



Medical image data augmentation: techniques, comparisons and interpretations

Evgin Goceri¹

Accepted: 27 February 2023 / Published online: 20 March 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract

Designing deep learning based methods with medical images has always been an attractive area of research to assist clinicians in rapid examination and accurate diagnosis. Those methods need a large number of datasets including all variations in their training stages. On the other hand, medical images are always scarce due to several reasons, such as not enough patients for some diseases, patients do not want to allow their images to be used, lack of medical equipment or equipment, inability to obtain images that meet the desired criteria. This issue leads to bias in datasets, overfitting, and inaccurate results. Data augmentation is a common solution to overcome this issue and various augmentation techniques have been applied to different types of images in the literature. However, it is not clear which data augmentation technique provides more efficient results for which image type since different diseases are handled, different network architectures are used, and these architectures are trained and tested with different numbers of data sets in the literature. Therefore, in this work, the augmentation techniques used to improve performances of deep learning based diagnosis of the diseases in different organs (brain, lung, breast, and eye) from different imaging modalities (MR, CT, mammography, and funduscopy) have been examined. Also, the most commonly used augmentation methods have been implemented, and their effectiveness in classifications with a deep network has been discussed based on quantitative performance evaluations. Experiments indicated that augmentation techniques should be chosen carefully according to image types.

Keywords Data augmentation · GAN · Medical images · Synthesis

1 Introduction

Medical image interpretations are mostly performed by medical professionals like clinicians and radiologists. However, the variations among different experts and complexities of medical images make it very difficult for the experts to diagnose diseases accurately all the time. Thanks to computerized techniques, the tedious image analysis task can be performed by semi-/fully-automatically, and they help the experts to make objective, rapid

✉ Evgin Goceri
evgin@akdeniz.edu.tr

¹ Department of Biomedical Engineering, Engineering Faculty, Akdeniz University, Antalya, Turkey

and accurate diagnoses. Therefore, designing deep learning based methods with medical images has always been an attractive area of research (Tsuneki 2022; van der Velden et al. 2022; Chen et al. 2022a). Particularly, deep Convolutional Neural Network (CNN) architectures have been used for several processes (e.g., image segmentation, classification, registration, content-based image retrieval, identification of disease severity, and lesion/tumor/tissue detection, etc.) in some medical diagnosis fields like eye, breast, brain, and lung (Yu et al. 2021). Although CNNs can produce results with higher performance compared to the results of traditional machine learning based methods, improving their generalization abilities and developing robust models are still significant challenging issues. Because training data sets should be constructed with a large number of images including all variations (i.e., heterogeneous images) to provide the high generalization ability and robustness of those architectures. However, the number of medical images is usually limited. The data scarcity problem can be due to not enough patients for some diseases, patients do not want to allow their images to be used, lack of medical equipment or equipment, inability to obtain images that meet the desired criteria, inability to access images of patients living in different geographical regions or of different races.

Another important issue is that, even if enough images are acquired, labeling them (unlike natural image labeling which is relatively easy) is not easy, requires certain domain knowledge of medical professionals, and is time-consuming. Also, legal issues and privacy are other concerns in the labeling of medical images. Therefore, imbalanced data causing overfitting, biased and inaccurate results is a significant challenge that should be handled while developing deep network based methods.

To overcome these challenges by using balanced number of images and expanding training datasets automatically, image augmentation methods are used. Thanks to the augmentation methods, deep network architectures' learning and generalization ability can be improved, and the desired robustness property of the networks can be provided (Sect. 2). However, various augmentation techniques have been applied with different types of images in the literature. It is not clear which data augmentation technique provides more efficient results for which image type since different diseases are handled, different network architectures are used, and these architectures are trained and tested with different numbers of data sets in the literature (Sect. 2). According to the segmentation or classification results presented in the literature, it is not possible to understand or compare the impacts or contributions of the used augmentation approaches on the results, and to decide the most appropriate augmentation method.

In the literature, researchers have presented some surveys about augmentation methods. However, they have focused on the augmentation methods applied either;

- (i) with a specific type of images like natural images (Khosla and Saini 2020), mammography images (Oza et al. 2022), Computed Tomography (CT), and Magnetic Resonance (MR) images (Chlap et al. 2021) acquired with different imaging techniques and have different properties.
- (ii) to improve the performance of specific operations like polyp segmentation (Dorizza 2021) and text classification (Bayer et al. 2021).
- (iii) with a specific technique, such as Generative Adversarial Network (GAN) (Chen et al. 2022b), erasing and mixing (Naveed 2021).

In this work, a more comprehensive range of image types has been handled. Also, a more in-depth analysis has been performed through the implementation of common augmentation methods with the same data sets for objective comparisons, and

evaluations based on quantitative results. According to our knowledge, there is no (i) comprehensive review that covers a greater range of medical imaging modalities and applications than articles in the literature, and (ii) any work on quantitative evaluations of the augmentation methods applied with different medical images to determine the most appropriate augmentation method.

To fill this gap in the literature, in this paper, the augmentation techniques that have been applied to improve performances of deep learning-based diagnosis of diseases in different organs (brain, lung, breast, and eye) using different imaging modalities (MR, CT, mammography, and funduscopy) have been examined. The chosen imaging modalities are widely used in medicine for many applications such as classification of brain tumors, lung nodules, and breast lesions (Meijering 2020). Also, in this study, the most commonly used augmentation methods have been implemented using the same datasets. Additionally, to evaluate the effectiveness of those augmentation methods in classifications from the four types of images, a classifier model has been designed and classifications with the augmented images for each augmentation method have been performed with the same classifier model. The effectiveness of the augmentation methods in the classifications has been discussed based on quantitative results. This paper presents a novel comprehensive insight into the augmentation methods proposed for diverse medical images by considering popular medical research problems altogether.

The key features of this paper are as follows:

1. The techniques used for augmentation of brain MR images, lung CT images, breast mammography images, and eye fundus images have been reviewed.
2. The papers published between 2020 and 2022 and mainly collected by Springer, IEEE Xplore, and ELSEVIER, have been reviewed, and a systematic and comprehensive review and analyses of the augmentation methods have been carried out.
3. Commonly used augmentations in the literature have been implemented with four different medical image modalities.
4. For comparisons of the effectiveness of the augmentation methods in deep learning-based classifications, the same network architecture and expanded data sets with the augmented images from each augmentation methods have been used.
5. Five evaluation metrics have been used, and the results provided from each classification have been evaluated with the same evaluation metrics.
6. The best augmentation method for different imaging modalities showing different organs has been determined according to quantitative values.

We believe that the results and conclusions presented in this work will be helpful for researchers to choose the most appropriate augmentation technique in developing deep learning-based approaches to diagnose diseases with a limited number of image data. Also, this work can be helpful for researchers interested in developing new augmentation methods for different types of images. This manuscript has been structured as follows: A detailed review of the augmentation approaches applied with the medical images handled in this work between 2020 and 2022 is presented in the second section. The data sets and methods implemented in this work are explained in the third section. Evaluation metrics and quantitative results are given in the fourth section. Discussions are given in the fifth section, and conclusions are presented in the sixth section.

2 Related works

In this section, the methods used in the literature for augmentation of brain MR images, lung CT images, breast mammography images, and eye fundus images are presented.

2.1 Augmentation of brain MR images

In the literature, augmentation of brain MR images have been performed with various methods such as rotation, noise addition, shearing, and translation to increase number of images and improve performances of several tasks (Table 1). For example, augmentation has been provided by noise addition and sharpening to increase the accuracy of tumor segmentation and classification in a study (Khan et al. 2021). In another study, augmented images have been obtained by translation, rotation, random cropping, blurring, and noise addition to obtain high performance from the method applied for age prediction, schizophrenia diagnosis, and sex classification (Dufumier et al. 2021). In different studies, random scaling, rotation, and elastic deformation have been applied to increase tumor segmentation accuracy (Isensee et al. 2020; Fidon et al. 2020). Although mostly those techniques have been preferred in the literature, augmentation has also been provided by producing synthetic images. To generate synthetic images, Mixup method, which combines two randomly selected images and their labels, or modified Mixup, or GAN structures have been used (Table 1).

For instance, a modified Mixup is a voxel-based approach called TensorMixup, which combines two image patches using a tensor, and has been applied (Wang et al. 2022) to boost the accuracy of tumor segmentation. In a recent study, a GAN based augmentation has been used for cerebrovascular segmentation (Kossen et al. 2021).

All methods applied for augmentation of brain MR images in the publications reviewed in this work and the details in those publications are presented in Table 1. Those augmentation methods have been used in the applications developed for different purposes. In those applications, different network architectures, datasets, and numbers of images have been used, and also their performances have been evaluated with different metrics. High performances have been obtained from those applications according to the reported results. However, the most effective and appropriate augmentation method used in them with the brain MR images is not clear due to those differences. Also, it is not possible to determine augmentation methods' contributions to those performances according to the presented results in those publications.

2.2 Augmentation of lung CT images

In computerized methods being developed to assist medical professionals in various tasks [e.g., nodule or lung parenchyma segmentation, nodule classification, prediction of malignancy level, disease diagnosis like coronavirus (covid-19), etc.], augmentations of CT lung images are commonly provided by traditional techniques like rotation, flipping, and scaling (Table 2). For example, image augmentation has been provided using flipping, translation, rotation, and brightness changing to improve automated diagnosis of covid-19 by Hu et al. (2020). Similarly, a set of augmentations (Gaussian noise addition, cropping, flipping, blurring, brightness changing, shearing, and rotation) has been applied to increase the performance of covid-19 diagnosis by Alshazly et al. (2021). Also, several GAN models have been used to augment lung CT images by generating new synthetic images (Table 2). While

Table 1 Augmentation methods applied with brain MR images and details given in publications

Augmentation Method	Application	Evaluation Metrics	Results	Conclusions	References
Random rotation, noise addition, zooming and sharpening	Tumor segmentation and classification (as benign and malignant)	Accuracy, sensitivity and specificity	Overall accuracy is 94.06%. Sensitivity and specificity for malignant cases are 89.09%, 90.01%, and for benign cases are 89.09% and 90.01%, respectively	The proposed segmentation approach can assist medical experts and radiologists in detection malignant and benign tumors	Khan et al. (2021)
Translation, rotation, random cropping, blurring, flipping and noise addition	Age prediction, schizophrenia diagnosis, and sex classification	Area Under the receiver operating characteristic Curve (AUC), Mean Absolute Error (MAE), pearson correlation coefficient (R), Root Mean Square Error (RMSE), and accuracy	MAE, RMSE, R for age prediction with voxel-based morphometry are 6.02 ± 0.24 , 7.25 ± 0.34 , 0.83 ± 0.008 , respectively. AUC is 0.93 ± 0.01 and accuracy is 0.83 ± 0.01 for sex classification. AUC is 0.79 ± 0.01 , and accuracy is 0.72 ± 0.01 for schizophrenia diagnosis	Deep ensemble learning improves accuracy. Data augmentation is task and dataset-specific, and a more in-depth study is required	Dufumier et al. (2021)
Random scaling, rotation and elastic deformation	Tumor segmentation (HGG and LGG tumors)	Hausdorff distance and dice similarity coefficient	Hausdorff distance value and dice score calculated by the results of the whole tumor segmentation are 8.49 millimeters and 88.95% respectively	The proposed technique achieved high performance in segmentation of tumorous regions	Isensee et al. (2020)
Image to image translation based on GAN	Tumor segmentation (HGG and LGG tumors)	Dice similarity coefficient	The dice score computed by using the results obtained from FLAIR images is 0.76	Segmentation was provided with high generalization performance	Li et al. (2020)

Table 1 (continued)

Augmentation Method	Application	Evaluation Metrics	Results	Conclusions	References
End-to-end GAN based framework	Tumor segmentation (HGG and LGG tumors)	Dice similarity coefficient	Tumor core and whole tumor segmentation was improved by 0.16 and 0.17 in terms of the dice similarity coefficient, respectively	The effectiveness of tumor segmentation increased with the presented approach	Sun et al. (2020)
Random rotation, cropping, zooming, Gaussian noise addition, blurring	Tumor segmentation (HGG and LGG tumors)	Hausdorff distance, and dice similarity coefficient	The dice score value and Hausdorff distance value calculated by using the results of the segmentation of whole tumors are 88.95% and 6.4 millimeters, respectively	High performance was obtained from the segmentation of tumorous regions	Fidon et al. (2020)
Conditional GAN	Stroke lesion segmentation	Intersection over Union (IoU), dice similarity coefficient, precision, and recall	The dice score, IoU, precision, and recall values are 0.549, 0.410, 0.721, 0.5213, respectively	The proposed method uses contextual information and is effective in increasing the performance of stroke lesion segmentation	Wu et al. (2020a)
Multi-scale gradient GAN	Tumor classification into three classes (meningioma, glioma, pituitary)	Accuracy	The accuracy computed by the results of the classification to classify tumors (as meningioma, glioma, pituitary tumor) is 93.10%	The presented approach classified images with high accuracy by using augmented images	Deepak and Ameer (2020)

Table 1 (continued)

