Computational Toxicology 2 (2017) 45–49

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

Autoencoder Predicting Estrogenic Chemical Substances (APECS): An improved approach for screening potentially estrogenic chemicals using in vitro assays and deep learning

Lyle D. Burgoon

US Army Engineer Research and Development Center, RTP, NC 27711, United States

ARTICLE INFO

Article history:
Received 28 January 2017
Received in revised form 7 March 2017
Accepted 12 March 2017
Available online 14 March 2017

Keywords:
Deep learning
Autoencoder
Endocrine disruptor
Estrogenic
Predictive toxicology
Computational toxicology
High throughput screening

ABSTRACT

In 2015 the US Environmental Protection Agency published a computational toxicology approach to screen chemicals for potential estrogenic activity. This complicated approach requires several steps, including concentration-response modeling (which includes fitting several different models and identifying the best model), application of a multi-factor mathematical model that attempts to model the concentration-response data, calculation of the area under the concentration-modeled response curve, and finally standardizing the area under the concentration-modeled response curve to that of 17-beta estradiol. Toxicologists will find it difficult to implement this approach on their own, creating a need for a more straightforward tool. Recently, it has been shown that deep learning approaches lead to less complicated approaches, that can run faster than more complicated approaches, while maintaining or improving overall algorithmic performance. In this paper we examine the Autoencoder Predicting Estrogenic Chemical Substances (APECS). APECS is two deep autoencoder models that achieve at least the same performance while being less complicated for an average toxicologist to use than the US EPA's approach. Our deep autoencoders achieved accuracies of 91% vs 86% and 93% vs 93% on the in vivo and in vitro datasets used by the US EPA in validating their approach. Users can use our deep autoencoder models to make predictions of assay data by using our open source Java desktop applications. APECS has a simple push-button interface and was written in Java.

Introduction

Governments around the world want to protect their citizens and environments from endocrine disrupting chemicals. These chemicals can either act as mimics of endocrine active substances, or disrupt endocrine signaling [4,5]. Depending upon the timing of exposure, the impacts of endocrine disrupting chemicals may be permanent or transient [5]. At the same time, there is interest in replacing animal models with lower cost and higher throughput in vitro assays.

The US Environmental Protection Agency recently developed a complicated, multistep algorithm and mathematical model to predict if a chemical is an endocrine disruptor using data from the ToxCast program [1]. Beyond the complicated nature of the algorithm, the approach is also somewhat subjective in nature. For instance, the approach uses either the Hill model or the Gain-Loss model. However, the Hill model is known to not fit all sigmoidal shapes well, and a generalized sigmoidal model may perform better generally [3]. In addition, non-sigmoidal relationships may exist in assay concentration-response data, which are best fit with other models, such as exponential or linear models [3]. Thus, a data-driven non-parametric approach to curve fitting is likely more appropriate [2].

In addition, the EPA's approach uses the area under the concentration-response curve (AUC) to calculate similarity between a chemical's concentration-response curve and 17-beta estradiol. The problem is that curves with very different shapes can all share the same AUC. For instance, a chemical with a sigmoidal concentration-response curve with an AUC of 75 units would be called similar to another chemical that is best fit with a quartic equation and an AUC of 75, or a chemical with an exponential concentration-response curve and an AUC of 75. These shapes are all very different, but yield the same AUC, and have been seen in Tox21 data before. A more robust alternative is to use Pearson correlation, which is sensitive to shape.

The United States Army has interests in developing predictive computational toxicology models that use in vitro high throughput assays to identify promising new chemicals of military interest

E-mail address: lyle.d.burgoon@usace.army.mil

http://dx.doi.org/10.1016/j.comtox.2017.03.002
2468-1113/Published by Elsevier B.V.

faster, that are less toxic to people and the environment, yet still acceptable to regulators. The Autoencoder Predicting Estrogenic Chemical Substances (APECS) are deep [learning] autoencoders that use in vitro high throughput screening assay data to predict if a chemical is estrogenic. Deep learning and autoencoder methods have been used previously to create reduced order models – models that are less computationally intense, yet still yield good predictions [9,10]. Deep autoencoders have also been used as a nonlinear replacement for principal components analysis or singular value decomposition, especially as an un-supervised or semi-supervised pattern recognition approach.

This paper will examine the APECS models and demonstrate that they perform generally at least as well as the US EPA's method for predicting if a chemical is estrogenic, while also having the advantage of being simpler to implement and use.

Materials and methods

Data

Data were obtained from the US EPA's ToxCast invitrodb_v2 database (https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data). We obtained the in vitro and in vivo estrogenicity "ground truth" calls for chemicals from Browne et al. [1].

