IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 23, 2020, accepted July 11, 2020, date of publication July 20, 2020, date of current version August 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3010326

An Accurate and Fast Cardio-Views Classification System Based on Fused Deep Features and LSTM

AHMED I. SHAHIN AND SULTAN ALMOTAIRI
Department of Natural and Applied Sciences, Community College, Majmaah University, Al-majmaah 11952, Saudi Arabia
Corresponding author: Sultan Almotairi (almotairi@mu.edu.sa)

This work was supported by a Grant of the Deanship of Scientific Research at Majmaah University under Project RGP-2019-29.

ABSTRACT Echocardiography is an ultrasound-based imaging modality that helps the physician to visualize heart chambers and valves motion activity. Recently, deep learning plays an important role in several clinical computer-assisted diagnostic systems. There is a real need to employ deep learning methodologies to increase such systems. In this paper, we proposed a deep learning system to classify several echocardiography views and identify its physiological location. Firstly, the spatial CNN features are extracted from each frame in the echo-motion. Secondly, we proposed novel temporal features based on neutrosophic sets. The neutrosophic temporal motion features are extracted from echo-motion activity. To extract the deep CNN features, we activated a pre-trained deep ResNet model. Then, both spatial and neutrosophic temporal CNN features were fused based on features concatenation technique. Finally, the fused CNN features were fed into deep long short-term memory network to classify echo-cardio views and identify their location. During our experiments, we employed a public echocardiography dataset that consisted of 432 videos for eight cardio-views. We have investigated several pre-trained network activation performance. ResNet architecture activation achieved the best accuracy score among several pre-trained networks. The Proposed system based on fused spatial neutrosophic temporal deep features achieved 96.3% accuracy and 95.75% sensitivity. For the classification of cardio-views location, the proposed system achieved 99.1% accuracy. The proposed system achieved more accuracy than previous deep learning methods with a significant decrease in the training time cost. The experimental results showed promising results for our proposed approach.

INDEX TERMS Ultrasound, echocardiography, cardio-views, deep learning, neutrosophic temporal desriptors, CNN features fusion, LSTM.

I. INTRODUCTION

Echocardiography is an ultrasound modality, which captures the cardiac activity during its motion based on M-Mode imaging and provides the physicians with more details about the blood supply [1]. In echocardiography, the physiological cardiac motion is recorded inside consequential frames, which represent a 3D structure. These dimensions are as follows: frame width, frame height, and time. Echocardiography imaging has several views for the heart while moving the transducer with different angles to capture heart motion activity [2]. After the physician manually recognizes the view, several anatomical structures can be detected and analyzed. The most considered eight views as shown in Fig. 1 are changing according to the transducer position into three locations. Location A consists of 4 views as the following: apical 2 chambers (A2C), apical 3 chambers (A3C), apical 4 chambers (A4C), and apical 5 chambers (A5C). Location B consists of a single view, which is parasternal long axis (PLA). Location C consists of 3 views, which are parasternal short axis of aorta (PSAA), parasternal short axis of papillary (PSAP) and parasternal short axis of mitral (PSAM) [3]. These views fundamentally obtain a discriminative information based on spatial and temporal perspective. Therefore, the accurate classification of such cardio-views aims to analyze and diagnose several cardio-diseases.

Traditional artificial intelligence systems are based on pre-Computer aided diagnostic (CAD) systems helps the physician to improve the diagnostic quality for several soft-tissue examination tasks [4]–[13].

CAD systems based on traditional artificial intelligence approaches consists of pre-processing, hand crafted-features extraction, features processing, and classification. Classical features extraction techniques are based on spatial features,

The associate editor coordinating the review of this manuscript and approving it for publication was Xiping Hu.

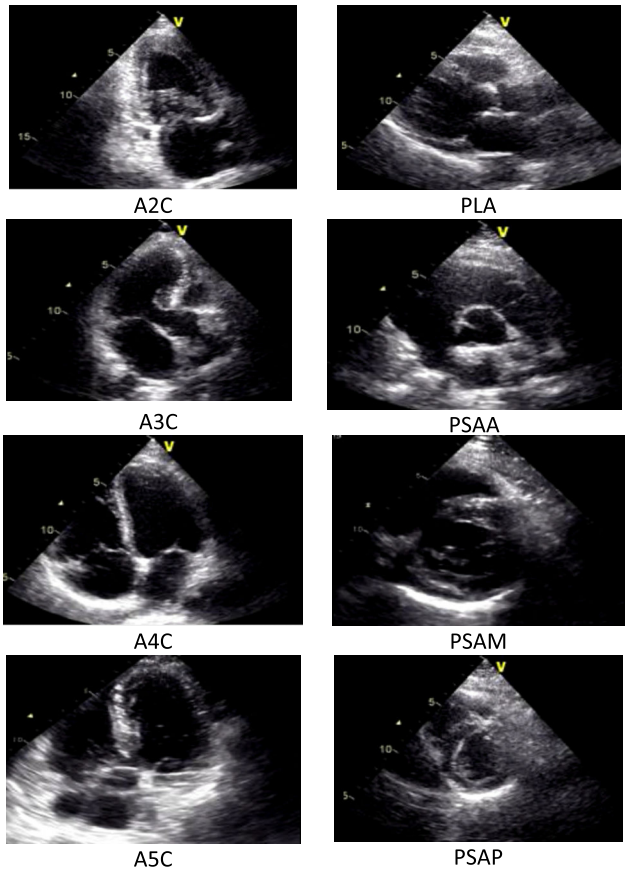

FIGURE 1. Samples for eight cardio-views of echocardiography.

morphological features, and temporal features [14]. There are real challenges in echocardiography to extract such classical features due to its sensitivity to motion artifacts [15].

In [16], LeCun introduced CNN as a new machine learning methodology to tackle the handcrafted features extraction procedure, which is commonly named with deep learning. CNN employed separable learnable convolutional filters to extract deep CNN features automatically without a need for handcrafted features extraction. Deep learning frameworks have been extended to several network architectures after plain CNN architecture such as deep inception CNN architecture [17], residual CNN architecture [18], deep generative adversarial architecture (DGAN) [19], deep belief network (DBN) architecture [20], and deep LSTM architecture [21].

Due to the exponential growth of hardware resources, deep learning was employed in several multi-class general classification tasks [22], [23]. On the other hand, deep learning has been proven as an excellent tool for several video classification tasks [24]–[35]. Recently, deep learning has been employed for several medical image modalities, dimensions, and applications. Deep learning has been applied to several images modalities such as x-ray [36], CT [37], MR [38], microscopic pathology [39], and ultrasound [40]. Deep learning has been applied to several two-dimensional medical images such as [41] and several 3D medical imaging systems [42]. Deep learning has several applications in automated medical image assessment (AMIA) systems such as de-noising [43], segmentation [44], classification [37], and detection [45].

The echocardioviews classification systems were based on traditional features extraction or even spatial CNN features extraction, lack of accuracy, and consumed a lot of processing time [46]–[48]. Therefore, the employment of several deep learning architectures that have successfully increased the video recognition systems is very important to enhance echocardiography views classification systems. On the other hand, it is important to decrease the processing time of such systems. In this paper, we aim to increase the state of art echocardoviews classification systems. Therefore, the integration of physician interpretation with accurate information extracted from CAD systems provides predictive information that cannot be detected due to human error and increase the diagnostic quality.

The rest of the paper is organized as the following. In section II, we cover the previous work for echocardiography computer assessment systems. In section III, we introduce the proposed classification system for echo cardio-views. In section IV, the results with its discussion are presented. Finally, in section V, the conclusion for our work is presented.

