Nat. Volatiles & Essent. Oils, 2021; 8(5): 2000 - 2007

NVEO Natural Volatiles & Essential Oils

# Development of medical aid artificial intelligence through detection of musculoskeletal X-ray abnormalities using artificial neural networks in medical image data

**Su Jeong Han[1], Choong lyeol Lee[2], Sung-Jun Kim *[3]**

[1]Ceo, Business Data Analysis Strategy Department, Tensor, 14215, South Korea
[2]PhD candidate, The Department of Business Administration, Namseoul University Graduate School, 06724, South Korea
[*3]Professor, The Department of Big Data Industry Security, Namseoul University Graduate School, 06724, South Korea
0001entropy@gmail.com[1], chllee@dnews.co.kr[2], mvstar@hanmail.net[*3]

**Abstract**

**Background / Object :** Recently, the development of medical support services using artificial intelligence is in the spotlight. In this study, DICOM (Digital Imaging and Communications in Medicine) is processed and combined with patient characteristic information to perform more accurate diagnosis in disease diagnosis.

**Methods / Statistical Analysis** : In this study, an artificial neural network algorithm can be used to analyze musculoskeletal X-rays to quickly and simply identify patient abnormalities. Classification techniques were used to identify musculoskeletal abnormalities, and ensemble models were applied to improve the model's performance.

**Findings :** The abnormality in the musculoskeletal structure was classified for abnormality/normality using the DenseNet algorithm. After that, the classified data was constructed as a classification model for new data using ResNet and FusionNet algorithms. By combining the ensemble model to improve performance, we were able to classify the data with a high accuracy of 85.43%.

**Improvements / Applications :** Using neural network algorithms, it is possible to quickly and accurately diagnose musculoskeletal abnormalities from DICOM data, which provides powerful medical support to doctors and patients.

**Keywords :** Radiology, DenseNet, ResNet, Fusion Net, Machine Learning, DICOM Preprocessing

## 1. Introduction

Our bodies are so precise and delicate that it is difficult to detect abnormalities in the body without an accurate examination by a doctor. In addition to being large-capacity, biomedical data in the form of image and video information as well as text are connected in a complex manner. Therefore, when a process for extracting and merging knowledge from this data or automatically finding and resolving conflicts between information is developed, it will be possible to do things that were not possible in the past. In everyday life, we face a variety of problems related to various musculoskeletal structures. This happens in many forms, from minor to very complex symptoms. The important thing is to detect abnormalities early and treat them quickly. [1] However, doctors have to take care of many patients. Therefore, artificial intelligence can be applied to the process of diagnosing or predicting diseases. Especially, it takes a lot of unnecessary time to accurately check and process DICOM data. In this study, an artificial neural network algorithm is used to learn problems related to the musculoskeletal system of normal and abnormal data. When new DICOM data is received, bone abnormalities can be quickly and easily identified in advance. Based on the CNN (Convolution Neural Network) algorithm used in computer vision, neural network algorithms such as Dense Net, ResNet, and Fusion Net were utilized. Through this, research was conducted to establish a data analysis process according to musculoskeletal symptoms and to provide test results. It is hoped that if the study is carried out, doctors will be able to get a lot of help when diagnosing patients. This is expected to be an important technology of artificial intelligence for medical assistance. This study analyzed data based on 40,561 data on fingers, elbows, and shoulders for 12,173 persons from the Picture Archive and Communication System of Stanford Hospital. This study conducted research using the following three key topics. 1) DICOM data preprocessing and analysis using Python programming-based neural network algorithm 2) Constructing a process of classifying data by composing various neural network algorithms into one framework 3) Ensemble model to improve the performance of classification model build

## 2. Materials and Methods

Deep learning is currently the most widely used technology in the field of artificial intelligence. Deep learning technology is mainly used when processing video or image data. In 2006, Professor Geoffrey Hinton of the University of Toronto published a paper on a deep trust neural network (DBN), a very effective algorithm for deep learning. Since then, research on deep learning learning techniques for neural network algorithms has been actively conducted, and unlike conventional methods of analyzing structured data, it is possible to analyze unstructured data such as images and sounds. In the field, research on deep learning-based data mining has been actively conducted. In particular, in the medical field, a doctor reads DICOM (Digital Imaging and Communications in Medicine) data to diagnose a patient's disease, which can be used as a medical assistance system using a deep learning algorithm. The following describes the data analysis techniques and data processing methods used in this study.

