (mammogram) images [31], CT chest images [32], and MRI brain images [33]), and each dataset contains different numbers of benign and malignant images. Each data set is divided into 70% training set and 30% test set. So that the training set is used to train the proposed model. The test set is also used to verify the model training results. A sample from all datasets is shown in Fig. 1.

Image processing data augmentation

One of the most important problems facing any training process is the lack of training data, which is the key to achieving the best classification accuracy. Therefore, a data augmentation technique was used, where the images of the model are entered in each epoch differently. The possible five augmentation techniques employed here are resizing and rotation followed by adding speckle and Gaussian noise, blur, sharpening, and filtering [4–21, 34]. Table 2 shows the dataset's image numbers and their specifications before and after data augmentation.

Deep features extraction and classification via transfer learning

The main function of using pre-trained networks is to transfer the values of weights, which is called transfer learning. Most of the pre-trained networks are trained on the ImageNet dataset, which contains various types of images. Then, the trained weights are transferred to be applied to smaller data sets to take advantage of the previously trained weights. Then, the last layers of the previously trained model, FC, are changed only to adjust the model for the task of classifying tumors [35–40]. The learning transfer procedure is shown in Fig. 2.

The transfer learning models used in this research are VGG16-19 and AlexNet. Transfer learning networks are also applied to the proposed datasets with and without data augmentation, and the results are compared. These networks are explained as follows:

Fig. 1 Samples of Us, X-ray, CT, and MRI tumors images: **a** Benign tumor images; **b** Malignant tumor images
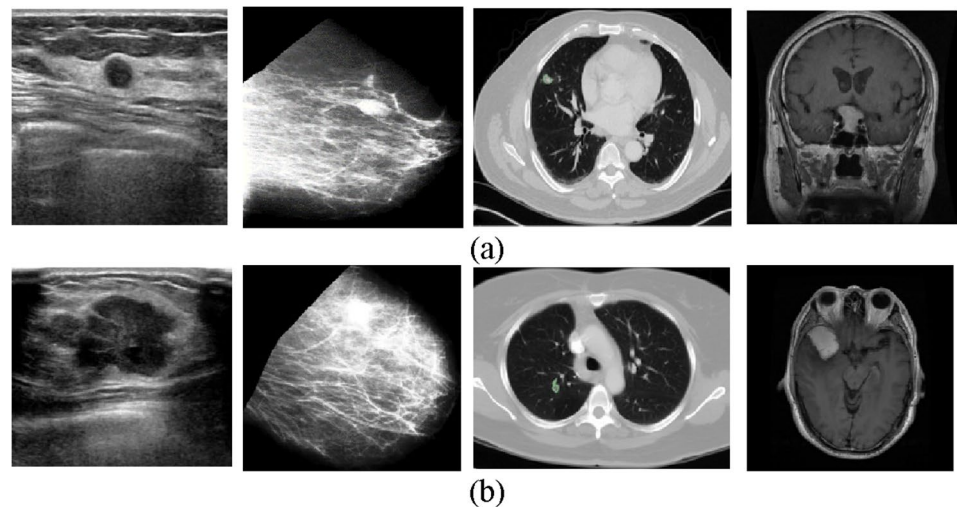

(a)

(b)

Table 2 Medical image datasets and their specifications before and after data augmentation

Datasets	Us		X-ray		CT		MRI	
	Before	After	Before	After	Before	After	Before	After
Number of Images	130	1056	100	1200	70	563	160	1160
Images Per Class benign/malignant	40/90	418/638	50/50	600/600	35/35	318/245	80/80	580/580
Images Size	227×227							
Images Type	JPG		BMP		DCM		BMP	

VGG-16 architecture

VGG16 [35, 36, 39, 40] is a primitive CNN consisting of 13 convolutional layers and three FC layers interspersed with five max-pooling layers, ending with a soft-max output layer. The VGG16 network contains about 138 million parameters, which is a very large number, but it guarantees a stable and relatively high classification accuracy. The VGG16 network was used to win the ImageNet competition in 2014. Figure 3 contains all the details of the VGG16 network.

VGG-19 architecture

The VGG19 [35, 37–45] is similar to the VGG16 network in layer arrangement but differs in the increase in the number of convolutional layers which is 16 layers. Figure 4 contains all details of the VGG19 network. The VGG19 network contains about 143 million parameters.

AlexNet architecture

The AlexNet architecture [35, 37–40] consists of five bypass layers and three FC layers. AlexNet also contains about 60 million parameters. Figure 5 contains all AlexNet details.

AlexNet is significant because it is the first neural network to win the 2012 ImageNet competition. AlexNet is also the first to use ReLU instead of the sigmoid or hyperbolic tangent function. AlexNet also provided a solution to the overfitting problem by applying a dropout between FC layers.
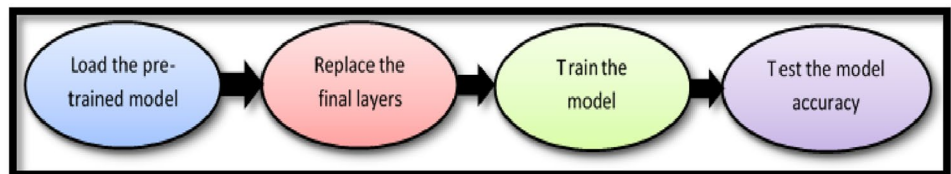

Fig. 2 The procedure of transfer learning (TL)

