IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 18, 2019, accepted May 29, 2019, date of publication June 14, 2019, date of current version July 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2923008

Abnormality Detection in Musculoskeletal Radiographs Using Capsule Network

A. F. M. SAIF[1], (Member, IEEE), CELIA SHAHNAZ[1], (Senior Member, IEEE), WEI-PING ZHU[2], (Senior Member, IEEE), AND M. O. AHMAD[2], (Fellow, IEEE)

[1]Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[2]Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada

Corresponding author: Wei-Ping Zhu (weiping@ece.concordia.ca)

ABSTRACT To treat the diseases or injuries of the joints, bones, muscles, and spine in both adult and pediatric imaging the musculoskeletal radiographs bring a significant depth of expertise. Abnormality detection in the musculoskeletal study is backbreaking as more than 1.7 billion people are affected by musculoskeletal condition (BMU, 2017). Hence if we want to create enough opportunity to treat a maximum amount of patients, machine learning and deep learning can play a crucial role. CNN is an excellent deep learning method for image classification and other computer vision tasks. But CNN has exhibited some serious limitations when the images are rotated and deformed. Hence capsule network architecture is introduced in this paper for musculoskeletal radiographs abnormality detection and this capsnet architecture has shown very promising features that can help to vanquish the limitations of CNN. In addition, this capsule network has scored 10% higher kappa score than 169 layer densenet using less training data in the case of musculoskeletal radiographs abnormality detection. This feature of capsule network can help to use deep learning in such cases where an aggregate of a large amount of data is not possible. For image quality investigation, blind image spatial quality evaluator (BRISQUE) and naturalness image quality evaluator (NIQE) scores are measured and it is found that when the pixel size of the resized images are more close to the pixel size of the original images, we get a better approximation. Hence in the case of musculoskeletal radiographs abnormality detection, our method outperforms state-of-the-art method using a less amount of training data.

INDEX TERMS Capsule network, routing-by-agreement, squashing, margin loss, Cohen's kappa statistic.

I. INTRODUCTION

Circumstances, ranging from work accidents and sports injuries to genetics and lifestyle choices can be the cause of Musculoskeletal problem. Injuries, osteoarthritis of the knee, osteoporosis of the bones and many other joint or muscle issues are the result of Musculoskeletal problem. Proper diagnosis and abnormality detection are very important for further treatment. But a large number of patients has made this task very difficult and time-consuming and so, computer-based automatic detection of abnormality can become very handy as well as time-saving.

Various machine learning processes have played a significant role in medical image classification. Decision Forests [1] has shown significant results in image classification. Support vector machine [2], [3] is another approach for medical image classification. K-Means clustering [4], Integrating spatial fuzzy clustering [6], Gaussian model [7] are also very popular algorithm used for medical image classification. In contrast in the field of deep learning, Convolutional Neural Network (CNN or ConvNet) [8], [9] has been used extensively in classifying images and segmentation problems. But CNN is not any state of the art architecture. It has shown some major drawbacks. If CNN is trained with data sets of images having an orientation to identify whether it contains a specific type and if it is not trained with images having orientation similar to that type, then for images rotated and deformed, CNN classifier does not produce a correct classification. These limitations of CNN come from the fact that its neurons are activated based on the chances of detecting specific feature. Properties of a feature such as orientation, size, velocity, color etc are not considered by neurons. Determining the spatial relationship in CNN requires precise location of features in the input image. To achieve translation invariance

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

MaxPooling is performed. Translation Invariance indicates that CNN will classify the input image in the same way regardless of how the information within the image is shifted and the features location information is lost at the Pooling layer. Again the performance of the neural network depends on the depth of the architecture. Adding more layers will reserve more information and improve performance. But that also increases the computational complexity and computational cost.

In contrast information at the neuron level is stored as vectors in capsule network rather than scalars like neural networks. The vector output of a capsule uses a powerful mechanism, dynamic routing [10]. These vectors contain information about: spatial orientation, magnitude/prevalence, and other attributes of the extracted feature. The main objective of this paper is to:

- Find an improved architecture for estimating abnormality in the musculoskeletal condition which can maximize the abnormality detection rate
- Examine the ability of capsule network in the case of image classification and processing and also investigate the over-fitting problem in capsnet.
- And also want to find out whether capsnet can outperform the convolutional neural network

MURA [12] dataset is used for training and testing the capsule network. In MURA dataset a 169 layer neural network(densenet) provides better performance in comparison to the best radiologist performance in detecting abnormalities on finger and wrist studies. This model performs lower than best radiologist performance in case of detecting the abnormality in elbow, forearm, hand, humerus and shoulder studies. Hence the aim of this paper is to find a more suitable architecture that can identify the abnormality in all of the cases more accurately than the previous method.

II. BACKGROUND AND RELATED WORKS

A. PREVIOUS WORKS

Medical image classification is an attractive topic in the field of computer vision and biomedical image processing. Machine learning algorithms and deep learning have been used for medical image processing and they deliver impressive result in such cases. These machine learning methods have achieved reasonable accuracy in many domains but they generally require very specific hand engineered features to work which greatly diminishes their ability to generalize to related problems. In contrast, neural network extracts feature by its own.

Capsule network also extracts features by its own but conserves spatial information through vector output. In ''BRAIN TUMOR TYPE CLASSIFICATION VIA CAPSULE NETWORKS'' [13] capsule network is used to classify brain tumor. It is found that capsnet performs better than CNN with less training data. In ''Capsules for Object Segmentation'' [14] a convolutional-deconvolutional capsule network called SegCaps based on capsule network is proposed which shows strong results for the task of object segmentation.

B. PROBLEM DEFINITION

Finding abnormality in musculoskeletal radiographs is a very difficult task and if automatic detection of abnormality can be introduced, it would be very helpful for further diagnosis and treatment. In this regards, MURA dataset was published with 40,561 images from 14,863 studies. In this paper, a capsule network is designed to classify normal and abnormal condition and compared the result with densenet architecture.