### 2.1. *DICOM Data Preprocessing with Python*

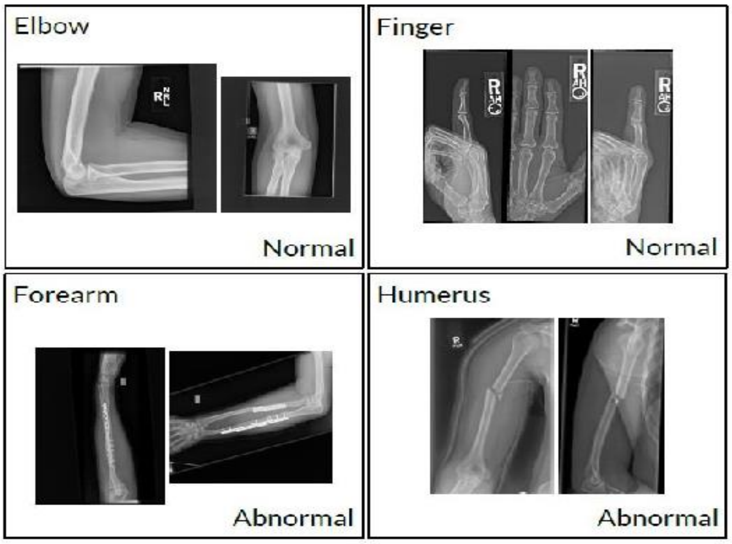


**Figure1. Normal/abnormal DICOM pictures for each musculoskeletal area.[2]**

DICOM (Digital Imaging and Communications in Medicine) refers to imaging data that occurs in a number of medical devices, such as MRI, CT or X-ray. Medical information contains a lot of information that can be fatal, such as patient personal information, not just data. In order to use medical images efficiently, it is necessary to store unique patient information in addition to the image itself. For this reason, DICOM (Digital Imaging and Communications in Medicine) has been developed, which is standard data for medical digital images that occur in various medical devices such as MRI, CT or X-ray. Through DICOM, doctors check and make a diagnosis of physical characteristics such as gender, age, and height of an individual. Since doctors have different symptoms or phenomena according to the patient's medical history and physical characteristics, it is necessary to closely examine medical image data in order to make an accurate diagnosis. By using deep learning technology, you can save a lot of time and identify areas that doctors may miss. For this reason, a technology that can process and analyze DICOM data for a computer to understand is important. And one of the powerful tools you can use to build these services is the Python programming language. The Python programming language is an object-oriented language that can easily implement many complex statistical and machine learning functions. Recently, as neural network algorithms are widely used, libraries are provided so that they can be easily used in the Python programming language. A library is a set of functions that perform a specific function. With several powerful libraries suitable for processing unstructured data in the desired form, DICOM files can be easily processed and analyzed.

### 2.2. *Convolution Neural Network (CNN)*

Unlike structured data, unstructured data is very important to find and learn the edge of one data. Since it is different from structured data analysis that grasps the relationship between columns, it is difficult to analyze it with the traditional machine learning technique used in the past. For this reason, data is often analyzed using neural network algorithms for unstructured data analysis.