Augmentation Method	Application	Evaluation Metrics	Results	Conclusions	References
TensorMixup	Tumor segmentation (glioma and its subregions)	Hausdorff distance, dice similarity coefficients, specificity, and sensitivity	Specificity, dice score, sensitivity, and Hausdorff distance values for whole tumor segmentation are 99.93%, 91.32%, 90.23%, and 10.88 millimeters, respectively.	Segmentation of tumorous tissues from brain MR images by using the TensorMixup is feasible and effective	Wang et al. (2022)
Conditional GAN	Tumor classification into four classes (as meningioma, glioma, pituitary and normal)	Accuracy, Matthews Correlation Coefficient (MCC), sensitivity, specificity, F1-score, and precision	The results in terms of MCC, specificity, F1-score, accuracy, precision, and sensitivity values are 0.9846%, 99.85%, 98.72%, 98.79%, 98.77%, and 98.77%, respectively	The proposed classifier provides considerable improvement when its performance is compared to the performances of the classifiers based on machine learning	Jha et al. (2022)
Flipping, contrast changing	Tumor segmentation (glioma and its subregions)	Dice similarity coefficient	The dice score computed with the results of the ensemble of models is 0.88	Segmentation with the ensemble of models has improved segmentation performance	Yuan (2020)
GAN	Tumor classification into two classes (relapsing-remitting and secondary-progressive multiple sclerosis)	F1-score	The classification result in terms of dice score is 81%	The results demonstrated high performance of the classification from images showing relapsing-remitting and secondary-progressive multiple sclerosis	Barile et al. (2021)

Table 1 (continued)

Augmentation Method	Application	Evaluation Metrics	Results	Conclusions	References
Wasserstein GAN by using spectral normalization and gradient penalty	Vessel segmentation	Hausdorff distance and dice similarity coefficient	Hausdorff distance and dice similarity values are 24.66 mm and 0.91, respectively by transfer learning with patch-label pairs from the augmentation	The proposed augmentation yielded images with high quality and they were used in a transfer learning approach successfully	Kossen et al. (2021)
Rotation with 90°, 180° and 270°, and flipping along the axial axis	Vessel segmentation	Jaccard index, dice similarity coefficient, recall, and Hausdorff distance	Dice score, Jaccard index, recall, and Hausdorff distance are 87.35%, 77.57%, 87.59%, and 4.86 millimeters respectively	The proposed cerebrovascular segmentation approach provided high performance	Xie et al. (2022)
Flipping and zooming	Vessel segmentation	Precision and dice similarity coefficient	Precision and dice score values are 89.52%, and 87.23% respectively	The proposed segmentation achieved to segment small veins better than other recent segmentation models	Quintana-Quintana et al. (2022)
Cropping and flipping	Vessel segmentation	Precision, specificity, dice similarity coefficient, and recall	Precision, specificity, dice similarity coefficient, and recall values are 93.92%, 99.96%, 93.84%, and 91.35%, respectively	The proposed segmentation approach achieved the best performance among the other compared approaches for vessel segmentation	Liu et al. (2022)
Rotation, shearing, flipping, cropping	Tumor classification (as benign and malignant)	Accuracy	Accuracy values computed in the validation stage are promising and tend to increase up to 0.96% at every epoch	VGG-16 network architecture produced high accuracy in classification of benign and malignant tumors	Srinivas et al. (2022)

Table 1 (continued)

Augmentation Method	Application	Evaluation Metrics	Results	Conclusions	References
Flipping, scaling, and rotation	Tumor segmentation (meningioma, glioma, and pituitary tumors)	Accuracy, error rate-1 (tumorous pixels predicted wrongly as non-tumorous), error rate-2 (non-tumorous pixels wrongly predicted as tumorous), recall, precision, paired t-test	For meningioma, glioma, and pituitary tumors, accuracy is 94.1%, 94.6%, 96.7%; error rate-1 is 2.8%, 1.7%, 0.5%; error rate-2 is 3.1%, 3.7%, 2.8%; precision is 96.9%, 96.3%, 97.2%; recall is 97.2%, 98.3%, 99.5%; partial t test is 0.17, 0.17, and 0.11 respectively	Experimental works indicated the efficiency and good performance of the presented segmentation approach with the augmented image data sets	Chaki (2022)
Rotation, zooming, brightness	Image classification (as tumorous, non-tumorous)	Accuracy, F1 score, sensitivity, precision, specificity, and MCC	The results provided with dataset are 99.93%, 99.32%, 95.36%, and 98.76% in terms of accuracy, specificity, sensitivity, and precision, respectively. F1-score is 99.09% and MCC is 98.10%	The results of the experimental works indicated that the diagnosis method can provide high performance in the diagnosis of brain tumors	Haq et al. (2022)
Translation, flipping, rotation, contrast adjustment	Tumor segmentation and image classification (as tumorous, non-tumorous)	Accuracy, specificity, sensitivity	The highest segmentation specificity, sensitivity, and accuracy are 92.5%, 92.8%, and 90.0%, respectively	The proposed approach produced enhanced performance	Neelima et al. (2022)

Table 1 (continued)

Augmentation Method	Application	Evaluation Metrics	Results	Conclusions	References
Rotation, flipping	Image classification (as tumorous, non-tumorous)	Accuracy, precision, F1 score, and recall	Recall, accuracy, F1 score, and precision values are 96.1%, 98.6%, 98.2%, and 97.8% respectively	The proposed network and the processed datasets have performed tumor detection with higher performance than the recent methods	Nayan et al. (2022)
Translation, flipping, shearing, scaling, rotation	Tumor segmentation (glioma and its subregions)	Dice similarity coefficient	Dice similarity coefficients for whole tumor, tumor core, and enhancing tumor are 0.905, 0.903, and 0.822, respectively	The presented framework almost yielded state-of-the-art results	Hashemi et al. (2022)

Table 2 Augmentation methods applied with lung CT images and details given in publications

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Flipping, cropping rotation, and shifting	Lung cancer detection	Precision, recall, F1 score, and accuracy	Accuracy is 96.5%, and cancer detection has been improved in terms of precision, recall, and F1 score metrics	The highest accuracy among the pretrained models used in this work was obtained by the proposed approach	Tandon et al. (2022a)
Rescaling, rotation, horizontal and vertical flipping	Nodule classification (as normal and malignant)	Specificity, sensitivity, accuracy, and F1-score	Accuracy, specificity, sensitivity, F1-score are 99.49%, 99.25%, 99.25%, and 99.12%, respectively	The proposed framework is successful in detection of lung carcinoma	Tandon et al. (2022b)
Style-based GAN	Pulmonary nodule segmentation	Positive predicted value, sensitivity, dice similarity coefficient	Positive predicted value, sensitivity, and dice score are 86.68%, 86.22%, and 85.21, respectively	The augmentation method improved the robustness, efficiency, and accuracy of the segmentation model	Shi et al. (2020)
Flipping and rotation	Nodule segmentation	Dice similarity coefficient, F1 score, and MCC	Dice score, F1 score, MCC are 0.8472, 0.6843, 0.6337, respectively	The proposed approach produced segmented nodules with high performance	Farheen et al. (2022)
Wasserstein GAN	Pulmonary nodule classification	Specificity and sensitivity	Specificity is 77.8% and the sensitivity is 93.9%.	Classification accuracy improved by using generated images	Onishi et al. (2020)
Flipping, translation, rotation, and brightness changing	Covid-19 diagnosis	Sensitivity, specificity, accuracy, and AUC	Sensitivity, specificity, and accuracy are 90.52%, 91.58%, and 91.21%, respectively. The AUC is 96.89%	The model developed with insufficient images can help medical experts in rapid and correct diagnosis of Covid-19	Hu et al. (2020)
GAN	Nodule classification into five classes based on nodule size	Accuracy	Average accuracy is 87.5% (range 84.2–90.1%)	Three-dimensional images of lung nodules can be generated and used in the classification of lung nodules	Nishio et al. (2020)

Table 2 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation, translation, shearing, and zooming	Lung parenchyma and infection segmentation	Accuracy, specificity, sensitivity, dice similarity coefficient	For lung parenchyma segmentation, sensitivity, specificity, accuracy, dice score are 93.47%, 98.64%, 98.07%, 92.46%, respectively. For lung infection segmentation, they are 83.33%, 99.84%, 99.20%, 82%, respectively	Accurate segmentation of lung parenchyma and infection was provided from 3 dimensional CT volumes by the proposed framework. It can be used as a tool for clinicians in analyzing contaminated regions	Aswathy and Vinod Chandra (2022)
Co-guided GAN	Prediction of malignancy level	Fréchet Inception Distance (FID), accuracy, sensitivity, F1 score	FID score is 89.35. Accuracy, sensitivity and F1 score are 90.43%, 80.55%, 84.03%, respectively	The accuracy for prediction of malignancy was improved with generated data by 5.6%	Wang et al. (2021)
Semi-conditional InfoGAN	Disease classification as (Small Cell Cancer (SCLC), ADenoCarcinoma (ADC), Squamous Cell Carcinoma (SCC))	FID, accuracy	FID values indicating similarities between the produced images and real images are 183.5, 198.1, 167.5 for ADC, SCC, SCLC, respectively. Classification accuracy is 57.7% when the network is pre-trained with additional augmented images	Classification accuracy was improved by 23.4% compared to the accuracy that was obtained with the results of the network trained only with real images. Therefore, the augmentation has the potential to generate images	Toda et al. (2021)
Translation, scaling randomly into different sizes, rotation, and flipping	Covid-19 detection	F1 score, sensitivity, accuracy, precision, specificity	Sensitivity, specificity, precision, accuracy, and F1-score are 95.71%, 100%, 100%, 98.00%, 97.81%, respectively	Experimental works indicated high performance of the proposed classification method	Woan et al. (2022)

Table 2 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation, shifting, flipping	Covid-19 detection	Accuracy, F1-score	Accuracy values are 97.30% and 98.87% on the two datasets. F1 scores on two modes are 89.10% and 91.59%	The proposed framework improved the classification performance	Bast et al. (2022)
Rotation, shifting, flipping	Covid-19 detection	Accuracy, AUC	Accuracy is 97%, AUC is 0.99	The proposed architecture is efficient to classify Covid-19	Halder and Datta (2021)
Noise addition, cropping, flipping, blurring, brightness changing, shearing, and rotation	Covid-19 detection	F1 score, sensitivity, accuracy, precision, specificity	F1 score, sensitivity, accuracy, precision, and specificity values are given as 92.5 ± 2.4 , 93.7 ± 3.4 , 92.9 ± 2.2 , $91.3\% \pm 2.2$, and 92.2 ± 2.2 , respectively	The results indicated the potential of the proposed approach for fast and automated diagnosis of Covid-19	Alshazly et al. (2021)
Scaling, mirroring, elastic deformations, rotations, contrast changing, Gaussian noise addition	Lung segmentation, and Covid-19 lesion segmentation	Dice similarity coefficient, IoU, sensitivity, and specificity	Dice score, IoU, sensitivity specificity: 0.804, 0.672, 0.778, 0.999. Dice score for lung and Covid-19 lesion segmentation after augmentation are 0.956, 0.761 respectively	Preprocessing and augmentation stages improved the performance by 0.630 and 0.647 for Covid-19 lesion and lung segmentation, respectively, in terms of dice score	Müller et al. (2021)
Scaling, flipping, shearing	Covid-19 detection	F1 score, recall, precision, accuracy, and AUC	Accuracy of VGG16, VGG19, and Xception is 98.83%, 98.05%, and 97.4%, respectively	Transfer learning techniques performed well in detection of lung nodules	Humayun et al. (2022)

Table 2 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Translation, scaling, flipping, brightness changing	Image classification (as a nodule or non-nodule image)	Dice similarity coefficient and IoU	Dice similarity coefficient and IoU values are 95.01% and 0.83, respectively	The proposed network resulted in considerable improvements over existing computer-assisted systems	Dodia et al. (2022)
Rotation, flipping	Pulmonary nodule classification	AUC, specificity, sensitivity, and accuracy	AUC, specificity, sensitivity, and accuracy values are 0.971, 0.956, 0.919, and 0.936, respectively	The presented approach with structural data outperformed several state-of-the-art nodule classification approaches	Tang et al. (2021)

doing a GAN based augmentation of lung CT scans, researchers usually tend to generate the area showing lung nodules in the whole image. Because the lung nodules are small and therefore it is more cost-effective than generating a larger whole image or background area. For instance, guided GAN structures have been implemented to produce synthetic images showing the area with lung nodules by Wang et al. (2021). Nodules' size information and guided GANs have been used to produce images with diverse sizes of lung nodules by Nishio et al. (2020).

All methods applied for augmentation of breast mammography images in the publications reviewed in this study and the details in those publications are presented in Table 2. Those augmentation methods have been applied in the applications developed for different purposes. In those applications, different network models, data sets and number of images have been used, and also their performances have been evaluated with different metrics. High performances have been obtained from those applications according to the reported results. However, the most effective and appropriate augmentation method used in them with the lung CT images is not clear due to those differences. Also, it is not possible to determine augmentation methods' contributions to those performances according to the presented results in those publications.