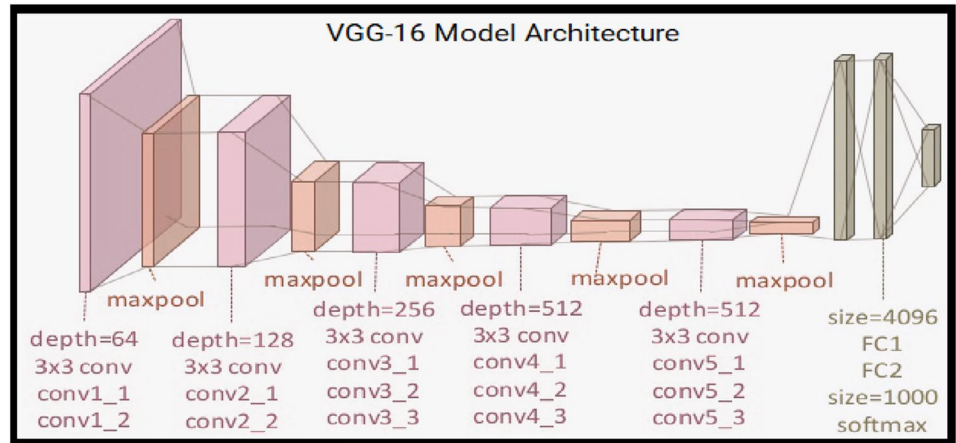**Fig. 3** VGG-16 architecture
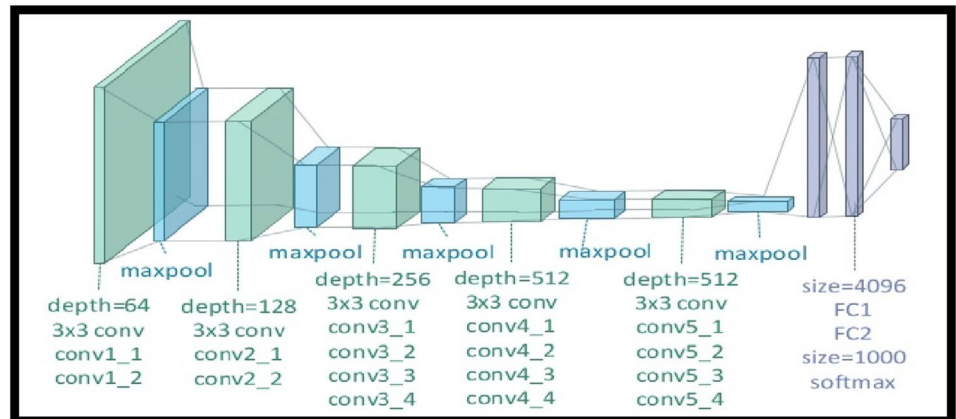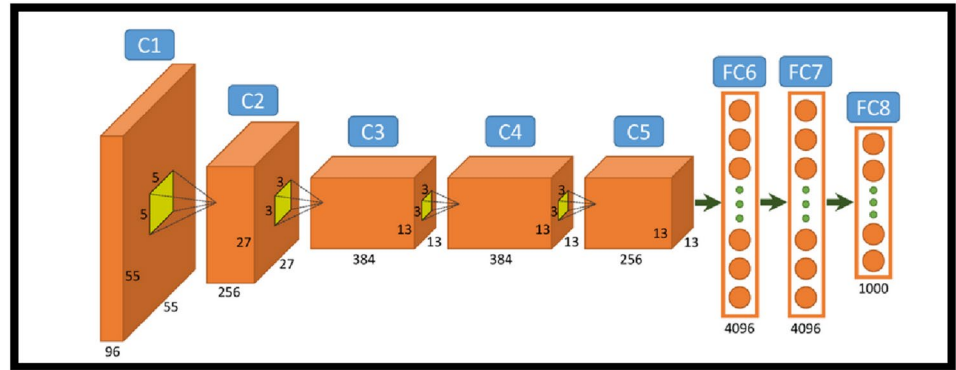

Fig. 4 VGG-19 architecture

Fig. 5 AlexNet architecture

Proposed CNN model

In the proposed model, we intend to use a smaller number of parameters than the rest of the pre-trained models that were previously explained. It is known that if the depth of a neural network increases, the accuracy, and performance of the neural network increase. However, memory and GPU consumption increase, and sometimes neural network performance does not improve.

As shown in Fig. 6, our proposed model consists of three convolutional layers and one FC layer interspersed with three batch normalization layers and two max-pooling layers. You can know the details of each layer in Fig. 6. The major advantage of the proposed model is the use of batch normalization layers to speed up the training process while reducing the number of parameters.

The training option and the hyperparameters can be specified as follows: We will employ the stochastic gradient descent with momentum (SGDM) optimizer with 0.9 Momentum while training the VGG-16, VGG-19, AlexNet, and the proposed model. A relatively average learning rate of 10^{-5} was used, with 10 epochs during training. The number of iterations is one step that was taken in the gradient descent algorithm toward minimizing the loss function using a mini-batch based on the dataset images number, i.e., 740, 740, 440, and 700 for Us, X-ray, CT, and MRI datasets, respectively. The training data are shuffled before each training epoch. The mini-batch size used here for each training iteration is 10. The L2 regularization is the contribution of the gradient step from the previous iteration to the current iteration of the training, specified as a scalar value from 0 to 1.

Experimental results and discussion

In this section, the experimental results of the proposed CNN model, as well as the pre-trained models, are presented for the different medical image datasets classification. The effect and benefits of using augmentation techniques are discussed.

