Discovering and interpreting transcriptomic drivers of imaging traits using neural networks

Nova F. Smedley [1 2] **Suzie El-Saden** [1 3] **William Hsu** [1 3 2]

Abstract

Cancer heterogeneity is observed at multiple biological levels. Approaches to link organ- and tissue-level information from diagnostic images and cellular-level information from genomics are needed. However, current "radiogenomic" studies often use linear, shallow models, depend on feature selection, or consider one gene at a time when mapping images to genes.

We present a neural network-based approach that takes high-dimensional gene expressions as input and performs nonlinear mapping to an imaging trait. We propose gene masking and gene saliency to extract learned relationships from radiogenomic neural networks. We demonstrate that neural networks can model transcriptomic heterogeneity to reflect differences in imaging and can be used to derive radiogenomic traits with clinical value.

Manuscript published in Bioinformatics: https://doi.org/10.1093/bioinformatics/btaa126.

Availability and implementation: https://github.com/novasmedley/deepRadiogenomics.

1. Introduction

Radiogenomic mapping is the integration of traits observed on medical images and traits found at the molecular level, such as gene expression profiling (Diehn et al., 2008). As such, the study of radiogenomics plays a role in precision medicine, where associations can describe prognosis or therapy response. Neural networks, with their ability to automatically learn nonlinear, hierarchical representations of large input spaces, are alternate approaches for radiogenomics. The models can combine low-level features into a structure of complex features to create a new, abstracted and transformed representation better suited for learning than the original input. Towards understanding the biological basis of imaging traits, we thus present an approach using the representational power of neural networks to model tumor transcriptomes and nonlinearly map genes to tumor imaging traits. No *a priori* selection is used on the transcriptome.

Moreover, a limitation of neural networks is that they are considered "black boxes," which makes it difficult to interpret the learned relationships. We provide approaches, called gene masking and gene saliency, for understanding radiogenomic neural networks. Our model interpretation methods can identify cohort-level relationships, which we refer to as *radiogenomic associations*, and patient-level relationships, which we refer to as *radiogenomic traits*.

2. Methods

We analyzed publicly available gene expression and medical imaging data to discover radiogenomic associations using a deep neural network. Transcriptomes were available for 528 GBM patients as part of The Cancer Genome Atlas (TCGA). Samples were previously analyzed by the Broad Institute on Affymetrix arrays, quantile normalized, and background corrected. Medical images for 262 GBM patients were downloaded from The Cancer Imaging Archive (TCIA) and matched to TCGA samples. A board-certified neuroradiologist (Dr. El-Saden, 26 years of experience) evaluated images using the Osirix medical image viewer. An electronic form was used to record MRI traits according to the Visually Accessible Rembrandt Images (VASARI) feature guide. 175 patients had pre-operative (pre-op) MRIs and transcriptomes. Traits were binarized given the small sample sizes.

To map relationships between gene expression profiles and MRI traits, feed-forward neural networks were used (Fig. 1a–b). We performed transfer learning by pretraining an autoencoder using a separate subset of 353 patients from the TCGA-GBM cohort with gene expression data only. The radiogenomic neural networks were pretrained using weights transferred from the autoencoder.

[1]Medical & Imaging Informatics, University of California, Los Angeles, USA [2]Department of Bioengineering, University of California, Los Angeles, USA [3]Department of Radiological Sciences, David Geffen School of Medicine at UCLA, Los Angeles, USA. Correspondence to: William Hsu <whsu@mednet.ucla.edu>.

Presented at the ICML 2020 Workshop on Computational Biology (WCB).

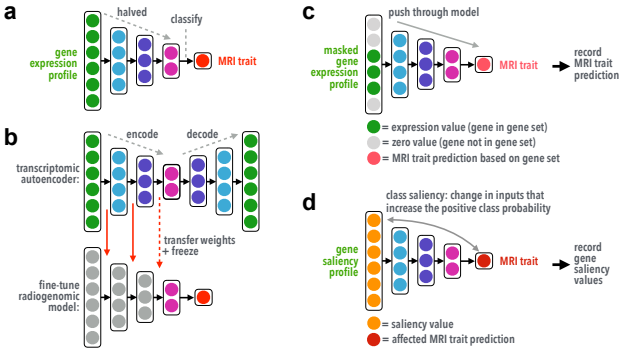

Figure 1. The radiogenomic neural network's (**a**) architecture, (**b**) transfer learning with autoencoder, and interpretation methods through (**c**) gene masking and (**d**) gene saliency.