2.3 Augmentation of breast mammography images

Generally, augmentation methods with mammography images have been applied to improve automated recognition, segmentation, and detection of breast lesions. Because failures in the detection or identification of the lesions lead to unnecessary biopsies or inaccurate diagnoses. In some works, in the literature, augmentation methods have been applied on extracted positive and negative patches instead of the whole image (Karthiga et al. 2022; Zeiser et al. 2020; Miller et al. 2022; Kim and Kim 2022; Wu et al. 2020b). Positive patches are extracted based on radiologist marks while negative patches are extracted as random regions from the same image. Also, refinement and segmentation of the lesion regions before the application of an augmentation method is the most preferred way to provide good performance from the detection and classification of the lesions. For instance, image refinement has been provided by noise reduction and resizing steps, and then segmentation and augmentation steps have been applied (Zeiser et al. 2020). Similarly, noise and artifacts have been eliminated before the segmentation of the images into non-overlapping parts and augmentation (Karthiga et al. 2022). Augmentation of patches or whole mammography images have usually been performed by scaling, noise addition, mirroring, shearing, and rotation techniques or using several combinations of them (Table 3). For instance, a set of augmentation methods (namely mirroring, zooming, and resizing) has been applied to improve automated classification of images into three categories (as normal, malignant, and benign) (Zeiser et al. 2020). Also, those techniques have been used together with GANs to provide augmentations with new synthetic images. For instance, flipping and deep convolutional GAN has been combined for binary classifications of images (as normal and with a mass) by Alyafi et al. (2020). Only GAN based augmentations have also been used. For instance, contextual GAN, which generates new images by synthesizing lesions according to the context of surrounding tissue, has been applied to improve binary classification (as normal and malignant) by Wu et al. (2020b). In another work (Shen et al. 2021), both contextual and deep convolutional GAN have been used to generate synthetic lesion images with margin and texture information and achieve improvement in the detection and segmentation of lesions.

Table 3 Augmentation methods applied with mammography images and details given in publications

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation, flipping, and shearing	Lesion classification (as benign, and malignant)	AUC, F1 score, accuracy, specificity, sensitivity, and precision	AUC, F1 score, accuracy, specificity, sensitivity, and precision are 0.9463, 0.8680, 0.867, 0.8110, 0.9302, and 0.8136, respectively	Training the system by using augmented data was improved classification accuracy of the proposed network model	Razali et al. (2021)
Rotations, flipping	Lesion classification (as benign, and malignant)	Accuracy	Accuracy using three datasets are 95.95%, 99.39%, and 96.53%, respectively	The suggested model can allow radiologists to obtain accurate results while minimizing their workload	Karthiga et al. (2022)
Mirroring, zooming, and resizing	Lesion classification (as normal, benign, and malignant)	Sensitivity, specificity, dice similarity coefficient, and accuracy	Specificity, sensitivity, accuracy, and dice score are 80.47%, 92.32%, 85.95%, and 79.39% respectively	The proposed system is successful in the identification of breast masses with high performance	Zeiser et al. (2020)
Deep convolutional GAN	Image classification (as normal, and cancer)	F1 score, specificity, accuracy, and sensitivity	F1 score, specificity, accuracy, and sensitivity are 88.0%, 86.1%, 87.0%, and 85.6%, respectively	Image classification performance was improved with the proposed model	Desai et al. (2020)
Rotation, shifting, shearing, scaling, and flipping	Lesion classification (as malignant, and non-malignant)	Accuracy, sensitivity, specificity, AUC, and mean False Positive Indications (mFPI) per image	The mFPIs, accuracy, specificity are 0.45, 0.86, 0.88 for hospital datasets, and they are 0.47, 0.85, 0.85 for clinical datasets respectively. AUC, sensitivity are 0.84 for both datasets	The proposed model can provide detection of breast cancers regardless of tissue density or type	Ueda et al. (2022)
Shifting	Lesion detection and segmentation	AUC	AUC is 0.897	The presented system can be used to detect and segment lesions with high accuracy	Kim and Kim (2022)

Table 3 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Contextual GAN	Image classification (as normal, and malignant case)	AUC	AUC is 0.846	Malignancy classification performance was improved	Wu et al. (2020b)
Resizing, cropping, flipping, rotation, and brightness adjustment	Lesion classification (as benign and malignant)	AUC	The AUCs of whole mammogram after finetuning with global-average-pooling and multiple instance pooling are 0.732, and 0.730 respectively	The proposed approach is promising in both distinction of patches and classification of images	Miller et al. (2022)
Rotation, shifting, and flipping	Lesion classification (as benign and malignant)	Accuracy, AUC, F1 score, sensitivity, and specificity	Accuracy, AUC, F1 score, sensitivity, and specificity are 0.9998, 0.9998, 0.9998, 1, and 0.9997 respectively	This method can be used as a tool for the detection of breast cancer in its early phases	Ayana et al. (2022)
Flipping and deep convolutional GAN	Image classification (as normal, and with a mass)	F1 score	F1 score is 0.09	Augmentation by appending synthetic images and flipping provides improvement in classification	Alyafi et al. (2020)
Deep convolutional and contextual GAN	Lesion detection	Learned Perceptual Image Patch Similarity (LPIPS), recall, and AUC	Mean LPIPS distances on two datasets are 0.02021, 0.15342; recall values on two datasets are 0.8273, 0.8182; AUCs on two datasets are 0.172 and 0.144	The proposed method achieves lesion detection and can be updated to detect lesions from various medical image modalities	Shen et al. (2021)

Table 3 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation	Lesion classification (as benign and malignant)	Accuracy, AUC, F1 score, sensitivity, and precision	Accuracy, AUC, sensitivity, precision, F1 score are 98.87%, 98.88%, 98.98%, 98.79%, and 97.99% respectively	The obtained results revealed high performance of the presented method in classification of breast lesions	Salama and Aly (2021)
Rotation, shrinkage, shifting, flipping, cropping, contrast and brightness changing	Lesion classification (as benign and malignant)	Accuracy	Classification accuracy is 97%	The presented network model performed better classification compared to common deep learning schemes	Mahmood et al. (2021)
Flipping, rotating, translations, and cropping	Lesion classification (as malignant and benign)	AUC, accuracy, precision, sensitivity, and F1 score	AUC, accuracy, precision, sensitivity, and F1 score are 0.99, 0.98, 0.98, 0.99, and 0.98, respectively	The model performed well in classifying breast lesions and so it has great potential for clinical procedures	Mahmood et al. (2022)
Rotation and flipping	Lesion classification (as benign and malignant)	Precision, recall	The system detects 89.4% of the lesions with an average precision of 94.2%, and 84.6% for benign and malignant lesions respectively	You Only Look Once (YOLO) helped to detect lesions in one shot, and lesion detection improved by the proposed system	Aly et al. (2021)
Cropping and flipping	Lesion segmentation and classification (as benign, and malignant)	Dice similarity coefficient, and AUC	Dice score is 93.69% and AUC value is 0.93	The presented model outperformed other related works in segmentation and yielded competitive classification performance	Li et al. (2021)

Table 3 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation	Lesion classification (as healthy, benign, and malignant)	Specificity, accuracy, precision, and sensitivity	Specificity, accuracy, precision, and sensitivity are 99.30%, 97.14%, 98.30%, and 98.67%, respectively	The presented network performed better than other network architectures because of the extensive preprocessing	Padalia et al. (2022)
Rotation, flipping	Lesion classification (as healthy, benign, and malignant)	Accuracy	The accuracy values achieved on three datasets are 99.7%, 99.8%, and 93.8% respectively	Comparisons indicated better performance of the proposed classifier than other existing methods	Zahoor et al. (2022)
Rotation, flipping	Lesion classification (as healthy, malignant, benign)	AUC, F1 score, accuracy, sensitivity, recall, specificity, and precision	AUC, F1 score, accuracy, sensitivity, recall, specificity, and precision are 0.9, 77.45%, 77.78%, 83.15%, 84.54%, and 83.15%, respectively	The presented model outperformed the 18 most popular deep-learning based approaches in classification of breast lesions	Soulami et al. (2022)
Shearing, zooming, rotating, flipping, and brightness changing	Classification of soft tissue opacity (as no opacity, probably benign opacity, and suspicious opacity)	Accuracy, specificity, and sensitivity	Accuracy values are ranged between 73.8% and 89.8%. Specificity and sensitivity values are 87.9% and 92% respectively	The presented network structure can be used for automated classifications of soft tissue opacities from mammography images	Sabani et al. (2022)

All methods applied for augmentation of breast mammography images in the publications reviewed in this work and the details in those publications are presented in Table 3. Those augmentation methods have been applied in the applications developed for different purposes. In those applications, different network architectures, datasets and number of images have been used, and also their performances have been evaluated with different metrics. High performances have been obtained from those applications according to the reported results. However, the most effective and appropriate augmentation method used in them with breast mammography images is not clear due to those differences. Also, it is not possible to determine augmentation methods' contributions to those performances according to the presented results in those publications.

2.4 Augmentation of eye fundus images

Eye fundus images show various important features, which are symptoms of eye diseases, such as abnormal blood veins (neovascularization), red/reddish spots (hemorrhages, microaneurysms), and bright lesions (soft or hard exudates). Those features are used in deep learning based methods being developed for different purposes like glaucoma identification, classification of diabetic retinopathy (DR), and identification of DR severity. Commonly used augmentation techniques with eye fundus images to boost the performance of deep learning based approaches are rotation, shearing, flipping, and translation (Table 4). For example, in a study (Shyamalee and Meedeniya 2022), a combined method including rotation, shearing, zooming, flipping, and shifting has been applied for glaucoma identification. Another combined augmentation including shifting and blurring has been used by Tufail et al. (2021) before binary classification (as healthy and nonhealthy) and multi-class (5 classes) classification of images to differentiate different stages of DR. In a different study (Kurup et al. 2020), combination of rotation, translation, mirroring, and zooming has been used to achieve automated detection of malaria retinopathy. Also, pixel intensity values have been used in some augmentation techniques since fundus images are colored images. For instance, to achieve robust segmentation of retinal vessels, augmentation has been provided by a channel-wise random gamma correction (Sun et al. 2021). The authors have also applied a channel-wise vessel augmentation using morphological transformations. In another work (Agustin et al. 2020), to improve the robustness of DR detection performance, image augmentation has been performed by zooming and a contrast enhancement process known as Contrast Limited Adaptive Histogram Equalization (CLAHE). Augmentations have also been provided by generating synthetic images using GAN models. For instance, conditional GAN (Zhou et al. 2020) and deep convolutional GAN (Balasubramanian et al. 2020) based augmentations have been used to increase the performance of the methods developed for automatic classifications of DR according to its severity level. In some works, retinal features (such as lesion and vascular features) have been developed and added to new images. For instance, NeoVessel (NV)-like structures have been synthesized in a heuristic image augmentation (Araújo et al. 2020) to improve detection of proliferative DR which is an advanced DR stage characterized by neovascularization. In this augmentation, different NV kinds (trees, wheels, brooms) have been generated depending on the expected shape and location of NVs to synthesize new images.

All methods applied for augmentation of eye fundus images in the publications reviewed in this study and the details in those publications are presented in Table 4. Those augmentation methods have been applied in the applications developed for different purposes. In

Table 4 Augmentation methods applied with eye fundus images and details given in publications

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation, shearing, zooming, and flipping, shifting	Glaucoma identification	Accuracy, F1-score, recall, precision, sensitivity, and specificity	Accuracy, F1-score, recall, precision, sensitivity, and specificity are 98.52%, 99.17%, 99.50%, 98.84%, 99.50%, and 90.90%, respectively	The method predicted glaucoma cases from eye fundus images with high performance	Shyamalee and Meedeniya (2022)
Channel-wise gamma correction	Segmentation of retinal vessels	F1 score, AUC, accuracy, specificity, and sensitivity	AUC, accuracy, F1 score, specificity, sensitivity, are 0.9893, 0.9683, 0.8082, 0.9745, and 0.8908, respectively	The segmentation can increase the robustness and performance of a traditional CNN model	Sun et al. (2021)
Zooming and CLAHE	DR detection	Accuracy, specificity, and sensitivity	Accuracy, sensitivity, and specificity values are 98%, 96%, 100% respectively	The proposed approach is able to detect DR lesions with high performance	Agustin et al. (2020)
Blurring and shifting	Binary classification (as healthy and nonhealthy), multi-class (5 classes) classification DR stages	Precision, specificity, MCC, accuracy, F1 score, sensitivity, relative classifier information (RCI), confusion entropy (CEN), geometric mean (GM), and index of balanced accuracy (IBA)	Accuracy, F1 score, MCC, sensitivity, specificity, precision are 59.89%, 63.32%, 0.1908, 63.64%, 55.42%, 63%, respectively, for binary classification, Overall classification, RCI, average accuracy, GM, MCC are 33.12%, 0.0647, 0.7772, 0.1463, 0.5209, 0.1642 respectively, for multi-class classification.	Combined augmentation improved the performance of binary classification while reducing the performance of multi-class classification	Tufail et al. (2021)
Deep convolutional GAN	DR severity classification	F1 score, recall, and precision	F1 score, recall, and precision values improved by 0.06, 0.05, and 0.04 respectively after augmentation	Classification performance improved after augmentation	Balasubramanian et al. (2020)

