



Corrigendum: Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data

Darcie A. P. Delzell^{1*}, Sara Magnuson¹, Tabitha Peter¹, Michelle Smith¹ and Brian J. Smith²

¹ Department of Mathematics and Computer Science, Wheaton College, Wheaton, IL, United States, ² Department of Biostatistics, University of Iowa, Iowa City, IA, United States

Keywords: radiomics, machine learning, CT image, biomarkers, lung cancer

A Corrigendum on

Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data

by Delzell, D. A. P., Magnuson, S., Peter, T., Smith, M., and Smith, B. J. (2019). *Front. Oncol.* 9:1393. doi: 10.3389/fonc.2019.01393

OPEN ACCESS

Edited and reviewed by:

Rong Tian,
Sichuan University, China

*Correspondence:

Darcie A. P. Delzell
darcie.delzell@wheaton.edu

Specialty section:

This article was submitted to
Cancer Imaging and Image-directed
Interventions,
a section of the journal
Frontiers in Oncology

Received: 03 April 2020

Accepted: 01 May 2020

Published: 05 June 2020

Citation:

Delzell DAP, Magnuson S, Peter T,
Smith M and Smith BJ (2020)
Corrigendum: Machine Learning and
Feature Selection Methods for
Disease Classification With Application
to Lung Cancer Screening Image
Data. *Front. Oncol.* 10:866.
doi: 10.3389/fonc.2020.00866

The data analyzed for this study were generated by Samantha Dilger, Ph.D and Jessica Sieren, Ph.D (Departments of Radiology and Biomedical Engineering, University of Iowa, Iowa City, IA, United States) who control the rights to the data and do not intend for the data to be shared publicly. Accordingly, this data which was included as Supplementary Material in the original article is being removed. In addition, the data were taken from a mix of low and high-dose CT scans, which were incorrectly referred to in the original article as low-dose scans.

The corrections below have been made to the **Methods**, subsection **Dataset**, paragraph 1.

“This retrospective study analyzed data originally taken from 200 CT scans of the lungs of patients at the University of Iowa Hospital. Pathology and radiology reports were reviewed to identify an analysis set of patients who met eligibility criteria of having (a) a solitary lung nodule (5–30 mm) and (b) a malignant nodule confirmed on histopathology or a benign nodule confirmed on histopathology or by size stability for at least 24 months. Manual segmentations were performed by a graduate student trained in medical image analysis in order to define a region of interest (ROI) around each nodule. The ROIs were defined to include amounts of parenchyma approximately proportional to the nodule sizes. Individual ROI voxels were labeled as belonging to either the nodule or the parenchyma, with radiomic features calculated separately for each to produce the complete set of 416 (approximately half nodule and half parenchyma) quantitative imaging biomarkers. These biomarkers measured features such as intensity, shape, and texture of the ROI (15). This study is a secondary analysis of de-identified data originally collected with approval from the University of Iowa institutional review board. Demographic information can be found in Table 1.”

The dataset has been removed from the online Supplementary Material and replaced with R code implementing the feature selection and classification models described in Methods Sections 2.3 and 2.4 of the article. The **Methods** section, subsection **Classifiers and Performance Metrics**, paragraph 2 has been updated to include a reference to the supplementary code as follows:

“The quality of model performance in most machine learning algorithms is dependent upon the choice of various tuning parameters. Some tuning parameters take into account the number of predictors after feature selection. For example, the `mtry` tuning parameter for `rf`, which determines the number of candidate variables at each branch, is equal to the square root of the number of predictors. Other tuning parameters were chosen based on standard practice (22, 23). For example, the decay tuning parameter for `nnet`, which helps prevent overfitting,

generally takes the values of 0.1, 0.01, and 0.001. All models were fit using the `caret` R package (24). Our R code implementing the feature selection and classification models is presented as Supplementary Material.”

The authors apologize for the inclusion of the data in the Supplementary Material and misstatement of “low-dose” CT. We state that these do not change the scientific conclusions of the article in any way. The original article has been updated.

Copyright © 2020 Delzell, Magnuson, Peter, Smith and Smith. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.