II. LITERATURE REVIEW

In this section, firstly, we present literature for recently artificial intelligent (AI) systems that had been employed to enhance echocardiography clinical examination. Then, we introduce several articles that were applied to move forward the echocardiography AI systems based on traditional machine learning or even deep learning techniques.

AI generally aims to increase the diagnostic capabilities of echocardiography computer assisted-systems such as detection of pathological cardio-diseases, quantification of cardio-motion [46], and computing echo image quality [47]. AI also helps the physicians automatically to classify several cardio-views [48], [49]. AI detected several cardio-pathological diseases such as wall motion disorders [50], detection of left ventricle disorders [51], mitral regurgitation [52]. AI also helps the physicians to quantify several cardiac-motion parameters such as: MV (Myocardial velocity) [53], EF (ejection fraction) [54], and LS (longitudinal strain) [55].

In [52], the authors presented a mitral regurgitation heart disease classification system. They utilized gradient local binary pattern descriptors. The system achieved 99.5% accuracy based on linear discriminant analysis combined with template matching algorithm for about 5000 image frames distributed between normal, mild, moderate, and severe cases. In [54], the authors proposed an automated system for heart failure with preserved ejection detection fraction under stress. They utilized high-dimensional descriptors, then they employed supervised dictionary learning and it achieved an average accuracy of 95% for only 70 echo-clips. In [53], the

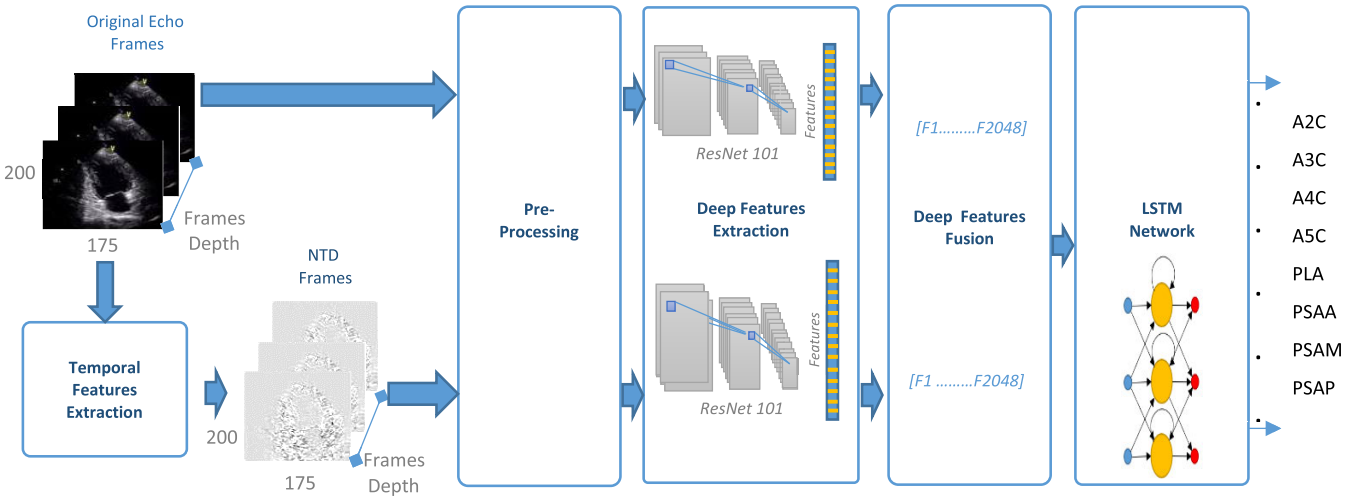

FIGURE 2. The proposed system architecture to classify 8 cardio-views.

authors proposed a system to detect cardio-wall motion disorders based on dual-tree discrete wavelet transform descriptors. The system employed the linear discriminant classifier and it achieved 96% accuracy with 279 images. In [56], the authors proposed an automated system for only three apical views of echocardiography. They utilized spatial-temporal cuboid descriptors, then they employed supervised dictionary learning and it achieved average accuracy of 95% for only 70 echo-clips. In [57], the authors presented an automated system for pathological cardio-diseases. They utilized high-morphological descriptors, then they employed support vector machine and it achieved 87% sensitivity, and 82% specificity for 139 patients included with their patient history. In [51], the authors presented an automated system to detect the left ventricle based on the active contour algorithm and random forest classifier. The system achieved 90% accuracy for only 85 images. In [58], the author presented an automated system to quantify wall motion stress. They employed morphological descriptors, then they utilized the hidden Markova model to classify stress echocardiography and the system achieved an improvement in classification (84.17%).

In the literature, a few numbers of researches had been proposed to classify cardio-views based on deep learning [48], [49]. In [48], a system to classify 8 cardio-views had been presented based on CNN training from scratch. The system achieved 92.1% accuracy after the fusion of spatial and acceleration features. In [49], the deep learning had been employed to classify 15 cardio-views based on CNN training from scratch. It achieved 96% accuracy based on spatial deep CNN features.

The previous studies showed that traditional machine learning methods consume a lot of time to extract handcrafted features and are very sensitive to motion artifacts [51]–[58]. On the other hand, deep learning methods based on training CNN from scratch consume a lot of time, which reached to a few days, and there is still a challenge to increase its accuracy [48]. In this paper, we propose a robust automated system to classify eight views of echocardiography imaging based on CNN activation combined with the LSTM network. We propose new descriptors that based on CNN features fusion between spatial and temporal descriptors. Our proposed system consumes less significant processing time compared to other methods in the literature. The proposed system achieved higher performance than traditional machine learning systems or even the state of the art one [48].

III. PROPOSED METHOD

In this paper, we apply a new methodology to classify cardio-views based on deep learning framework, which combined between convolutional neural network and LSTM architectures as shown in Fig.2. We utilize our proposed system to classify 3 cardio-locations. Moreover, we extract novel temporal descriptors based on neutrosophic sets domain. In this paper, we combine spatial and neutrosophic temporal descriptors. We extract both deep CNN features by employing the pre-trained networks as a deep features extractor. After spatial and temporal deep features extraction, we fuse both features types. Finally, we employ LSTM classifier to classify each echo-clip into 8 cardio-views.

A. NEUTROSOPHIC TEMPORAL FEATURES EXTRACTION

The temporal descriptors contain the motion features between each two consequences frames. We propose novel temporal descriptors based on neutrosophic subsets as described in Algorithm 1.

Echocardiography clips are usually stored in DICOM format with 4-D (height, width, channels, and frames depth). Each frame contains the spatial descriptors, and each consequence frames contain the temporal descriptors. We extract the temporal features by dividing each frame into different N blocks, which is set here by 8 blocks. The temporal features represent the difference of pixel values between each block

Algorithm 1 Temporal Neutrosophic Sets Approach

Read Each two echo-cardio consequences frames.
Divide each frame into blocks.
Get Temporal features descriptors f(x,y).
For z =1: frames depth-1
Calculate T, I, F subsets for each pixel inside f(x,y) based on Eq.1, 2,and 3.
End
Obtain the final neutrosophic sets temporal descriptors (NTD) based on Eq.4

in the current frame and the opposite block in the next frame. Each pixel in the neutrosophic domain has three values membership as follows: Truth (T), Indeterminacy (I), and False (F) [59]. The study of similarity between neutrosophic subsets provides more useful information than the standalone sunsets [60]. For this reason, we utilize the similarity score algorithm between both truth and indeterminacy subsets and neglect the false subset value. The neutrosophic subsets can be given by Eq.1, 2 and 3 [59], [60]:

$$T(x,y) = \frac{ft(x,y) - ft_{min}}{ft_{max} - ft_{min}} \tag{1}$$

$$I(x,y) = 1 - \frac{ftd(x,y) - ftd_{min}}{ftd_{max} - ftd_{min}} \tag{2}$$

$$F(x,y) = 1 - T(x,y) \tag{3}$$

where ft (x,y) represents the input temporal pixel and ftd(x,y) represents the gradient value on the temporal pixels values. The neutrosophic temporal descriptors (NTD) can be derived from the similarity degree between three neutrosophic (T, I, and F) subsets as in Eq.4, shown at the bottom of the page, [59], [60], where A^* represents the ideal alternative. T_{C_j}, I_{C_j} and F_{C_j} represent neutrosophic at specific criteria (Cj).