Table 4 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Conditional GAN	DR severity classification	Accuracy, and kappa coefficient	Augmented images increased the mean accuracy by 1.08% and the kappa by 1.15% AUC is 0.988	The method improved the performance of DR severity grading	Zhou et al. (2020)
Style based GAN	DR severity classification	AUC	AUC is 0.988	Using a sufficient number of synthetic and real images increases the robustness of the classifier	Lim et al. (2020)
Cycle GAN	DR classification, tessellated fundus segmentation, lesion detection (hemorrhages, soft and hard exudate lesions)	Accuracy, F1 score	Classification accuracy increased from 68.28–70.57%, segmentation accuracy increased from 52.27% to 68.18 after training the model with augmented images, F1 score for detection is 66.27%	The proposed model improved the performance of classification, segmentation and detection tasks	Ju et al. (2021)
Heuristic augmentation with NV-like structures	Proliferative DR detection	Kappa	Kappa values for three datasets are 0.78, 0.74, and 0.70	Proliferative DR detection performance improved with the expanded datasets	Ara'ujo et al. (2020)
Rotation, translation, mirroring, zooming	Malarial retinopathy detection	Sensitivity, specificity, and accuracy	Sensitivity, specificity, accuracy values are 90%, 100%, and 81% respectively	The presented method demonstrates the potential of accurate malarial retinopathy detection	Kurup et al. (2020)

Table 4 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Rotation, flipping, scaling, clipping, and translation	DR classification (symptomatic and asymptomatic)	AUC, accuracy, sensitivity, F1 Score, and specificity	AUC, accuracy, sensitivity, F1 Score, and specificity are 99.42%, 99.42%, 99.45%, 99.41%, and 99.38% respectively	The proposed model is efficient in classification of DR with high performance	Asia et al. (2020)
Rotation, cropping, brightness changing	Exudates and drusen detection	AUC	AUC is 0.994, 0.972, and 0.988 in three datasets	The system achieved high sensitivity and specificity	Li et al. (2022)
Flipping, rotation, style GAN	Detection of chronic ocular diseases	F1 score, AUC, and Kappa	F1 score, kappa, and AUC values are 0.9070, 0.6081, and 0.8806, respectively	The presented model is superior than the recent techniques in detection of chronic ocular diseases	Mayya et al. (2022)
Translation, stretching, rotation, and flipping	Exudates classification	Accuracy, specificity, and sensitivity	Sensitivity, specificity, accuracy values for one dataset are 98.56%, 92.54%, 97.20%, and for another dataset are 98.99%, 91.87%, and 97.67%, respectively	The proposed approach outperforms several existing classifiers such as random forest	Ramya et al. (2022)
Rotation, cropping	Glaucoma prediction	Sensitivity, specificity, F1 score, precision, AUC, and accuracy	Sensitivity, specificity, F1 score, precision, AUC, accuracy are 93.78%, 97.89%, 95.44%, 97.17%, 95.84%, and 96.11% respectively	The proposed model has clinical potential in automatic glaucoma prediction	Singh et al. (2022)
Rotation, flipping, cropping, and shifting	Cataract diagnosis	Sensitivity, specificity, F1 score, precision, AUC, and accuracy	Sensitivity, specificity, F1 score, precision, AUC, and accuracy are 96.25%, 98.74%, 97.47%, 96.24%, 96.25%, and 97.96%, respectively	The suggested model outperformed existing models applied for cataract diagnosis	Yadav and Yadav (2022)

Table 4 (continued)

Augmentation method	Application	Evaluation metrics	Results	Conclusions	References
Translation, rotation, and flipping	Glaucoma diagnosis	Sensitivity, specificity, F1 score, precision, AUC, and accuracy	Sensitivity, specificity, F1 score, precision, AUC, and accuracy are 0.538, 0.956, 0.490, 0.457, 0.904, and 0.930, respectively	The proposed model produced a good performance in Glaucoma diagnosis	Lin et al. (2022)
Rotation, flipping	Detection of DR severity level	Accuracy, precision, recall, F1-score, and AUC	Accuracy, precision, recall, F1-score, and AUC are 84.36%, 73.84%, 70.51%, 70.49%, and 93.82%, respectively	The presented approach achieved better performance compared to the existing methods	Islam et al. (2022)

those applications, different network structures, data sets and number of images have been used, and also their performances have been evaluated with different metrics. High performances have been obtained from those applications according to the reported results. However, the most effective and appropriate augmentation method used in them with eye fundus images is not clear due to those differences. Also, it is not possible to determine augmentation methods' contributions to those performances according to the presented results in those publications.

2.5 Commonly applied augmentation methods

2.5.1 Rotation

Rotation-based image augmentations are provided by rotating an image by concerning its original position. The rotation uses a new coordinate system and retains the same relative positions of the pixels of an image. It can be rightward or leftward across an axis within the range of $[1^\circ, 359^\circ]$. Image labels may not always be preserved if the degree of rotation increases. Therefore, the safety of this augmentation technique depends on the rotation degree. Although rotation transformation is not safe on images showing 6 and 9 in digit recognition applications, it is generally safe on medical images.

2.5.2 Flipping

Flipping is a technique generating a mirror image from an image. The pixel's positions are inverted by concerning one of the two axes (for a two-dimensional image). Although it can be applied with vertical or/and horizontal axes, vertical flipping is rarely preferred since the bottom and top regions of an image may not always be interchangeable (Nalepa et al. 2019).

2.5.3 Translation

Images are translated along an axis using a translation vector. This technique preserves the relative positions between pixels. Therefore, translated images provide prevention of positional bias (Shorten and Khoshgoftaar 2019) and the models do not focus on properties in a single spatial location (Nalepa et al. 2019).

2.5.4 Scaling

Images are scaled along different axes with a scaling factor, which can be different or the same for each axis. Especially, scaling changings can be interpreted as zoom out (when the scaling factor is less than 1) or zoom in (when the scaling factor is greater than 1).

2.5.5 Shearing

This transformation slides one edge of an image along the vertical or horizontal axis, creating a parallelogram. A vertical direction shear slides an edge along the vertical axis,

while a horizontal direction shear slides an edge along the horizontal axis. The amount of the shear is controlled by a shear angle.

2.5.6 Augmentation using intensities

This augmentation is provided by modification of contrast or brightness, blurring, intensity normalization, histogram equalization, sharpening, and addition of noise. If the preferred type of noise is Gaussian, intensity values are modified by sampling a Gaussian distribution randomly. If it is salt-and-pepper type of noise, pixel values are set randomly to white and black. Uniform noise addition is performed by modification of pixel values using randomly sampling a uniform distribution.

2.5.7 Random cropping and random erasing

Augmentation by random cropping is applied by taking a small region from an original image and resizing it to match the dimensions of the original image. Therefore, this augmentation can also be called scaling or zooming. Augmentation by random erasing is performed by randomly eliminating image regions.

2.5.8 Color modification

Image augmentation by color modification is performed in several ways. For example, an RGB-colored image is stored as arrays, which correspond to 3 color channels representing levels of red, green, and blue intensity values. Because of this, the color of the image can be changed without losing its spatial properties. This operation is called color space shifting, and it is possible to perform the operation in any number of spaces because the original color channels are combined in different ratios in every space. Therefore, a way to augment an RGB-colored image by color modification is to put the pixel values in the color channels of the image into a histogram and then manipulate them using filters to generate new images by changing the color space characteristics. Another way is to take 3 shifts (integer numbers) from 3 RGB filters and add each shift into a channel in the original image. Also, image augmentation can be provided by updating values of hue, saturation, and value components, isolating a channel (such as a blue, red, or green color channel), and converting color spaces into one another to augment images. However, converting a color image to its grayscale version should be performed carefully because this conversion reduces performance by up 3% (Chatfield et al. 2014).

2.5.9 GAN based augmentation

GANs are generative architectures constructed with a discriminator to separate synthetic and true images and a generator to generate synthetic realistic images. The main challenging issues of GAN based image augmentations are generating images with high quality (i.e., high-resolution, clear) and maintaining training stability. To overcome those issues, several GAN variants were developed such as conditional, deep convolutional, and cycle GAN. Conditional GAN adds parameters (e.g., class labels) to the input of the generator to control generated images and allows conditional image generation from the generator. Deep convolutional GAN uses deep convolution networks in the GAN structure.

Cycle GAN is a conditional GAN and translates images from one domain to another (i.e., image-to-image translation), which modifies input images into novel synthetic images that meet specific conditions. Wasserstein GAN is the structure that is constructed with the loss function designed with the earth-mover distance to increase the stability of the model. Style GAN and progressively trained GAN add coarse to fine details and learn features from training images to synthesize new images.

3 Materials and methods

3.1 Materials

In this work, publicly available brain MR, lung CT, breast mammography, and eye fundus images have been used.

Brain MR images have been provided from Multimodal Brain Tumor Image Segmentation Challenge (BRaTS) (Bakas et al. 2018). The data for each patient includes four diverse models of MR images which are Fluid-Attenuated Inversion Recovery (FLAIR) images, T2-weighted MR images, T1-weighted MR images with Contrast Enhancement (T1-CE), and T1-weighted MR images. Different modal images provide different information about tumors. In this work, FLAIR images have been used. Example FLAIR images showing High-Grade Glioma (HGG) (Fig. 1a, b) and Low-Grade Glioma (Fig. 1c, d) (LGG) are presented in Fig. 1. The BRaTS database contains 285 images including HGG (210 images) and LGG (75 images) cases.

Lung images have been provided from Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database (Armato et al. 2011; McNitt-Gray et al. 2007; Wang et al. 2015) by the LUNA16 challenge (Setio et al. 2017). In the database, there are 1018 CT scans which belong to 1010 people. Those scans were annotated by four experienced radiologists. The nodules that were marked by those radiologists were divided into three groups which are “non-nodule ≥ 3 millimeters”, “nodule < 3 millimeters”, and “nodule ≥ 3 millimeters” (Armato et al. 2011). The images having a slice thickness greater than 3 mm were removed due to their usefulness according to the suggestions (Manos et al. 2014; Naidich et al. 2013). Also, the images having missing slices or not consistent slice spacing were discarded. Therefore, there are a total of 888 images extracted from 1018 cases. Example images with nodules (Fig. 2a) and without nodules (Fig. 2b) have been presented in Fig. 2.

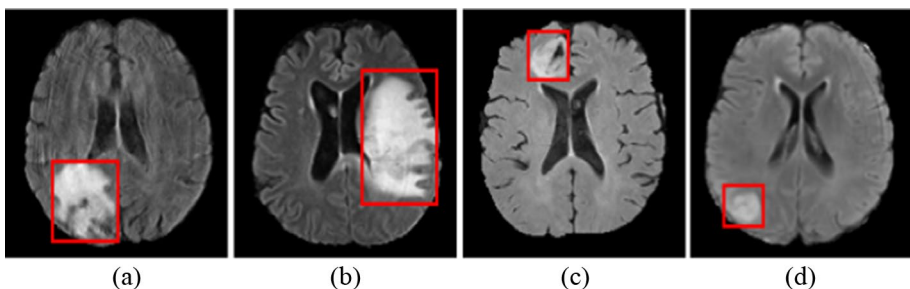


Fig. 1 Brain MR scans showing HGG (a, b) and LGG (c, d) (Bakas et al. 2018)

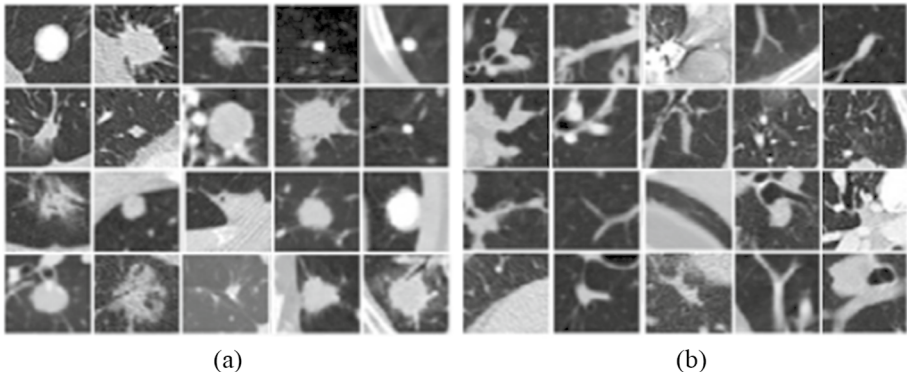


Fig. 2 Example images showing nodules (a) and non-nodule regions (b) (Armato et al. 2011)

Breast mammography images have been provided from the INbreast database (Moreira et al. 2012). It contains 410 images, of which only 107 images show breast masses, and the total number of mass lesions is 115 since an image may contain more than one lesion. The number of benign and malignant masses is 52 and 63, respectively. Example mammography images without any mass (Fig. 3a), with a benign mass (Fig. 3b), and malignant mass (Fig. 3c) lesions have been presented in Fig. 3.

