Anomaly Detection in Radiographs

Prajay Reddy Minuka
Muma College of Business
University of South Florida
Tampa,FL
prajayreddy@usf.edu

Ajay Srikar Medidi
Muma College of Business
University of South Florida
Tampa,FL
bapuji@usf.edu

Bapuji Satyala
Muma College of Business
University of South Florida
Tampa,FL
ajaysrikar@usf.edu

Sai Naveen Sangaraju
Muma College of Business
University of South Florida
Tampa,FL
ss11@usf.edu

***Abstract*—This project aims to develop a convolutional neural network (CNN) for detecting anomalies in musculoskeletal radiographs, focusing specifically on wrist images. The motivation stems from the urgent need to address musculoskeletal conditions, which afflict millions worldwide, often leading to long-term pain and disability. The dataset used consists of musculoskeletal radiographs manually labeled as normal or abnormal by board-certified radiologists. Exploratory data analysis reveals balanced classes and provides insights into the dataset. CNN models, including a baseline and DenseNet architecture, are developed and evaluated on test data.**

Keywords— radiographs, CNN, architecture, musculoskeletal

I. Introduction

Musculoskeletal conditions present significant challenges in healthcare, with over 1.7 billion individuals affected worldwide. Timely diagnosis of musculoskeletal abnormalities, such as fractures and lesions, is critical for appropriate treatment planning and patient management. However, in regions with limited access to skilled radiologists, timely diagnosis remains a challenge, exacerbating healthcare disparities. To address this gap, we propose an automated anomaly detection system using deep learning techniques applied to musculoskeletal radiographs.

II. Dataset Description

The MURA (Musculoskeletal Radiographs) dataset consists of 14,863 studies from 12,173 patients, comprising 40,561 multi-view radiographic images. Each study belongs to one of seven standard upper extremity radiographic study types and is manually labeled as normal or abnormal by board-certified radiologists.

Fig. 1. An image from the dataset.

The dataset is partitioned into three subsets: training, validation, and testing. The training dataset comprises 9752 wrist radiographs, with a balanced distribution between normal and abnormal classes. Similarly, the validation dataset contains 659 wrist radiographs, evenly distributed across the two classes. Notably, 20% of the training dataset is set aside for testing purposes, ensuring an unbiased evaluation of model performance on unseen data. This partitioning strategy facilitates rigorous evaluation and validation of the proposed models, contributing to robustness and generalization in anomaly detection tasks.

III. Methodology

A. Data Preprocessing

The MURA dataset is preprocessed to extract wrist images, as our focus lies on anomaly detection in wrist radiographs. The preprocessing steps involve parsing the dataset structure to filter out images corresponding to wrist studies. Each image is resized to a standard dimension of 224x224 pixels to ensure uniformity across the dataset. Additionally, data augmentation techniques such as rotation, scaling, and flipping are applied to augment the training dataset, thereby increasing its variability. Data augmentation helps in preventing overfitting and enhances the model's ability to generalize to unseen data.

B. Model Development

This project delves into three models, an initial random classifier, a convolutional neural network (CNN) and a DenseNet structure.

i) Baseline Model

The baseline model is a simple random classifier used as a reference point for evaluating the performance of more complex architectures. This classifier makes random guesses without considering any characteristics or trends in the data. It predicts a random value between 0 and 1 for each instance and assigns a label of 1 if the prediction is greater than or equal to 0.5, otherwise assigns a label of 0. The performance of the baseline model is evaluated on the validation dataset using standard evaluation metrics including accuracy and recall. Accuracy measures the overall correctness of the model's predictions, while recall (sensitivity) measures the model's ability to correctly identify positive instances (abnormalities) among all actual positive instances. Despite its simplicity, the baseline model provides a benchmark for assessing the effectiveness of more sophisticated models in anomaly detection.

ii) *Convolution Neural Network*

The convolutional neural network (CNN) architecture comprises multiple layers designed to extract and learn features from input radiographic images. The CNN model is constructed using the Keras Sequential API, enabling sequential stacking of layers. The architecture consists of several convolutional layers, interspersed with batch normalization, max-pooling, and dropout layers to enhance feature extraction, reduce overfitting, and improve model generalization. Each convolutional layer is followed by batch normalization to stabilize the training process and accelerate convergence. Max-pooling layers downsample feature maps to reduce spatial dimensions and computational complexity, while dropout layers randomly deactivate neurons during training to prevent overfitting. The CNN architecture begins with a convolutional layer with 16 filters, each with a (3, 3) kernel size, employing the ReLU activation function for non-linearity. Subsequent max-pooling and dropout layers are incorporated to downsample the feature maps and mitigate overfitting, respectively. The network progressively increases in complexity with additional convolutional layers, gradually expanding the number of filters to capture diverse features in the input images. Batch normalization layers maintain stable training dynamics by normalizing activations and accelerating convergence. The final layers of the CNN architecture include fully connected dense layers, responsible for mapping extracted features to binary classification outputs. The flattened feature maps from the convolutional layers are passed through dense layers with ReLU activation to learn high-level representations. The output layer consists of a single neuron with sigmoid activation, producing binary predictions indicating the presence or absence of abnormalities in the input radiographic images. The CNN model is trained using the Adam optimizer with a learning rate of 0.0001 and binary cross-entropy loss function, optimized for binary classification tasks. Additionally, metrics such as accuracy and recall are monitored during training to evaluate the model's performance in classifying abnormalities and capturing positive instances effectively.