III. CONVOLUTIONAL NEURAL NETWORKS AND ITS LIMITATIONS

The convolutional neural network works on three steps: convolution, relu, pooling. These steps create each layer and adding more layers create a deep convolutional neural network. Each layer receives input from the previous layer. Convolution the main building blocks of a CNN is a mathematical combination of two functions to produce a third function and merges two sets of information. This convolution is computed between input data and filter. They produce third information known as feature map. For adding non-linearity relu is used. Then comes the pooling layer which is used to decrease the parameter and for convolutional neural network maxpooling is extensively used. In the end, they all connected to a fully connected layer.

Comparing CNN with human brain some serious shortcomings have been found. CNN suffers from poor translational invariance and lack of information about orientation. CNN faces a problem when objects are rotated or when lighting conditions are changed. Again, pooling is an important part of CNN structure. Pooling which is introduced to reduce redundancy of representation and reduce the number of parameters, recognizing that precise location is not important for object classification. Maxpooling helps to speed up the CNN. Maxpooling just picks the neuron with the highest activation but it causes some serious labyrinth by losing the feature location information. For this reason, CNN cannot identify any kind of deformation or rotation in the image and so they need enormous training data and rotation of data.

All these aforementioned drawbacks are the inspiration behind the capsule network which is more robust to translation and rotation. In capsule network scaler output is replaced by vector and max-pooling is replaced by more effective routing-by-agreement which prevents data information lose.

IV. PROPOSED METHOD FOR MUSCULOSKELETAL RADIOGRAPHS ABNORMALITY DETECTION

In this paper, the capability of capsnet in classifying musculoskeletal radiographs is studied by designing a capsnet for detecting the abnormality in musculoskeletal radiographs. Data is preprocessed and then the processed data is fed in the designed network.

A. PREPROCESSING DATA

As a part of preprocessing, image data is resized and normalized. As the radiographs are of different size, resizing is needed to get an equal image size. At first they are resized

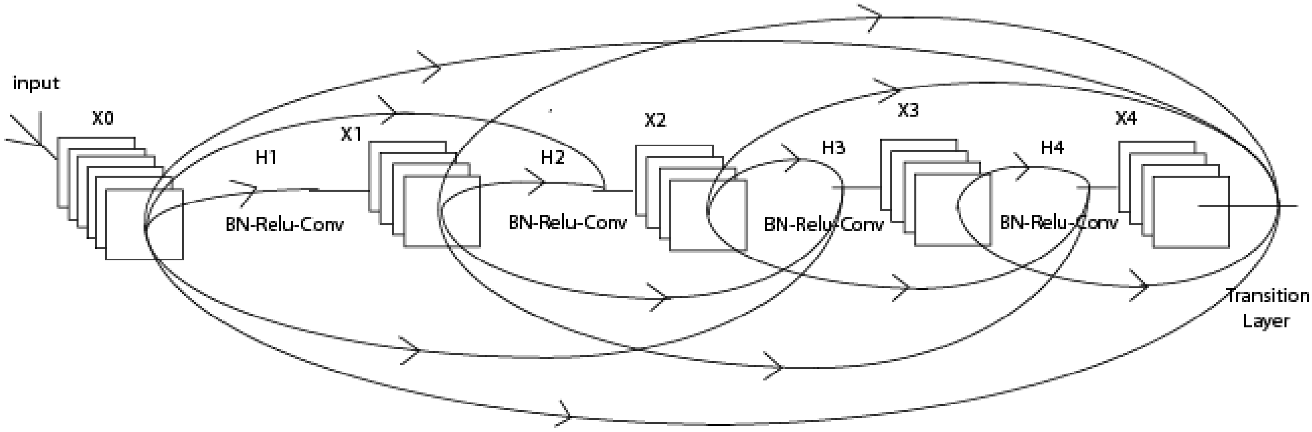
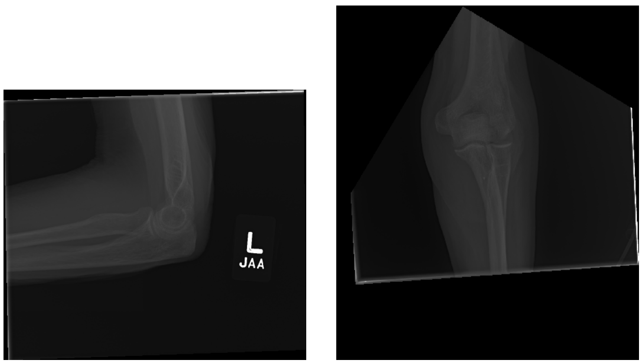


FIGURE 1. A 5-layer densely connected convolutional neural network(densenet).



(a) (b)

FIGURE 2. Original images from the dataset showing variability in image size. (a) Abnormal elbow. (b) Normal elbow.

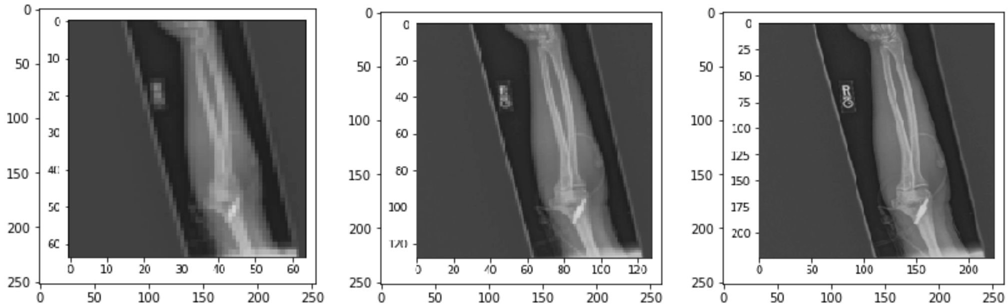
to 64 × 64 pixel size. Then they are resized to 128 × 128 pixel size and at last all the images are resized to 224 × 224 pixel size. All these three different types of resized images have some merits and demerits when they are used to train the network. When the image size is 64 × 64, most of the features are lost and as a result very poor image quality is obtained which provides very low accuracy and very high loss. But when the feature maps are reduced, it takes very small amount of time to train the network. In the second case, when the image size is 128 × 128 accuracy is improved and loss is reduced. But as the feature maps are increased it takes almost three times more time to train the network. And for the last case when the image size is 224 × 224 the best accuracy is obtained and loss is also very small. But it takes more time than the first two types mentioned above to train. From fig.3 the discussion mentioned above can be related. Table 1 shows the Blind Image Spatial Quality Evaluator (BRISQUE) and Naturalness Image Quality Evaluator (NIQE) score [15], [16]. The BRISQUE and the NIQE algorithms calculate the quality score of an image with computational efficiency after the model is trained and in this case MURA dataset is used for training purpose. Lower BRISQUE and NIQE score indicate that the resized images conserve more features or we lose less features after resizing. Images are also normalized as a part of preprocessing. Hence fig.3 is a presentation of resized and normalized images. To get a better result in CNN, data augmentation is needed in the training phase but capsule network can perform better without data augmentation in the training phase and for that reason, we do not apply data augmentation while training capsule network but we have used data augmentation while training densenet. In the result section, we have compared the accuracy between capsule network and densenet and the effect of resizing and normalizing of images with their respective accuracy is also described.