Eye fundus images have been provided from the Messidor database (Decenci re et al. 2014) including 1200 images categorized into three classes (0-risk, 1-risk, and 2-risk) according to the diabetic macular edema risk level. The number of images in the

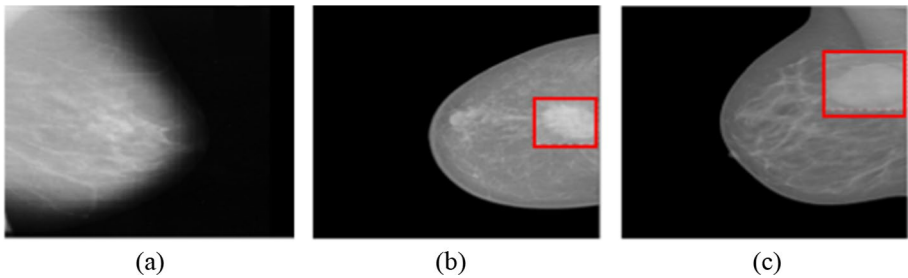


Fig. 3 Mammography images without mass (a), a benign mass (b), and a malignant mass (c) (Moreira et al. 2012)

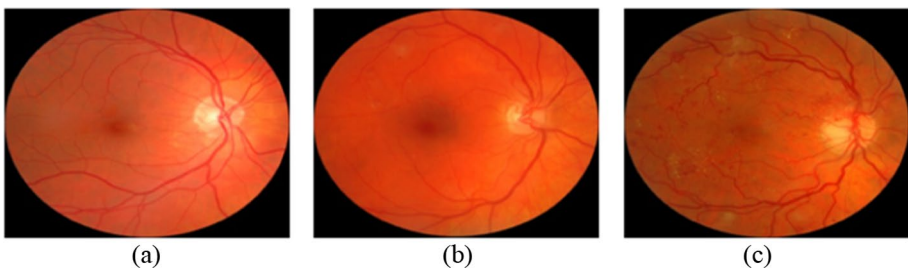


Fig. 4 Eye fundus images showing macular edema with 0-risk (a), 1-risk (b), and 2-risk (c) (Decenci re et al. 2014)

categories is 974, 75, and 151, respectively. Example images showing macular edema with 0-risk (Fig. 4a), 1-risk (Fig. 4b), and 2-risk (Fig. 4c) have been presented in Fig. 4.

3.2 Implemented augmentation methods

In this work, commonly used augmentation methods based on transformations and intensity modifications have been implemented to see and compare their effects on the performance of deep learning based classifications from MR, CT, and mammography images. Also, a method based on a color modification to augment colored eye fundus images has been applied. Those methods are explained below.

1st Method In this method, a shearing angle, which is selected randomly within the range $[-15^\circ, 15^\circ]$, is used and a shearing operation is applied within the x and y-axis. This operation is repeated 10 times with 10 different shearing angles. By this method, 10 images are generated from an original image.

2nd Method In this method, translation operation is applied on the x and y-axis. This operation is repeated 10 times by using different values sampled randomly within the range $[-15, 15]$. By this method, 10 images are generated from an original image.

3rd Method Rotation angles are selected randomly within the range $[-25^\circ, 25^\circ]$ and clockwise rotations are applied using 10 different angle values. 10 images are generated from an original image by this method.

4th Method In this method, noise addition is applied by $I = I + (\text{rand}(\text{size}(I)) - FV) \times N$, where the term FV refers to a fixed value, N refers to noise and I refers to an input image. The noise to be added is generated using a Gaussian distribution. The variance of the distribution is computed from the input images while the mean value is set to zero. In this study, three different fixed values (0.3, 0.4, and 0.5) are used and 3 images are generated from an input image.

5th Method In this method, noise is added to each image with 3 density values (0.01, 0.02, and 0.03) and a salt-and-pepper type noise is used. Therefore, 3 images are generated from an input image by this method.

6th Method In this method, augmentation is provided by salt-and-pepper type noise addition (5th method) followed by shearing (1st method). 30 images are generated from an input image by this method.

7th Method In this method, augmentation is provided by Gaussian noise addition (4th method) followed by rotation (3rd method). 30 images are generated from an input image by this method.

8th Method In this augmentation method, clockwise rotation and then translation operations are applied. Rotation angles are selected randomly within the range $[-25^\circ, 25^\circ]$ and translation is applied on the x and y-axis by using different values sampled randomly within the range $[-15, 15]$. This successive operation is applied 10 times, and 10 images are generated from an input image by this method.

9th Method In this method, translation is applied followed by the shearing. The translation step is applied on the x and y-axis with different values sampled randomly within the range $[-15, 15]$. In the shearing step, an angle is selected randomly within the range $[-15^\circ, 15^\circ]$, and shearing is applied within the x and y-axis. The subsequent operations are applied 10 times and 10 images are generated from an input image by this method.

10th Method In this method, three operations (translation, shearing, and rotation) are applied subsequently. Implementation of the translation and shearing operations are performed as explained in the 9th method. In the rotation step, a random angle is selected within the range $[-25^\circ, 25^\circ]$ clockwise rotation is applied. The successive operations are applied 10 times, and 10 images are generated from an input image by this method.

11th Method Color shifting, sharpening, and contrast changing have been combined in this method. Color shifting has been applied by taking 3 shifts (integer numbers) from 3 RGB filters and adding each shift into one of the 3 color channels in the original images. Sharpening has been changed by blurring the original images and subtracting the blurred images from the original images. A Gaussian filter, whose variance value is 1, has been used for the blurring process. Contrast changing has been applied by scaling the original images linearly between i_1 and i_2 ($i_1 < i_2$), which will be the maximum and minimum intensities in the augmented images, and then mapping the original images' pixels with intensity values higher than i_2 (or lower than i_1) to 255 (or 0). Therefore, 3 images are generated from an input image by this method.

3.3 Classification

In this work, to compare the effectiveness of the augmentation approaches in classifications, a CNN-based classifier has been used. In this step, the original images taken from the public databases (Sect. 3.1) have been used in construction of training and testing datasets. However, the databases have imbalanced class distributions. Therefore, to obtain unbiased classification results, an equal number of images have been taken from the databases for the classes. Those balanced numbers of images have been used with the augmentation methods and the generated images from them have been added into the data sets. 80% of the total data has been used in the training stage, the rest 20% of the data has been used in the testing stage. It should be reminded here that the goal in this study is to evaluate the augmentation techniques used with different types of images in improving the classification rather than evaluations of classifier models. Therefore, ResNet101 architecture has been chosen and implemented in this work. The reason to choose this architecture is to utilize its advantage in addressing the gradient vanishing issue and high learning ability (He et al. 2016). Pre-training of networks by using large datasets [such as ImageNet (Russakovsky et al. 2015)] and then fine-tuning the networks for target tasks having less training data has been commonly applied in recent years. Pre-training has provided superior results on many tasks, such as action recognition (Simonyan and Zisserman 2014; Carreira and Zisserman 2017), image segmentation (He et al. 2017; Long et al. 2015), and object detection (Ren et al. 2015; Girshick 2015; Girshick et al. 2014). Therefore, transfer learning from a model trained using ImageNet has been used in this work for fine-tuning of the network. Optimization has been provided by Adam, and activation has been provided by ReLU. The initial learning rate has been chosen as 0.0003, and cross-entropy loss has been computed to update the weights in the training phase. The number of epochs and mini-batch size have been set as 6 and 10, respectively.

4 Results

Classification results have been evaluated by using these five measurements: specificity, accuracy, sensitivity, Matthew Correlation Coefficient (MCC), and F1-score computed by:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$F1_{score} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

where FP, FN, TN, and TP refer to false positives, false negatives, true negatives, and true positives, respectively. The number of images obtained from each augmentation method for each augmentation approach is given in Table 5.

The augmentation methods and quantitative results obtained by classifications of HGG and LGG cases from brain MR images have been presented in Table 6.

Lung CT scans have been classified as a pulmonary nodule or a non-nodule with the same classifier. 1004 lung nodules of which 450 positive candidates provided by the LIDC-IDRI have been used. After generation more positive candidate nodules with the augmentation methods, equal number of negative and positive candidates have been used in the classifications. The augmentation methods and quantitative results of classifications have been presented in Table 7.

Table 5 The number of images from each augmentation method

Image augmentation method	Operations	Number of augmented images from 1 image
1st augmentation method	Shearing	10
2nd augmentation Method	Translation	
3rd augmentation method	Rotation	
4th augmentation method	Gaussian noise addition	3
5th augmentation method	Salt and pepper noise addition	
6th augmentation method	Salt and pepper noise addition and shearing	30
7th augmentation method	Gaussian noise addition and rotation	30
8th augmentation method	Rotation and translation	10
9th augmentation method	Translation and shearing	10
10th augmentation method	Translation, shearing and rotation	10
11th augmentation method	Color shifting, sharpening, contrast changing	3

Table 6 Results of classifications from brain MR images

Augmentation method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	MCC (%)
1st augmentation method	83.33	89.18	76.19	84.44	67.36
2nd augmentation method	80.38	80.57	93.19	80.44	70.67
3rd augmentation method	84.31	84.16	92.24	84.19	76.43
4th augmentation method	79.52	79.34	96.68	78.65	78.81
5th augmentation method	78.57	85.71	71.43	80.00	58.74
6th augmentation method	62.38	65.16	70.06	62.51	56.08
7th augmentation method	85.24	86.10	96.64	84.35	81.41
8th augmentation method	85.69	80.95	90.78	85.00	71.46
9th augmentation method	83.81	84.82	94.38	83.06	80.22
10th augmentation method	88.24	90.48	97.13	88.52	82.79

Table 7 Results of classifications from lung CT images

Augmentation method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	MCC (%)
1st augmentation method	80.48	80.07	95.81	78.72	76.52
2nd augmentation method	60.95	63.13	73.57	60.28	51.97
3rd augmentation method	74.51	74.44	87.30	74.52	61.74
4th augmentation method	84.12	85.10	96.11	83.15	81.62
5th augmentation method	72.55	74.21	86.29	72.86	59.61
6th augmentation method	62.38	65.16	93.76	65.52	57.08
7th augmentation method	75.00	83.38	76.67	76.92	54.71
8th augmentation method	76.19	85.71	77.48	78.26	53.36
9th augmentation method	85.74	86.16	97.65	84.35	82.81
10th augmentation method	83.82	84.83	96.29	82.24	81.24

Table 8 Results from classifications from mammography images

Augmentation method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	MCC (%)
1st augmentation method	81.70	82.72	96.09	81.08	75.93
2nd augmentation method	70.00	73.30	84.92	70.33	55.88
3rd augmentation method	73.33	74.72	86.75	73.50	60.68
4th augmentation method	76.58	76.84	88.40	76.62	65.10
5th augmentation method	80.00	84.19	91.27	79.74	72.86
6th augmentation method	66.67	71.48	83.70	65.93	53.41
7th augmentation method	78.57	78.45	91.52	77.72	74.73
8th augmentation method	76.19	84.71	66.67	78.26	56.36
9th augmentation method	76.67	77.50	88.60	76.53	65.49
10th augmentation method	83.34	85.29	97.21	83.60	78.95

Table 9 Results of classifications from eye fundus images

Augmentation method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	MCC (%)
1st augmentation method	79.52	79.34	93.68	78.65	78.81
2nd augmentation method	73.81	76.15	66.67	75.56	48.11
3rd augmentation method	84.31	84.16	92.24	84.19	76.43
4th augmentation method	86.24	86.81	93.32	86.22	79.73
5th augmentation method	82.17	82.43	91.14	82.34	73.46
6th augmentation method	78.43	78.59	89.24	78.32	67.58
7th augmentation method	80.39	80.81	90.21	80.47	70.79
8th augmentation method	82.35	82.67	91.19	82.56	73.68
9th augmentation method	78.43	80.16	89.31	78.45	68.01
10th augmentation method	82.35	82.59	91.25	82.32	73.67
11th augmentation method	88.24	89.18	94.13	88.52	82.79

The breast mammography images have been classified into three groups as malignant, benign, and normal. The augmentation methods and quantitative results of classifications have been presented in Table 8.

Unlike from MR, CT, and mammography images, the eye fundus images are colored. Therefore, in addition to those 10 augmentation methods applied for grayscale images, the 11th method has also been used for augmentation of this images. The augmentation methods and quantitative results of the classifications into three classes according to macular edema risk level have been presented in Table 9.

