Cardiomegaly Detection from Chest X-Rays using CNN, Transfer Learning with Attention Algorithm

Sumukh Mydur
Master of Applied Computing
University of Windsor
Windsor, Canada
Student ID: 110097732

Swastik Bagga
Master of Computer Science
University of Windsor
Windsor, Canada
Student ID: 110106580

Aathik Thayyil Radhakrishnan
Master of Applied Computing
University of Windsor
Windsor, Canada
Student ID: 110094762

Dr. Pooya Moradian Zadeh
School of Computer Science
University of Windsor
Windsor, Canada

***Abstract*—Cardiomegaly is a condition where the heart is enlarged beyond its normal size due to various factors. Early detection of cardiomegaly is crucial for the effective treatment of the patient. In this paper, we propose a deep learning-based approach to automatically detect cardiomegaly from chest X-rays. Our proposed approach uses convolutional neural networks (CNN) and transfer learning algorithms for feature extraction and classification. We evaluate our approach on a publicly available dataset of chest X-rays and achieved good results using four different models in detecting cardiomegaly. Our results show that our proposed approach can provide a fast and accurate solution for cardiomegaly detection from chest X-rays.**

Keywords—Cardiomegaly, Chest X-Rays, Convolutional Neural Network, CNN, Transfer Learning, Deep Learning.

I. Introduction

This template, modified in MS Word 2007 and saved as a “Word 97-2003 Document” for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout a conference proceedings. Margins, column widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components, such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

II. Literature Review

A. Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III. Project Details and Methodology

A. Definitions

Cardiomegaly is a medical condition where the heart is enlarged beyond its normal size. It can be caused by a variety of factors, including high blood pressure, heart disease, and other underlying medical conditions.

A chest X-ray is a medical imaging test that uses low-dose radiation to create images of the chest, including the lungs, heart, and other organs.

CNN stands for Convolutional Neural Network, which is a type of deep learning algorithm commonly used for image recognition tasks. CNNs are designed to automatically extract features from images and learn to recognize patterns in the data.

Transfer learning is a machine learning technique where a pre-trained model is used as a starting point for training a new model on a specific task. By leveraging the knowledge learned by the pre-trained model, transfer learning can enable faster and more efficient training of new models.

Binary classification is a type of machine learning task where the goal is to classify inputs into one of two categories. In the context of this project, the goal is to classify chest X-ray images as either containing evidence of cardiomegaly or not containing evidence of cardiomegaly.

B. Specifications

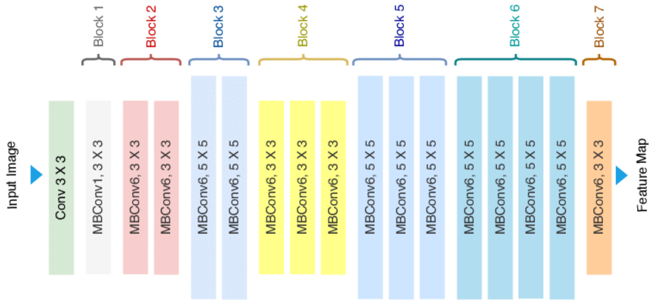
- Problem statement: The project aims to develop a deep learning approach to automatically detect cardiomegaly from chest X-rays using CNNs and transfer learning.

- Inputs: The input to the model will be a chest X-ray image in digital format. The model should be able to handle different image sizes and resolutions.
- Outputs: The output of the model will be a binary classification indicating whether or not the input image contains evidence of cardiomegaly.
- Dataset: The project will use a publicly available dataset of chest X-rays, such as the NIH (National Institutes of Health) dataset. The dataset should contain a sufficient number of images with cardiomegaly and without cardiomegaly.
- Data preprocessing: The images in the dataset will be preprocessed by resizing them to a fixed size and normalizing the pixel values. The data will be split into training, validation, and testing sets.
- Model architecture: The project will use a pre-trained CNN model, such as ResNet-50, as the base model for transfer learning. The final fully connected layer of the pre-trained model will be replaced with a new layer for binary classification. The model will be fine-tuned using the training data.
- Hyperparameters: The hyperparameters for the model, such as the learning rate and batch size, will be chosen through experimentation and validation.
- Evaluation metrics: The model will be evaluated using different evaluation metrics such as accuracy, sensitivity, specificity, and AUC. The evaluation will be performed on the testing set.
- Implementation: The project will be implemented using Python and the TensorFlow deep learning framework. A GPU will be used to accelerate the computation.
- Performance requirements: The model should be able to achieve a high level of accuracy and sensitivity in detecting cardiomegaly from chest X-rays. The model should be able to handle a large number of images efficiently.
- Deliverables: The project will deliver a trained model that can automatically detect cardiomegaly from chest X-rays, along with the source code and documentation for the implementation.