B. GENERAL STRUCTURE OF CAPSULE NETWORK

Many small groups of neurons called capsules create each layer in a capsule network. At any capsule layer n, a set of capsule types consists of $a^n \times b^n$ grid of z^n dimensional child capsule and this is the output of layer n-1. At the next $(n+1)^{th}$ layer of the network, a set of capsule types is an $a^{n+1} \times b^{n+1}$ grid of z^{n+1} dimensional parent capsules, where $a^{n+1} \times b^{n+1}$ is the spatial dimension of the output of layer n. Output vector of a capsule represent the probability that the entity represented by the capsule is present in the current input. Then non-linear ''squashing'' function is used to ensure that short vectors get shrunk to almost zero length so that it does not route in the mother capsule and long vectors get shrunk to a length below 1. If v_j is the vector output of capsule j and s_j is its total output then they used the following non-linear function

$$v_j = \frac{\| s_j \|^2}{1 + \| s_j \|^2} \frac{s_j}{\| s_j \|} \tag{1}$$

The input to a mother capsule s_j is a weighted sum of prediction vectors $\hat{\boldsymbol{u}}_{ji}$ is calculated from the output of child capsules by multiplying the output u_i of a child capsule output by a



(a) (b) (c)

FIGURE 3. Down sampling degrade image quality and as the pixel size is increased more features are included. (a) image size 64 × 64. (b) image size 128 × 128. (c) image size 224 × 224.

weighted matrix $\boldsymbol{W}_{ij}$

$$s_j = \sum_i c_{ij} \hat{\boldsymbol{u}}_{ji} \tag{2}$$

$$\hat{\boldsymbol{u}}_{ji} = \boldsymbol{W}_{ij} \boldsymbol{u}_i \tag{3}$$

Coupling coefficients c_{ij} indicates the coupling or bonding between the mother or higher capsule and the child or lower capsule and is calculated by using softmax function

$$c_{ij} = \frac{\exp b_{ij}}{\sum_k \exp b_{ik}} \tag{4}$$

where b_{ij} is the log probability that indicates whether capsule i should be coupled with capsule j and it's initial value is set to 0 at the beginning of the routing by agreement process. The log probabilities is updated in the routing process based on the agreement between $\boldsymbol{v}_j$ and $\hat{\boldsymbol{u}}_{ji}$ and so they produce a large inner product as

$$a_{ij} = \boldsymbol{v}_j . \hat{\boldsymbol{u}}_i \tag{5}$$

This agreement is treated as if it is a log likelihood and is added to the initial logit, b_{ij} before computing the new values for all the coupling coefficients linking capsule i to higher level capsules.

C. MARGIN LOSS

Margin loss is introduced to ensure intraclass compactness and interclass separability. In capsnet margin loss is used to describe the presence of any class in corresponding capsule. Proper identification of classes reduce the margin loss. As the capsules outputs are vector so the margin loss depend on the length of the vector. If any entity is present in any capsule output then its instantiation vector must have the largest length means the top-level capsule for digit class k to have a long instantiation vector if and only if that entity is present in the image. So a separate margin loss, L_k is introduced for each digit capsule k:

$$L_k = T_k max(0, m^+ - || v_k ||)^2 + \lambda(1 - T_k) max(0, || v_k || - m^-)^2 \tag{6}$$

where T_k is 1 whenever class k is actually present, and is 0 otherwise. Terms m^+, m^-, and λ are hyper parameters to be indicated before the learning process.

D. PROPOSED CAPSNET ARCHITECTURE FOR MUSCULOSKELETAL RADIOGRAPHS ABNORMALITY DETECTION

The structure is designed for three types of image size, they are 64 × 64, 128 × 128 or 224 × 224. Here image size of 224 × 224 is only considered for describing the structure. The summary of the layers of the proposed model which is illustrated in Fig.5 is as follows:

- The dataset contain images of variable sizes hence they are resized to 224 × 224 image size. This is the input in the input layer.
- After the input layer the first convolutional layer, conv1 which has 256, 9 × 9 convolutional kernels with stride of 1 and RELU activition function is activated. The output of this layer is the input to the primary capsules. Pixel intensities are converted to the activities of local feature detectors in this layer.
- Second layer is a convolutional capsule layer having 32 channels of convolutional 8D capsules. Primary capsule 32 × 104 × 104 capsule output and each output is an 8D vector. Each capsule is connected like a 104 × 104 grid and each is sharing weights with each other.
- The final capsule layer is a "class capsule" includes 2 capsule and each capsule has a dimension of 16.
- Additional reconstruction loss is used to encourage the class capsules to encode the instantiation parameters of the input. The output of the class capsule is fed into a decoder consisting of 3 fully connected layers that model the pixel intensities. This decoder will learn to reconstruct the input images based on the output of the capsule network. This will force the capsule network to preserve all the information required to reconstruct the digits, across the whole network. This constraint regularizes the model: it reduces the risk of overfitting the training set, and it helps generalize to new types. The decoder part is composed of three fully connected

