Convolutional Neural Network Based Abnormality Detection in X-Ray Images

Osman Onur KUZUCU, N21131025, n21131025@cs.hacettepe.edu.tr

Hacettepe University, Computer Engineering Department, Turkey

Abstract: Every year, a lot of people go to hospital to check their bone structure. Some countries have enough radiologists to understand whether the bone structure is normal or abnormal. However some countries do not have enough radiologists to check if the bone structure is normal or abnormal. In this paper, we built neural network models that can detect abnormality using x-ray images. Some image enhancement techniques are applied to images in preprocessing. Some models are used to feature extraction. Some classification algorithms are used to classify features. Pre-trained models are used to detect abnormality. The results are promising.

Keywords: Convolutional Neural Network, Transfer Learning, Pre-trained Network, Image Enhancement, Feature Extraction, Radiological Images

1 Introductions

A lot of people have bone abnormality diseases every year. Detecting abnormality in bone requires a lot of information and experience in the radiology domain. Because of that, a few radiologists can detect abnormalities clearly. There is a problem that some hospitals have enough radiologists to detect abnormalities clearly. However some hospitals do not have enough radiologists to detect abnormalities clearly. Our proposed model helps radiologists to detect abnormalities from bones' x-ray images.

Some radiologists easily detect abnormalities. Some radiologists detect abnormalities with too much effort. This is a problem. Our model can solve this problem to help radiologists. Our model can guide the radiologists.

Abnormality detection from radiological images can be time-consuming. Some cases of the radiological images can be detected easily, but some cases consume a lot of time to detect. In addition, abnormality detection can require too much effort. This is another problem. Proposed models can help radiologists to reduce effort and consumed-time.

2 Related Work

In "In MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs", researchers built a model to detect abnormalities on bones. In this model 169-layer convolutional neural network is used. The model gets data from the Mura Dataset that has seven extremities. The model achieved good results that can help radiologists. [1]

In "X-Ray bone abnormalities detection using MURA dataset", researchers built a convolutional neural network model. Researchers used a 169 layer convolutional neural network. Researchers did data augmentation using random flipping of the images horizontally, random rotation of the images up to 30 degrees, scaling randomly in the range 95-130 per cent, randomly changing brightness in the range 80-120 per cent. With these techniques, performance is better than only 169 layer convolutional neural network models.[2]

In “Improving convolutional neural networks performance for image classification using test time augmentation: a case study using MURA dataset”, researchers built a model to detect abnormality in bone structure. Nine data augmentation methods is used that are rotation, zooming, horizontal flipping, vertical flipping, horizontal fip with vertical fip, horizontal fip with rotation, vertical fip with rotation, horizontal fip, vertical fip, and rotation, horizontal fip, vertical fip, rotation, and zooming. In this model, test train augmentation improves the performance. [3]

In “Ensemble Based Neural Network for the Classification of MURA Dataset”, researchers built a model to detect abnormality in bone structure. Proposed method has ten hidden layers.Each hidden layer is used Adaboost algorithm.Model used validation error and Cohen’s kappa coefficients. This method performs as well as the Stanford Machine Learning team’s model in a shorter time.In this model, when epoch number increases, performance of the model increases. [4]

In “Anomaly Detection of Arm X-Ray Based on Deep Learning”, a researcher built a model that detects arm abnormality. In this model, NASNetMobile is used. The Mura dataset is used in this model to train and test. The model accuracy is %73. The model is good enough. [5]

In “Image enhancement on digital x-ray images using N-CLAHE” researchers compared some image enhancement techniques. N-CLAHE, CLAHE, HE and USM are the x-ray image enhancement techniques. N-CLAHE performs best to enhance x-ray images. CLAHE is second best in these techniques. Researchers achieved these results using a radiographic chest phantom. [6]

3 Methodology

In this project, the proposed models trained and tested with x-ray images of hand and wrist. These images are taken from Mura dataset that was built by Stanford Machine Learning Group. In these models, different classification algorithms are used. Detailed information is in Section 3.3. In this project, some image enhancement techniques are used to enhance the image.Detailed information in Section 3.2.