5 Discussions

To overcome the imbalance distribution and data scarcity problems in deep learning based approaches with medical images, data augmentations are commonly applied with various techniques in the literature (Sect. 2). In those approaches, different network architectures, different types of images, parameters, and functions have been used. Although they achieved good performances with augmented images (Tables 1, 2, 3 and 4), the results presented in the publications do not indicate on which aspects of the approaches contributed the most. Although augmentation methods are applied as pre-processing steps and therefore have significant impacts on the remaining steps and overall performance, their contributions are not clear. Because of this, it is not clear which data augmentation technique provides more efficient results for which image type. Therefore, in this study, the effects of augmentation methods have been investigated to determine the most appropriate method for automated diagnosis from different types of medical images. This has been performed by these steps: (i) The augmentation techniques used to improve the performance of deep learning based diagnosis of diseases in different organs (brain, lung, breast, and eye) have been reviewed. (ii) The most commonly used augmentation methods have been implemented with four types of images (MR, CT, mammography, and funduscopy). (iii) To evaluate the effectiveness of those augmentation methods in classifications, a classifier model has been designed. (iv) Classifications using the augmented images for each augmentation method have been performed with the same classifier and datasets for fair

Table 10 The pros and cons of traditional augmentation approaches

Approach	Pros	Cons
Augmentations with geometric transformations	Augmentations with geometric transformations (e.g., scaling, rotation, shearing, and flipping) are effective in increasing the amount and the diversity of the datasets by adding modified versions of existing images. Because the generated new images are identifiable and resembling	The new images may not represent enough variations in the location, shape and pathology of a lesion since the contents of the real and the newly produced images are similar
Scaling, rotation	Since lesions can appear at different positions and scales, adding rotated and scaled images into the training sets help the models in learning	
Scaling	Image augmentation using scaling can produce realistic images since lesions can vary in size. Also, the usage of augmented images obtained by scaling allows network models to learn properties without depending on the original scale	
Translation	Augmentation using translation provides the prevention of positional bias. In other words, training a network model with augmented images using translation enables the network to learn geographically invariant features. The network model do not focus on properties in a single spatial location	
Flipping	They can preserve the labels and features of the reconstructed images. The label-preserving techniques are helpful in classification tasks. For example, after the translation of an image showing HGG in a classification problem, the image remains intact	Label-preserving property may not always be provided since profile information can be on the left or right side of the images. For example, breast profiles are mostly on the leftward of mammography images in some datasets. Therefore, flipping may not conserve the labels for the images in those datasets

Table 10 (continued)

Approach	Pros	Cons
Modification of contrast or brightness, intensity normalization, histogram equalization, noise addition and blurring, and sharpening	Medical images are acquired with various imaging modalities and technologies, and they can be diversified in pixel intensities. Therefore, generating new images by changing the intensity values of existing images (e.g., modification of contrast or brightness, intensity normalization, histogram equalization, noise addition and blurring, and sharpening) can help to augment medical images and provide an improvement in the diversity of datasets	Augmentation by changing the intensity values, particularly noise addition and blurring techniques may disturb the quality of the images (i.e., disintegrate images' features) and change the lesions' characteristics, causing the generation of unnatural images and low accuracy
Random erasing, random cropping	Augmentations with random erasing or cropping are to improve the robustness of a network. Also, they are applied without depending on any parameter	Augmentations with random erasing, cropping, and color space transformations may lead to the loss of significant information
Color space transformations	Through color space transformations, biases in an image due to various illumination can be prevented	

comparisons of those augmentation methods. (v) The effectiveness of the augmentation methods in classifications have been discussed based on the quantitative results.

It has been observed from the experiments in this work that transformation-based augmentation methods are easy to implement with medical images. Also, their effectiveness depends on the characteristics (e.g., color, intensity, etc.) of the images, and whether all significant visual features according to medical conditions exist in the training sets. If most of the images in the training sets have similar features, then there is a risk of constructing a classifier model with overfitting and less generalization capability. In general, the pros and cons of the widely used traditional augmentation approaches are presented in Table 10.

Generating synthetic images by GAN based methods can increase diversity. However, GANs have their own challenging problems. A significant problem is vanishing gradient and mode collapsing (Yi et al. 2019). Since the collapse of the mode restricts the capacity of GAN architecture to be varied, this interconnection is detrimental for augmentations of medical images (Zhang 2021). Another significant problem is that it is hard to acquire satisfactory training results if the training procedure does not assure the symmetry and alignment of both generator and discriminator networks. Artifacts can be added while synthesizing images, and therefore the quality of the produced images can be very low. Also, the trained model can be unpredictable since GANs are complex structures and managing the coordination of the generator and discriminator is difficult. Furthermore, they share similar flaws with neural networks (such as, poor interpretability). Additionally, they require strong computer resources, a long processing time, and hamper the quality of the generated images in case of running on computers having low computational power with constrained resources.

Quality and diversity are two critical criteria in the evaluation of synthetically generated images from GANs. Quality criterion indicates the level of similarity between the synthetic and real images. In other words, it shows how representative synthetic images are of that class. A synthetic image's quality is characterized by a low level of distortions, fuzziness, noise (Thung and Raveendran 2009), and its feature distribution matching with the class label (Zhou et al. 2020; Costa et al. 2017; Yu et al. 2019).

Diversity criteria indicates the level of dissimilarity of the synthetic and real images. In other words, it shows how uniform or wide the feature distribution of the synthetic image is (Shmelkov et al. 2018). When the training datasets are expanded with the images lacking diversity, the datasets can provide only limited coverage of the target domain and cause low classification performance due to incorrect classifications of the images containing features that belong to the less represented regions. When the training datasets are expanded by adding synthetic images with low quality, the classifier cannot learn the features representing different classes which leads to low classification performance. Therefore, if synthetic images are used in the training sets, they should be sufficiently diverse to represent features of each class and have high quality.

The diversity and quality of the new images produced by GANs are evaluated manually by a physician or quantitatively by using similarity evaluation measurements, which are generally FID, structural similarity index, and peak signal-to-noise ratio (Borji 2019; Xu et al. 2018). The validity of these measurements for medical image data sets is still under investigation and there is no accepted consensus. Manual evaluations are subjective as well as time-consuming (Xu et al. 2018) since they are based on the domain knowledge of the physician. Also, there is no commonly accepted metric for evaluating the diversity and quality of synthesized images.

About Brain MR Image Augmentation techniques The methods used for brain MR image augmentation have some limitations or drawbacks. For instance, in a study

(Isensee et al. 2020), the augmentation approach uses elastic deformations, which add shape variations. However, the deformations can bring lots of damage and noise when the deformation field is varied seriously. Also, the generated images seem not to be realistic and natural. It has been shown in the literature that widely used elastic deformations produce unrealistic brain MR images (Mok and Chung 2018). If the simulated tumors are not in realistic locations, then the classifier model can focus on the lesion's appearance features and be invariant to contextual information. In another study (Kossen et al. 2021), although the generated images yielded a high dice score in the transfer learning approach, performance only slightly improved when training the network with real images and additional augmented images according to the results presented by the authors. This might be because of the less blurry and noisy appearance of the images produced by the used augmentation. The discriminator observed results of the generator network at intermediate levels in the augmentation with multi-scale gradient GAN (Deepak and Ameer 2020). Because the proposed GAN structure included a single discriminator and generator with multiple connections between them. Although augmentation with TensorMixup improved the diversity of the dataset (Wang et al. 2022), its performance should be further verified with more datasets including medical images with complicated patterns as samples generated with this augmentation may not satisfy the clinical characterizations of the images. Li et al. (2020) has observed that an unseen tumor label cannot be provided with the augmentation method and therefore the virtual semantic labels' diversity is limited.

It has been observed that usage of the augmented images obtained by the combination of rotation with shearing, and translation provides higher performance than the other augmentation techniques in the classifications of HGG and LGG cases from FLAIR types of brain images. On the other hand, augmentation with the combination of shearing and salt-and-pepper noise addition is the least efficient approach for augmentation in improving classification performance (Table 6).

About Lung CT Image Augmentation techniques Further studies on augmentation techniques for lung CT images are still needed since current methods still suffer from some issues. For instance, in a study (Onishi et al. 2020), although performance in the classification of isolated nodules and nodules having pleural tails increased, it did not increase in the classification of nodules connected to blood vessels or pleural. The reason might be due to the heavy usage of isolated nodule images (rather than the images with the nodules adjacent to other wide tissues like the pleural) for the training of the GAN. Also, the Wasserstein GAN structure leads to gradient vanishing problems due to small weight clipping and a long time to connect because of huge clipping. In another study (Nishio et al. 2020), the quality of the generated 3-dimensional CT images that show nodules is not low. On the other hand, in some of those images, the lung parenchyma surrounding the nodules does not seem natural. For instance, the CT values around the lung parenchyma in the generated images are relatively higher than the CT values around the lung parenchyma in the real images. Also, lung vessels, chest walls, and bronchi in some of those generated images are not regular. The radiologists easily distinguish those generated images according to the irregular and unnatural structures. The augmentation method in a different work (Nishio et al. 2020) generate only lung nodules' 3-dimensional CT images. However, there exist various radiological findings (e.g., ground glass, consolidation, and cavity) and it is not clear whether those findings are generated or not. Also, the application can classify nodules only according to their sizes rather than other properties (e.g., absence or presence of spicules, margin characteristics, etc.). Therefore, further evaluation should be performed to see whether nodule classification performance can be increased with the new lung

nodules for other classifications, such as malignant or benign cases. A constant coefficient was used in the loss function to synthesize lung nodules by Wang et al. (2021). It affects the training performance and should be chosen carefully. Therefore, the performance of the augmentation method should be evaluated with increased a number of images. Although the method proposed by Toda et al. (2021) has the potential to generate images, spicula-like features are obscure in the generated images and it does not include the distribution of true images. Also, the generation of some features contained in the images (e.g., cavities around or inside the tumor, prominent pleural tail signs, non-circular shapes) are difficult with the proposed augmentation. Therefore, images with those features tend to be misclassified. In a study, augmentation with elastic deformations, which add shape variations, brings noise and damage when the deformation field is varied seriously (Müller et al. 2021). In a different approach, as it is identified by the authors, the predicted tumor regions are prone to imperfections (Farheen et al. 2022).

It has been observed that usage of the augmented images obtained by the combination of translation and shearing techniques in the classification of images with nodules and without nodules provides the highest performance than the other augmentation methods. On the other hand, augmentation only by shearing is the least efficient approach (Table 7).

About breast Mammography Image Augmentation techniques Augmentation of breast mammography images is one of the significant and fundamental directions that need to be focused on for further investigations and future research efforts. Although scaling, translation, rotation, and flipping are widely used, they are not suitable enough for augmentation since the additional information provided by them is not sufficient to make variations in the images, and so the diversity of the resulting dataset is limited. For instance, in a study performed by Zeiser et al. (2020), although the classifier classifies pixels having high intensity values as masses as it is desired, the network ends up producing FPs in dense breasts. Therefore, the augmentation technique applied as a pre-processing step should be modified to expand the datasets more efficiently and to increase the generalization skill of the proposed classifier. Besides, the performance of the applications should be tested with not only virtual images but also real-world images. Although the results of the applications that use generated images from a GAN based augmentation are promising (Table 3), their performances should be evaluated with increased numbers and variations of images. Since augmentations by synthesizing existing mammography images are difficult because of the variations of masses in terms of shape and texture as well as the presence of diverse and intricate breast tissues around the masses. The common problems in those GAN based augmentations are mode collapsing and saddle point optimization (Yadav et al. 2017). In the optimization problem, there exists almost no guarantee of equilibrium between the training of the discriminator and generator functions, causing one network to inevitably become stronger than the other network, which is generally the discriminator function. In the collapsing mode problem, the generator part focuses on limited data distribution modes and causes the generation of images with limited diversity.

Experiments in this study indicated that the combined technique consisting of translation, shearing, and rotation is the most appropriate approach for the augmentation of breast mammography images in order to improve the classification of the images as normal, benign, and malignant. On the other hand, the combined technique consisting of salt and pepper noise addition and shearing is the least appropriate approach (Table 8).

About Eye Fundus Image Augmentation techniques Further research on the augmentation techniques for fundus images is still needed to improve the reliability and robustness of computer-assisted applications. Because the tone qualities of fundus images are affected by the properties of fundus cameras (Tyler et al. 2009) and the images used

in the literature have been acquired from different types of fundus cameras. Therefore, the presented applications fitting well to images obtained from a fundus camera may not generalize images from other kinds of fundus camera systems. Also, images can be affected by pathological alterations. Because of this reason, the applications should be robust on those alterations, particularly on the pathological changes that do not exist in the images used in the training steps. Therefore, although the results of the applications in the current literature indicated high performances (Table 4), their robustness should be evaluated using other data sets with increased numbers and variations of images and taken from different types of fundus cameras. Also, the contributions of the applied augmentation techniques to the presented performances are not clear. In a study performed by Zhou et al. (2020), although the GAN structure is able to synthesize high-quality images in most cases, the lesion and structural masks used as inputs are not real ground truth images. Therefore, the generator's performance depends on the quality of those masks. Also, the applied GAN architecture fails to synthesize some lesions, such as microaneurysms. In another study performed by Ju et al. (2021), the GAN based augmentation might lead to biased results because of matching the generated images to the distribution of the target domain.

Experiments in this study indicated that usage of the augmented images obtained by the combination of color shifting with sharpening, and contrast changing provides higher performance than the other augmentation techniques in the classifications of eye fundus images. On the other hand, augmentation with translation is the least efficient approach for augmentation to improve classification performance (Table 9).

6 Conclusion

Transformation-based augmentation methods are easy to implement with medical images. Also, their effectiveness depends on the characteristics (e.g., color, intensity, etc.) of the images, and whether all significant visual features according to medical conditions exist in the training sets.

GAN based augmentation methods can increase diversity. On the other hand, GANs have vanishing gradient and mode collapsing problems. Also, obtaining satisfactory training results is not easy if the training procedure does not assure the symmetry and alignment of both generator and discriminator networks. Besides, GANs are complex structures and managing the coordination of the generator and discriminator is difficult.

Combination of rotation with shearing, and translation provides higher performance than the other augmentation techniques in the classifications of HGG and LGG cases from FLAIR types of brain images. However, augmentation with the combination of shearing and salt-and-pepper noise addition is the least efficient approach for augmentation in improving classification performance.