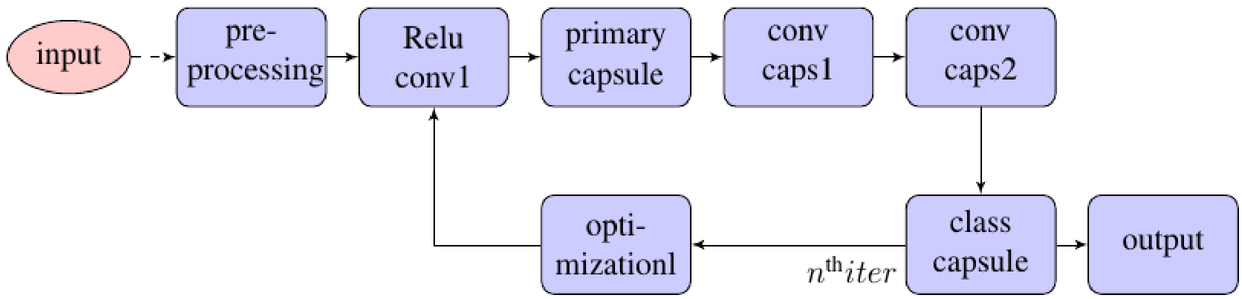


FIGURE 4. A flowchart indicating the whole process of Musculoskeletal abnormality detection using capsule network.

layers having 512, 1024 and 50176 neurons respectively. The sum of squared differences between the outputs of the logistic units and the pixel intensities is minimized. We scale down reconstruction loss by 0.0005 so that it does not dominate the margin loss during training.

- As mentioned in the aforementioned discussion, a squashing function is needed to squash the output but a problem may arise as in some cases $|| s_j ||$ can be zero and then we will get a undefined v_j. Hence a very small value ϵ is added with it so that v_j does not become undefined. In fig.4 a flowchart is introduced for better understanding of the process.

V. EXPERIMENTAL SETUP

A. DATASET

MURA dataset is used to feed the network and the result is compared with the result obtained by using 169 layer densenet. The dataset contains seven types of upper extrem musculoskeletal radiographs and classified into normal and abnormal class. In total, MURA dataset contains 14656 images. Total 13457 training images and 1199 validation images. Among them 8941 images are normal and 5715 images are abnormal. Cohen's kappa statistic is used to compare the results.

B. TRAINING AND TESTING CONDITION

Different types of resized images are used to train the network. Only 50% of data is used for training and validation and rest of the data is used for testing purpose. Randomly selecting the training and validation data and performing the training and testing task for several times the average of the results is taken. Routing is an important feature of capsnet. Different number of routing is used to see the effect. 10 iterations are used. m^+ is set equal to. 9 and m^- is set equal to. 1. λ is set to. 5. The whole training data set has not been used to train the network. As a result training time has reduced significantly.

VI. RESULTS

The final quantitative results of these experiments are discussed in this section.

TABLE 1. Accuracy obtained using variable image size and BRISQUE and NIQE score.

Image size	Average BRISQUE score	Average NIQE score
64 × 64	26.06	9.55
128 × 128	22.82	6.24
224 × 224	17.034	6

A. OBJECTIVE EVALUATION OF IMAGE SIZE

In table 1 BRISQUE and NIQE score for different image sizes have been compared. BRISQUE score and NIQE score both indicate the image quality by comparing them with reference images. In our used dataset, the reference images are the original images having a variable pixel size and from the table 1, it is clear that as we decrease the image size, the BRISQUE and NIQE score increase. And when we increase the image size, BRISQUE and NIQE score decrease. Lower BRISQUE and NIQE score indicate that the transformed images are more close to the original images.

As mentioned earlier, decreasing the image size lose information means when decreased the image size or downsample the image, image loses important features. Table 1 also justifies the reason for taking 224 × 224 pixel size of the images as this image size has the lowest BRISQUE and NIQE score.

B. COMPARISON OF RESULTS OBTAINED BY VARYING ROUTING NUMBER

Routing by agreement is an iterative process where routing algorithm acts like an orientation-popularity filter. Input from lower level capsule will send to the higher level capsule that agrees with its input.

From table 2 to 5 it can be observed that the number of routing effects the accuracy. As the number of routing is increased, accuracy gets better. But a problem arises with the increasing number of routing and this increased number of routing makes the network overfit for hand and humerus data. From fig.6 it can be seen that hand and humerus data suffer from overfitting problem. Theoretically, routing helps the child capsule's output to find the best route that agrees with its input, more and more routing can overfit

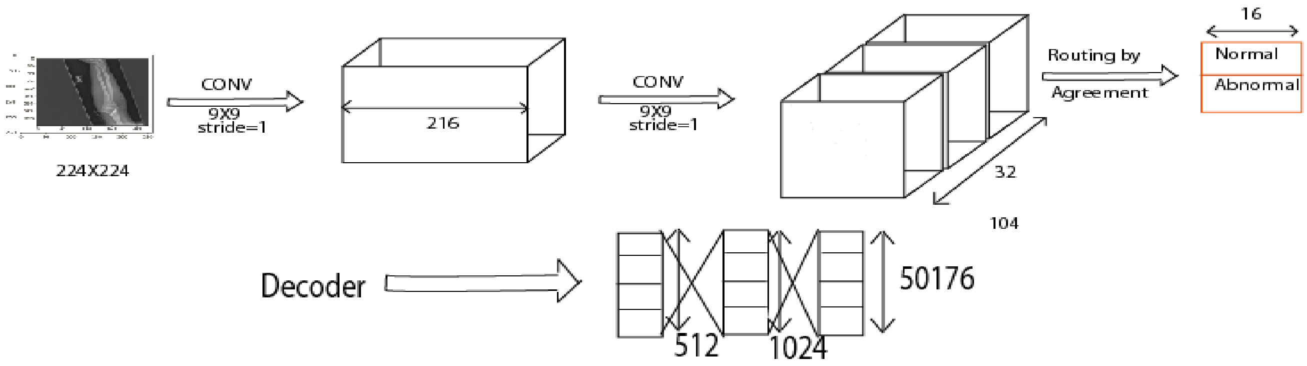


FIGURE 5. Proposed Capsule Network architecture.

TABLE 2. Accuracy obtain using one routing.