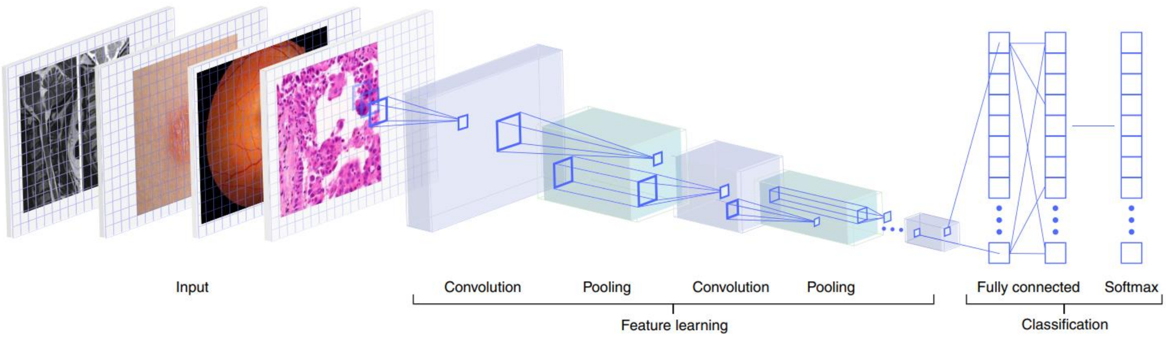


**Figure 2. Structure of CNN that processes image data [3]**

The neural network algorithm composes several layers, passes the data through the activation function in each layer, and can accurately learn the edge derived from each data. Unstructured data such as images go through a convolution process and a pooling process that reduces the dimension to a form that can be analyzed by a computer while maintaining the edge of the data. Through the process, a learning model is constructed. This structured model is called a Convolution Neural Network (CNN). It categorizes images by some criteria or finds specific patterns within them. In the future, this CNN algorithm has been studied in detail to analyze unstructured data such as more diverse images and images. In this study, the CNN-based ResNet algorithm and the DenseNet algorithm were used. In the neural network algorithm, through a function called Loss Function, the weight in the layer is updated in the direction in which the actual value and the predicted value by the model decrease. Weight is the coefficient of the active function for the model to properly learn the edge, and according to the learning step, the learning proceeds in the direction in which the residual value of the loss function is minimized. However, if the step is continuously increased and learned, the optimum weight value cannot be found, and thus the error value increases, and an overfitting phenomenon occurs. Because of these problems, a new algorithm was developed based on CNN's model.