Without data augmentation

The validation accuracies, validation loss, and training time, for each dataset without data augmentation using VGG-16, VGG19, AlexNet models, and the proposed model are shown in Table 3. It can be observed that the accuracies when applying the VGG16 model are 52.6%, 50%, 100%, and 100% on Us, X-ray, CT, and MRI datasets, respectively, and the accuracies of applying the VGG19 model are 68.4%, 53.3%, 100% and 100% on Us, X-ray, CT and MRI datasets, respectively, and the accuracies of applying AlexNet model are 89.47%, 60.00%, 100% and 100% on Us, X-Ray, CT and MRI datasets, respectively. When applying the proposed model to Us, X-ray, CT, and MRI datasets, the accuracies are 75%, 63.3%, 100%, and 100%, respectively. Also, it can achieve the lowest validation loss.

The results for the Us and X-ray datasets are not convincing enough due to the low image quality of the Us and X-ray;
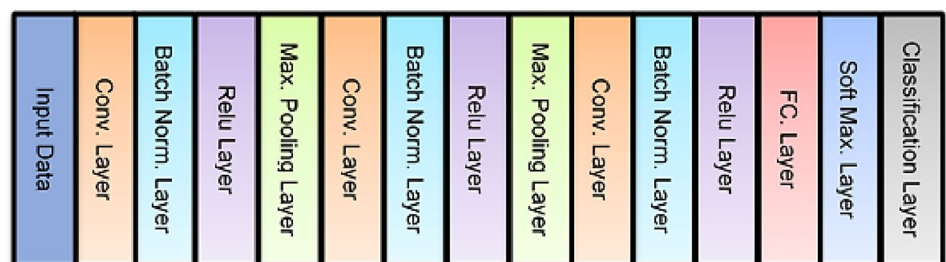
Fig. 6 The proposed CNN model

Table 3 The accuracy (Acc.) and loss of the pre-trained models and the proposed model for Us, X-ray, CT, MRI datasets without data augmentation

Dataset	VGG16		VGG19		AlexNet		Proposed	
	Acc	CPU (min)	Acc	CPU (min)	Acc	CPU (min)	Acc	CPU (min)
Us	52.6%	27:32	68.4%	32:52	83.47%	8:43	75%	2:50
X-Ray	50%	36:41	53.3%	44:01	60.00%	9:11	63.3%	2:46
CT	100%	24:53	100%	36:45	100%	7:31	100%	2:09
MRI	100%	33:34	100%	42:20	100%	11:34	100%	3:20

also, the properties of the two classes are very similar and the number of images in each dataset is small. On the other hand, the four models perform very well with CT and MRI datasets due to the high variance between the two classes and the good image quality. The CPU time illustrates the complexity of the models. As can be shown in Table 3, the proposed model can obtain the lowest system complexity due to its simple architecture.

With data augmentation

In this section, the effect of using data augmentation on transfer learning models and the proposed model are discussed. As a result, evaluation results are more compelling for use in real-world applications. When applying the VGG16 model, the accuracy for the Us dataset is increased from 52.6% to 58.3% and for the X-ray dataset from 50% to 54.7%. Similarly, the accuracy for Us and X-ray for the VGG19 model is increased from 68.4% to 75%, and 53.3% to 63.8%, respectively.

For the AlexNet, the accuracies increased from 83.47% to 89.9% for the Us database and from 60% to 70.6%. The overall accuracy obtained for the proposed model is 92.7% and 91.1% for Us and X-ray datasets which are much greater than the accuracies achieved without data augmentation. Thus, the experiments illustrate that the data augmentation techniques have an apparent effect on classification accuracy. In Figs. 7 and 8, the training and validation curves and confusion matrices of the proposed CNN model are shown with data augmentation.

There is a difference between medical images and natural images, which is the gradation of colors. From this point of view, the results of training the pre-trained models may not be the most suitable choice for classifying medical images (Table 4).

Initially, the proposed model achieved relatively low training accuracy for the Us and X-ray datasets due to the low image quality and the small number of images.

However, after using the data augmentation technique, the accuracy began to gradually increase.

From the simulation results, this paper presents a comprehensive comparison between the use of transfer learning models, which are VGG16, VGG19, and AlexNet, and between the use of a simple proposed model with a few parameters. The results showed that the proposed model outperformed the previously trained models in classifying various medical images, but after using the data augmentation technique. The proposed model also helps the radiologist make an accurate decision to classify different medical images.