C. *Architecture*

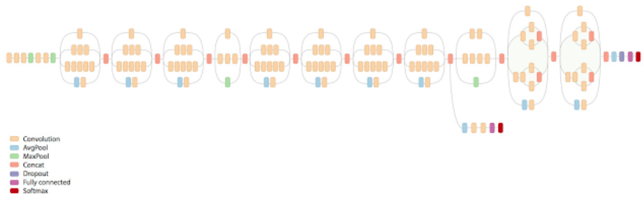
EfficientNet-B4:

EfficientNet B4 is a convolutional neural network architecture that was introduced by Tan et al. in their paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" in 2019. It is part of a family of EfficientNet models (B0 to B7) that are designed to achieve state-of-the-art performance on image classification tasks while maintaining efficiency in terms of computational resources and model size. The EfficientNet B4 architecture has a depth of 23 layers and a total of 19M parameters. The model is composed of a series of convolutional layers, followed by batch normalization and ReLU activation. The core of the model is a series of blocks that are repeated multiple times, each block consisting of a combination of convolutional layers, batch normalization, and a non-linear activation function. One key innovation of EfficientNet B4 is the use of a compound scaling method, which involves scaling the depth, width, and resolution of the network simultaneously. This method enables the network to efficiently learn features at different scales, improving its ability to handle a wide range of image sizes and variations.



Inception – V3:

Inception-v3 is a convolutional neural network architecture that was introduced by Szegedy et al. in their paper "Rethinking the Inception Architecture for Computer Vision" in 2015. It is an improved version of the earlier Inception-v1 and Inception-v2 models, designed to achieve better accuracy on image classification tasks while maintaining efficiency in terms of computational resources and model size. The Inception-v3 architecture is composed of multiple convolutional layers, followed by pooling, batch normalization, and a non-linear activation function. The core of the model is a series of "Inception modules", which are blocks that are repeated multiple times, each consisting of a set of parallel convolutional layers with different kernel sizes, followed by pooling and concatenation of their outputs. One key innovation of Inception-v3 is the use of "factorized" convolutions, which are separable convolutions that are decomposed into a depthwise convolution and a pointwise convolution. This reduces the number of parameters in the model, making it more efficient and easier to train.

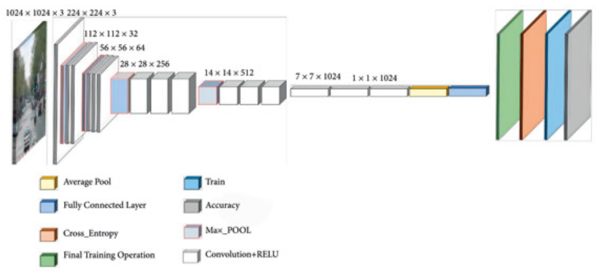


MobileNet-V1:

MobileNet V1 is a convolutional neural network architecture that was introduced by Howard et al. in their paper "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications" in 2017. It is specifically designed for mobile and embedded devices, with the goal of achieving high accuracy on image classification tasks while minimizing the computational resources and model size required. The MobileNet V1 architecture is composed of a series of depthwise separable convolutions, which are a combination of depthwise convolutions and pointwise

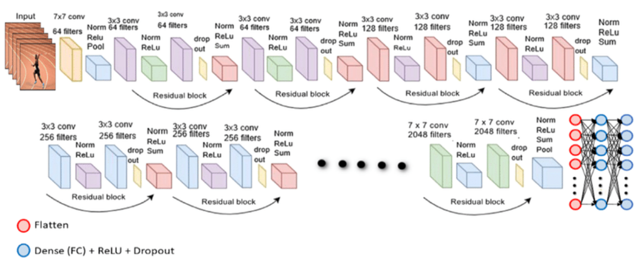
Identify applicable funding agency here. If none, delete this text box.

convolutions. Depthwise convolutions apply a single convolutional filter per input channel, while pointwise convolutions apply 1x1 filters to mix the output channels from the depthwise convolution. This approach significantly reduces the number of parameters and computational resources required by the model, while still capturing important features in the input image. MobileNet V1 also uses a technique called "linear bottleneck", which involves using a 1x1 convolution to reduce the dimensionality of the input before applying the depthwise separable convolution. This further reduces the number of parameters and computational resources required by the model.