#### *2.1.1. ResNet Model*

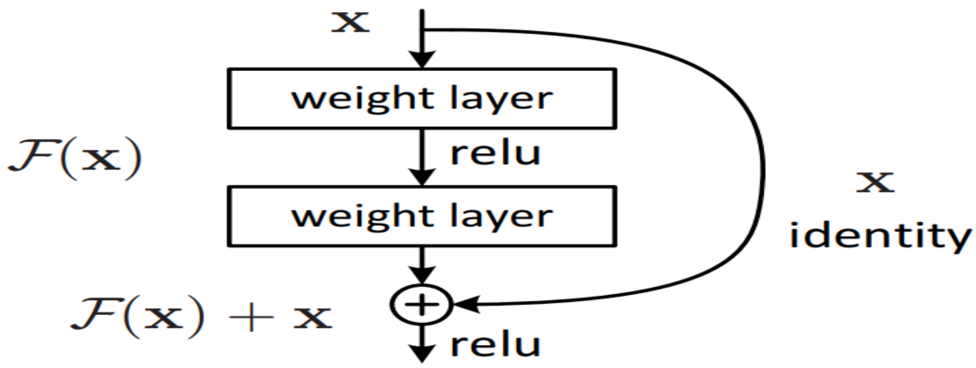


**Figure 3. ResNet algorithm structure[4]**

| layer name | output size | 18-layer | 34-layer | 50-layer | 101-layer | 152-layer |
|---|---|---|---|---|---|---|
| conv1 | 112×112 | 7×7, 64, stride 2 | | | | |
| conv2_x | 56×56 | 3×3 max pool, stride 2 | | | | |
| | | $\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times2$ | $\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$ | $\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$ | $\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$ | $\begin{bmatrix}1\times1, 64\\3\times3, 64\\1\times1, 256\end{bmatrix}\times3$ |
| conv3_x | 28×28 | $\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times2$ | $\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$ | $\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$ | $\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times4$ | $\begin{bmatrix}1\times1, 128\\3\times3, 128\\1\times1, 512\end{bmatrix}\times8$ |
| conv4_x | 14×14 | $\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times2$ | $\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$ | $\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times6$ | $\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times23$ | $\begin{bmatrix}1\times1, 256\\3\times3, 256\\1\times1, 1024\end{bmatrix}\times36$ |
| conv5_x | 7×7 | $\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times2$ | $\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$ | $\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$ | $\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$ | $\begin{bmatrix}1\times1, 512\\3\times3, 512\\1\times1, 2048\end{bmatrix}\times3$ |
| | 1×1 | average pool, 1000-d fc, softmax | | | | |
| FLOPs | | $1.8\times10^9$ | $3.6\times10^9$ | $3.8\times10^9$ | $7.6\times10^9$ | $11.3\times10^9$ |

**Figure 4. ResNet algorithm structure[5]**

The ResNet (Residual Networks) algorithm was designed based on the VGGNet algorithm to solve the problem that the performance of the algorithm rapidly deteriorates when more than a certain number of layers in the CNN Model are crossed. For networks prior to ResNet, the output of the previous layer is affected. On the other hand, in the case of the ResNet algorithm, each output is output to a different network. Input This model adjusts the number of filter layers according to the size of the feature to prevent residual overfitting that occurs in the existing learning process, and performs convolution instead of pooling in the process of reducing the feature, reducing the amount of computation. And at the same time, it is an algorithm that improves the performance of the model in the same layer.

### *2.1.2. DenseNet Model*

DenseNet is an architecture configured to track evaluation indicators to optimize model performance. It is one of the CNN methods that are configured by connecting each layer by the feed-forward method. In DenseNet, since the weight is transmitted through the layer composed of feed-forward of the previously learned features, there is no need to repeatedly update the newly incoming data. Because of this characteristic, the parameter value is small in the model itself, and easy and fast learning is possible.

| Layers | Output Size | DenseNet-121 | DenseNet-169 | DenseNet-201 | DenseNet-264 |
|---|---|---|---|---|---|
| Convolution | 112 × 112 | 7 × 7 conv, stride 2 | | | |
| Pooling | 56 × 56 | 3 × 3 max pool, stride 2 | | | |
| Dense Block (1) | 56 × 56 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 | [1 × 1 conv, 3 × 3 conv] × 6 |
| Transition Layer (1) | 56 × 56 | 1 × 1 conv | | | |
| | 28 × 28 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (2) | 28 × 28 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 | [1 × 1 conv, 3 × 3 conv] × 12 |
| Transition Layer (2) | 28 × 28 | 1 × 1 conv | | | |
| | 14 × 14 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (3) | 14 × 14 | [1 × 1 conv, 3 × 3 conv] × 24 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 48 | [1 × 1 conv, 3 × 3 conv] × 64 |
| Transition Layer (3) | 14 × 14 | 1 × 1 conv | | | |
| | 7 × 7 | 2 × 2 average pool, stride 2 | | | |
| Dense Block (4) | 7 × 7 | [1 × 1 conv, 3 × 3 conv] × 16 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 32 | [1 × 1 conv, 3 × 3 conv] × 48 |
| Classification Layer | 1 × 1 | 7 × 7 global average pool | | | |
| | | 1000D fully-connected, softmax | | | |

**Figure 4. DenseNet Network Architecture [6]**

This model uses Weight Binary Cross Entropy as the Loss function, and is an algorithm constructed using the Sigmoid activity function in the dual classification model.

### *2.1.3. FusionNet Model*

Data processing and analysis can be composed of one neural network framework. The FusionNet algorithm consists of three parts and proceeds from data processing to analysis at once. FusionNet is a network that can run Encoder and Decoder at the same time. It is a structure that can bring the Encoder Layer and connect it to the Decoder Layer. This network was developed to segment the neural structure of Connectomics Data.

