Utilizing Chest X-rays for Age Prediction and Gender Classification

Chris Solomou
Dept. of Computer Science
University of York
York, UK
https://orcid.org/0000-0002-7034-5755

Dimitar Kazakov
Dept. of Computer Science
University of York
York, UK
https://orcid.org/0000-0002-0637-8106

Abstract - **In this paper we present a framework for automatically predicting the gender and age of a patient using chest x-rays (CXRs). The work of this paper derives from common situations in medical imaging where the gender/age of a patient might be missing or in situations where the x-ray is of poor quality, thus leaving the medical practitioner unable to treat the patient appropriately. The proposed framework comprises of training a large CNN which jointly outputs the gender/age of a CXR. For feature extraction, transfer learning was employed using the EfficientNetB0 architecture, with a custom trainable top layer for both classification and prediction. This framework was applied to a combination of publicly available data, which collectively represent a heterogeneous dataset showing a variation in terms of race, location, patient's health, and quality of image. Our results are robust with respect to these factors, as none of them was used as input to improve the results. In conclusion, Deep Learning can be implemented in the medical imaging domain for automatically predicting characteristics of a patient.**

I. INTRODUCTION

Medical Imaging is the process of creating visual representations of the interior parts of the body such as organs or tissues. Its applications include clinical purposes to diagnose and monitor health conditions, and for treating diseases or injuries. Amongst medical images, chest x-rays (CXR) are a useful tool used to diagnose potential health issues around the thorax. The application of deep learning to the medical imaging domain has seen a significant increase over the past few years [1]. Specifically, the application of DCN (Deep Convolutional Networks) in medical imaging allows for an immediate and unbiased diagnosis since its trained on diagnoses of past patients with similar symptoms.

However, automatic labelling of a patient's info using CXRs, is an exciting application of modern ML algorithms, and one that we believe has not been explored to its full potential. Until recently there was not much work done, but the release of multiple large CXRs datasets has boosted the interest in this area. By being able to predict the age and gender of a patient (especially when this information is not available) using a CXR, a medical practitioner can make a more informed diagnosis regarding the condition of a patient, and how he/she should be treated. For instance, an older patient would need different medical treatment than a younger patient.

II. BACKGROUND

Past work utilizing Machine Learning techniques for age prediction includes bone-age studies, with most of them based on hand and wrist radiographs. The samples used in these studies were mostly comprehensive or bordered the age of 18 [2]. For example, Zulkifley et al. [3] using a transfer learning regressor (Xception-41) managed to predict the age of a child within a MAE of 8.2 months.

Age prediction from chest x-rays however, has not extensively being researched and there are only a few attempts in predicting a patient's biological age from CXRs. First piece of research comes from Karargyris et.al. [4] where the authors of this paper use a large CXRs dataset, and employing transfer-learning (Dense Net), trained a network in regression fashion that allowed to predict the age of patient within a ± 9 years window for 93% of the data. All the available image-data were used, and the authors did not distinguish between gender or the condition of a patient.

Gender prediction has been more extensively researched, and generally can be considered easier to estimate than age. Specifically, Zhiyun et.al [6] managed to automatically identify the gender of a patient based on the patient's CXR. The authors applied transfer learning and employed a range of large pretrained models for extracting features from the CXRs. Thereafter, SVM was used for classifying the gender of each x-ray. Best performance comes from the VggNet-16 feature extractor with an SVM classifier, with accuracy of 86.6%. The dataset used was a combination of multiple datasets containing images obtained from different countries, institutions and x-ray machines.

Past work for both gender classification and age prediction using CXRs, to our knowledge has only been done very recently by Chung-Yi Yang et al [7]. In this work the authors managed to predict age and classify gender with very high accuracy, however their methodology was based on a specific dataset that contained only healthy patients, of the same race and place (location and institution). Additionally, the dataset contained only "high-quality chest radiographs" that came from the same x-ray machines. It is also noteworthy that the standard deviation of the participant's age was fairly low in the used dataset.

Although there is no clear indication as to which parts of the CXRs yield the highest information for predicting age/gender, as proven by past work [2,5,7], there are certain features present in a CXR that produce this information. Currently, and to our knowledge there is not significant past work for predicting both age and gender in an arbitrary setting, meaning for a network to be able to extrapolate and perform well, without any confounding variables affecting its performance. This paper attempts to utilize a sample of public datasets and apply a network architecture that can generalize well and achieve good results irrespectively of the quality of the CXRs or the condition / characteristics of the patients. In conclusion, we devised and implemented a network that combines automatic age and gender prediction and can generalize well using the CXRs of various patients with different attributes.