Image	Training accuracy	Validation accuracy
Finger	35.3%	29.7%
Humerus	36.27%	31.43%
Elbow	32.8%	30.1%
Forearm	38%	40.5%
Hand	39.75%	35.8%
Shoulder	42.73%	34.4%
Wrist	37.77%	33.83%

TABLE 3. Accuracy obtain using two routing.

Image	Training accuracy	Validation accuracy
Finger	65.9%	56.7%
Humerus	70.5%	67.3%
Elbow	68.7%	51.1%
Forearm	75.83%	59.8%
Hand	81.1%	77.2%
Shoulder	80.35%	69.88%
Wrist	86.6%	71.5%

TABLE 4. Accuracy obtain using three routing.

Image	Training accuracy	Validation accuracy
Finger	89.2%	72.4%
Humerus	92.43%	83.7%
Elbow	84.3%	78.87%
Forearm	87.6%	88.5%
Hand	93.6%	92.2%
Shoulder	91.7%	83.4%
Wrist	85.5%	79.87%

the data by increasing the c_{ij} value for a specific mother capsule. So to avoid overfitting we have decided to take four routing.

C. COHEN KAPPA SCORE COMPARISON

Cohen-kappa statistic is considered more robust in health care studies and in the case of musculoskeletal studies kappa statistic gives more valuable information [17], [18]. It measures inter-rater agreement for qualitative or categorical items. If there are two raters whom each classify N items into C mutually exclusive categories we can use Cohen's kappa

TABLE 5. Accuracy obtain using four routing.

Image	Training accuracy	Validation accuracy
Finger	98.3%	75.5%
Humerus	94.8%	81.2%
Elbow	88.12%	86.12%
Forearm	90.1%	89.11%
Hand	93.07%	91.09%
Shoulder	94.06%	92.08%
Wrist	96.04%	95.15%

TABLE 6. Kappa statistic score for MURA dataset using capsnet.

Image	Kappa score
Finger	.7351 (.9586,.5116)
Humerus	.754(.896,.61177)
Elbow	.7334(.7538,.713)
Forearm	.785(.795,.775)
Hand	.8351(.8557,.8145)
Shoulder	.8558(.87582,.8358)
Wrist	.90775(.9172,.8983)
Average	.80115(.8646,.7377)

TABLE 7. Kappa statistic score for MURA dataset using densenet.

Image	Kappa score
Finger	.389 (.446,.332)
Humerus	.6(.642,.558)
Elbow	.71(.745,.674)
Forearm	.737(.766,.707)
Hand	.851(.871,.83)
Shoulder	.729(.76,.697)
Wrist	.931(.94,.922)
Average	.705(.7,.71)

to measure the agreement between them. If P_o represents the relative observed agreement among raters which is identical to accuracy and P_e represents the hypothetical probability of chance agreement then the formula to calculate Cohen's kappa for two raters is:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{7}$$

Table 7 demonstrated the kappa score of densenet trained on MURA dataset [19]

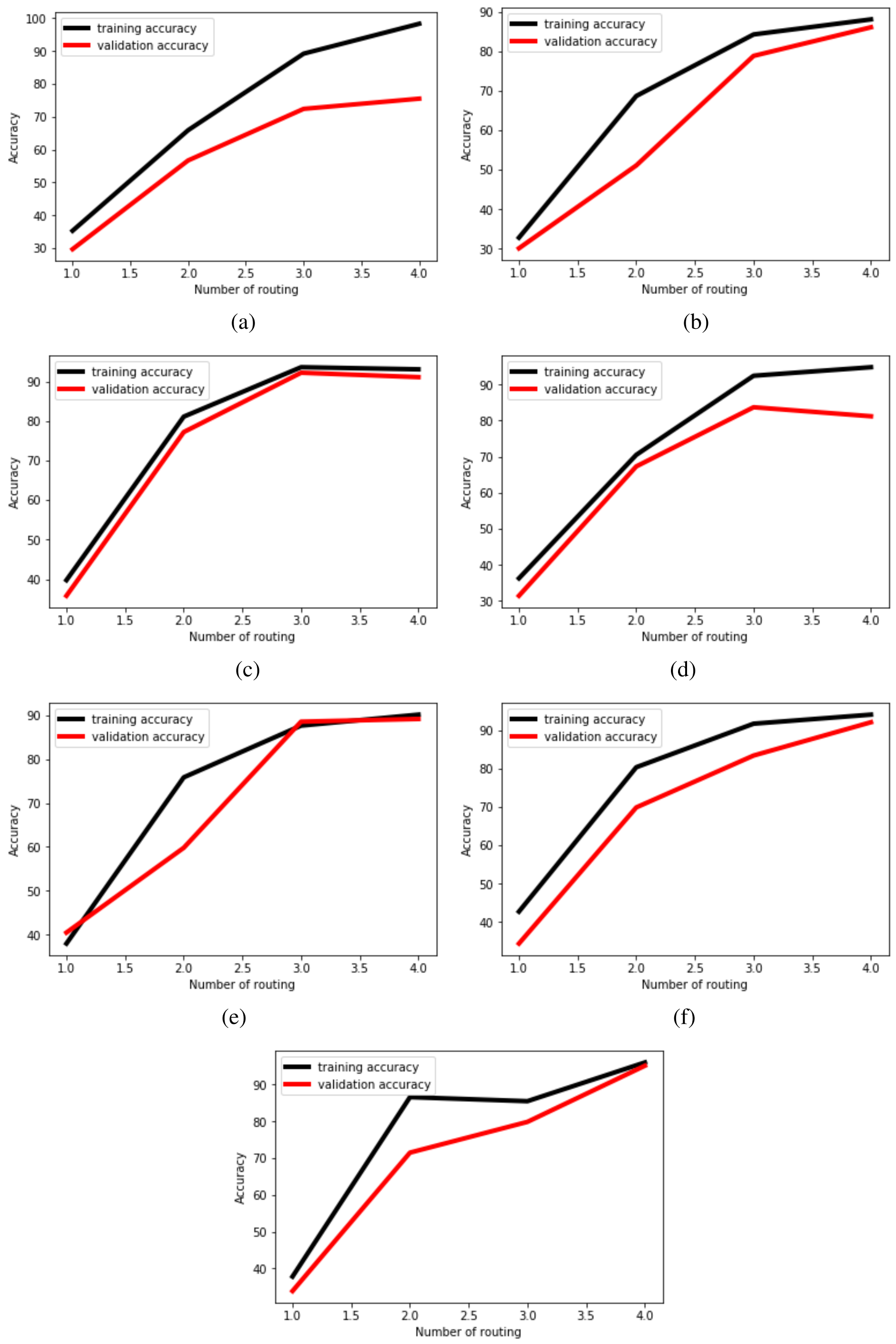


FIGURE 6. **Training-Validation accuracy vs. Number of routing graph indicating that increasing the number of routing causes overfitting in the case of hand and humerus. (a) Finger. (b) Elbow. (c) Hand. (d) Humerus. (e) Forearm. (f) Shoulder. (g) Wrist.**