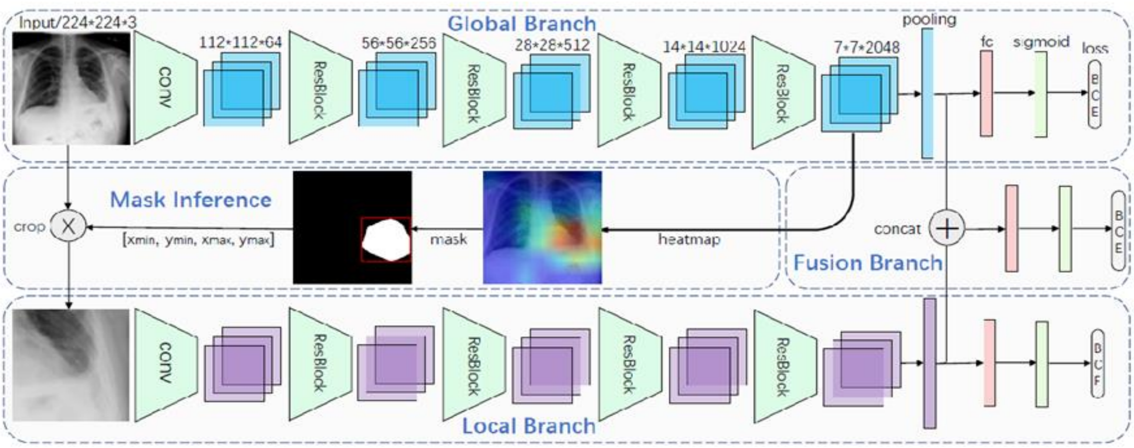


**Figure 5. FusionNet Framework [7]**

FusionNet is composed of Global Branch, Local Branch, and Fusion Branch. New images are processed through the Convolution Layer and Pooling Layer for data in the Global Branch. And it divides Normal Image and Target Image and sends Target Image to Local Branch. Local Branch handles the same as Global Branch. And it is a structure that finally performs Binary Classification by passing the image obtained through Global Branch and Local Branch through the Fully Connected Layer.

### *2.1.4. Ensemble Model*

In supervised learning, it is very important to increase the performance through generalization of the model. Previously, prediction or classification learning was carried out using a single algorithm, but with the development of the algorithm, several models were combined and used. Ensemble Model combines two or more algorithms in multiple classifications and regressions.