III. DESIGN FRAMEWORK

NETWORK ARCHITECTURE:

For the basic network architecture and feature extraction from the CXRs the EfficientNetB0 (pre-trained on ImageNet) was utilized that offers high accuracy and good computational performance [9]. When using other versions of the same model the accuracy did not show any significance increase, and the computation load was much higher. As the final layer of EfficientNetB0 is used for classification, the top layer was removed and replaced by the following additional layers. An average global pooling layer was applied which aggregates all the important information and reduces the computation time.

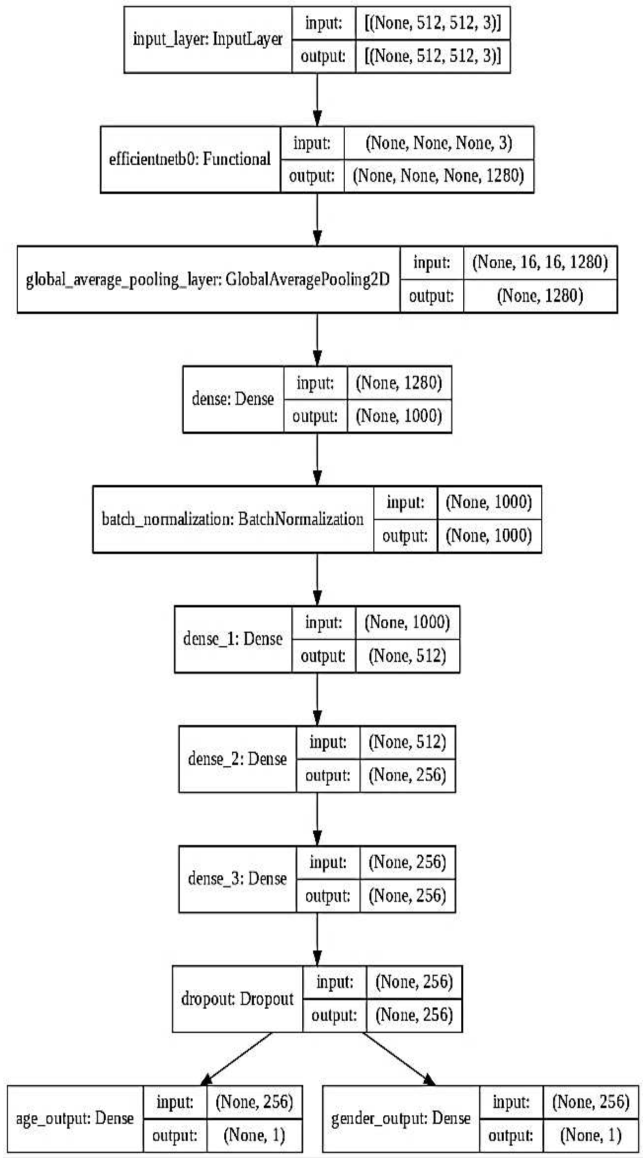

Fig. 1: Network Architecture

The resulting output from EfficientNetB0 is a flatten layer containing 1280 features from each image.
Thereafter a fully connected layer with 1000 neurons was added, with Batch Normalization applied to its output values. Thereafter another 3 fully-connected dense layers containing 512, 256 and 256 neurons respectively. Leaky Relu activation functions were used in all the layers after the

features were extracted (apart from the final output layer).

Finally, a Dropout layer with a parameter of 0.1 was applied to the final layer so the network can learn to generalize. For age prediction the output layer contains a single neuron with no activation function since it is performing a regression-style prediction. For gender-classification, the output layer contains a single neuron with sigmoid activation function, which is then rounded to give the class-prediction.

DATASET METHODOLOGY:
For training the model two publicly available datasets were utilized [10,11]. Both image-datasets come with metadata that contain additional information including the age and gender of a patient. The CXRs were filtered to include adults aged between 18-85. Furthermore, among the two datasets there is a mix of healthy patient x-rays, and patients that were in fact diagnosed from suffering from a chest-condition. After these conditions have been accounted the resulting dataset encompasses 52842 examples with an average age of 48.6 ± 14.7.

For predicting the age and the gender of a patient we will not be distinguishing between the conditions that a patient suffers from, since it would be reasonable that the model will be able to predict these target variables regardless of the patient's current health situation. Additionally, using CXRs of patients with different chest-conditions could benefit the model from generalizing better and not overfitting.