3.1 Dataset

In this project, the Mura dataset is used. Mura Dataset is built by STanford Machine Learning Group. Mura dataset is public, Everybody can access the dataset. This dataset contains x-ray images.These x-ray images are taken from 7 human extremities. Elbow, finger, hand,humerus, forearm, shoulder and wrist are the extremities. These radiological images are collected to detect abnormality in the images. In the Mura dataset, 14863 musculoskeletal studies are stated. Every study has one or more images. The Mura dataset has 40561 images. Detailed information about Mura dataset is in Figure 1.[1]

Study	Train		Validation		Total
	Normal	Abnormal	Normal	Abnormal	
Elbow	1094	660	92	66	1912
Finger	1280	655	92	83	2110
Hand	1497	521	101	66	2185
Humerus	321	271	68	67	727
Forearm	590	287	69	64	1010
Shoulder	1364	1457	99	95	3015
Wrist	2134	1326	140	97	3697
Total No. of Studies	8280	5177	661	538	14656

Figure 1: Distribution of radiological studies [1]

The train part of the dataset has 1497 normal, 521 abnormal hand studies. There are 4995 images in the train part. The validation part of the dataset has 101 normal, 66 abnormal hand studies. There are 460 images in the validation part. Model is tested by %10 percent of the train hand studies. There are 548 images in the test part.

The train part of the dataset has 2134 normal, 1326 abnormal wrist studies. There are 8775 images in the train part. The validation part of the dataset has 140 normal, 97 abnormal wrist studies. There are 658 images in the validation part. Model is tested by %10 percent of the train wrist studies. There are 977 images in the test part.

3.2 Features

In this project, some image enhancement techniques are used to enhance images. First, you can see the original image in Figure 2. In Figure 2, image size is 512x352.

Figure 2: Original Image

Sharper filter is applied to the original image. Sharper filter applied image is in Figure 3. After applying sharper filter, the image is resized to 224x224. Sharper filter matrix is in Figure 4.

Figure 3: Sharper Filter applied on Original Image

$$A = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Figure 4: Sharper filter Matrix

Sobel edge filters applied to sharper filter applied images. The output of the sobel filters is added to a sharper filter applied image. After these operations, the enhanced image is in Figure 5. Sobel edge filters matrixes are in Figure 6

Figure 5: Sobel edge filters applied image.

$$X = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad Y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Figure 6: Sobel edge filters for x and y axis

The CLAHE algorithm is applied to the enhanced image. CLAHE applied image is in Figure 7.

Figure 7: CLAHE algorithm applied image

The GLCM contrast algorithm is applied to the image that is in Figure 7. New image is in Figure 8.

Figure 8: GLCM contrast algorithm applied image

All enhancement techniques are used to enhance the details in the image. Abnormality detection algorithm must focus on the edges and the contrast of the image. Using edges and contrast, detection algorithms perform well.

3.3 Classification Methods

In this project, different methods are used. First method is the convolutional neural network that is similar to the Lenet-5 structure. Lenet-5 structure is in Figure 9.The only difference is the input shape of the model. First method's input shape is 224x224.

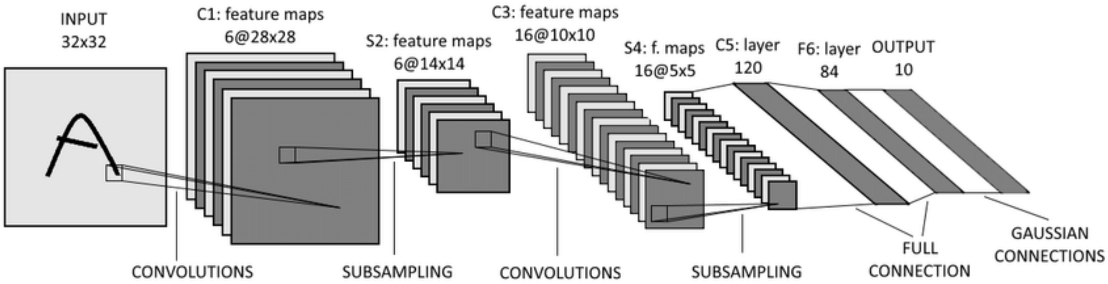

Fİgure 9: Lenet-5 Structure

Second method is VGG16. VGG16 has 16 layers, 5 max-pooling layers and 3 fully connected layers. VGG16 is a pre-trained convolutional neural network. Generally VGG16 is used for transfer learning. VGG16 is trained with the images. VGG16's input shape is 224x224x3. VGG16's weight is ImageNet. In VGG16, adam optimizer and sparse categorical Cross Entropy are used.VGG16 structure is in Figure 10.