Conclusions and future work

In the scientific community, it has become necessary to use static models that work on different data sets. So in this paper, a relatively simple model with few parameters based on neural networks is proposed. Where the proposed model is used to classify various data sets such as Us, X-ray, CT, and MRI. The proposed model also achieved a classification accuracy of 92.7%, 91.1%, 100%, and 100% for the datasets Us, X-ray, CT, and MRI, respectively. A comparison was made between the proposed simple model with transfer learning models such as VGG16, VGG19, and AlexNet. It is also possible to use the proposed model on the simplest available resources due to the small number of layers with a small number of parameters. Medical diagnosis in developing countries is one of the easiest and most important factors for epidemic prevention. Therefore, our proposed model can be used on mobile platforms because of the small neural network used with high efficiency on different data sets. The proposed model can also be used in real-time discovery applications. In future, the performance of the proposed model will be tested on recent data sets with better improvements in accuracy and complexity.
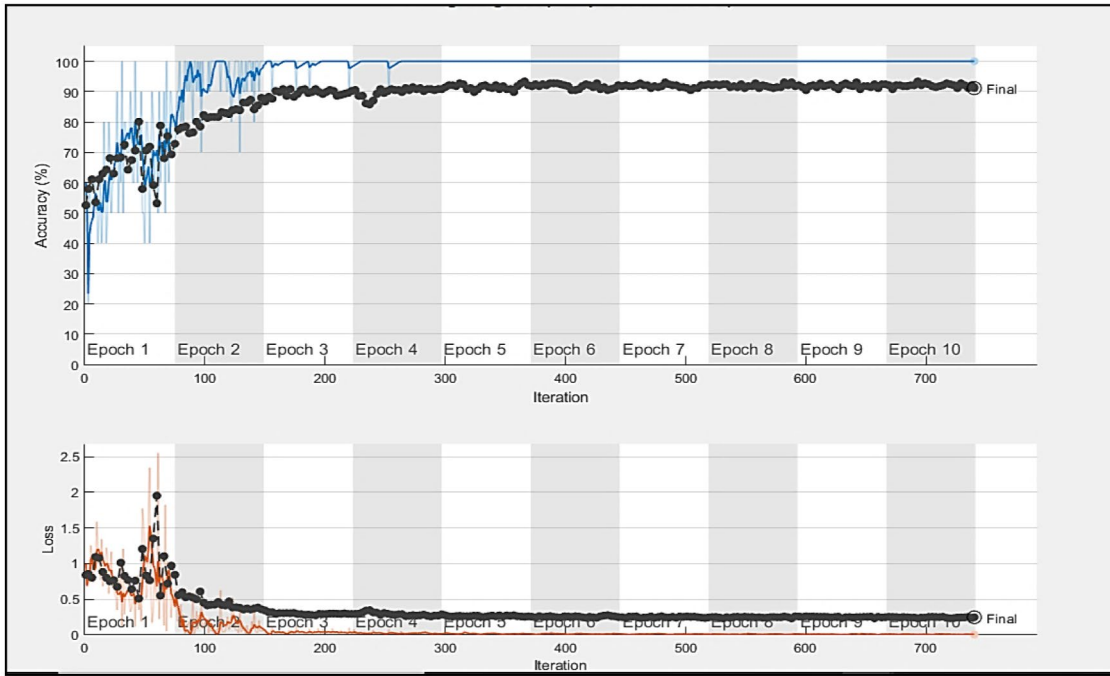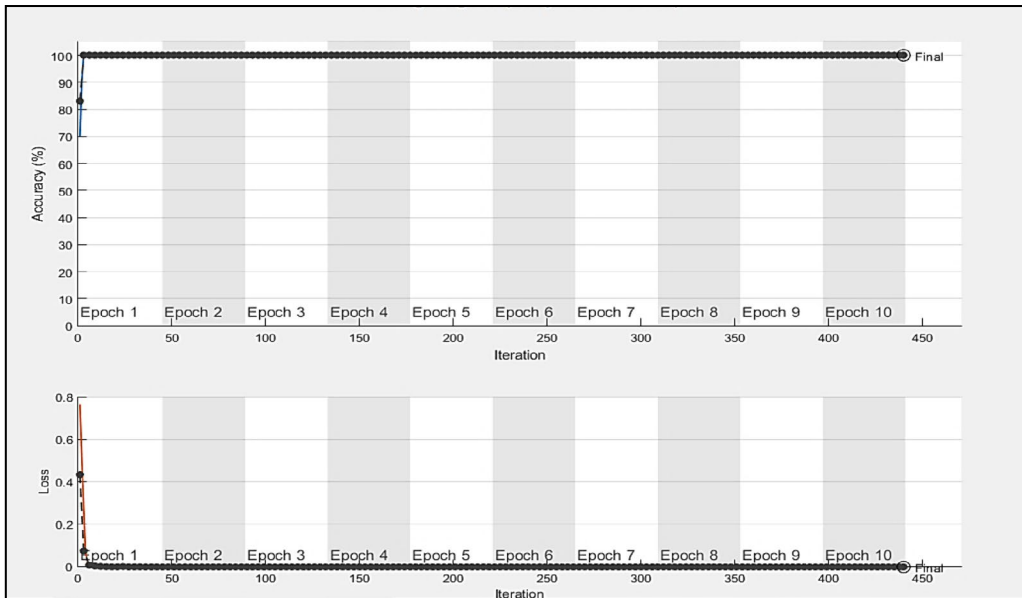
(a) Training and validation curves for accuracy and loss using the proposed model for the Us dataset..
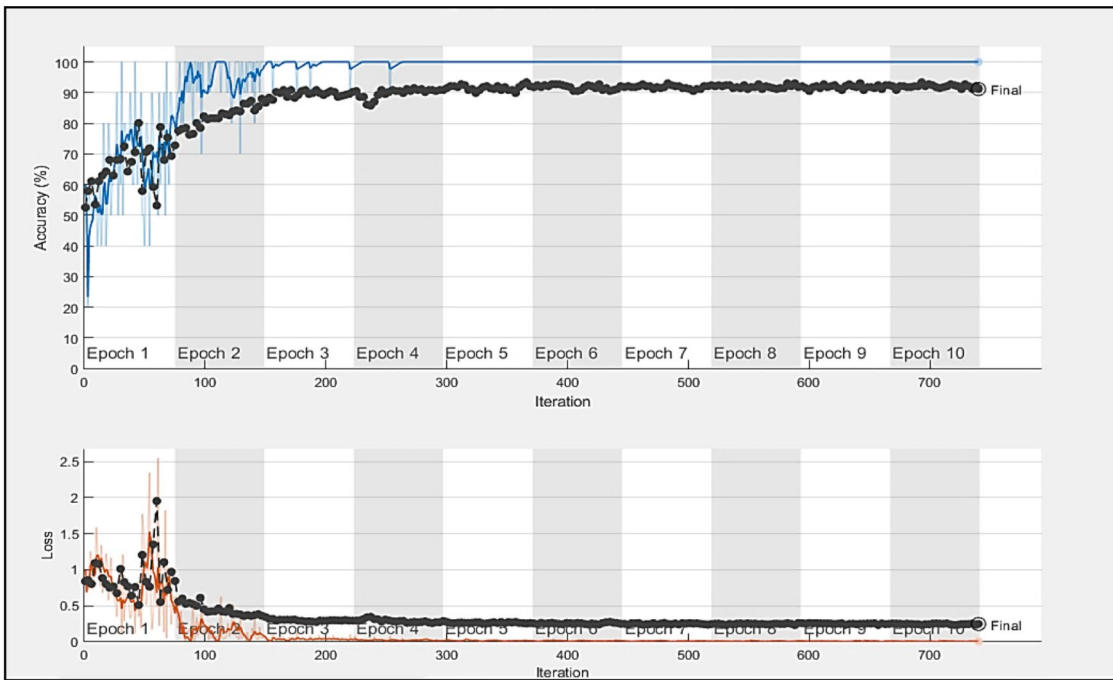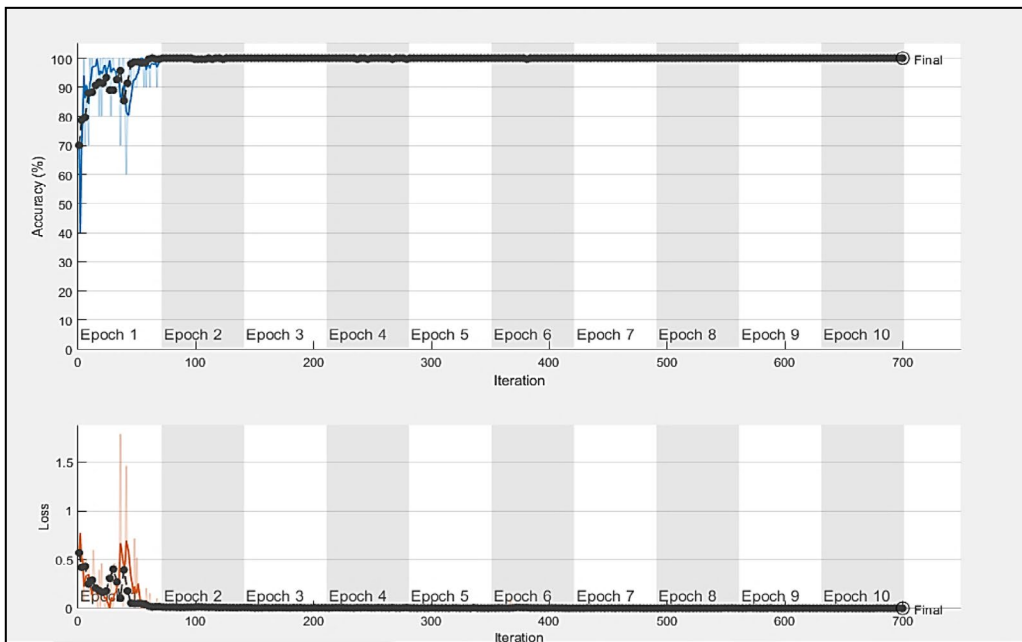