A sample of original cardio-view is shown in Fig.3.a, predicated temporal feature map is shown in Fig.3.b, and the NTD features map is shown in Fig.3.c.

B. PREPROCESSING

DICOM-formatted echocardiogram clips, which are used in our paper was stored in RGB format with two different resolutions (434 × 636 pixels × 26 and 341 × 415) with 26 frames depth. All pre-trained networks employed in our study have the following sizes: AlexNet input layer is (227 × 227 × 3), VGGNet architectures, GoogleNet, DenseNet, three ResNet architectures input layer size are (224 × 224 × 3). Therefore, we resize both spatial and temporal frames to fit each pre-trained network input layer.

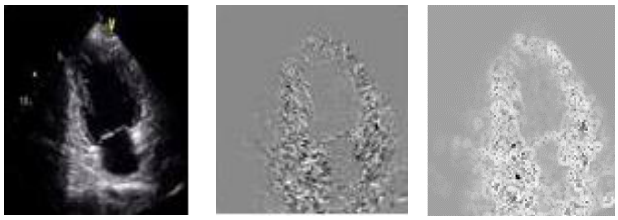

FIGURE 3. (a) An example of an original cardio-view frame, (b) temporal features map, and (c) NTD map.

C. CNN FEATURES EXTRACTION

In image task classification, CNN can be used based on three methods, which are training from scratch method, pre-trained network activation method, and fine-tuning of pre-trained network method [39]. As introduced in the literature, the CNN training from scratch or even fine-tuning of pre-trained networks still consume a lot of processing time. Therefore, we employ the pre-trained networks as CNN features extractor and transfer learning based on pre-trained networks will be more efficient. These networks had been trained previously and acquired their learned parameters to distinguish between different general images datasets. These datasets are such as CIFAR10 / CIFAR100, Caltech 101/ Caltech 256, ImageNet. These pre-trained networks are Alexnet, VGG16Net, VGG19Net, GoogleNet, densenet, ResNet18, ResNet50, and ResNet 101. In this paper, we evaluate each pre-trained network performance related to its classification accuracy. Deep activation features can be extracted from each convolutional features map inside CNN. However, in [61], the author proved that the latent fully connected layer activation features achieved the best performance. In this paper, we extract the last deep CNN features from the latent fully connected layer in the pre-trained network.

D. DEEP FEATURES FUSION

The fusion procedure helps to collect the latest information of concatenated spatial- temporal descriptors from both fully connected layers (FC) of the two model's streams. In AlexNet and VGG16/19, we have two features pole with a size 4096. In GoogleNet, we have two features pole with a size 1024. In DenseNet, we have two features pole with a size 1920. In ResNet 18, we have two features pole with a size 512 In ResNet 50/101 architectures, we have two features pole with a size 2048. As followed in [62], we employ the concatenation fusion function that achieved the best performance.

E. LSTM CLASSIFICATION

In our proposed system, we employ the LSTM network to perform the classification task for the fused deep CNN features.

$$NTD(f(x,y),A^*) = \frac{\left[T_{C_j}(x,y)\,T_{C_j}(A^*) + I_{C_j}(x,y)\,I_{C_j}(A^*) + F_{C_j}(x,y)\,F_{C_j}(A^*)\right]}{\sqrt{T^2_{C_j}(x,y) + I^2_{C_j}(x,y) + F^2_{C_j}(x,y))}\sqrt{(T^2_{C_j}(A^*) + I^2_{C_j}(A^*) + F^2_{C_j}(A^*))}} \tag{4}$$

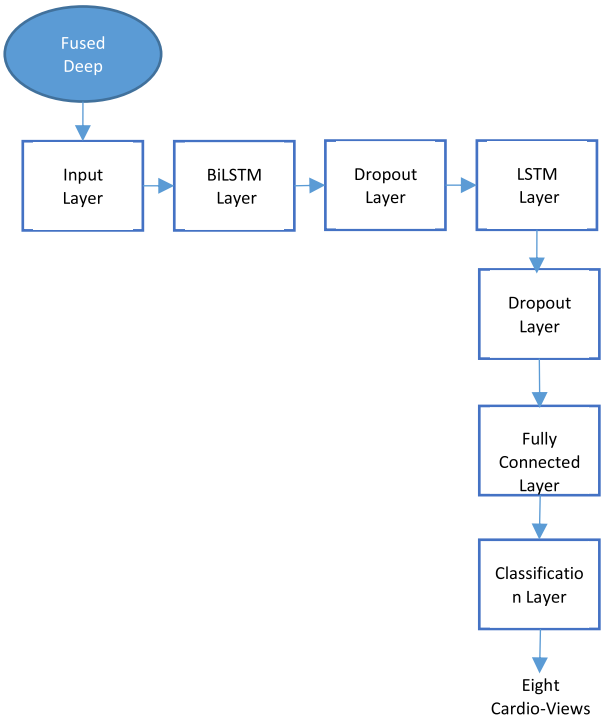

FIGURE 4. The proposed LSTM architecture.

Our proposed LSTM architecture as shown in Fig.4 consists of seven layers as follows: an input layer, bidirectional LSTM (BiLSTM) layer, dropout layer, LSTM layer, fully connected layer, and a classification layer.

The input layer receives the deep features pole, which consists of spatial-temporal descriptors. The input layer is followed by a BiLSTM layer. The traditional LSTM receives its information only from old values. The BiLSTM layer advantage is that it learns between the start of input sequences data to the end in bidirectional form. Therefore, it helps the network to get effective and faster learning. After the first BiLSTM, we increase the depth of LSTM architecture by adding another unidirectional LSTM layer. In this paper, the input layer size is set to fit the fused deep features. The no. of hidden units inside the first BiLSTM layer is 64 units. The no. of hidden units inside the second LSTM layer is 128 units. To achieve the best performance with the lowest training time, we insert a dropout layer after both LSTM and BiLSTM layer to prevent overfitting after BiLSTM layer. We set the two dropout neurons to 0.5 inside each dropout layer. Finally, the classification layer based on a softmax classifier is applied to classify a given echo-view and its cardio-location. In our experiments, we optimize the best optimize to train the proposed LSTM classifier.

IV. RESULTS AND DISCUSSION

In this paper, we employ an echocardiography public dataset, which contains eight cardio-views [48]. The dataset contains 432 echocardiography clips. The data was collected from 2 different hospitals in China provided with their ground truth and from different 93 patients. Each echo clip was acquired using GE-Vivid 7 ultrasound equipment for only 1 sec. The recorded frame rate was 26 frames/sec. In our paper, we prevent overfitting and make our proposed system more robustness by randomly splitting the dataset into 3 sets as follows: training set (70%), validation set (15%), and test set (15%).

The proposed deep learning architecture is implemented using the Matlab 2019 a. During our algorithm training, we utilize Quad-Core 2.9 GHz Intel i5 with 16 GB of memory, and moderate graphic processing unit NVIDIA TITAN-Xp GPU with 12 GB RAM.