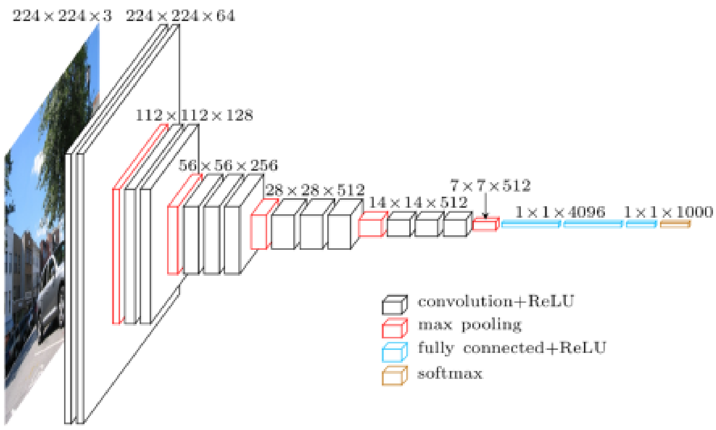

Figure 10: VGG16 Structure

Third method is based on VGG16. In this method, feature extraction is done from the VGG16 “flatten” layer. After extraction, features are classified with logistic regression and basic neural network that has 3 layers.

Fourth method is Densenet-169. Densenet-169 has 169 layers. Densenet-169 is a pre-trained convolutional neural network. Generally Densenet-169 is used for transfer learning. Densenet-169 is trained with the images. Densenet-169’s input shape is 224x224x3. Densenet-169’s weight is ImageNet. In Densenet-169, adam optimizer and sparse categorical Cross Entropy are used. Densenet-169 structure is in Figure 11.

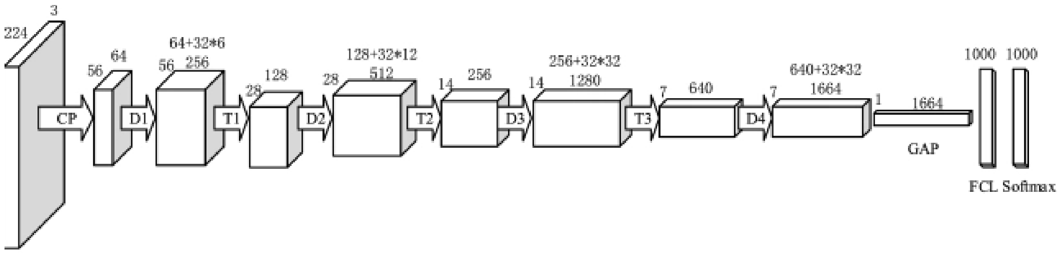

Figure 11: Densenet-169 Structure

Fiveth method is based on Densenet-169. In this method, feature extraction is done from the Densenet-169 “avg_pool” layer. After extraction, features are classified with logistic regression and basic neural network that has 3 layers.

Sixth method is Resnet-50 v2. Resnet-50 v2 has 50 layers. Resnet-50 v2 is a pre-trained convolutional neural network. Generally Resnet-50 v2 is used for transfer learning. Resnet-50 v2 is trained with the images. Resnet-50 v2’s input shape is 224x224x3. Resnet-50 v2’s weight is ImageNet. In Resnet-50 v2, adam optimizer and sparse categorical Cross Entropy are used. Resnet-50 v2 structure is in Figure 12.

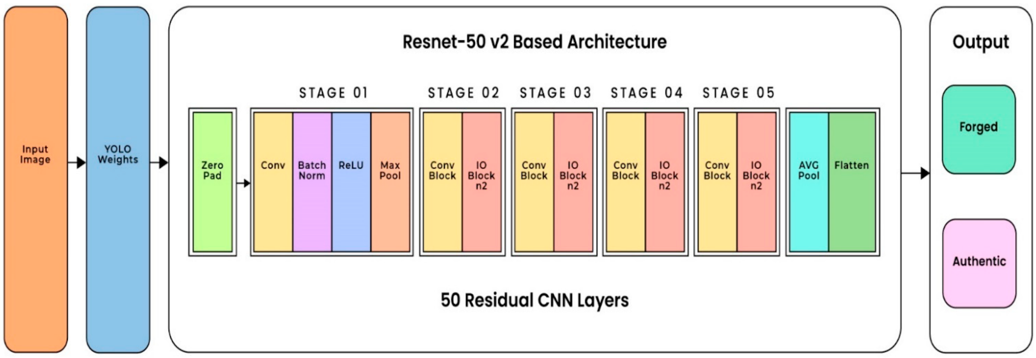

Figure 12: Resnet-50 v2 Structure

Seventh method is based on Resnet-50 v2. In this method, feature extraction is done from the Resnet-50 v2 “avg_pool” layer. After extraction, features are classified with logistic regression and basic neural network that has 3 layers.

4 Results

Hand images and wrist images run separately. The result of the hand images is mentioned in Table 1. In Table 1, ten methods were compared. Sensitivity, specificity and accuracy results are compared.

The result of the wrist images is mentioned in Table 2. In Table 1, ten methods were compared. Sensitivity, specificity and accuracy results are compared.
The best model for hand images is Logistic regression with Resnet-50 v2's feature. The best model for wrist images is Logistic regression with Densenet-169 's feature based on accuracy metric.