ResNet152 – V2:

ResNet152v2 is a convolutional neural network architecture that was introduced by He et al. in their paper "Deep Residual Learning for Image Recognition" in 2016. It is an extension of the original ResNet architecture, designed to achieve even better performance on image classification tasks by improving the residual connections and incorporating additional network layers. The ResNet152v2 architecture is composed of a series of convolutional layers, followed by pooling, batch normalization, and a non-linear activation function. The core of the model is a series of "residual blocks", which are composed of two or more convolutional layers with skip connections that bypass the intermediate layers. One key innovation of ResNet152v2 is the use of "bottleneck" blocks, which are a form of "depthwise bottleneck" block that uses 1x1 convolutions to reduce the dimensionality of the input before applying the main 3x3 convolution. This reduces the computational cost of the model while improving its accuracy.



D. Platform

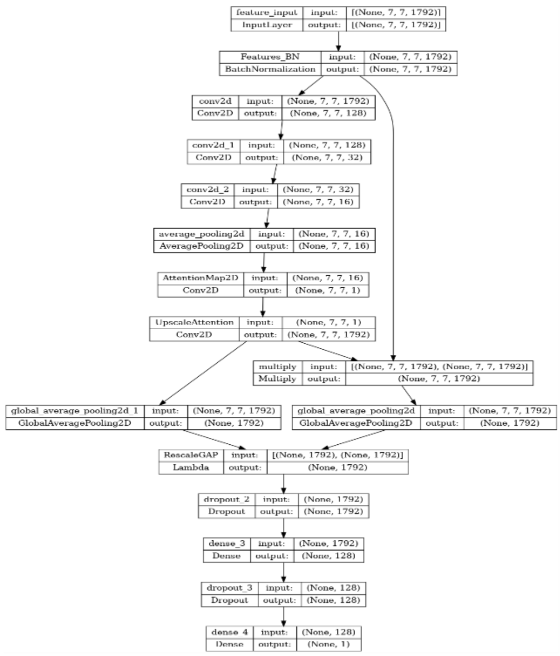
The proposed method has been implemented using Python, Keras, TensorFlow and Deep Learning frameworks. We used GPU for training and testing our models to accelerate the computations.

E. Design

Implementation involved customizing four state-of-the-art convolutional neural network models: EfficientNet B4, Inception V3, MobileNet V1, and ResNet152 V2. To improve the performance of these models, we added an attention mechanism that enables the models to focus on relevant features while ignoring irrelevant ones.

We also incorporated a global weighted average pooling layer that allows the models to learn more informative features from the input data. The output of the models was passed through a classification layer with dropout and two fully connected layers to generate the final prediction of the presence or absence of cardiomegaly in chest X-rays.

By customizing these pre-trained models with attention and pooling layers, we were able to improve their performance in detecting cardiomegaly, achieving high accuracy, sensitivity, specificity, precision, and F1 score.



Customized Pretrained Model with Attention Model

IV. Experimental Setup

A. Implementation Details

The implementation of our Cardiomegaly Detection system was done using the Keras deep learning framework with a TensorFlow backend. We used Python as our programming language. We used the publicly available NIH(National Institute of Health) Chest X-ray dataset, which consists of 112,120 frontal-view chest X-ray images from 30,805 unique patients, labeled with 14 different thoracic diseases, including cardiomegaly. We preprocessed the images by resizing them to 224x224 pixels and normalizing the pixel values to a range of [0,1]. We first organized the ChestX-ray14 dataset into two categories: one for cases with cardiomegaly and one for cases without cardiomegaly. We then performed exploratory data analysis techniques to examine the distributions and check the balance of the dataset.