(b) Training and validation curves for accuracy and loss using the proposed model for the X-Ray dataset.

Fig. 7 Training and validation curves for accuracy and loss using the proposed model for different datasets after data augmentation

(c) Training and validation curves for accuracy and loss using the proposed model for the CT dataset.

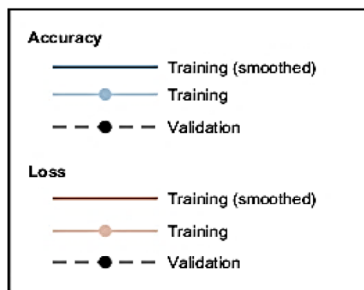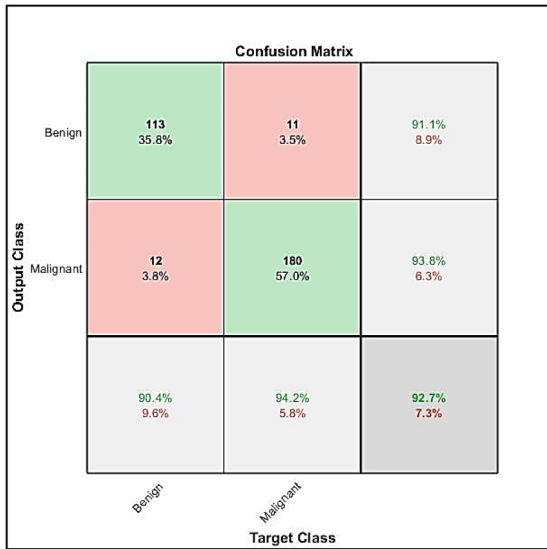(d) Training and validation curves for accuracy and loss using the proposed model for the MRI dataset.
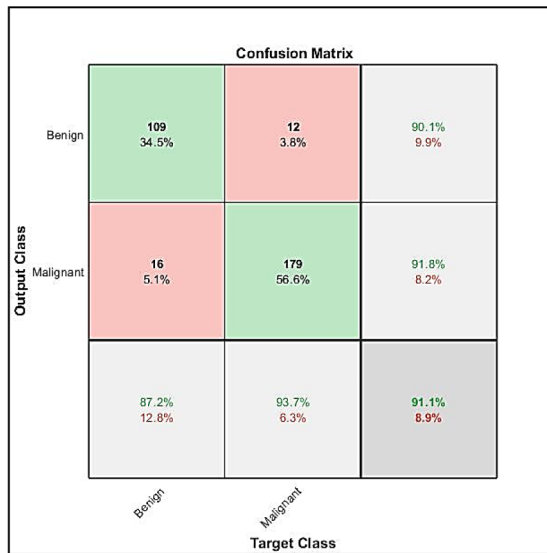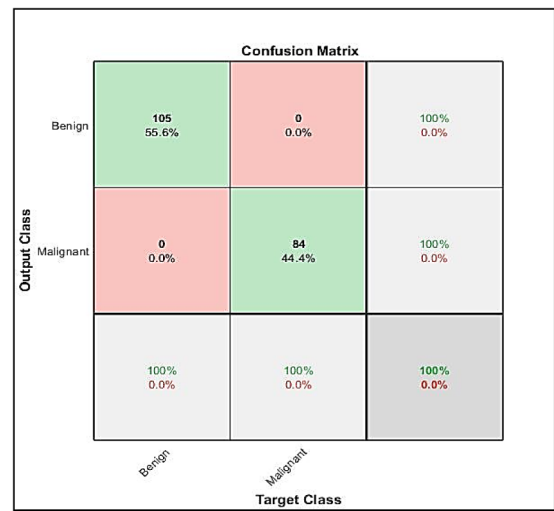

Fig. 7 (continued)

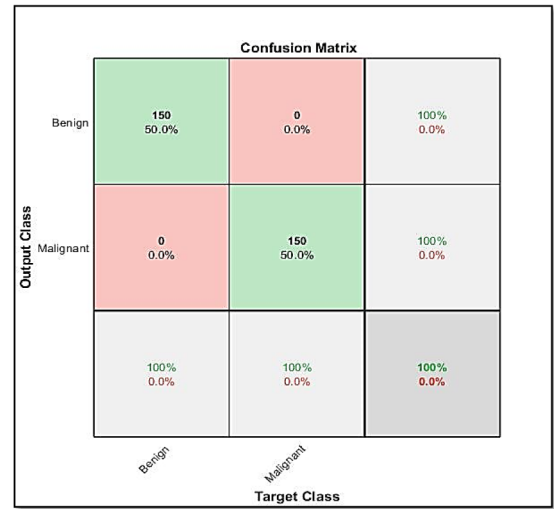(a) The confusion matrices of the proposed CNN model for Us dataset.

(c) The confusion matrices of the proposed CNN model for CT dataset.

(b) The confusion matrices of the proposed CNN model for X-Ray dataset.

(d) The confusion matrices of the proposed CNN model for MRI dataset.

Fig. 8 (continued)

Fig. 8 The confusion matrices of the proposed architecture for different datasets

Table 4 The accuracy (Acc.) and loss of the pre-trained models and the proposed model for Us, X-ray, CT, MRI datasets with data augmentation

Dataset	VGG-16		VGG-19		AlexNet		Proposed	
	Acc	CPU (min)	Acc	CPU (min)	Acc	CPU (min)	Acc	CPU (min)
Us	58.3%	≅ 840	75%	≅ 1140	89.9%	110:35	92.7%	79:17
X-Ray	54.7%	≅ 1130	63.8	≅ 1460	70.6%	67:24	91.1%	43:23
CT	100%	≅ 430	100%	≅ 840	100%	33:17	100%	19:38
MRI	100%	≅ 990	100%	≅ 1290	100%	69:31	100%	41:10

Authors' contributions All authors are equally contributed.

Funding The authors did not receive support from any organization for the submitted work.

Data availability All data are available upon request from the corresponding author.

Declarations

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

Ethical approval All authors are contributing and accepting to submit the current work.

Consent for publication All authors are accepting to submit and publish the submitted work.

References

1. W. El-Shafai, I. Almomani, A. Alkhayer, Optical bit-plane-based 3D-JST cryptography algorithm with cascaded 2D-FrFT encryption for efficient and secure HEVC communication. IEEE Access **2**(9), 35004–35026 (2021)

2. A. Alarifi, M. Amoon, M. Aly, W. El-Shafai, Optical PTFT asymmetric cryptosystem-based secure and efficient cancelable biometric recognition system. IEEE Access **3**(8), 221246–221268 (2020)

3. O. Faragallah, A. Afifi, H. El-Sayed, M. Alzain, J. Al-Amri, F. Abd El-Samie, W. El-Shafai, Efficient HEVC integrity verification scheme for multimedia cybersecurity applications. IEEE Access **7**(8), 167069–167089 (2020)

4. I. Elashry, W. El-Shafai, E. Hasan, S. El-Rabaie, A. Abbas, A. El-Samie, O. Faragallah, Efficient chaotic-based image cryptosystem with different modes of operation. Multimed. Tools Appl. **79**(29), 20665–20687 (2020)