In Table 1 and Table 2, some metrics' results could not be calculated. These results are represented with x. In the proposed model metrics are changed to the "sensitivity" and "specificity". But the model can not run these metrics. Feature extraction in the last dense layer is applied to calculate "sensitivity" and "specificity" manually. The manual calculations are used with logistic regression. In some cells, two results are displayed.

	Accuracy	Sensitivity	Specificity
Lenet-5/Logistic Regression	0.759/0.757	x/0.52	x/0.759
VGG16/Logistic Regression	0.75/0.763	x/0.762	x/1
Logistic regression with VGG16's feature	0.77	0.55	0.84
Multilayer Perceptron with VGG16's feature	0.74	x	x
Densenet-169/Logistic Regression	0.45/0.75	x/0.5	x/0.76
Logistic regression with Densenet-169's feature	0.786	0.60	0.81
Multilayer Perceptron with Densenet-169's feature	0.76	x	x
Resnet-50 v2/Logistic Regression	0.63/0.75	x/0.51	x/0.75
Logistic regression with Resnet-50 v2's feature	0.77	0.69	0.77
Multilayer Perceptron with Resnet-50 v2's feature	0.76	x	x

Table 1: Results of the model using hand images

	Accuracy	Sensitivity	Specificity
Lenet-5/Logistic Regression	0.62/0.64	x/0.66	x/0.54
VGG16/Logistic Regression	0.62/0.68	x/0.699	x/0.64
Logistic regression with VGG16's feature	0.77	0.82	0.68
Multilayer Perceptron with VGG16's feature	0.37	x	x
Densenet-169/Logistic Regression	0.80/0.62	x/0.66	x/0.51
Logistic regression with Densenet-169s's feature	0.62	0.66	0.52
Multilayer Perceptron with Densenet-169's feature	0.37	x	x
Resnet-50 v2/Logistic Regression	0.778/0.62	x/0.62	x/0.66
Logistic regression with Resnet-50 v2's feature	0.68	0.75	0.58
Multilayer Perceptron with Resnet-50 v2's feature	0.375	x	x

Table 2: Results of the model using wrist images

5 Conclusion

In this project, our aim is to help radiologists to detect abnormalities in radiological images. Because detecting abnormalities is not easy sometimes. In some cases detecting abnormalities can be time consuming, hard to detect and require too much effort. Our project can help these problems that are mentioned before. In this project, Hand and wrist studies from the Mura dataset are used to make experiments that models are good enough to help radiologists.

In hand studies, experiments show that Densenet-169 feature-extraction with logistic regression performs best. The Resnet-50 v2 feature-extraction with logistic regression performs second best. The VGG16 feature-extraction with logistic regression performs third best.

In wrist studies, experiments show that the original Densenet-169 performs best based on accuracy. However sensitivity and specificity results of the original Densenet-169 cannot be calculated. The original Resnet-50 v2 performs second best based on accuracy. But sensitivity and specificity results of the original Resnet-50 v2 cannot be calculated. The VGG16 feature-extraction with logistic regression performs third best. Majority of these models can guide radiologists, that is this study can be normal or abnormal.

In the future of this project, some image enhancement techniques can be improved. Some models' parameters can be modified. Another dataset can be used. Some studies can be cleared to train the model accurately.

References

1. Pranav Rajpurkar, Jeremy Irvin, Aarti Bagul , Daisy Ding1 , Tony Duan1 , Hershel Mehta1 , Brandon Yang1 , Kaylie Zhu1 , Dillon Laird1 , Robyn L. Ball , Curtis Langlotz , Katie Shpanskaya , Matthew P. Lungren, † , Andrew Y. Ng1 : MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. 1st Conference on Medical Imaging with Deep Learning (MIDL 2018)

2. A. Solovyova, Igor Solovyov: X-Ray bone abnormalities detection using MURA dataset. (2020)

3. Ibrahem Kandel, Mauro Castelli: Improving convolutional neural networks performance for image classification using test time augmentation: a case study using MURA dataset, Kandeland Castelli Health Inf Sci Syst (2021)

4. Mithun Ghosh, Sahil Hassan , Prama Debnath: Ensemble Based Neural Network for the Classification of MURA Dataset. Journal of Nature, Science & Technology, 3(2021), 25-61.

5. Jingchen Wang: Anomaly Detection of Arm X-Ray Based on Deep Learning, IOP Conference Series Earth and Environmental Science 440(4):042056(2020)

6. Kitti Koonsanlt, Saowapak Thongvigitmanee, Napapong Pongnapang, Palrash Thajchayapong.: Image enhancement on digital x-ray images using N-CLAHE. 10th Biomedical Engineering International Conference (BMEiCON)