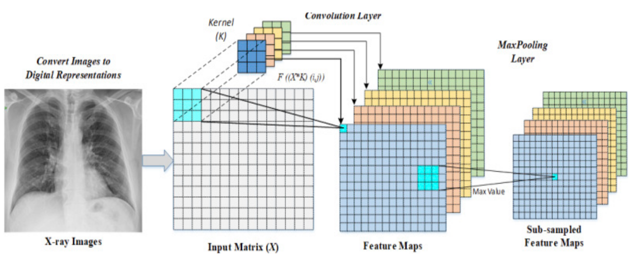
We split the dataset into training, validation, and test datasets with a ratio of 70:15:15. To balance the distribution in the dataset, we used oversampling to increase the number of cases with cardiomegaly. We used the ImageDataGenerator function in Keras to perform data augmentation on the training dataset, including horizontal and vertical flipping, random rotation, and zooming.

The implementation of the dataset on the entire model without any pretrained weights, as well as on the model with an appended layer of fully connected layers, did not give satisfactory accuracy. This approach was tried initially to train the model from scratch without any prior knowledge or weights. However, due to the lack of training data and the complexity of the problem, the accuracy obtained was not up to the mark. Hence, the transfer learning approach was utilized, where a pre-trained model was used as a starting point for training the model on the given dataset. Although the accuracy improved significantly, the performance was still not optimal.

Therefore, we added an attention model to the pre-trained models to improve their performance. We also applied global weighted average pooling to capture the most important features of the images and output them to a classification layer with dropout and two fully connected layers.

We also applied these techniques on both unbalanced and balanced datasets. Using an unbalanced dataset can be problematic because it can lead to biased and inaccurate model performance. In an unbalanced dataset, the number of examples in each class is not evenly distributed. This means that the model is more likely to be trained on one class over another, which can result in a biased model that performs poorly on the underrepresented class. On the other hand, a balanced dataset ensures that the model is trained on an equal number of examples from each class, resulting in a more accurate and unbiased model. The model is forced to learn the features that are common to all classes rather than relying on the bias towards one class.

We trained the models using transfer learning, fine-tuning pre-trained models on the ImageNet dataset, including Efficient Net B4, InceptionV3, MobileNetV1, and ResNet152-V2. We initialized the models with their pre-trained weights and then trained them on the balanced NIH dataset with the added attention model and global weighted average pooling. We used the binary cross-entropy loss function and the Adam optimizer with a learning rate of 0.0001. We used early stopping with patience of 10 to prevent overfitting.



$$\mathrm{GWAP}(x,y,d) = \frac{\sum_x \sum_y \mathrm{Attention}(x,y,d)\mathrm{Feature}(x,y,d)}{\sum_x \sum_y \mathrm{Attention}(x,y,d)}$$

Global Weighted Average

B. Testing

To evaluate the performance of our "Cardiomegaly Detection from Chest X-Rays using CNN and Transfer Learning Algorithm" implementation, we tested our models on the previously split test dataset. We used a confusion matrix to visualize the performance of our models in predicting true positive, false positive, true negative, and false negative cases.

We also calculated the accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score for each model. Our results indicate that the proposed approach is effective in detecting cardiomegaly from chest X-rays. However, there is still room for improvement, and future work can focus on addressing the imbalanced nature of the dataset and exploring other deep-learning models to further enhance the performance of the system.

TP = True Positive	TN = True Negative
FP = False Positive	FN = False Negative

Accuracy = (TP+TN)/(TP+FP+FN+TN)
Precision = TP/(TP+FP)
Sensitivity/Recall = TP/(TP+FN)
Specificity = TN/(TN+FP)
Positive Predictive Value = TP/(TP+FP)
Negative Predictive Value = TN/(TN+FN)
F1 Score = 2*(Recall * Precision) / (Recall + Precision)

C. Discussions on Finding and Challenges

TABLE I. BALANCED DATASET

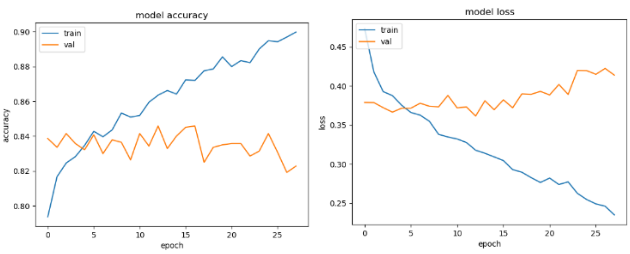
Models	**Results (BALANCED DATASET)**		
	Metrics used to measure	***Validation***	***Test***
EfficientNetB4	Accuracy	0.7457	0.7349
	Positive Predictive Value	0.4187	0.4004
	Negative Predictive Value	0.9083	0.8964
	Sensitivity	0.6942	0.6511
	Specificity	0.7586	0.7559
	F Score	0.77	0.76
InceptionV3	Accuracy	0.7579	0.7586
	Positive Predictive Value	0.4378	0.4376
	Negative Predictive Value	0.9197	0.9162
	Sensitivity	0.7338	0.7194
	Specificity	0.764	0.7685