5. N. Soliman, M. Khalil, A. Algarni, S. Ismail, R. Marzouk, W. El-Shafai, Efficient HEVC steganography approach based on audio compression and encryption in QFFT domain for secure multimedia communication. Multimed. Tools Appl. **80**(3), 4789–4823 (2021)

6. W. El-Shafai, Joint adaptive pre-processing resilience and post-processing concealment schemes for 3D video transmission. 3D Res. **6**(1), 1–13 (2015)

7. K. Abdelwahab, A. El-atty, M. Saied, W. El-Shafai, S. El-Rabaie, A. El-Samie, Efficient SVD-based audio watermarking technique in FRT domain. Multimed. Tools Appl. **79**(9), 5617–5648 (2020)

8. A. Algarni, G. El Banby, S. Ismail, W. El-Shafai, F. El-Samie, N.F. Soliman, Discrete transforms and matrix rotation based cancelable face and fingerprint recognition for biometric security applications. Entropy **22**(12), 1361 (2020)

9. N. El-Hag, A. Sedik, W. El-Shafai, H. El-Hoseny, A. Khalaf, A. El-Fishawy, G. El-Banby, Classification of retinal images based on convolutional neural network. Microsc. Res. Tech. **84**(3), 394–414 (2021)

10. L. Abou Elazm, S. Ibrahim, M. Egila, H. Shawky, M. Elsaid, W. El-Shafai, F. Abd El-Samie, Cancellable face and fingerprint recognition based on the 3D jigsaw transform and optical encryption. Multimed. Tools Appl. **7**(9), 14053–14078 (2020)

11. W. El-Shafai, S. El-Rabaie, M. El-Halawany, F. El-Samie, Enhancement of wireless 3d video communication using

color-plus-depth error restoration algorithms and Bayesian Kalman filtering. Wireless Pers. Commun. **97**(1), 245–268 (2017)

12. O. Faragallah, W. El-Shafai, A. Sallam, I, Elashry, E. EL-Rabaie, A. Afifi, H. El-sayed, Cybersecurity framework of hybrid water-marking and selective encryption for secure HEVC communication. J. Ambient. Intell. Humaniz. Comput. **13**(2), 1215–1239 (2022)

13. I. Badr, A. Radwan, E. El-Rabaie, L. Said, G. El Banby, W. El-Shafai, F. Abd El-Samie, Cancellable face recognition based on fractional-order Lorenz chaotic system and Haar wavelet fusion. Digital Sign Process. **11**(6), 103103 (2021)

14. W. El-Shafai, F. Mohamed, H. Elkamchouchi, M. Abd-Elnaby, A. Elshafee, Efficient and secure cancelable biometric authentication framework based on genetic encryption algorithm. IEEE Access **2**(9), 77675–77692 (2021)

15. F. Abd El-Samie, R. Nassar, M. Safan, M. Abdelhamed, A. Khalaf, G. El Banby, W. El-Shafai, Efficient implementation of optical scanning holography in cancelable biometrics. Appl. Opt. **60**(13), 3659–3667 (2021)

16. S. Ibrahim, M. Egila, H. Shawkey, M. Elsaid, W. El-Shafai, F. Abd El-Samie, Hardware implementation of cancellable biometric systems, In Fourth IEEE International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 1145–1152, (2020)

17. H. El-Hameed, N. Ramadan, W. El-Shafai, A. Khalaf, H. Ahmed, S. Elkhamy, F. El-Samie, Cancelable biometric security system based on advanced chaotic maps. Vis. Comput. **38**(6), 2171–2187 (2022)

18. I. Almomani, W. El-Shafai, A. AlKhayer, A. Alsumayt, S. Alja-meel, Proposed biometric security system based on deep learning and chaos algorithms. Comput. Mater.Continua **74**(2), 3515–3537 (2023)

19. L. Elazm, W. El-Shafai, S. Ibrahim, M. Egila, H. Shawkey, Efficient hardware design of a secure cancellable biometric crypto-system. Intell. Autom. Soft Comput. **36**(1), 929–955 (2023)

20. W. El-Shafai, M. Elsayed, M. Rashwan, M. Dessouky, A. El-Fishawy, Optical ciphering scheme for cancellable speaker iden-tification system. Comput. Syst. Sci. Eng. **45**(1), 563–578 (2023)

21. A. Ayoup, A. Khalaf, W. El-Shafai, F. Abd El-Samie, F. Alraddady, S. Eldin, Cancellable multi-biometric template genera-tion based on arnold cat map and aliasing. CMC-Comput. Mater. Continua **72**(2), 3687–3703 (2022)

22. S. El-Gazar, W. El Shafai, G. El Banby, H. Hamed, G. Salama, M. Abd-Elnaby, F. Abd El-Samie, Cancelable speaker identification system based on optical-like encryption algorithms. Comput. Syst. Sci. Eng. **43**(1), 87–102 (2022)

23. A. Ayoup, A. Khalaf, F. Alraddady, F. Abd El-Samie, W. El-Shafai, Cancelable multi-biometric template generation based on dual-tree complex wavelet transform". Intell. Autom. Soft Com-put. **33**(2), 1289–1304 (2022)

24. I. Almomani, A. AlKhayer, W. El-Shafai, Novel ransomware hid-ing model using HEVC steganography approach. CMC Comput. Mater. Contin **70**(2), 1209–1228 (2021)

25. A. Alarifi, S. Sankar, T. Altameem, K. Jithin, M. Amoon, W. El-Shafai, A novel hybrid cryptosystem for secure streaming of high efficiency H. 265 compressed videos in IoT multimedia applica-tions. IEEE Access **8**(2), 128548–128573 (2020)