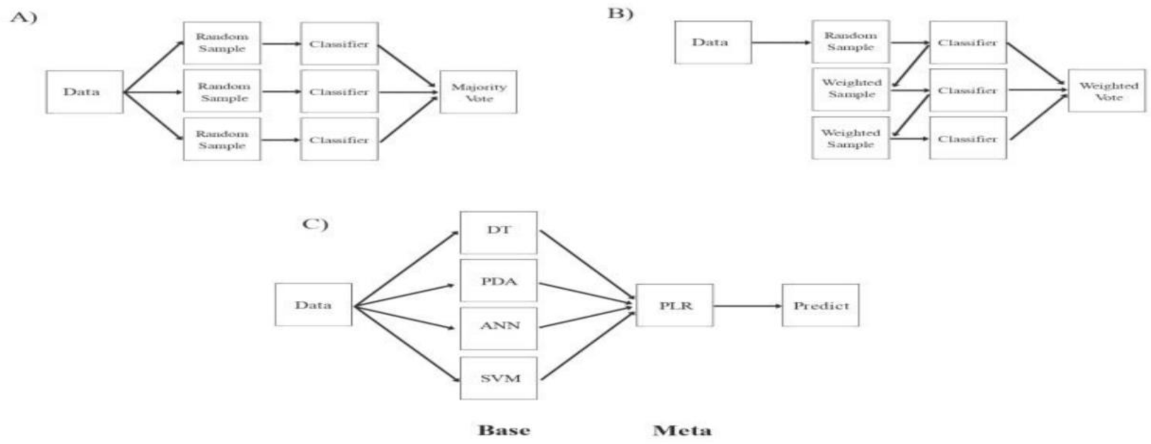


**Figure 6. Representative algorithm of the ensemble model [8]**

As above, the Ensemble Model is largely classified into three categories. Voting method to select the best model among several models, Bagging method to select the best model by performing random sampling from data, and model by reflecting the residual of the previous model in the modeling step There is a boosting technique that composes and selects the best model. In machine learning, the Ensemble algorithm based on the Tree Model has been developed and used, but recently, the neural network algorithm also uses the Ensemble Model using Bagging or Boosting techniques.

## 3. Result and Discussion

The doctor proposed in this study can detect symptoms regardless of the patient's basic physical characteristics by using a neural network algorithm. In this study, the data were analyzed by processing X-ray images of 7 areas including elbows, fingers, forearms, hands, humerus, shoulders, and wrists.[9]

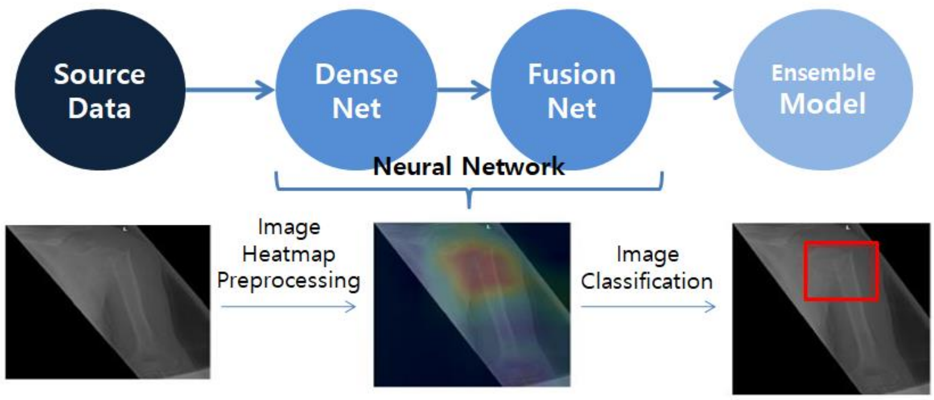


**Figure 6. Design DICOM Data Processing Process**

Classify the target image by passing the X-ray through Dense Net and Fusion Ne as above, converting it into Heatmap format.

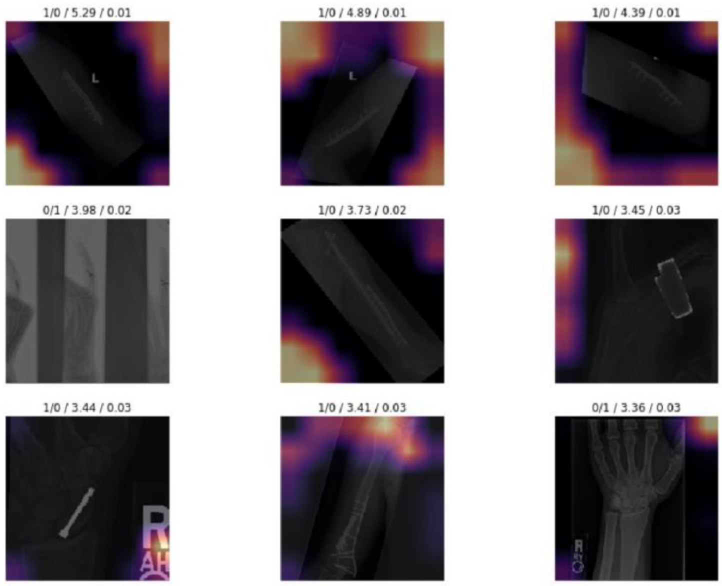


**Figure 7. Photos in the order of the greatest loss among misclassified photos**

After that, through the Ensemble Model, the target data and normal data that have anomalies in the X-ray are learned to construct a classification model. The performance of the model was derived as shown in the table below. Accuracy was measured for each part, and the problem of sampling unbalance in Binary Classification was evaluated through the AUC value.[10]