The images were resized to 512*512, since it yielded the highest accuracy for this network architecture, with no data augmentation being applied. The pixel-values were not normalized since EfficientNetB0 performs better with regular pixel values. The full dataset was randomly split into 85% for training the model and 15% for model-validation. Finally, the model was trained for 400 epochs or until the relevant metrics stopped improving.

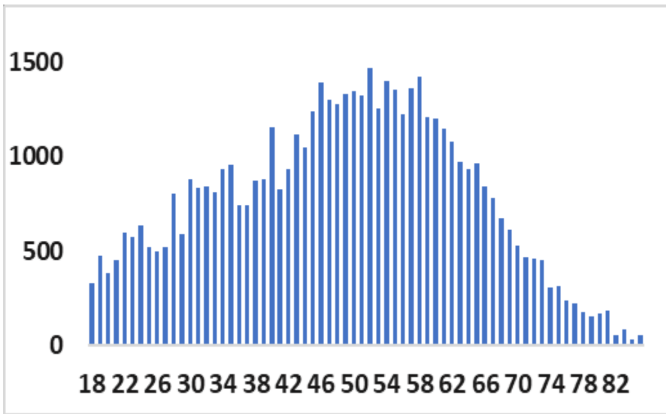

Fig. 2: Distribution among Age Groups

CHEST-CONDITIONS DISTRIBUTION			
finding	***count***	***finding***	***count***
Chlamydophila	*3*	*Pneumonia*	*186*
ARDS	*4*	*COVID-19*	*322*
Bacterial	*4*	*Edema*	*653*
Coli	*4*	*Pleural Thickening*	*711*
Influenza	*6*	*Emphysema*	*718*
Varicella	*6*	*Fibrosis*	*731*
Klebsiella	*8*	*Pneumothorax*	*1048*
Nocardia	*8*	*Mass*	*1259*
Legionella	*8*	*Cardiomegaly*	*1269*
Tuberculosis	*9*	*Consolidation*	*1401*
Lipoid	*13*	*Nodule*	*1433*
SARS	*16*	*Effusion*	*3466*
Streptococcus	*17*	*Infiltration*	*4984*
Pneumocystis	*25*	*Atelectasis*	*5183*
Hernia	*83*	*No Finding*	*29264*

GENDER DISTRIBUTION	
Gender	***Count***
Female	23421
Male	29421

IV. EXPERIMENTS AND RESULTS

For evaluating the success of our network, apart from the network's loss, we have implemented the following metrics:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}} \quad (1)$$

RMSD for determining the standard deviation from the mean age,

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (2)$$

the coefficient of determination for evaluating the fit of the predicted age with actual age,

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{(\boldsymbol{TP+FP+TN+FN})} \quad (3)$$

accuracy for gender classification compared to the ground truth,

$$\boldsymbol{Recall} = \frac{\boldsymbol{TP}}{(\boldsymbol{TP+FN})} \quad (4)$$

and recall for evaluating explicitly the accuracy among different age and gender groups.

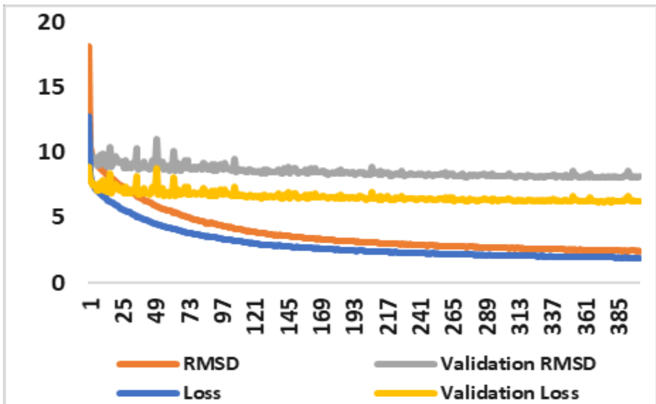

Fig. 3: Accuracy - Loss curves for age prediction

AGE PREDICTION RESULTS

Training Loss	***1.94***
Training RMSD	***2.46***
Validation Loss	***6.19***
Validation RMSD	***8.04***
Recall ± *5 years*	***0.581***
Recall ± *10 years*	***0.8511***
R^2	***0.71***