A. EVALUATION CRITERIA

To evaluate our proposed system results, at first, we evaluate the performance of LSTM through different optimizers. Secondly, we compare our proposed fused features with the previous features in the literature. Thirdly, we compare different pre-trained networks that we utilize for CNN features extraction procedure.

To compare our proposed system results versus the previous deep learning system in the literature [48], we utilize the confusion matrix, accuracy, precision, sensitivity, and specificity as quantified metrics to evaluate as follows:

$$Accuracy = \frac{TP + TN + FP + FN}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Specifity = \frac{TN}{TN + FP} \tag{8}$$

Moreover, we compare our proposed system with the previous traditional systems and deep learning systems [48] that classified the same dataset into 8 cardio-views and 3 cardio-locations. Finally, we compare our feature extraction and training time cost vs. the previous deep learning system in the literature [48].

B. OUR PROPOSED SYSTEM RESULTS

In our experiments, firstly, we investigate the following points: the training and validation accuracy curve, the training and validation loss curve through different network optimizers. We select the best optimizer based on the lowest epoch's numbers and the highest accuracy score. Secondly, we investigate, which pre-trained network activation will work better as a feature extractor. We utilize 8 different network architectures for our classification task. Thirdly, we discuss the evaluation metrics for our proposed system. Finally, we discuss the confusion matrix results for both cardio-views and cardio-locations.

The Optimization algorithm plays a crucial role during the training process to increase the performance of the LSTM network [63]. To select the best optimizer in our proposed method; we compare the performance of root mean square

(rmsprop), stochastic gradient descent (SGD), and adaptive moment estimation algorithm (Adam) optimizers. During the training process, we utilize 500 epochs to ensure that the training phase will be converged with min-batch size 16. The initial learning rate setup is 0.001.

Learning curve represents an efficient tool to evaluate the performance of the LSTM classifier during its training process through a mathematical representation of the learning process that occurs during iterations. For the training set, the performance of Adam optimizer learning curve appears to be more robust with lower required training time. The rmsprop optimizer achieved similar performance to Adam optimizer. SGD optimizer achieved the lowest performance during the training process. The performance of training process related to the three examined optimizers is shown in Fig.5.

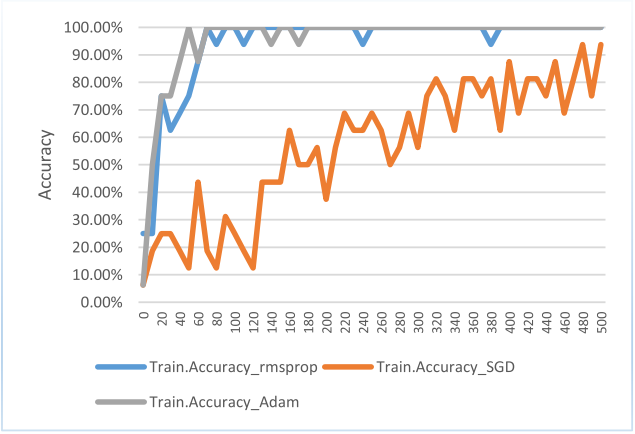

FIGURE 5. The proposed system training accuracy performance through different optimizers.

For the validation set, the performance of Adam optimizer achieved the highest training accuracy score of 87.5% with lower required training time. The rmsprop optimizer achieved similar performance to Adam optimizer with lower training accuracy score of 86.05%. SGD optimizer achieved the lowest performance of training accuracy score 83.72%. The performance related to the three examined optimizers of the give validation set is shown in Fig.6.

In Fig. 7 and 8, both training and validation sets loss are shown. The performance of Adam optimizer learning curve appears to be more robust with lower required training time. The rmsprop optimizer achieved similar performance to Adam optimizer with higher loss. SGD optimizer achieved the highest loss performance during training and validation. From the previous experiment, we prove that Adam optimizer is more efficient and robust during our echo-cardio views classification task.

For the following experiments, we utilized the test set to visualize our system robustness as followed in Gao et.al [48].

In this experiment, we investigate the most discriminant powerful features pole suitable for our classification task. As shown in Fig.9, we compare several hand crafted features, deep CNN features, CNN spatial-temporal

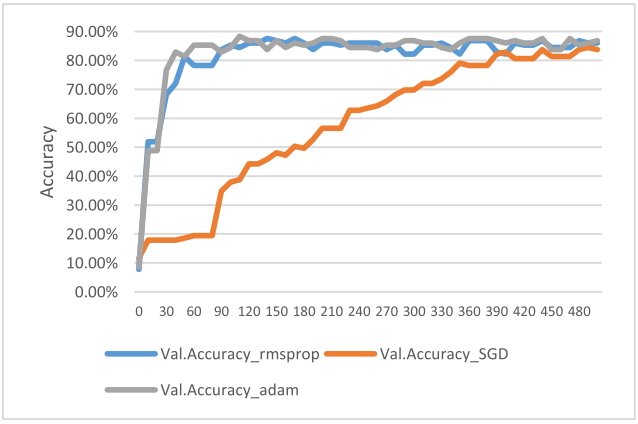

FIGURE 6. The proposed system validation accuracy performance through different optimizers.

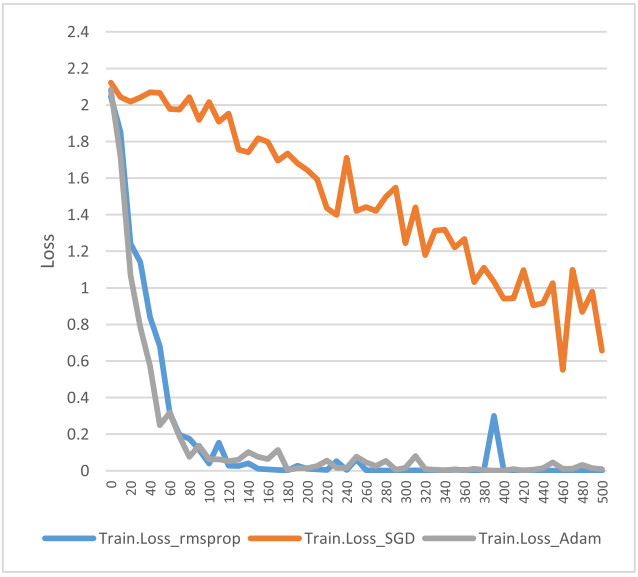

FIGURE 7. The proposed system training loss performance through different optimizers.

features fusion, our proposed spatial features, NTD features, and our proposed fused features. In [48], deep features achieved higher accuracy than traditional handcrafted features. On the other hand, CNN features based on training from scratch achieved accuracy of 89.5% and increased to 92.1% after spatial-temporal features fusion. Our proposed system based on pre-trained network activation and LSTM network achieved the following accuracies: spatial features achieved 90.5% accuracy, NTD features achieved 93.1%, and both features fusion achieved 96.3%, which is better than the previous handcrafted features or even deep CNN features.

As shown in Fig.9, a significant improvement related to the proposed system accuracy reached 2.6% has been noticed based on NTD descriptors, which reflect the robustness of the proposed neutrosophic temporal features. On the other hand, it has been noticed that the previous CNN features fusion based on training CNN from scratch achieved higher accuracy than our proposed spatial features and lower accuracy than our proposed NTD descriptors.

FIGURE 8. The proposed system validation loss performance through different optimizers.