From table 6 and 7 it is cleared that the proposed capsnet architecture provides almost 10% better kappa score than the 169 layers of densenet while using 50% less training data.

In fig.7 simultaneous plot of loss and accuracy for each epoch is given.

VII. FUTURE RESEARCH AREA

Capsule network has a very high potential in overcoming the limitations of CNN. Still now CNN is considered as one of the highly capable deep learning algorithm and is applied in various fields though it has some serious issues. To overcome

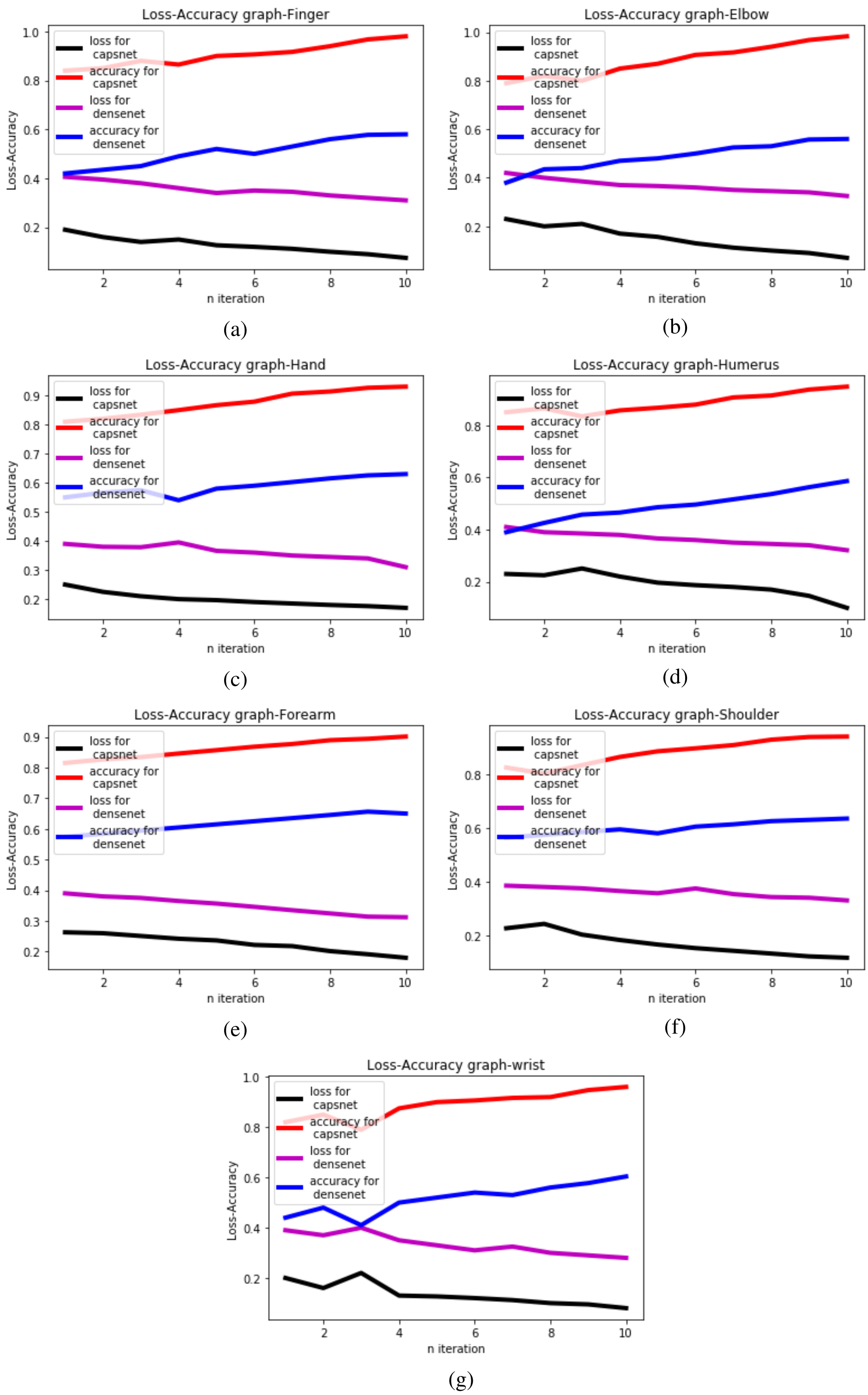


FIGURE 7. **Loss-Accuracy curve plotted against the number of iteration is showing the comparison between the densenet and capsnet and also indicating the improvement in accuracy and minimization of loss in every iteration. To compare both models randomly chosen 50% data were used to train both networks. (a) Finger. (b) Elbow. (c) Hand. (d) Humerus. (e) Forearm. (f) Shoulder. (g) Wrist.**

these issues of CNN, capsule network is a very good alternative architecture. Hence research in capsule network is very promising. As capsule network is a very recent idea, it has been used in very limitate areas like classification problem. Also musculoskeletal problem is already a matter of concern as more than 1.7 billion people are currently suffering from it. Hence accurate detection will help to go further diagnosis and treatment. Hence this is also a great area for future research.