Combination of translation and shearing techniques in the classification of lung CT images with nodules and without nodules provides the highest performance than the other augmentation methods. However, augmentation only by shearing is the least efficient approach.

Combination of translation, shearing, and rotation is the most appropriate approach for the augmentation of breast mammography images in order to improve the classification of the images as normal, benign, and malignant. On the other hand, the combined technique consisting of salt and pepper noise addition and shearing is the least appropriate approach.

Combination of color shifting with sharpening, and contrast changing provides higher performance than the other augmentation techniques in the classifications of eye fundus images. On the other hand, augmentation with translation is the least efficient approach for augmentation to improve classification performance.

As an extension of this work, the effectiveness of the augmentation methods will be evaluated in the diagnosis of diseases from positron emission tomography, ultrasonography images, other types of MR sequences (e.g., T2, T1, and proton density-weighted), and also in the classification of other types of images such as satellite or natural images. Also, GAN based augmentations will be applied, quantitative and qualitative analyses of the generated images will be performed to ensure their diversity and realness. In this study, ResNet101 has been used due to its advantage based on residual connections and efficiency in classification. Therefore, the implementation of other convolutional network models will be performed in our future works.

Author contributions It is a single-authored article.

Declarations

Competing Interests The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Agustin T, Utami E, Al Fatta H (2020) Implementation of data augmentation to improve performance cnn method for detecting diabetic retinopathy. In: 3rd International conference on information and communications technology (ICOIACT), Indonesia, Yogyakarta, pp 83–88
- Alshazly H, Linse C, Barth E et al (2021) Explainable covid-19 detection using chest ct scans and deep learning. *Sensors* 21:1–22
- Aly GH, Marey M, El-Sayed SA, Tolba MF (2021) Yolo based breast masses detection and classification in full-field digital mammograms. *Comput Methods Programs Biomed* 200:105823
- Alyafi B, Diaz O, Marti R (2020) DCGANs for realistic breast mass augmentation in X-ray mammography. IN: Medical imaging 2020: computer-aided diagnosis, International Society for Optics and Photonics, pp 1–4. <https://doi.org/10.1117/12.2543506>
- Araújo T, Aresta G, Mendonça L et al (2020) Data augmentation for improving proliferative diabetic retinopathy detection in eye fundus images. *IEEE Access* 8:462–474
- Armato IIIISG, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA (2011) The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys* 38:915–931
- Asia AO, Zhu CZ, Althubiti SA, Al-Alimi D, Xiao YL, Ouyang PB, Al-Qaness MA (2020) Detection of diabetic retinopathy in retinal fundus images using cnn classification models. *Electronics* 11:1–20
- Aswathy AL, Vinod Chandra SS (2022) Cascaded 3D UNet architecture for segmenting the COVID-19 infection from lung CT volume. *Sci Rep*. <https://doi.org/10.1038/s41598-022-06931-z>
- Ayana G, Park J, Choe SW (2022) Patchless multi-stage transfer learning for improved mammographic breast mass classification. *Cancers*. <https://doi.org/10.3390/cancers14051280>
- Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, Shinohara RT, Berger C, Ha SM, Rozycki M, Prastawa M (2018) Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint*, pp 1–49. [arXiv:1811.02629](https://arxiv.org/abs/1811.02629)
- Balasubramanian R, Sowmya V, Gopalakrishnan EA, Menon VK, Variyar VS, Soman KP (2020) Analysis of adversarial based augmentation for diabetic retinopathy disease grading. In: 11th International

- conference on computing, communication and networking technologies (ICCCNT), India, Kharagpur, pp 1–5
- Barile B, Marzullo A, Stamile C, Durand-Dubief F, Sappey-Marinier D (2021) Data augmentation using generative adversarial neural networks on brain structural connectivity in multiple sclerosis. *Comput Methods Programs Biomed* 206:1–12
- Basu A, Sheikh KH, Cuevas E, Sarkar R (2022) Covid-19 detection from CT scans using a two-stage framework. *Expert Syst Appl* 193:1–14
- Bayer M, Kaufhold MA, Reuter C (2021) A survey on data augmentation for text classification. *ACM-CSUR*. <https://doi.org/10.1145/3544558>
- Borji A (2019) Pros and cons of gan evaluation measures. *Comput Vis Image Underst* 179:41–65
- Carreira J, Zisserman A (2017) Quo vadis, action recognition? a new model and the kinetics dataset. In: *Conference on computer vision and pattern recognition*, Hawaii, Honolulu, pp 6299–6308
- Chaki J (2022) Two-fold brain tumor segmentation using fuzzy image enhancement and DeepBrainet2.0. *Multimed Tools Appl* 81:30705–30731
- Chatfield K, Simonyan K, Vedaldi A, Zisserman A (2014) Return of the devil in the details: delving deep into convolutional nets, pp 1–11. arXiv preprint. <https://doi.org/10.48550/arXiv.1405.3531>
- Chen X, Wang X, Zhang K, Fung KM, Thai TC, Moore K, Mannel RS, Liu H, Zheng B, Qiu Y (2022a) Recent advances and clinical applications of deep learning in medical image analysis. *Med Image Anal*. <https://doi.org/10.1016/j.media.2022.102444>
- Chen Y, Yang X, Wei Z, Heidari AA et al (2022b) Generative adversarial networks in medical image augmentation: a review. *Comput Biol Med*. <https://doi.org/10.1016/j.combiomed.2022.105382>
- Chlap P, Min H, Vandenberg N et al (2021) A review of medical image data augmentation techniques for deep learning applications. *Med Imaging Radiat Oncol* 65:545–563
- Costa P, Galdran A, Meyer MI, Niemeijer M, Abràmoff M, Mendonça AM, Campilho A (2017) End-to-end adversarial retinal image synthesis. *IEEE Trans Med Imaging* 37:781–791
- Decencière E, Zhang X, Cazuguel G, Lay B, Cochener B, Trone C, Gain P, Ordenez R, Massin P, Erginay A et al (2014) Feedback on a publicly distributed image database: the messidor database. *Image Anal Stereol* 33:231–234
- Deepak S, Ameer P (2020) MSG-GAN based synthesis of brain mri with meningioma for data augmentation. In: *IEEE international conference on electronics, computing and communication technologies (CONECCT)*, India, Bangalore, pp 1–6
- Desai SD, Giraddi S, Verma N, Gupta P, Ramya S (2020) Breast cancer detection using gan for limited labeled dataset. In: *12th International conference on computational intelligence and communication networks*, India, Bhimtal, pp 34–39
- Dodia S, Basava A, Padukudru Anand M (2022) A novel receptive field-regularized V-net and nodule classification network for lung nodule detection. *Int J Imaging Syst Technol* 32:88–101. <https://doi.org/10.1002/ima.22636>
- Dorizza A (2021) Data augmentation approaches for polyp segmentation. Dissertation, Università Degli Studi Di Padova
- Dufumier B, Gori P, Battaglia I, Victor J, Grigis A, Duchesnay E (2021) Benchmarking cnn on 3d anatomical brain mri: architectures, data augmentation and deep ensemble learning. arXiv preprint, pp 1–25. [arXiv:2106.01132](https://arxiv.org/abs/2106.01132)
- Farheen F, Shamil MS, Ibtehad N, Rahman MS (2022) Revisiting segmentation of lung tumors from CT images. *Comput Biol Med* 144:1–12
- Fidon L, Ourselin S, Vercauteren T (2020) Generalized wasserstein dice score, distributionally robust deep learning, and ranger for brain tumor segmentation: brats 2020 challenge. in: *International MICCAI brain lesion workshop*, Lima, Peru, pp 200–214
- Girshick R (2015) Fast R-CNN. In: *IEEE international conference on computer vision*, Santiago, USA, pp 1440–1448
- Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE conference on computer vision and pattern recognition*, Columbus, USA, pp 580–587
- Halder A, Datta B (2021) COVID-19 detection from lung CT-scan images using transfer learning approach. *Mach Learn: Sci Technol* 2:1–12
- Haq AU, Li JP, Agbley BLY et al (2022) IIMFCBM: Intelligent integrated model for feature extraction and classification of brain tumors using mri clinical imaging data in IoT-Healthcare. *IEEE J Biomed Health Inf* 26:5004–5012

- Hashemi N, Masoudnia S, Nejad A, Nazem-Zadeh MR (2022) A memory-efficient deep framework for multi-modal mri-based brain tumor segmentation. In: 2022 44th annual international conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Scotland, Glasgow, UK, pp 3749–3752
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, USA, pp 770–778
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. In: Proceedings of the IEEE international conference on computer vision, Venice, Italy, pp 2961–2969
- Hu R, Ruan G, Xiang S, Huang M, Liang Q, Li J (2020) Automated diagnosis of covid-19 using deep learning and data augmentation on chest CT. medRxiv, pp 1–11
- Humayun M, Sujatha R, Almuayqil SN, Jhanjhi NZ (2022) A transfer learning approach with a convolutional neural network for the classification of lung carcinoma. Healthcare. <https://doi.org/10.3390/healthcare10061058>
- Isensee F, Jäger PF, Full PM et al (2020) nnu-Net for brain tumor segmentation. In: International MICCAI brainlesion workshop. Springer, Cham, pp 118–132
- Islam MR, Abdulrazak LF, Nahiduzzaman M, Goni MO, Anower MS, Ahsan M, Haider J, Kowalski M (2022) Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images. Comput Biol Med. <https://doi.org/10.1016/j.compbimed.2022.105602>
- Jha M, Gupta R, Saxena R (2022) A framework for in-vivo human brain tumor detection using image augmentation and hybrid features. Health Inf Sci Syst 10:1–12
- Ju L, Wang X, Zhao X, Bonnington P, Drummond T, Ge Z (2021) Leveraging regular fundus images for training UWF fundus diagnosis models via adversarial learning and pseudo-labeling. IEEE Trans Med Imaging 40:2911–2925
- Karthiga R, Narasimhan K, Amirtharajan R (2022) Diagnosis of breast cancer for modern mammography using artificial intelligence. Math Comput Simul 202:316–330
- Khan AR, Khan S, Harouni M, Abbasi R, Iqbal S, Mehmood Z (2021) Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification. Microsc Res Tech 84:1389–1399
- Khosla C, Saini BS (2020) Enhancing performance of deep learning models with different data augmentation techniques: a survey. In: International conference on intelligent engineering and management (ICIEM), London, UK, pp 79–85
- Kim YJ, Kim KG (2022) Detection and weak segmentation of masses in gray-scale breast mammogram images using deep learning. Yonsei Med J 63:S63
- Kossen T, Subramaniam P, Madai VI, Hennemuth A, Hildebrand K, Hilbert A, Sobesky J, Livne M, Galinovic I, Khalil AA, Fiebach JB (2021) Synthesizing anonymized and labeled TOF-MRA patches for brain vessel segmentation using generative adversarial networks. Comput Biol Med 131:1–9
- Kurup A, Soliz P, Nemeth S, Joshi V (2020) Automated detection of malarial retinopathy using transfer learning. In: IEEE southwest symposium on image analysis and interpretation (SSIAI), Albuquerque, USA, pp 18–21
- Li Q, Yu Z, Wang Y et al (2020) Tumorgan: a multi-modal data augmentation framework for brain tumor segmentation. Sensors 20:1–16
- Li H, Chen D, Nailon WH, Davies ME, Laurenson DI (2021) Dual convolutional neural networks for breast mass segmentation and diagnosis in mammography. IEEE Trans Med Imaging 41:3–13
- Li Z, Guo C, Nie D, Lin D, Cui T, Zhu Y, Chen C, Zhao L, Zhang X, Dongye M, Wang D (2022) Automated detection of retinal exudates and drusen in ultra-widefield fundus images based on deep learning. Eye 36:1681–1686
- Lim G, Thombre P, Lee ML, Hsu W (2020) Generative data augmentation for diabetic retinopathy classification. In: IEEE 32nd international conference on tools with artificial intelligence (ICTAI), Baltimore, USA, pp 1096–1103
- Lin M, Hou B, Liu L, Gordon M, Kass M, Wang F, Van Tassel SH, Peng Y (2022) Automated diagnosing primary open-angle glaucoma from fundus image by simulating human's grading with deep learning. Sci Rep. <https://doi.org/10.1038/s41598-022-17753-4>
- Liu Y, Kwak HS, Oh IS (2022) Cerebrovascular segmentation model based on spatial attention-guided 3D inception U-Net with multi-directional MIPs. Appl Sci. <https://doi.org/10.3390/app12052288>
- Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, USA, pp 3431–3440