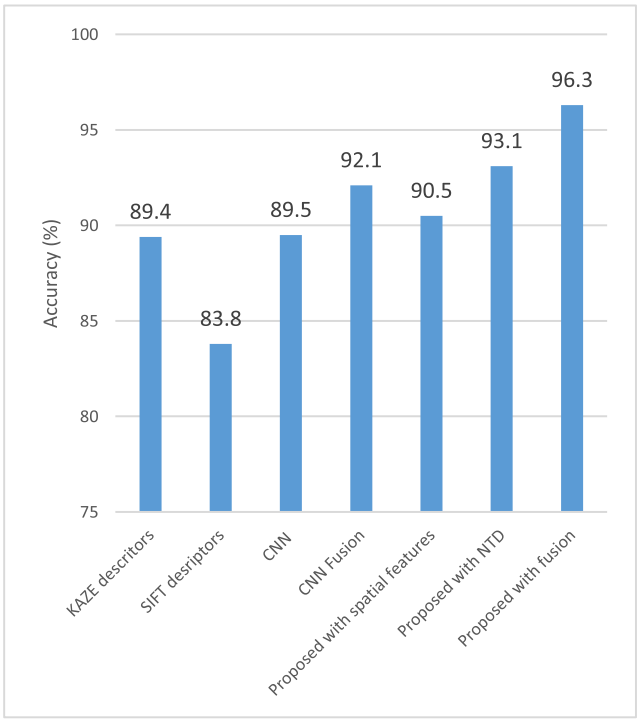

FIGURE 9. Proposed system accuracy vs. previous traditional and deep systems based on proposed fused features.

In this experiment, we investigate the accuracy of each pre-trained network activation with input features (spatial features, NTD features, and fused features) as shown in Fig. 10. We employ several pre-trained networks such as AlexNet, GoogleNet, DenseNet, ResNet, architectures, and VGGNet architectures. ResNet50/101 architectures achieved the highest accuracies 91.2%, and 96.3% respectively with the proposed fused features. On the other hand, GoogleNet achieved the lowest accuracy of 76% with the input spatial features. We have also noticed that NTD features increase the classification accuracy through all pre-trained network activation. Moreover, the fusion of both spatial and NTD features helps to increase the classification accuracy through all pre-trained network activation. ResNet 101 architecture achieved the highest accuracy score (90.5%) for the spatial features, the highest accuracy score (93.1%) for the NTD features and the highest accuracy score (96.3%) for the fusion of spatial features with NTD features.

We compare our proposed system with the state of the art [48] based on the achieved accuracy, sensitivity, specificity, and precision for 8 cardio-views classification. For the 3 cardio-locations, we compare our proposed system based on the achieved accuracy. Moreover, we compare between our proposed system and the previous method [48] for each cardio-view classification accuracy and each cardio-location classification accuracy. Finally, we investigate the time cost of features extraction procedure and classifier training time.

The confusion matrix of the 8 cardio-view classification system is shown in Fig. 11. It is noticed that the high true positive value of A2C and PSAA cardio-views classification with 100 % accuracy. A3C cardio-view achieved the lowest accuracy of 87%. The misclassification between A3C, and (A2C, A4C) cardio-views has been noticed. A4C, A5C, and PLA cardio views achieved accuracy above 95%. PSAM and PSAP achieved 91.7% and 92.9% respectively. The overall system accuracy is 96.3% for 8 cardio-views classification.

As followed in [48], we evaluate our proposed system to classify 3 cardio-views locations (Location A, Location B, and Location C). Location A represents the apical angle, location B represents the parasternal long axis, and location C represents the parasternal short axis. In Fig. 12, the confusion matrix to for 3 cardio-locations classification is shown. Location B achieved the highest classification accuracy of 100 %. Location A achieved intermediate accuracy score of 99.5%. Location C achieved the lowest classification accuracy of 98%.

In this experiment, we compare our proposed system with the state of the art based on the evaluation criteria introduced in the evaluation criteria section as shown in Fig.13. The proposed system achieved the highest performance through several metrics. It achieved 96.3% accuracy greater than the state of the art accuracy with a significant increase of 4.2%. The sensitivity of our proposed system achieved 95.75% greater than the state of the art sensitivity with a significant increase of 4.2%. The precision of our proposed system achieved 96.41 % greater than the state of the art precision with a significant increase of 4.4%. On the other hand, our proposed system achieved a little impact more on the specificity performance more than the state of art specificity with increase of 0.6%.

In this experiment, we compare between our proposed system and the state of art based on each cardio view classification accuracy as shown in Fig.14. For A2C cardio-view classification accuracy, our proposed system achieved equal performance classification accuracy with 100 % accuracy

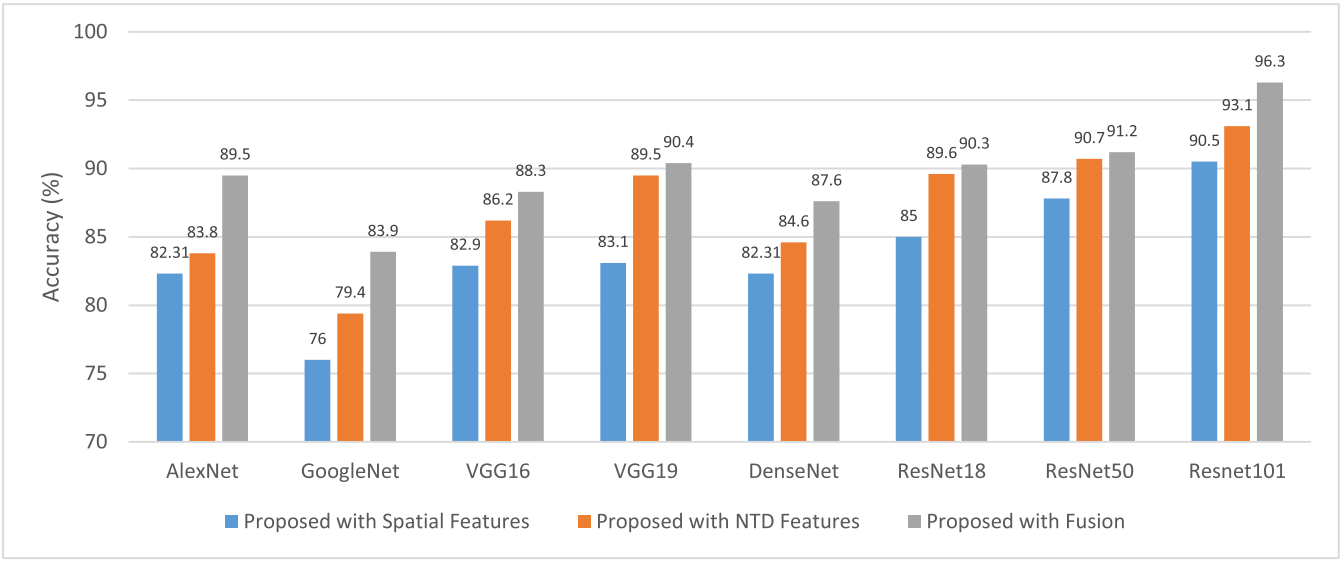

FIGURE 10. The proposed system accuracy based on several pre-trained network activation with (spatial features, NTD features, and fused CNN features).

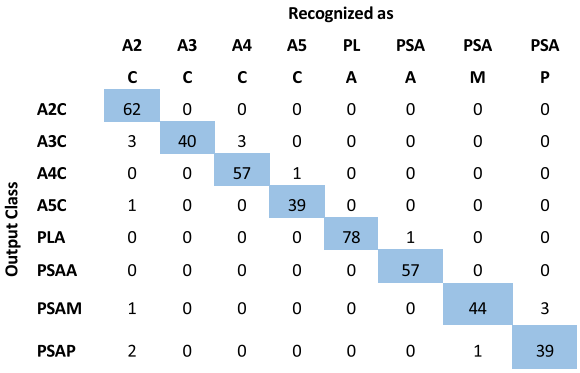

FIGURE 11. Confusion matrix for our proposed system to classify 8 cardio-views.