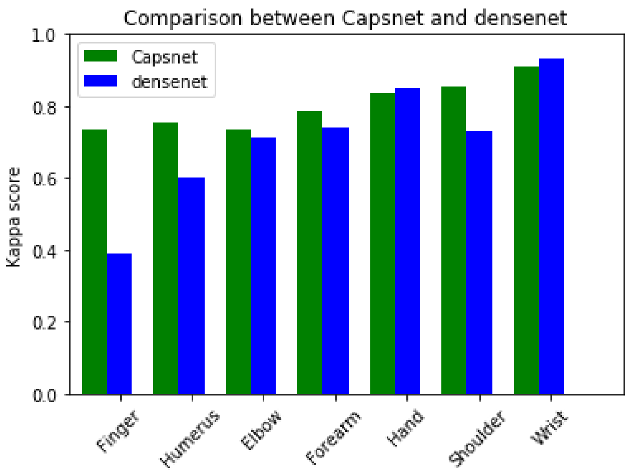


FIGURE 8. Comparison of Kappa score between Capsnet and densenet.

As we have used only 224×224 pixel size and only 50% of data for training the capsule network, we have a plan to use higher pixel size and more training data to train the model and analyze the results. Moreover, we want to use a deeper network and test the outcomes.

VIII. CONCLUSION

Proposed capsule network is capable of determining the abnormality in musculoskeletal radiography more accurately than 169 layer densenet. Also from the BRISQUE score and NIQE score it is cleared that accuracy increases when the resized image size is more close to the actual image size. Not only that, 50% less training data is used to train the network. As this network can perform well with a small number of data, it can also outperform CNN in such cases where a large amount of data cannot be provided.

ACKNOWLEDGMENT

Special appreciation for BUET IICT department for providing computational facilities.

REFERENCES

[1] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Found. Trends Comput. Graph. Vis.*, vol. 7, nos. 2–3, pp. 81–227, Feb. 2012. doi: 10.1561/0600000035.

[2] E. Ricci and R. Perfetti, "Retinal blood vessel segmentation using line operators and support vector classification," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1357–1365, Oct. 2007. doi: 10.1109/TMI.2007.898551.

[3] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1552–1563, Dec. 2002.

[4] H. P. Ng, S. H. Ong, K. W. C. Foong, P. S. Goh, and W. L. Nowinski, "Medical image segmentation using k-means clustering and improved watershed algorithm," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, Denver, CO, USA, Mar. 2006, pp. 61–65. doi: 10.1109/SSIAI.2006.1633722.

[5] E. I. Zacharaki, Sumei Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme," *Magn. Reson. Med.*, vol. 62, no. 6, pp. 1609–1618, Dec. 2009. doi: 10.1002/mrm.22147.

[6] B. N. Li, C. K. Chui, S. Chang, and S. H. Ong, "Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation," *Comput. Biol. Med.*, vol. 41, no. 1, pp. 1–10, 2011.

[7] R. J. Martis, C. Chakraborty, and A. K. Ray, "A two-stage mechanism for registration and classification of ECG using Gaussian mixture model," *Pattern Recognit.*, vol. 42, no. 11, pp. 2979–2988, Nov. 2009.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, Dec. 2012, pp. 1097–1105.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Jun. 2016, pp. 770–778.

[10] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.

[11] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," in *Artificial Neural Networks and Machine Learning*, T. Honkela, W. Duch, M. Girolami, and S. Kaski, Eds. Berlin, Germany: Springer, 2011, pp. 44–51.

[12] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R. L. Ball, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "MURA: Large dataset for abnormality detection in musculoskeletal radiographs," in *Proc. 1st Conf. Med. Imag. Deep Learn. (MIDL)*, 2017, pp. 1–10. [Online]. Available: https://arxiv.org/abs/1712.06957

[13] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," Mar. 2018, *arXiv:1802.10200*. [Online]. Available: https://arxiv.org/abs/1802.10200

[14] R. LaLonde and U. Bagci, "Capsules for object segmentation," Apr. 2018, *arXiv:1804.04241*. [Online]. Available: https://arxiv.org/abs/1804.04241

[15] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[16] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Letters.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[17] J. Sim and C. C. Wright, "The kappa statistic in reliability studies: Use, interpretation, and sample size requirements," *Phys. Therapy*, vol. 85, no. 3, pp. 257–268, Mar. 2005.

[18] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The kappa statistic," *Fam Med*, vol. 37, no. 5, pp. 360–363, 2005.

[19] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Sep. 2016, *arXiv:1608.06993*. [Online]. Available: https://arxiv.org/abs/1608.06993

[20] L. Berlin, "Liability of interpreting too many radiographs," *Amer. J. Roentgenol.*, vol. 175, no. 1, pp. 17–22, 2000.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

[22] W. Gale, L. Oakden-Rayner, G. Carneiro, A. P. Bradley, and L. J. Palmer, "Detecting hip fractures with radiologist-level performance using deep neural networks," Nov. 2017, *arXiv:1711.06504*. [Online]. Available: https://arxiv.org/abs/1711.06504

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[24] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2921–2929.

[25] A. D. Woolf and B. Pflege, "Burden of major musculoskeletal conditions," *Bull. World Health Org.*, vol. 81, no. 9, pp. 646–656, 2003.

[26] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

A. F. M. SAIF received the B.Sc. degree from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, in 2018. He is currently doing his research under Dr. C. Shahnaz. His research interests include machine learning and deep learning, bio-medical signal processing, and image processing.

CELIA SHAHNAZ received the B.Sc. and M.Sc. degrees from the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), in 2000 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2009. She is currently serving as a Professor with Department of Electrical and Electronic Engineering, BUET.

She has published more than 100 international journal and conference papers. Her research interests include speech analysis, speech enhancement, digital watermarking, biomedical signal processing, audio-visual recognition for biometric security, pattern recognition and machine learning, multimedia communication, control systems, robotics and signal processing, and pattern recognition for power signals.

