Automated Detection of Sex in Mutant Mice Using Convolutional Neural Networks Applied to X-ray Images

Kola Babalola
Hamed Haseli Mashhadi
Violeta Muñoz-Fuentes
Jeremy Mason
Terry Meehan
On behalf of the International Mouse Phenotyping Consortium

European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Editors: Under Review for MIDL 2019

Abstract

In this study we train a convolutional neural network (CNN) using a dataset of 3,104 X-ray images of wildtype mice generated by the International Mouse Phenotyping Consortium (IMPC) to classify males and females. Our model achieves over 97% accuracy on the test set and on 10,400 images from knockout mice lines. By investigating class activation maps of the misclassified images, we observed potential abnormalities in the pelvis in one knockout line. This study suggests that CNNs can be trained for the purpose of quality control in addition to the discovery of novel genotype-phenotype associations.

1. Introduction

Recent developments in deep learning have led to the successful application of convolutional neural networks (CNNs) in medical image analysis (see for example (Litjens et al., 2017) ).

The International Mouse Phenotyping Consortium (IMPC) is a global infrastructure with over 20 centres on 5 continents seeking to produce and phenotype a knockout mouse strain for each of the ∼20K protein coding genes in the mouse genome. Over 250 biological parameters are assessed to associate gene to function. One procedure takes whole body X-rays of the mice followed by expert annotation, e.g., shape of vertebrae, no. of ribs, etc.

We leverage on the property that CNNs trained for classification inherently capture useful features of images (Zhou et al., 2016). Visualisation methods showing areas of importance in images classified by CNNs e.g. (Zhou et al., 2016), (Zeiler and Fergus, 2014), can provide insights into physical characteristics important for classification of image features.

High throughput image data generation will often have missing metadata that can make interpretation of results difficult. Training a CNN to classify mice by sex from X-ray images allows automated validation of crucial metadata that is processed manually. Moreover, we hypothesise that the model captures intrinsic properties of sexual dimorphism between males (M) and females (F). Hence grouping the scores generated by the model will provide a quantitative way to explore changes due to knockout (KO) genes. Visualisation of images...
using class activation maps (CAMs) will allow qualitative exploration of the associated sex-related changes.

2. Method

Construction and application of the model

Our dataset from IMPC included 3,014 wildtype (WT) mice (1,502M; 1,512F) - used for model building, and 10,400 knockout mice (5,211M; 5,189F), representing 729 knockout lines, with typically 14 knockout mice (7M; 7F) per line.

Our CNN was based on VGG-16 (Simonyan and Zisserman, 2015) where all the convolutional (feature) layers were frozen and the fully-connected (classifier) layers were retrained using our data. The number of outputs of the final layer was changed to 2 (from 1000) for the male/female classes. The constitution of the training, validation and test sets were: 1,808 training images (882M, 926F), 602 validation (299M, 303F) and 604 test (321M, 283F).

The output of the model are scores for each class. An image is put into the class with a higher score. After training, the model achieved a classification accuracy of over 97% on both the test set and the knockout images. Both scores for each image were stored.

Visualisation of the model

To investigate the features important in obtaining the classification scores, we used a variation of CAMs (Zhou et al., 2016). Whilst Zhou et al used a fully convolutional network, our network had three fully connected layers between the last convolutional layer and the output layer; thus, we did not have a global pooling layer as they did. However, their concept of mapping the activations back to the last convolutional layer still applies, except we had to backpropagate through two additional layers and perform an element-wise multiplication between the backpropagated activations and the features of the last convolutional layer ($512 \times 7 \times 7$) to give the CAM. The mean of this over all 512 layers gives a $7 \times 7$ image which was upsampled to the same dimensions as the input image and overlaid on it.

Figure 1: Box plots of male (blue) and female (green) scores for all WT ($n = 1502$) and KO ($n = 5211$) males, and for the three KO genes with lowest mean male scores and the three with highest mean male scores ($n = 7$ in each case)
3. Results

Here we present results for the male mice, grouping male and female classification scores by gene (Figure 1). Although there were no observable differences between wildtypes and knockouts when pooled, apparent differences were observed when each knockout line was individually considered. In particular, Duoxa2 and Tmem189 had female scores higher than male scores for images of male mice. CAMs for males and females indicated anatomical features that were relevant in sex classification (Figure 2) and, in the case of the Douxa2 individuals, hinted at either a misclassification issue or an abnormal phenotype associated with the knockout of the gene (Figure 3).

![Figure 2: Class activation maps of 3 random wildtype females (1st three) & males (last 3)](image)

![Figure 3: CAMs of 7 Duoxa2 KO male mice (all except the rightmost misclassified as female)](image)

4. Discussion and Conclusion

The accuracy of the model on both the wildtypes and the knockouts show that it can be used as an automated quality control tool to highlight images that may be mislabelled. Furthermore, the classification scores can be used to identify abnormal phenotypes in knockout mice. Interestingly, the class activation maps show the chest and pelvic regions are important in classification which seem intuitive to the human observer.

Results for two mutant lines, Duoxa2 and Tmem189 indicated a potential mislabelling of the images (a quality control issue) or phenotype changes due to the knockout genes. The images associated with these lines have been set aside for further manual inspection.

In future work we will apply the same analysis to female mice from the knockout lines. Additionally, this preliminary approach, focusing on the selection of sex as an attribute to classify images, will be expanded to be more general by using autoencoders. We also intend to quantify bone abnormalities in general to detect deviating phenotypes in knockouts.
Acknowledgments

This work was supported by the United States National Institutes of Health (NIH) Grant U54 HG006370.

References