26. O. Faragallah, H. El-Hoseny, W. El-Shafai, W. El-Rahman, H. El-sayed, S. El-Rabaie, G. Geweid, Optimized multimodal medical image fusion framework using multi-scale geometric and multi-resolution geometric analysis. Multimed. Tools Appl. **81**(10), 14379–14401 (2022)

27. R. Ali, F. El-Sayed, W. El-Shafai, T. Elsayed, Efficient fusion of medical images based on CNN. Menoufia J. Electron. Eng. Res. **30**(2), 79–83 (2021)

28. H. El-Hoseny, W. El-Rahman, W. El-Shafai, E. El-Rabaie, K. Mahmoud, F. Abd El-Samie, O. Faragallah, Optimal multi-scale geometric fusion based on non-subsampled contourlet transform and modified central force optimization. Int. J. Imaging Syst. Technol. **29**(1), 4–18 (2019)

29. H. El-Hoseny, W. El-Rahman, W. El-Shafai, G. El-Banby, S. El-Rabaie, F. Abd El-Samie, K. Mahmoud, Efficient multi-scale non-sub-sampled shearlet fusion system based on modified central force optimization and contrast enhancement. Infrared Phys. Technol. **102**, 102975 (2019)

30. W. El-Shafai, S. El-Rabaie, M. El-Halawany, F. Abd El-Samie, Security of 3D-HEVC transmission based on fusion and water-marking techniques. Multimed. Tools Appl. **78**(19), 27211–27244 (2019)

31. O. Faragallah, H. El-Hoseny, W. El-Shafai, W. Abd El-Rahman, H. El-Sayed, S. El-Rabaie, G. Geweid, A comprehensive survey analysis for present solutions of medical image fusion and future directions. IEEE Access. **9**, 11358–11371 (2020)

32. N. El-Hag, A. Sedik, G. El-Banby, W. El-Shafai, A. Khalaf, W. Al-Nuaimy, H. El-Hoseny, Utilization of image interpolation and fusion in brain tumor segmentation. Int. J. Numer. Methods Biomed. Eng. **37**(8), e3449 (2021)

33. C. Ghandour, W. El-Shafai, S. El-Rabaie, Comparative Study between Different Image Fusion Techniques Applied on Biomedical Images, In 2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC) (pp. 165–176). IEEE (2021).

34. A. Ayoup, A. Khalaf, F. Alraddady, F. Abd El-Samie, W. El-Safai, S. Eldin, Selective cancellable multi-biometric template genera-tion scheme based on multi-exposure feature fusion. Intell Autom Soft Comput **33**(1), 549–565 (2022)

35. W. El-Shafai, N. El-Hag, A. Sedik, G. Elbanby, F. Abd El-Samie, An efficient medical image deep fusion model based on convo-lutional neural networks. Comput. Mater. Continua **74**(2), 2905–2925 (2023)

36. W. El-Shafai, F. Khallaf, E. El-Rabaie, F. El-Samie, Robust medical image encryption based on DNA-chaos cryptosystem for secure telemedicine and healthcare applications. J. Ambient. Intell. Humaniz. Comput. **12**(10), 9007–9035 (2021)

37. W. El-Shafai, E. Hemdan, Robust and efficient multi-level security framework for color medical images in telehealthcare services. J. Ambient. Intell. Humaniz. Comput. **8**(5), 1–16 (2021)

38. W. El-Shafai, M. Aly, A. Algarni, F. Abd El-Samie, N. Soliman, Secure and robust optical multi-stage medical image cryptosys-tem. CMC-Comput. Mater. Continua **70**(1), 895–913 (2022)

39. N. Soliman, N. Ali, M. Aly, A. Algarni, W. El-Shafai, F. Abd El-Samie, An efficient breast cancer detection framework for medi-cal diagnosis applications. CMC-Comput. Mater. Continua **70**(1), 1315–1334 (2022)

40. F. Alqahtani, M. Amoon, W. El-Shafai, A fractional fourier based medical image authentication approach. CMC-Comput. Mater. Continua **70**(2), 3133–31504 (2022)

41. W. El-Shafai, S. El-Nabi, S. El-Rabaie, A. Ali, N. Soliman, Effi-cient deep-learning-based autoencoder denoising approach for medical image diagnosis. CMC-Comput. Mater. Continua **70**(3), 6107–6125 (2022)

42. A. Siam, M. Almaiah, A. Al-Zahrani, A. Elazm, G. El Banby, W. El-Shafai, N. El-Bahnasawy, Secure health monitoring com-munication systems based on IoT and cloud computing for medi-cal emergency applications. Comput. Intell. Neurosci. **8**(3), 1–20 (2021)

43. E. Hemdan, W. El-Shafai, A. Sayed, CR19: A framework for pre-liminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications. J. Ambient. Intell. Humaniz. Comput. **3**(6), 1–13 (2022)

44. W. El-Shafai, S. El-Rabaie, M. El-Halawany, F. El-Samie, Recur-sive Bayesian filtering-based error concealment scheme for 3D video communication over severely lossy wireless channels. Cir-cuits Systems Signal Process. **37**(11), 4810–4841 (2018)

45. C. Ghandour, W. El-Shafai, S. El-Rabaie, Medical image fusion based on weighted least square optimization and deep learning algorithm, In 2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC) (pp. 159–163). IEEE (2021)