Dr. Celia is a fellow of IEB. She was a recipient of the 2015 WIE Inspiring Member Award from the IEEE WIE and of the 2016 IEEE MGA Leadership Award with citation For leadership in engineering and technology driven innovative IEEE Women in Engineering activities for enhanced membership development and engagement in R10 and across the globe. Under her leadership, WIE BD AG has received 2015 WIE Affinity Group of the Year Award-Honorable Mention from the IEEE Global WIE. She was also a recipient of the 2013 IEEE R10 WIE Professional Volunteer award. While she was an Advisor, WIE BD AG also received the 2016 R10 section WIE AG of the Year Award and 2017 WIE AG of the Year Award from the IEEE WIE. BUET WIE SB has also received WIE SB AG of the Year Award-Honorable Mention from global IEEE WIE and 2018 R10 WIE SB AG of the year award, where she is serving as an Advisor of the group. She was a recipient of the Canadian Commonwealth Scholarship and Fellowship for pursuing Ph.D. study in Canada, in 2004. She is the mentor of 2nd Prize Winning Project in the IAS CMD Robotics Contest 2018 and that of First Prize Winning Project in the Category HEALTH FACILITY in IAS CMD Humanitarian Project Contest 2017. She is the Supervisor of the 5th Rank Winning Team, SPCUP Competition in ICASSP 2015, Australia. Recently, her papers have received best paper awards in biomedical Engineering tracks at TENCON 2017 and at IEEE WIECON-ECE 2016, in Humanitarian Challenge track at R10 HTC 2017, and the Best Interactive Poster Award at icIVPR 2017. Her papers have been selected for top ten best paper awards, student Paper Contests, 2014 MWSCAS, TX, USA, and the 2008 MWSCAS, TN, USA. She was also a recipient of the Best Student Paper Award, 2008 IEEE ICNNSP, China. She was selected as one of the finalists, Student Research Presentation Competition, 2009 SYTACOM Workshop, Montreal, Canada. She has been appointed as the 2017–2018 Chair of the IEEE WIE Workshops Subcommittee and the 2017–2018 IEEE PES Women in Power (WiP) R10 Representative, the 2017 Communications Chair, the IEEE SIGHT Steering Committee, and the 2017 Member, IEEE SSIT WIE, and SIT subcommittees. She has served as the IEEE R10 WIE Coordinator 2016. She is currently serving as a Chair of the IEEE Bangladesh Section (BDS), where she was the 2017 Vice Chair (activity) and the 2012–2013 membership Development Chair. She is the Founding Chair of the WIE Affinity Group, IEEE BDS, for which she is currently acting as an Advisor. She was the Founder and Technical Program Chair of the IEEE WIECON-ECE 2015, the General Chair, IEEE WIECON-ECE 2016, the General Co-Chair of the IEEE WIECON-ECE 2017 and IEEE R10 HTC 2017. She is the Founder or Co-Founder of the IEEE Signal Processing, the IEEE Industrial Applications, the IEEE Robotics and Automation societies, and the IEEE Society on Social Implications on Technology of Bangladesh Chapters.

WEI-PING ZHU (SM'97) received the B.E. and M.E. degrees from the Nanjing University of Posts and Telecommunications, and the Ph.D. degree from Southeast University, Nanjing, China, in 1982, 1985, and 1991, respectively, all in electrical engineering. He was a Postdoctoral Fellow, from 1991 to 1992 and a Research Associate, from 1996 to 1998 with the Department of Electrical and Computer Engineering, Concordia University, Montreal, Canada. From 1993 to 1996, he was an Associate Professor with the Department of Information Engineering, Nanjing University of Posts and Telecommunications. From 1998 to 2001, he was with telecommunication industries in Ottawa, Canada, including Nortel Networks and SR Telecom, Inc. Since 2001, he has been with the Electrical and Computer Engineering Department, Concordia University, as a full-time Faculty Member, where he is currently a Full Professor. He has served as a Technical Program Committee Member and/or a Session Chair for a large number of IEEE sponsored conferences. He was an Associate Editor of the IEEE Transactions on Circuits and Systems Part I: Fundamental Theory and Applications, from 2001 to 2003, and an Associate Editor of *Circuits, Systems and Signal Processing*, from 2006 to 2009. He currently serves as an Associate Editor for the IEEE Transactions on Circuits and Systems Part II: Express Briefs. He is also a Guest Editor for the IEEE Journal on Selected Areas in Communications: Broadband Wireless Communications for High Speed Vehicles, and the *EURASIP Journal on Advances in Signal Processing*: Special Issue on Compressive Sensing for Speech and Audio Signal Processing. He is a Frequent Reviewer of the IEEE Transactions on Circuits and Systems, the IEEE Transactions on Signal Processing, and the IEEE Transactions on Communications, and other journals, and of many conference submissions.

M. O. AHMAD received the B.Eng. degree in electrical engineering from Sir George Williams University, Montreal, QC, Canada, and the Ph.D. degree in electrical engineering from Concordia University, Montreal. From 1978 to 1979, he was a member of the Faculty of the New York University College, Buffalo. In 1979, he joined the Faculty of Concordia University, where he was an Assistant Professor of computer science, where he joined the Department of Electrical and Computer Engineering, he was the Chair of the Department, from 2002 to 2005, and he is currently a Professor. He holds the Tier I Concordia University Research Chair in multimedia signal processing. He has published extensively in the area of signal processing and holds four patents. His current research interests include the areas of multidimensional filter design, speech, image and video processing, nonlinear signal processing, communication DSP, artificial neural networks, and VLSI circuits for signal processing. He was a Founding Researcher in Micronet from its inception, in 1990 as a Canadian Network of Centres of Excellence until its expiration, in 2004. He was an Examiner of the Order of Engineers of Quebec.

Dr. Ahmad was a recipient of numerous honors and awards, including the Wighton Fellowship from the Sandford Fleming Foundation and the Excellence in Doctoral Supervision award from the Faculty of Engineering and Computer Science, Concordia University. He was the Local Arrangements Chairman of the 1984 IEEE International Symposium on Circuits and Systems. Since 1988, he has been a member of the Admission and Advancement Committee of the IEEE. He has also served as the Program Co-Chair for the 1995 IEEE International Conference on Neural Networks and Signal Processing, the 2003 IEEE International Conference on Neural Networks and Signal Processing, and the 2004 IEEE International Midwest Symposium on Circuits and Systems. He is currently the Chairman of the IEEE Circuits and Systems Chapter (Montreal Section). He was an Associate Editor of the IEEE Transactions on Circuits and Systems Part I: Fundamental Theory and Applications, from 1999 to 2001.