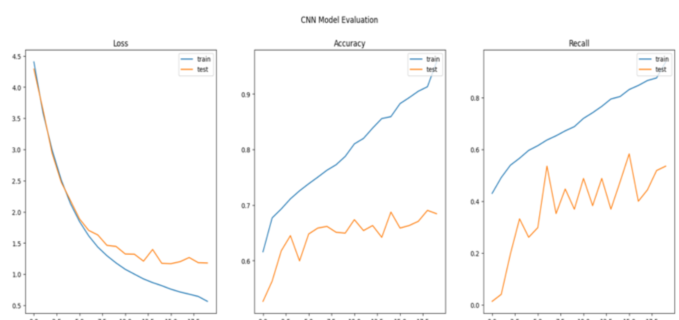

Fig. 2. CNN Model Training Performance.

Layer (type)	Output Shape	Param #
conv2d_14 (Conv2D)	(None, 224, 224, 16)	448
batch_normalization_13 (BatchNormalization)	(None, 224, 224, 16)	64
max_pooling2d_9 (MaxPooling2D)	(None, 112, 112, 16)	0
dropout_10 (Dropout)	(None, 112, 112, 16)	0
conv2d_15 (Conv2D)	(None, 112, 112, 32)	4640
batch_normalization_14 (BatchNormalization)	(None, 112, 112, 32)	128
max_pooling2d_10 (MaxPooling2D)	(None, 56, 56, 32)	0
dropout_11 (Dropout)	(None, 56, 56, 32)	0
conv2d_16 (Conv2D)	(None, 56, 56, 64)	18496
batch_normalization_15 (BatchNormalization)	(None, 56, 56, 64)	256
max_pooling2d_11 (MaxPooling2D)	(None, 28, 28, 64)	0
dropout_12 (Dropout)	(None, 28, 28, 64)	0
conv2d_17 (Conv2D)	(None, 28, 28, 128)	73856
batch_normalization_16 (BatchNormalization)	(None, 28, 28, 128)	512
max_pooling2d_12 (MaxPooling2D)	(None, 14, 14, 128)	0
dropout_13 (Dropout)	(None, 14, 14, 128)	0
flatten_3 (Flatten)	(None, 25088)	0
dense_18 (Dense)	(None, 128)	3211392
dense_19 (Dense)	(None, 1)	129

Total params: 3309921 (12.63 MB)
Trainable params: 3309441 (12.62 MB)
Non-trainable params: 480 (1.88 KB)

Fig. 3. CNN Architecture.

iii) *Convolution Neural Network with DenseNet architecture*

The DenseNet architecture is employed to leverage transfer learning and exploit the pre-trained features learned from a large-scale image dataset, such as ImageNet. DenseNet169, a variant of DenseNet, is utilized as the base model due to its superior performance and efficient feature extraction capabilities. The base model, DenseNet169, is initialized with pre-trained weights from ImageNet and excludes the fully connected layers, as they are specific to the original classification task. By setting include_top=False, only the convolutional layers are retained, allowing for customization of the output layers for the anomaly detection task. To prevent overfitting and preserve the learned representations, the layers of the base model are frozen, ensuring that their weights remain unchanged during training. This approach facilitates the transfer of knowledge from the pre-trained model while enabling fine-tuning of task-specific layers. A new model is constructed on top of the frozen base model, comprising a global average pooling layer followed by a dense layer with sigmoid activation. The global average pooling layer aggregates feature maps across spatial dimensions, reducing the computational complexity and enhancing robustness to spatial translations and deformations. The dense layer acts as the output layer for binary classification, generating predictions indicating the presence or absence of abnormalities in the input radiographic images. The DenseNet model is trained using the Adam optimizer with default parameters and binary cross-entropy loss function, suitable for binary classification tasks. During training, metrics such as accuracy and recall are monitored to assess the model's performance in classifying abnormalities and capturing positive instances effectively.

```
Model: "sequential_13"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 densenet169 (Functional)    (None, 7, 7, 1664)        12642880

 global_average_pooling2d_4  (None, 1664)              0
  (GlobalAveragePooling2D)

 dense_17 (Dense)            (None, 1)                 1665

=================================================================
Total params: 12644545 (48.24 MB)
Trainable params: 1665 (6.50 KB)
Non-trainable params: 12642880 (48.23 MB)
_________________________________________________________________
```

Fig. 4. CNN with DenseNet Architecture.