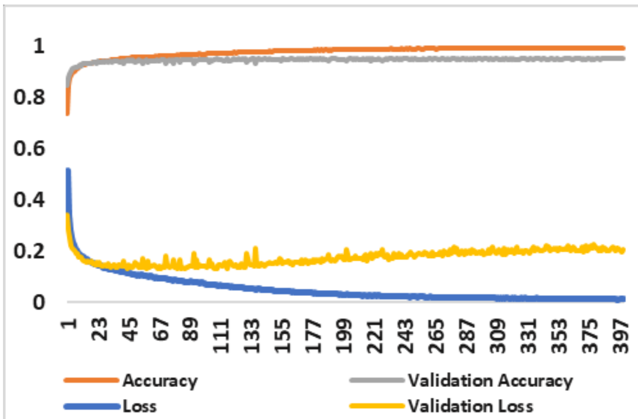

Fig. 4: Accuracy - Loss curves for gender classification

GENDER CLASSIFICATION RESULTS

Training Loss	***0.01252***
Training Accuracy	***0.996***
Validation Loss	***0.2066***
Validation Accuracy	***0.958***
Recall (both genders)	***0.95***

Generating a scatterplot of the actual and the predicted age (using the validation dataset), it can be seen that the age predictions of our network present a fairly good fit, with a coefficient of determination of almost 0.71. It is worth noting that the mean age from the validation dataset was very close to our network's predictions 48.38, 48.56. However, our network produced a lower standard deviation than the validation dataset 14.75, 12.65, signifying that our network succeeded in reducing the deviation from the mean age. Overall, the most outliers derive from the age group between 40-70, and it seems that our model fits best (less deviation from the mean) in the age group of 73+.

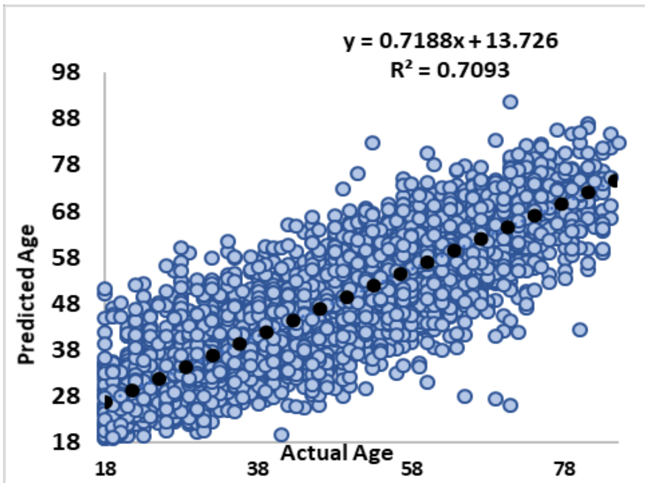

Fig. 5: Actual vs Predicted Age

Although we are not certain, there could be specific features present in these x-rays that make it easier for the model to distinguish than x-rays of patients with lower age. Also, there are some significant outliers present, which could be because of the poor quality of certain x-rays.

Generating a confusion matrix for assessing the gender-classification results it can be seen that the model performs equally well for both classes with approximately 96% accuracy for each class, and an overall 95.8% overall.
Generally, our network achieves higher performance for predicting gender than age and we believe this to be because of the different bone structures of the two genders.

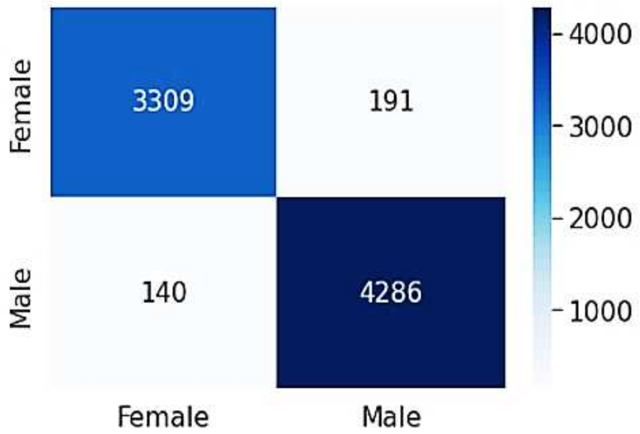

Fig. 6: Accuracy between genders

V. CONLCUSIONS

Because of missing data or modern data privacy regulations, sometimes the gender and age information of a patient are not available, leaving the medical practitioners unable to perform their work appropriately. In this paper we managed to combine past work that has been done for age-prediction and gender-classification using CXRs, by creating an architecture that achieves great performance, using a smaller and "noisier" dataset than previous work.