FIGURE 12. Confusion matrix for our proposed system to classify three cardio-Locations.

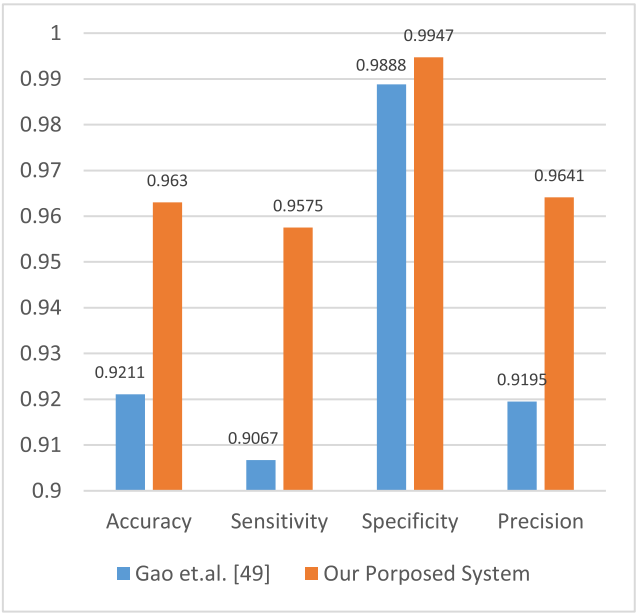

FIGURE 13. A comparison between our proposed system and the state of art for 8 cardio-views classification based on several metrics.

score. Our proposed system reflects the highest classification accuracy for PSAP, PSAM, PSAA, PLA, and A5C. However, the state of art achieved more accuracy than our proposed system for A3C and A4C.

In this experiment, we compare the state of the art [48], and our proposed method based on each cardio location classification accuracy as shown in Fig.15. The proposed system achieved 99.1 % accuracy greater than the state of the art accuracy with a significant increase of 1.1 %. The proposed system achieves more classification accuracy than the state of the art for both location A and location B. However, the state of the art achieved a little increase in the classification accuracy more than our proposed one with 0.5%. This can be explained by the low classification accuracy of A3C and A4C classification accuracy.

In this experiment, we compare between our proposed system and the previous state of art [48] based on the required processing time for both features extraction and training procedures. Table.1 shows the training processing time between

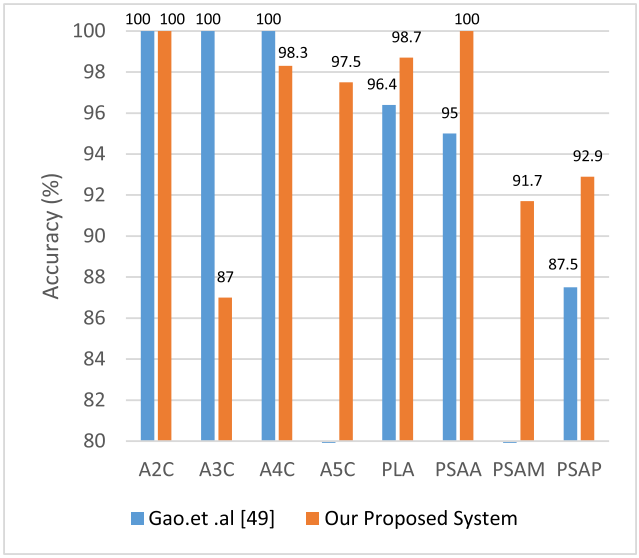

FIGURE 14. A comparison between our proposed system and the state of art to classify each cardio-view.

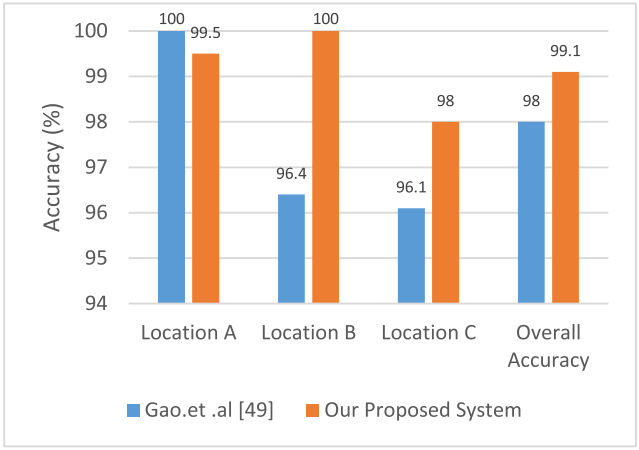

FIGURE 15. A comparison between our proposed system and the state of art for each cardio-location classification accuracy.

TABLE 1. Time cost of our proposed system vs. the state of the art [48].

	Features Extraction Time Cost	Classifier Training Time Cost
Gao.et .al [48]	Two Days	Two Days
Our Proposed System	29 min 30 sec	3 min 43 sec

our proposed system and Gao.et.al [48]. The state of the art is consumed about two days for offline handcrafted features extraction and two days for training CNN from scratch. One of the advantages of our proposed system is the lower processing time that it achieved. It only consumes about 14 min for CNN deep features extraction through deep activation. On the other hand, the LSTM classifier training consumed only about 4 min to achieve 96.3% accuracy.

V. CONCLUSION

In this paper, an echo-cardio views and locations classification system architecture have been proposed. Moreover, novel temporal features descriptors based on neutrosophic sets similarity have been proposed. We propose a novel spatial-temporal deep features pole based on concatenation fusion. The proposed neutrosophic temporal descriptors achieved the highest accuracy more than the previous temporal descriptors or even deep spatial features. ResNet 101 architecture achieved the best performance as a deep features extractor for both spatial, temporal features, and fused features pole. A significant improvement in the classification accuracy of cardio-view classification accuracy score 96.3% and 99.1% for cardio-location classification accuracy. Adam optimizer achieved the best performance for LSTM classifier. The proposed system has been saved the time cost required for training and features extraction procedure. Experimental results were very promising and we suggest generalizing our methods for other medical application recognition problems. In the future work, we suggest increasing the system performance by employing a robust optimization-training algorithm.

ACKNOWLEDGMENT

The authors would like to thank Gao, *et al.* [48] for providing their valuable echocardiography dataset.

REFERENCES

[1] B. Pinamonti, E. Abate, A. De Luca, G. Finocchiaro, and R. Korcova, "Role of cardiac imaging: Echocardiography," in *Dilated Cardiomyopathy: From Genetics to Clinical Management [Internet]*, G. Sinagra, M. Merlo, and B. Pinamonti, Eds. Cham, Switzerland: Springer, May 2020, pp. 83–111.

[2] H. Vaseli, Z. Liao, A. H. Abdi, H. Girgis, D. Behnami, C. Luong, F. T. Dezaki, N. Dhungel, R. Rohling, K. Gin, P. Abolmaesumi, and T. Tsang, "Designing lightweight deep learning models for echocardiography view classification," *Proc. SPIE*, vol. 10951, Mar. 2019, Art. no. 109510F.

[3] J. Verjans, W. B. Veldhuis, G. Carneiro, J. M. Wolterink, I. Išgum, and T. Leiner, "Cardiovascular diseases," in *Artificial Intelligence in Medical Imaging*, E. Ranschaert, S. Morozov, and P. Algra, Eds. Cham, Switzerland: Springer, 2019.