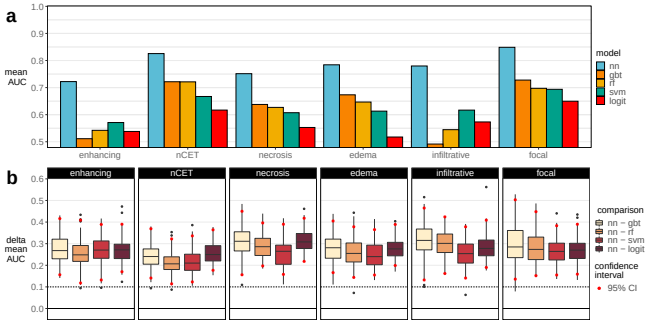

Figure 2. Radiogenomic model performance. (**a**) Observed 10-fold cross-validation performance. (**b**) Performance differences between a neural network and another model in 100 bootstrapped datasets. Notation: neural network (nn), gradient boosted trees (gbt), random forest (rf), support vector machines (svm), logistic regression (logit).

We trained a different neural network for each imaging trait. All neural networks were trained on NVIDIA Tesla K80 and V100 GPUs through Amazon Web Services using `Python 3.6`, `Keras 2.2.4`, and `TensorFlow 1.12.0` on a Ubuntu 16.04 machine. Comparison models (gradient boosted trees, random forest, support vector machine, and logistic regression) were implemented via `XgBoost 0.80` and `sklearn 0.20.0`.

Given the trained networks, we performed a sensitivity analysis where the value(s) of one or more components of the input are kept while all others are replaced with zeros. The goal is to determine the impact that the kept input components have on the end classification; this procedure was previously described in (Zeiler & Fergus, 2014). Here, we define "gene masking" to extract radiogenomic associations from a trained neural network (Fig. 1c). For each individual, the gene expression values of a particular gene set were kept while all other expressions were replaced with zeros. The masked profiles were pushed through a fully trained neural network and the output, a class probability based on using genes from the gene set, was recorded.

We also explored the concept of "gene saliency", which is based on class saliency, a visualization technique used to compute the gradient of an output class prediction with respect to an input via backpropagation (Simonyan et al., 2014). Here, we define "gene saliency" as the genes whose change in expression would increase the model's belief of the positive class label (Fig. 1d). In each model, salient genes are derived for each patient, ranked, and analyzed using gene set enrichment analysis to determine if a gene set is relevant to predicting his/her MRI trait. Subsequently, positive enrichment between a single patient's salient genes and a gene set is defined as a "radiogenomic trait."

3. Results

The deep neural networks were better at estimating MRI traits than all other classifiers (Fig. 2a). In predicting VASARI features, the first hidden layer used frozen pre-trained weights and only the last two hidden layers were used for fine-tuning. Bootstrapping showed neural networks had higher performances in Fig. 2b.

The full results including a detailed interpretation of the gene masking and gene saliency results can be found in our published paper, available here (Smedley et al., 2020).

References

Diehn, M., Nardini, C., Wang, D. S., McGovern, S., Jayaraman, M., Liang, Y., Aldape, K., Cha, S., and Kuo, M. D. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. *Proc. Natl. Acad. Sci. U. S. A.*, 105(13):5213–8, apr 2008. ISSN 1091-6490. doi: 10.1073/pnas.0801279105. URL https://doi.org/10.1073/pnas.0801279105.

Simonyan, K., Vedaldi, A., and Zisserman, A. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv Prepr. arXiv1312.6034*, 2014. URL https://arxiv.org/abs/1312.6034.

Smedley, N. F., El-Saden, S., and Hsu, W. Discovering and interpreting transcriptomic drivers of imaging traits using neural networks. *Bioinformatics*, 02 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa126. URL https://doi.org/10.1093/bioinformatics/btaa126. btaa126.

Zeiler, M. D. and Fergus, R. Visualizing and understanding convolutional networks. In *Eur. Conf. Comput. Vis.*, pp. 818–833. Springer, 2014. URL https://doi.org/10.1007/978-3-319-10590-1{_}53.