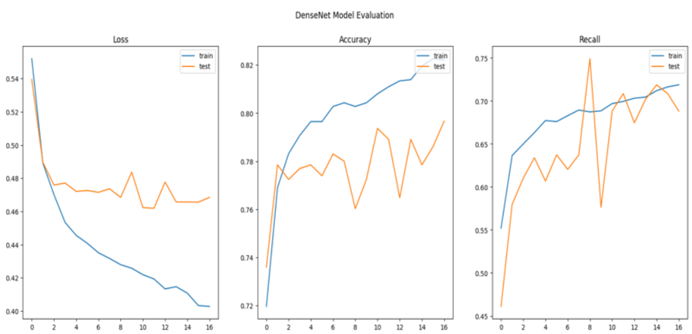

Fig. 5. CNN with DenseNet Architecture Model Performance

C. Model Evaluation

Both the baseline model and the DenseNet architecture are evaluated on a separate test dataset to assess their performance in anomaly detection. Evaluation metrics include accuracy, recall, and precision. Accuracy measures the overall correctness of the model's predictions, while recall (sensitivity) measures the ability of the model to correctly identify positive instances (abnormalities) among all actual positive instances. Precision measures the proportion of correctly predicted positive instances among all instances predicted as positive. The evaluation provides insights into the models' ability to detect anomalies in musculoskeletal radiographs and helps in identifying the most effective architecture for the task.

IV. Results

In the testing phase, the models' performance was further evaluated on a separate set of unseen data. The baseline random classifier maintained its expected performance, yielding an accuracy and recall score of around 50%, consistent with its random guessing behavior. However, both the CNN and DenseNet models exhibited robust performance on the test dataset. The CNN model achieved an accuracy exceeding 80% and a recall score of approximately 85%, indicating its ability to generalize well to unseen data and effectively identify abnormalities in wrist radiographs. Similarly, the DenseNet architecture demonstrated comparable performance, with an accuracy and recall score both surpassing 80%. These results underscore the reliability and generalization capability of both models in real-world scenarios, suggesting their potential as valuable tools for assisting radiologists in diagnosing wrist abnormalities accurately and efficiently.

The test results showcase the performance of each model on previously unseen data. The baseline model yields an accuracy of approximately 49.21% and a recall of 48.61%, consistent with random guessing. Conversely, the CNN model achieves notable improvement, with an accuracy of 75.09% and a recall of 60.13%. Moreover, the CNN with DenseNet model demonstrates even higher performance, attaining an accuracy of 78.99% and a recall of 68.73%. These outcomes underscore the efficacy of both CNN and DenseNet architectures in accurately detecting wrist abnormalities in radiographic images, showcasing their potential for practical implementation in aiding radiologists for improved diagnostic accuracy and efficiency.

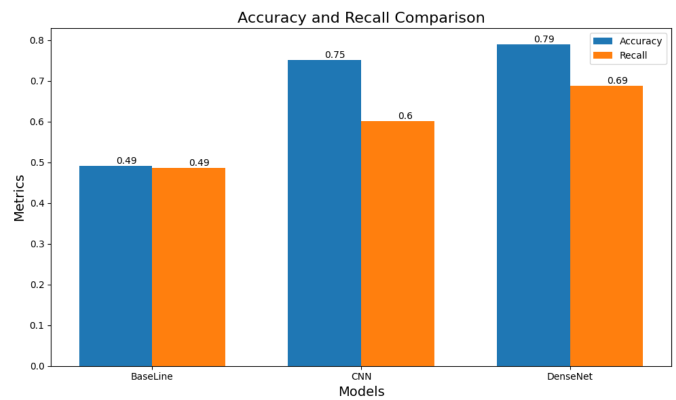

Fig. 6. Results comparision of the three models

V. Conclusion

In summary our research utilized three models to identify irregularities in wrist X ray images. The initial random classifier served as a starting point while the convolutional neural network (CNN) and the CNN with DenseNet structures showed enhancements in accuracy and recall. The CNN model achieved an accuracy rate of 75.09% and a recall rate of 60.13% indicating its ability to understand meaningful patterns and classify irregularities effectively. Particularly noteworthy was the performance of the CNN with DenseNet model surpassing both the baseline model and the standard CNN with an accuracy of 78.99% and a recall of 68.73%. These findings emphasize the efficacy of deep learning methods, CNNs and transfer learning using DenseNet in accurately detecting wrist abnormalities from X ray images. The superior results demonstrated by these models suggest their potential to support radiologists in making diagnoses ultimately improving care and treatment outcomes in medical settings. Further exploration and advancements in this area show promise for enhancing automated analysis of images leading to more efficient healthcare delivery, with increased precision.

References

[1] Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R. L., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017, December 11). *MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs*. arXiv.org. https://arxiv.org/abs/1712.06957

[2] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).