[4] P. J. Sheth, G. H. Danton, Y. Siegel, R. E. Kardon, J. C. Infante, E. Ghersin, and J. E. Fishman, "Cardiac physiology for radiologists: Review of relevant physiology for interpretation of cardiac MR imaging and CT," *RadioGraphics*, vol. 35, no. 5, pp. 1335–1351, Sep. 2015.

[5] J. K. Dave, M. E. Mc Donald, P. Mehrotra, A. R. Kohut, J. R. Eisenbrey, and F. Forsberg, "Recent technological advancements in cardiac ultrasound imaging," *Ultrasonics*, vol. 84, pp. 329–340, Mar. 2018.

[6] D. D. B. Carvalho, A. M. Arias Lorza, W. J. Niessen, M. de Bruijne, and S. Klein, "Automated registration of freehand B-Mode ultrasound and magnetic resonance imaging of the carotid arteries based on geometric features," *Ultrasound Med. Biol.*, vol. 43, no. 1, pp. 273–285, Jan. 2017.

[7] J. B. Liu, D. A. Merton, F. Forsberg, and, B. B Goldberg, "Contrast-enhanced ultrasound imaging," in *Diagnostic Ultrasound*. Boca Raton, FL, USA: CRC Press, 2019, pp. 51–74.

[8] D. S. Teyhen, C. E. Miltenberger, H. M. Deiters, Y. M. D. Toro, J. N. Pulliam, J. D. Childs, R. E. Boyles, and T. W. Flynn, "The use of ultrasound imaging of the abdominal drawing-in maneuver in subjects with low back pain," *J. Orthopaedic Sports Phys. Therapy*, vol. 35, no. 6, pp. 346–355, Jun. 2005.

[9] Y. Lu, J. Li, X. Zhao, J. Li, J. Feng, and E. Fan, "Breast cancer research and treatment reconstruction of unilateral breast structure using three-dimensional ultrasound imaging to assess breast neoplasm," *Breast Cancer Res. Treatment*, vol. 176, no. 1, pp. 87–94, Jul. 2019.

[10] D. W. Gould, E. L. Watson, T. J. Wilkinson, J. Wormleighton, S. Xenophontos, J. L. Viana, and A. C. Smith, "Ultrasound assessment of muscle mass in response to exercise training in chronic kidney disease: A comparison with MRI," *J. Cachexia, Sarcopenia Muscle*, vol. 10, no. 4, pp. 748–755, Aug. 2019.

[11] A. Ferrero, R. Lo Tesoriere, and N. Russolillo, "Ultrasound liver map technique for laparoscopic liver resections," *World J. Surg.*, vol. 43, no. 10, pp. 2607–2611, Oct. 2019.

[12] A. Karlapalem, A. H. Givan, M. Fernandez-Del-Valle, M. R. Fulton, and J. D. Klingensmith, "Classification of cardiac adipose tissue using spectral analysis of ultrasound radiofrequency backscatter," *Proc. SPIE*, vol. 10955, Mar. 2019, Art. no. 109550F.

[13] A. Lao, V. Sharma, M. Katz, and A. Alexandrov, "Diagnostic criteria for transcranial Doppler ultrasound," in *Diagnostic Ultrasound: Logical Approach*, J. P. McGahan and B. B. Goldberg, Eds., 2nd ed. New York, NY, USA: Informa Healthcare, 2008, pp. 552–554.

[14] A. Ghorbani, D. Ouyang, A. Abid, B. He, J. H. Chen, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Deep learning interpretation of echocardiograms," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–10, Dec. 2020.

[15] I. Ghori, D. Roy, R. John, and K. M. Chalavadi, "Echocardiogram analysis using motion profile modeling," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1767–1774, May 2020.

[16] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[17] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[19] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[20] N. Qiang, Q. Dong, W. Zhang, B. Ge, F. Ge, H. Liang, Y. Sun, J. Gao, and T. Liu, "Modeling task-based fMRI data via deep belief network with neural architecture search," *Computerized Med. Imag. Graph.*, vol. 83, Jul. 2020, Art. no. 101747.

[21] X. SHI, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[24] Y. Gu, Z. Shao, L. Qin, W. Lu, and M. Li, "A deep learning framework for cycling maneuvers classification," *IEEE Access*, vol. 7, pp. 28799–28809, 2019, doi: 10.1109/ACCESS.2019.2898852.

[25] Z. Liu, B. Huang, Y. Cui, Y. Xu, B. Zhang, L. Zhu, Y. Wang, L. Jin, and D. Wu, "Multi-task deep learning with dynamic programming for embryo early development stage classification from time-lapse videos," *IEEE Access*, vol. 7, pp. 122153–122163, 2019, doi: 10.1109/ACCESS.2019.2937765.

[26] Y. Tao, Z. Ling, and I. Patras, "Universal foreground segmentation based on deep feature fusion network for multi-scene videos," *IEEE Access*, vol. 7, pp. 158326–158337, 2019, doi: 10.1109/ACCESS.2019.2950639.

[27] T. Ogawa, Y. Sasaka, K. Maeda, and M. Haseyama, "Favorite video classification based on multimodal bidirectional LSTM," *IEEE Access*, vol. 6, pp. 61401–61409, 2018, doi: 10.1109/ACCESS.2018.2876710.

[28] I. U. Haq, K. Muhammad, A. Ullah, and S. W. Baik, "DeepStar: Detecting starring characters in movies," *IEEE Access*, vol. 7, pp. 9265–9272, 2019, doi: 10.1109/ACCESS.2018.2890560.

[29] P. Wang, W. Hao, Z. Sun, S. Wang, E. Tan, L. Li, and Y. Jin, "Regional detection of traffic congestion using in a large-scale surveillance system via deep residual TrafficNet," *IEEE Access*, vol. 6, pp. 68910–68919, 2018, doi: 10.1109/ACCESS.2018.2879809.

[30] M. A. Khan, T. Akram, M. Sharif, M. Y. Javed, N. Muhammad, and M. Yasmin, "An implementation of optimized framework for action classification using multilayers neural network on selected fused features," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1377–1397, Nov. 2019.

[31] E. E. Cust, A. J. Sweeting, K. Ball, and S. Robertson, "Machine and deep learning for sport-specific movement recognition: A systematic review of model development and performance," *J. Sports Sci.*, vol. 37, no. 5, pp. 568–600, Mar. 2019.

[32] K.-J. Kim, P.-K. Kim, Y.-S. Chung, and D.-H. Choi, "Multi-scale detector for accurate vehicle detection in traffic surveillance data," *IEEE Access*, vol. 7, pp. 78311–78319, 2019, doi: 10.1109/ACCESS.2019.2922479.

[33] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, A. Patsekin, J. Kindelsberger, L. Ding, S. Seaman, A. Mehler, A. Sipperley, A. Pettinato, B. D. Seppelt, L. Angell, B. Mehler, and B. Reimer, "MIT advanced vehicle technology study: large-scale naturalistic driving study of driver behavior and interaction with automation," *IEEE Access*, vol. 7, pp. 102021–102038, 2019, doi: 10.1109/ACCESS.2019.2926040.

[34] S. U. Khan, I. U. Haq, S. Rho, S. W. Baik, and M. Y. Lee, "Cover the violence: A novel Deep-Learning-Based approach towards violence-detection in movies," *Appl. Sci.*, vol. 9, no. 22, p. 4963, Nov. 2019.

[35] G. Sreenu and M. A. Saleem Durai, "Intelligent video surveillance: A review through deep learning techniques for crowd analysis," *J. Big Data*, vol. 6, no. 1, pp. 6–48, Dec. 2019, doi: 10.1186/s40537-019-0212-5.