- Mahmood T, Li J, Pei Y, Akhtar F, Jia Y, Khand ZH (2021) Breast mass detection and classification using deep convolutional neural networks for radiologist diagnosis assistance. In: 45th Annual computers, software, and applications conf (COMPSAC), Madrid, Spain, pp 1918–1923
- Mahmood T, Li J, Pei Y, Akhtar F, Rehman MU, Wasti SH (2022) Breast lesions classifications of mammographic images using a deep convolutional neural network-based approach. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0263126>
- Manos D, Seely JM, Taylor J, Borgaonkar J, Roberts HC, Mayo JR (2014) The lung reporting and data system (LU-RADS): a proposal for computed tomography screening. *Can Assoc Radiol J* 65:121–134
- Mayya V, Kulkarni U, Surya DK, Acharya UR (2022) An empirical study of preprocessing techniques with convolutional neural networks for accurate detection of chronic ocular diseases using fundus images. *Appl Intell* 1:1–19
- McNitt-Gray MF, Armato SG III, Meyer CR, Reeves AP, McLennan G, Pais RC, Freymann J, Brown MS, Engelmann RM, Bland PH et al (2007) The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. *Acad Radiol* 14:1464–1474
- Meijering E (2020) A bird's-eye view of deep learning in bioimage analysis. *Comput Struct Biotechnol J* 18:2312–2325
- Miller JD, Arasu VA, Pu AX, Margolies LR, Sieh W, Shen L (2022) Self-supervised deep learning to enhance breast cancer detection on screening mammography. *arXiv preprint*, pp 1–11. [arXiv:2203.08812](https://arxiv.org/abs/2203.08812)
- Mok TC, Chung A (2018) Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In: International MICCAI brain lesion workshop, Granada, Spain, pp 70–80
- Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) Inbreast: toward a full-field digital mammographic database. *Acad Radiol* 19:236–248
- Müller D, Soto-Rey I, Kramer F (2021) Robust chest CT image segmentation of COVID-19 lung infection based on limited data. *Inf Med Unlocked* 25:1–11
- Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, Macchiarini P, Crapo JD, Herold CJ, Austin JH, Travis WD (2013) Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 266:304–317
- Nalepa J, Marcinkiewicz M, Kawulok M (2019) Data augmentation for brain-tumor segmentation: a review. *Front Comput Neurosci* 13:1–18
- Naveed H (2021) Survey: Image mixing and deleting for data augmentation. *arXiv preprint*, pp 1–15. [arXiv:2106.07085](https://arxiv.org/abs/2106.07085)
- Nayan AA, Mozumder AN, Haque M, Sifat FH, Mahmud KR, Azad AK, Kibria MG (2022) A deep learning approach for brain tumor detection using magnetic resonance imaging. *arXiv preprint*, pp 1–9. [arXiv:2210.13882](https://arxiv.org/abs/2210.13882)
- Neelima G, Chigurukota DR, Maram B, Girirajan B (2022) Optimal DeepMRSeg based tumor segmentation with GAN for brain tumor classification. *Biomed Signal Process Control*. <https://doi.org/10.1016/j.bspc.2022.103537>
- Nishio M, Muramatsu C, Noguchi S, Nakai H, Fujimoto K, Sakamoto R, Fujita H (2020) Attribute-guided image generation of three-dimensional computed tomography images of lung nodules using a generative adversarial network. *Comput Biol Med*. <https://doi.org/10.1016/j.compbiomed.2020.104032>
- Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, Imaizumi K, Fujita H (2020) Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. *Int J Comput Assist Radiol Surg* 15:173–178
- Oza P, Sharma P, Patel S, Adedoyin F, Bruno A (2022) Image augmentation techniques for mammogram analysis. *J Imaging*. <https://doi.org/10.3390/jimaging8050141>
- Padalia D, Vora K, Mehta D, Mehendale N (2022) EEF-Net: an enhanced efficientnet for breast tumor classification in mammograms. *SSRN* 4220435. <https://doi.org/10.2139/ssrn.4220435>
- Quintana-Quintana OJ, De León-Cuevas A, Gonzalez-Gutierrez A, Gorrostieta-Hurtado E, Tovar-Arriaga S (2022) Dual U-Net-based conditional generative adversarial network for blood vessel segmentation with reduced cerebral MR training volumes. *Micromachines*. <https://doi.org/10.3390/mi13060823>
- Ramya J, Rajakumar MP, Maheswari BU (2022) Deep cnn with hybrid binary local search and particle swarm optimizer for exudates classification from fundus images. *J Digit Imaging* 35:56–67
- Razali NF, Isa IS, Sulaiman SN, Karim NK, Osman MK (2021) High-level features in deeper deep learning layers for breast cancer classification. In: 11th IEEE international conference on control system, computing and engineering (ICCSCE), Penang, Malaysia, pp 170–175
- Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* 28:1–9

- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC (2015) Imagenet large scale visual recognition challenge. *Int J Comput Vis* 115:211–252
- Sabani A, Landsmann A, Hejduk P, Schmidt C, Marcon M, Borkowski K, Rossi C, Ciritis A, Boss A (2022) BI-RADS-based classification of mammographic soft tissue opacities using a deep convolutional neural network. *Diagnostics*. <https://doi.org/10.3390/diagnostics12071564>
- Salama WM, Aly MH (2021) Deep learning in mammography images segmentation and classification: automated CNN approach. *Int Alexandria Eng J* 60:4701–4709. <https://doi.org/10.1016/j.aej.2021.03.048>
- Setio AA, Traverso A, De Bel T, Berens MS, Van Den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, van der Gugten R (2017) Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Med Image Anal* 42:1–3
- Shmelkov K, Schmid C, Alahari K (2018) How good is my GAN? In: *Proceedings of the European conference on computer vision (ECCV)*, Germany, Munich, pp 213–229
- Shen T, Hao K, Gou C, Wang FY (2021) Mass image synthesis in mammogram with contextual information based on GANS. *Comput Methods Programs Biomed*. <https://doi.org/10.1016/j.cmpb.2021.106019>
- Shi H, Lu J, Zhou Q (2020) A novel data augmentation method using style-based GAN for robust pulmonary nodule segmentation. In: *IEEE Chinese control and decision conference (CCDC)*, Hefei, China, pp 2486–2491
- Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. *J Big Data* 6:1–48
- Shyamalee T, Meedeniya (2022) D CNN based fundus images classification for glaucoma identification. In: *2nd International conference on advanced research in computing (ICARC)*, Belihuloya, Sri Lanka, pp 200–205
- Simonyan K, Zisserman A (2014) Two-stream convolutional networks for action recognition in videos. *Adv Neural Inf Process Syst* 27:1–9
- Singh LK, Garg H, Khanna M (2022) Deep learning system applicability for rapid glaucoma prediction from fundus images across various data sets. *Evol Syst* 13:807–836
- Soulami KB, Kaabouch N, Saidi MN (2022) Breast cancer: classification of suspicious regions in digital mammograms based on capsule network. *Biomed Signal Process Control*. <https://doi.org/10.1016/j.bsp.2022.103696>
- Srinivas C, KS NP, Zakariah M, Alothaibi YA, Shaikat K, Partibane B, Awal H (2022) Deep transfer learning approaches in performance analysis of brain tumor classification using MRI images. *J Healthc Eng*. <https://doi.org/10.1155/2022/3264367>
- Sun Y, Yuan P, Sun Y (2020) MM-GAN: 3D MRI data augmentation for medical image segmentation via generative adversarial networks. In: *IEEE international conference on knowledge graph (ICKG)*, Nanjing, China, pp 227–234
- Sun X, Fang H, Yang Y et al (2021) Robust retinal vessel segmentation from a data augmentation perspective. In: *International workshop on ophthalmic medical image analysis*, pp 189–198
- Tandon R, Agrawal S, Raghuvanshi R, Rathore NP, Prasad L, Jain V (2022a) Automatic lung carcinoma identification and classification in CT images using CNN deep learning model. In: *Augmented intelligence in healthcare: a pragmatic and integrated analysis*. Springer, Singapore. https://doi.org/10.1007/978-981-19-1076-0_9
- Tandon R, Agrawal S, Chang A, Band SS (2022b) VCNet: Hybrid deep learning model for detection and classification of lung carcinoma using chest radiographs. *Front Public Health*. <https://doi.org/10.3389/fpubh.2022.894920>
- Tang N, Zhang R, Wei Z, Chen X, Li G, Song Q, Yi D, Wu Y (2021) Improving the performance of lung nodule classification by fusing structured and unstructured data. *Inf Fusion* 88:161–174
- Thung KH, Raveendran P (2009) A survey of image quality measures. In: *IEEE international conference for technical postgraduates (TECHPOS)*, Kuala Lumpur, Malaysia, pp 1–4
- Toda R, Teramoto A, Tsujimoto M, Toyama H, Imaizumi K, Saito K, Fujita H (2021) Synthetic ct image generation of shape-controlled lung cancer using semi-conditional InfoGAN and its applicability for type classification. *Int J Comput Assist Radiol Surg* 16:241–251
- Tsuneki M (2022) Deep learning models in medical image analysis. *J Oral Biosci*. <https://doi.org/10.1016/j.job.2022.03.003>
- Tufail AB, Ullah I, Khan WU, Asif M, Ahmad I, Ma YK, Khan R, Ali M (2021) Diagnosis of diabetic retinopathy through retinal fundus images and 3D convolutional neural networks with limited number of samples. *Wirel Commun Mob Comput* 2021:1–15
- Tyler ME, Hubbard LD, Boydston K, Pugliese AJ (2009) Characteristics of digital fundus camera systems affecting tonal resolution in color retinal images. *J Ophthalmic Photogr* 31:1–9

- Ueda D, Yamamoto A, Onoda N, Takashima T, Noda S, Kashiwagi S, Morisaki T, Fukumoto S, Shiba M, Morimura M et al (2022) Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets. *PLoS ONE* 17:1–15
- van der Velden BH, Kuijff HJ, Gilhuijs KG, Viergever MA (2022) Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal.* <https://doi.org/10.1016/j.media.2022.102470>
- Wang W, Luo J, Yang X, Lin H (2015) Data analysis of the lung imaging database consortium and image database resource initiative. *Acad Radiol* 22:488–495
- Wang Q, Zhang X, Zhang W, Gao M, Huang S, Wang J, Zhang J, Yang D, Liu C (2021) Realistic lung nodule synthesis with multi-target co-guided adversarial mechanism. *IEEE Trans Med Imaging* 40:2343–2353
- Wang Y, Ji Y, Xiao H (2022) A Data Augmentation Method for Fully Automatic Brain Tumor Segmentation. *arXiv preprint*, pp 1–15. [arXiv:2202.06344](https://arxiv.org/abs/2202.06344)
- Woan SL, Lai KW, Chuah JH, Hasikin K, Khalil A, Qian P, Xia K, Jiang Y, Zhang Y, Dhanalakshmi S (2022) Multiclass convolution neural network for classification of COVID-19 CT images. *Comput Intell Neurosci.* <https://doi.org/10.1155/2022/9167707>
- Wu W, Lu Y, Mane R, Society et al (2020a) EMBC, Canada, Montreal, pp 1516–1519
- Wu E, Wu K, Lotter W (2020b) Synthesizing lesions using contextual gans improves breast cancer classification on mammograms. *arXiv preprint*, pp 1–10. [arXiv:2006.00086](https://arxiv.org/abs/2006.00086)
- Xie L, Chen Z, Sheng X et al (2022) Semi-supervised region-connectivity-based cerebrovascular segmentation for time-of-flight magnetic resonance angiography image. *Comput Biol Med.* <https://doi.org/10.1016/j.combiomed.2022.105972>
- Xu Q, Huang G, Yuan Y, Guo C, Sun Y, Wu F, Weinberger K (2018) An empirical study on evaluation metrics of generative adversarial networks. *arXiv preprint*, pp 1–14. [arXiv:1806.07755](https://arxiv.org/abs/1806.07755)
- Yadav JK, Yadav S (2022) Computer-aided diagnosis of cataract severity using retinal fundus images and deep learning. *Comput Intell* 38:1450–1473
- Yadav A, Shah S, Xu Z, Jacobs D, Goldstein T (2017) Stabilizing adversarial nets with prediction methods. *arXiv preprint*, pp 1–21. [arXiv:1705.07364](https://arxiv.org/abs/1705.07364)
- Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. *Med Image Anal.* <https://doi.org/10.1016/j.media.2019.101552>
- Yu Z, Xiang Q, Meng J, Kou C, Ren Q, Lu Y (2019) Retinal image synthesis from multiple-landmarks input with generative adversarial networks. *Biomed Eng Online.* <https://doi.org/10.1186/s12938-019-0682-x>
- Yu H, Yang LT, Zhang Q, Armstrong D, Deen MJ (2021) Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives. *Neurocomputing* 444:92–110. <https://doi.org/10.1016/j.neucom.2020.04.157>
- Yuan Y (2020) Automatic brain tumor segmentation with scale attention network. In: International MICCAI brain lesion workshop. *Lecture notes in computer science.* https://doi.org/10.1007/978-3-030-72084-1_26
- Zahoor S, Shoaib U, Lali IU (2022) Breast Cancer Mammograms classification using deep neural network and entropy-controlled Whale optimization Algorithm. *Diagnostics.* <https://doi.org/10.3390/diagnosics12020557>
- Zeiser FA, da Costa CA, Zonta T et al (2020) Segmentation of masses on mammograms using data augmentation and deep learning. *J Digit Imaging* 33:858–868
- Zhang K (2021) On mode collapse in generative adversarial networks. In: 30th International conference on artificial neural networks, Bratislava, Slovakia, pp 563–574
- Zhou Y, Wang B, He X, Cui S, Shao L (2020) DR-GAN: conditional generative adversarial network for fine-grained lesion synthesis on diabetic retinopathy images. *IEEE J Biomedical Health Inf* 26:56–66

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

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