Our methodology comprises of four main steps: image pre-processing, transfer learning for feature extraction, feature selection and prediction/classification. Using the same architecture for both target variables (age and gender) and modifying the top layer by changing the activation function, we are allowed to perform both a regression-style prediction and binary-label classification. This methodology is applied to a combined dataset that consists of two sources, that results to the final dataset that the model was trained on.

Using half of the available data for reducing computation cost we were able to implement a network that produces consistent results, in regardless of the patient's characteristics (country, race, health condition etc.) and the quality of the training data. Specifically, we managed to predict correctly the age of a patient within an average window of ± 8 years, and for 86% of the x-rays within a decade of their actual age. For gender classification, we managed to predict correctly the gender of patient with an accuracy of almost 96%. Although we did not quite match state of the art results for age prediction, we did succeed in developing a versatile model that can be used for automatic prediction (labeling) of both target-variables.

The work performed in this paper, sets the foundation for future work that we want to carry-out. In particular, setting-up an automatic system that can predict the age and gender of a patient using CXRs, and make a diagnosis as to if a patient suffers from a certain chest condition. Furthermore, if other datasets that contain more information about the patients' attributes become available (e.g. country, race) in the future, a model can be trained to predict these as well. Putting together this system, and achieving high-accuracies among these variables, could allow for medical practitioners to make better and more rounded diagnoses, and for the patients to be benefited with early and precise treatments.

Finally, a mapping system should also be set-up that will allow to map the features that correspond to a correct or incorrect prediction from the model for each target variable.

REFERENCES

[1] Erdi Çallı, EcemSogancioglu, Bram van Ginneken, Kicky G.van Leeuwen, Keelin Murphy, “Deep learning for chest X-ray analysis: A survey”, Medical Image Analysis, Volume 72 361-8415, https://doi.org/10.1016/j.media.2021.102125, 2021

[2] Ana Luiza Dallora, Peter Anderberg, Ola Kvist, Emilia Mendes, Sandra Diaz Ruiz, Johan Sanmartin Berglund, Bone age assessment with various machine learning techniques: A systematic literature review and meta-analysis, https://doi.org/10.1371/journal.pone.0220242,2019

[3] Mohd Asyraf Zulkifley, Siti Raihanah Abdani, Nuraisyah Hani Zulkifley, “Automated Bone Age Assessment with Image Registration Using Hand X-ray Images”, Appl. Sci. 2020, 10, 7233, https://www.mdpi.com/2076-3417/10/20/7233, 2020

[4] Alexandros Karargyris, Satyananda Kashyap, Joy T Wu, Arjun Sharma, Mehdi Moradi, Tanveer Syeda-Mahmood, “Age prediction using a large chest X-ray dataset”, Proceedings Volume 10950, Medical Imaging 2019: Computer-Aided Diagnosis, https://arxiv.org/ftp/arxiv/papers/1903/1903.06542.pdf, 2019

[5] François Bellemare, Alphonse Jeanneret, Jacques Couture “Sex differences in thoracic dimensions and configuration”, American Journal of Respiratory and Critical Care Medicine 168(3), 305–312,2003

[6] Xue, Zhiyun & Antani, Sameer & Long, L. & Thoma, George, “Using deep learning for detecting gender in adult chest radiographs”, Proceedings of the SPIE, Volume 10579, id. 105790D9pp, https://doi.org/10.1117/12.2293027, 2018

[7] Chung-Yi Yang, Yi-Ju Pan, Yen Chou, Chia-Jung Yang, Ching-Chung Kao, Kuan-Chieh Huang, Jing-Shan Chang, Hung-Chieh Chen, Kuei-Hong Kuo, “Using Deep Neural Networks for Predicting Age and Sex in Healthy Adult Chest Radiographs”, J. Clin. Med. 2021, 10, 4431. https://doi.org/10.3390/jcm10194431, 2021

[8] Allwyn Joshua, Lathika Shetty, Vidyashambahava Pare, “Variations in dimensions and shape of thoracic cage with aging: An anatomical review”, Anatomy Journal of Africa, 3 (2): 346 – 355, 2014

[9] Zebin T, Rezvy S. COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization [published online ahead of print, 2020 Sep 12]. Applied Intelligence. 2020;1-12. doi:10.1007/s10489-020-01867-1

[10] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald M. Summers, “ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases”, https://arxiv.org/abs/1705.02315, 2017

[11] Joseph Paul Cohen, Paul Morrison, Lan Dao, “COVID-19 Image Data Collection”, https://arxiv.org/abs/2003.11597, 2020