[36] F. Ciompi, K. Chung, S. J. van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. W. Wille, A. Marchianò, U. Pastorino, M. Prokop, and B. van Ginneken, "Erratum: Corrigendum: Towards automatic pulmonary nodule management in lung cancer screening with deep learning," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, Art. no. 46479.

[37] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in CT images," in *Proc. 12th Conf. Comput. Robot Vis.*, Jun. 2015, pp. 133–138.

[38] P. Moeskops, M. Veta, M. W. Lafarge, K. A. J. Eppenhof, and J. P. W. Pluim, "Adversarial training and dilated convolutions for brain MRI segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 2017, pp. 56–64.

[39] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Informat.*, vol. 7, no. 1, p. 29, 2016.

[40] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, and L. Li, "Discrimination of breast cancer with microcalcifications on mammography by deep learning," *Sci. Rep.*, vol. 6, no. 1, pp. 1–9, Jun. 2016.

[41] J. Ker, L. Wang, J. Rao, and T. Lim, "Deep learning applications in medical image analysis," *IEEE Access*, vol. 6, pp. 9375–9389, 2018, doi: 10.1109/ACCESS.2017.2788044.

[42] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Comput. Sci.*, vol. 102, pp. 317–324, Jan. 2016.

[43] S. Vaishali, K. K. Rao, and G. V. S. Rao, "A review on noise reduction methods for brain MRI images," in *Proc. Int. Conf. Signal Process. Commun. Eng. Syst.*, Jan. 2015, pp. 363–365.

[44] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, "Skin lesion segmentation in clinical images using deep learning," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 337–342.

[45] L. Duran-Lopez, F. Luna-Perejon, I. Amaya-Rodriguez, J. Civit-Masot, A. Civit-Balcells, S. Vicente-Diaz, and A. Linares-Barranco, "Polyp detection in gastrointestinal images using faster regional convolutional neural network," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 626–631.

[46] A. Karuzas, K. Sablauskas, L. Skrodenis, D. Verikas, E. Rumbinaite, D. Zaliaduonyte-Peksiene, K. Ziuteliene, J. J. Vaskelyte, R. Jurkevicius, and J. Plisiene, "P1465Artificial intelligence in echocardiography—Steps to automatic cardiac measurements in routine practice," *Eur. Heart J.*, vol. 40, no. 1, Oct. 2019, Art. no. ehz748-0230.

[47] A. H. Abdi, C. Luong, T. Tsang, J. Jue, K. Gin, D. Yeung, D. Hawley, R. Rohling, and P. Abolmaesumi, "Quality assessment of echocardiographic cine using recurrent neural networks: Feasibility on five standard view planes," in *Medical Image Computing and Computer Assisted Intervention MICCAI* (Lecture Notes in Computer Science), vol. 10435. Cham, Switzerland: Springer, 2017, pp. 302–310.

[48] X. Gao, W. Li, M. Loomes, and L. Wang, "A fused deep learning architecture for viewpoint classification of echocardiography," *Inf. Fusion*, vol. 36, pp. 103–113, Jul. 2017.

[49] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *NPJ Digit. Med.*, vol. 1, no. 1, pp. 1–8, Dec. 2018.

[50] K. Chykeyuk, D. A. Clifton, and J. A. Noble, "Feature extraction and wall motion classification of 2D stress echocardiography with support vector machines," *Proc. SPIE*, vol. 7963, Mar. 2011, Art. no. 79630H.

[51] G. Belous, A. Busch, and D. Rowlands, "Segmentation of the left ventricle from ultrasound using random forest with active shape model," in *Proc. 1st Int. Conf. Artif. Intell., Modeling Simulation*, Dec. 2013, pp. 315–319, doi: 10.1109/AIMS.2013.58.
[52] H. Moghaddasi and S. Nourian, "Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos," *Comput. Biol. Med.*, vol. 73, pp. 47–55, Jun. 2016.
[53] S. Sanchez-Martinez, N. Duchateau, T. Erdei, A. G. Fraser, B. H. Bijnens, and G. Piella, "Characterization of myocardial motion patterns by unsupervised multiple kernel learning," *Med. Image Anal.*, vol. 35, pp. 70–82, Jan. 2017.
[54] H. W. Rahmouni, B. Ky, T. Plappert, K. Duffy, S. E. Wiegers, V. A. Ferrari, M. G. Keane, J. N. Kirkpatrick, F. E. Silvestry, and M. St. John Sutton, "Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence-assisted border detection," *Amer. Heart J.*, vol. 155, no. 3, pp. 562–570, Mar. 2008.
[55] C. Knackstedt, S. C. A. M. Bekkers, G. Schummers, M. Schreckenberg, D. Muraru, L. P. Badano, A. Franke, C. Bavishi, A. M. S. Omar, and P. P. Sengupta, "Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain," *J. Amer. College Cardiol.*, vol. 66, no. 13, pp. 1456–1466, Sep. 2015.
[56] H. Khamis, G. Zurakhov, V. Azar, A. Raz, Z. Friedman, and D. Adam, "Automatic apical view classification of echocardiograms using a discriminative learning dictionary," *Med. Image Anal.*, vol. 36, pp. 15–21, Feb. 2017.
[57] S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, and P. P. Sengupta, "Machine-learning algorithms to automate morphological and functional assessments in 2D echocardiography," *J. Amer. College Cardiol.*, vol. 68, no. 21, pp. 2287–2295, Nov. 2016.
[58] H. A. Omar, J. S. Domingos, A. Patra, R. Upton, P. Leeson, and J. A. Noble, "Quantification of cardiac bull's-eye map based on principal strain analysis for myocardial wall motion assessment in stress echocardiography," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1195–1198.
[59] Y. Guo, A. Şengür, and J. Ye, "A novel image thresholding algorithm based on neutrosophic similarity score," *Measurement*, vol. 58, pp. 175–186, Dec. 2014.
[60] Y. Guo and A. Şengür, "A novel image edge detection algorithm based on neutrosophic set," *Comput. Electr. Eng.*, vol. 40, no. 8, pp. 3–25, Nov. 2014.
[61] M. Afifi, "11K hands: Gender recognition and biometric identification using a large dataset of hand images," *Multimedia Tools Appl.*, vol. 78, no. 15, pp. 20835–20854, Aug. 2019.
[62] M. Hammad, Y. Liu, and K. Wang, "Multimodal biometric authentication systems using convolution neural network based on different level fusion of ECG and fingerprint," *IEEE Access*, vol. 7, pp. 26527–26542, 2019.
[63] P. Li, M. Abdel-Aty, and J. Yuan, "Real-time crash risk prediction on arterials based on LSTM-CNN," *Accident Anal. Prevention*, vol. 135, Feb. 2020, Art. no. 105371.

AHMED I. SHAHIN received the M.Sc. degree in biomedical engineering from Helwan University, Cairo, in 2014, and the Ph.D. degree in deep learning and computer vision from the Faculty of Engineering, Cairo University, in 2018. He is currently an Associate Professor with the Department of Applied Sciences, Faculty of Community, Majmaah University, Saudi Arabia. His research interests include artificial intelligence, deep learning, image processing, and the IoT.

SULTAN ALMOTAIRI received the B.Sc., M.Sc., and Ph.D. degrees in computer science from the Florida Institute of Technology, Melbourne, USA, in 2010, 2012, and 2014, respectively. He has been the Dean of the Community College, Majmaah University, since June 2015. He is currently an Associate Professor with the Department of Natural and Applied Sciences, Community College, Majmaah University. In 2016, he was elected as the Chairman of the Municipality Council of Majmaah. His research interests include neural networks, deep learning, pattern recognition, machine learning, image processing, and computer vision.