**Table 1: Accuracy and AUC of the model used in the ensemble and the final ensemble model**

| Part | DenseNet | ResNet50(64) | ResNet50(32) | FusionNet | Ensemble |
|---|---|---|---|---|---|
| **Humerus** | 0.8235 | 0.8928 | 0.8792 | 0.9118 | 0.9191 |
| **Elbow** | 0.8498 | 0.8523 | 0.8302 | 0.8688 | 0.8688 |
| **Finger** | 0.7834 | 0.8592 | 0.8203 | 0.8171 | 0.8514 |
| **Hand** | 0.7298 | 0.7894 | 0.7928 | 0.8383 | 0.8024 |
| **Forearm** | 0.7621 | 0.8450 | 0.8409 | 0.8467 | 0.8467 |
| **Shoulder** | 0.7848 | 0.8201 | 0.7723 | 0.7897 | 0.8000 |
| **Wrist** | 0.8473 | 0.8752 | 0.8329 | 0.8908 | 0.8950 |
| **Average** | 0.7972 | 0.8478 | 0.8240 | 0.8510 | 0.8543 |
| **AUC** | 0.8506 | 0.9025 | 0.8759 | 0.9076 | 0.9153 |

In terms of error, to solve the problem of localization, which is a typical problem of neural network algorithms, Hyperparameter Tuning was performed, and in the evaluation step, we found the contact point between Recall and Precision and adjusted the threshold value to increase the model's performance.

## 4. Conclusion

In this study, anomalies in musculoskeletal X-ray images were detected using the CNN-based DenseNet algorithm. DICOM data was processed with neural networks and analyzed using an ensemble model. Unlike previous studies, this study utilized the ResNet-based Fusion Net algorithm. Finally, AUC 91.53% and accuracy 85.43%, leading to high sorting performance. This improved image processing performance, allowing us to build much higher performance classification models. In order to develop a generalized optimal model, there are many variables besides the algorithm. Recently, artificial intelligence technology has been introduced into the medical field, and research is actively progressing, and it is gradually appearing in the medical field. Therefore, we plan to proceed with pre-diagnosis and system design to provide clear diagnosis results to patients by helping doctors make decisions through future research.

## 5. Reference

[1] Thomas Davenport, Ravi Kalakota : The potential for artificial intelligence in healthcare : Future Healthc J. 2019

[2] Pranav Rajpurkar, Jeremy A. Irvin, Aarti Bagul, D. Ding, T. Duan, Hershel Mehta, Brandon J Yang, Kaylie Zhu, Dillon Laird, Robyn L. Ball, C. Langlotz, K. Shpanskaya, M. Lungren, A. Ng: MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. :2017

[3] Andre Esteva, Alexandre Robicquet, Bharath Ramsundar1, Volodymyr Kuleshov1, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun1 and Jeff Dean : A guide to deep learning in healthcare, Nature Medicine volume 25, 2019

[4] Gao Huang, Zhuang Liu, Laurens van der Maaten. Densely Connected Convolutional Net.: 2018

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun : Deep Residual Learning for Image recognition

[6] Gao Huang, Zhuang Liu, Laurens van der Maaten. Densely Connected Convolutional Net.: 2018

[7] Qingji Guan, Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng and Yi Yang. Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification

[8] Kyle Peterson : Resting Heart Rate Variability Can Predict Track and Field Sprint Performance : University of Iowa, Iowa City, IA, USA : Sports Science : 2018

[9] Ramprasaath R. Selvaraju ,Michael Cogswell , Abhishek Das ,Ramakrishna Vedantam ,Devi Parikh , Dhruv Batra. : Grad-CAM:Visual Explanations from Deep Networks via Gradient-based Localization

[10] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba. Learning Deep Features for Discriminative Localization