Models	Results (BALANCED DATASET)		
	Metrics used to measure	***Validation***	***Test***
	F Score	0.78	0.78
MobileNetV1	Accuracy	0.7968	0.781
	Positive Predictive Value	0.4945	0.4672
	Negative Predictive Value	0.9051	0.9062
	Sensitivity	0.6511	0.6655
	Sensitivity	0.8333	0.8099
	F Score	0.81	0.79
ResNet152V2	Accuracy	0.7767	0.7767
	Positive Predictive Value	0.4577	0.4581
	Negative Predictive Value	0.896	0.8976
	Sensitivity	0.6223	0.6295
	Sensitivity	0.8153	0.8135
	F Score	0.79	0.79

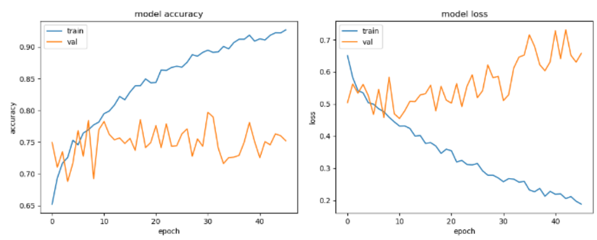
TABLE II. UN-BALANCED DATASET

Models	Results (DATASET NOT BALANCED)		
	Metrics used to measure	***Validation***	***Test***
EfficientNetB4	Accuracy	0.8314	0.8184
	Positive Predictive Value	0.6429	0.5903
	Negative Predictive Value	0.8549	0.8449
	Sensitivity	0.3561	0.3058
	Specificity	0.9505	0.9468
	F Score	0.81	0.79
InceptionV3	Accuracy	0.8343	0.8249
	Positive Predictive Value	0.6463	0.6115
	Negative Predictive Value	0.8595	0.8522
	Sensitivity	0.3813	0.3453
	Specificity	0.9477	0.945
	F Score	0.82	0.81
MobileNetV1	Accuracy	0.8458	0.8386
	Positive Predictive Value	0.7133	0.6688
	Negative Predictive Value	0.8619	0.8607
	Sensitivity	0.3849	0.3849
	Sensitivity	0.9613	0.9523
	F Score	0.83	0.82
ResNet152V2	Accuracy	0.835	0.8401
	Positive Predictive Value	0.6899	0.6944
	Negative Predictive Value	0.8499	0.8569
	Sensitivity	0.3201	0.3597
	Sensitivity	0.964	0.9604
	F Score	0.81	0.82

The table shows the results of four different pre-trained models, EfficientNetB4, InceptionV3, MobileNetV1, and ResNet152V2, on both balanced and unbalanced datasets for the task of cardiomegaly detection from chest X-rays. The balanced dataset results reveal that MobileNetV1 achieved the highest accuracy of 0.7968 in the validation set and 0.781 in the test set, followed by InceptionV3 with 0.7579 in the validation set and 0.7586 in the test set. The unbalanced dataset results show that MobileNetV1 achieved the highest accuracy of 0.8458 in the validation set and 0.8386 in the test set, followed by ResNet152V2 with 0.835 in the validation set and 0.8401 in the test set. However, the sensitivity scores of all models were low for both balanced and unbalanced datasets, indicating that the models had difficulty in detecting true positives. The results also show that the performance of the models varied significantly depending on the balance of the dataset. The findings suggest that more efforts are needed. to improve the sensitivity scores of the models for better cardiomegaly detection



MobileNet V1 Model Accuracy and Loss for UnBalanced Dataset



MobileNet V1 Model Accuracy and Loss for Balanced Dataset

V. CONCLUSION AND FUTURE WORK

REFERENCES

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title,

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

.

[4] K. Elissa, "Title of paper if known," unpublished.

[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from