Software

All analysis and model building was performed in R (v3.2.4)· H2O (v3.8.3.3) was used for model development. One model was built using the in vitro ground truth information (APECS-vitro) and another model was built using the in vivo ground truth information (APECS-vivo). Deep autoencoders are simply deep neural networks. Plain old java objects (POJOs) were exported from the H2O server. These POJOs contain the neuron weights, neural network architecture, and bias factors.

A JavaFX graphical user interface was built in Java (v1.8.0_05) that uses the POJOs to make predictions on user-supplied in vitro data. The open source graphical user interface code is available at GitHub (https://github.com/DataSciBurgoon/apecs_vivo and https://github.com/DataSciBurgoon/apecs_vitro). The executable desktop applications can be downloaded from GitHub (APECS-vivo: https://github.com/DataSciBurgoon/apecs_vivo/releases and APECS-vitro: https://github.com/DataSciBurgoon/apecs_vitro/releases).

Analysis

Data for 10 of the assays reported in Browne et al. [1] were used. These 10 assays were: 1) NVS_NR_hER, 2) OT_ER_ERaERa_0480, 3) OT_ER_ERaERa_1440, 4) OT_ER_ERaERb_0480, 5) OT_ER_ERaERb_1440, 6) OT_ER_ERbERb_0480, 7) OT_ER_ERbERb_1440, 8) TOX21_ERa_BLA_Agonist, 9)ATG_TRANS, 10) ATG_CIS. We had difficulty finding the other assays listed by Browne et al. [1] within the ToxCast database. This is not a concern given that APECS' performance surpassed that reported in Browne, et al. [1] and the aim was to develop a reduced order model, not to perform a direct reproduction of the Browne et al. [1] study.

Loess was used to fit a nonlinear model to the concentration-response data for each chemical and assay combination. Pearson correlation was used to measure the similarity of the concentration-response curves for each chemical and assay combination to the concentration-response curve for 17-beta estradiol in each assay. The resulting matrix (chemicals as rows, assays as columns, and correlation in each cell) was fed into the autoencoder function from H2O.

For the chemicals that are estrogenic in vitro and their negative controls from the Browne et al. study [1], the autoencoder had 3 hidden layers with 10, 2, and 10 neurons, respectively. For the estrogenic in vivo chemicals and their negative controls, also from the Browne et al. study [1], the autoencoder had 7 hidden layers with 43, 20, 5, 2, 5, 20, and 43 neurons, respectively. The number of neurons and the number of hidden layers in both cases was chosen using a grid search, with an eye toward optimal separation of the chemicals based on their classification as estrogenic or not.

The middle (2 neurons) hidden layer was projected into a Cartesian plane for each autoencoder. This 2-dimensional projection serves as a nonlinear unsupervised clustering of the chemical data. A Euclidean distance that results in the best classification accuracy was chosen for each autoencoder. This is similar to choosing a circular decision boundary centered on 17-beta estradiol. For the in vitro autoencoder the optimal distance was 1.35 units, and for the in vivo autoencoder the optimal distance was 1.50 units.

Results

The autoencoder approach achieved marginally higher accuracy than the ToxCast ER Model (Table 1). For the in vivo data, the autoencoder achieved 91% accuracy vs 86% for the ToxCast ER Model. The autoencoder did a better job at identifying true negatives, resulting in fewer false positives, while achieving the same performance for true positives. For the in vitro data, the autoencoder achieved the same accuracy as the ToxCast ER Model (93% accuracy for both). Here, the autoencoder did a better job of identifying true positives, resulting in no false negatives. The autoencoder misclassified three true negatives as false positives, versus the ToxCast ER Model which misclassified only one, resulting in a lower specificity for the autoencoder. Having higher sensitivity is something we typically want to achieve in a screening assay, even at the expense of specificity.

One of the advantages of the autoencoder approach is that we can generate visualizations from the autoencoder that help us see the results (Figs. 1 and 2). In Figs. 1 and 2, we can see the impact of moving the decision boundary to greater than or less than 1.50 units (plots generated in R). This also allows us to see which chemicals have the most similar and the most different behaviors in the ToxCast assays compared to 17-beta estradiol.

Table 1
In Vivo and In Vitro Autoencoder Model Performance vs ToxCast ER Model Performance.

Performance	In Vivo APECS*	ToxCast ER Model In Vivo	In Vitro APECS*	ToxCast ER Model In Vitro
True Positives	29	29	28	26
True Negatives	10	8	9	11
False Positives	3	5	3	1
False Negatives	1	1	0	2
Sensitivity	97%	97%	100%	93%
Specificity	80%	67%	75%	92%
Accuracy	91%	86%	93%	93%

* Results are using the JavaFX APECS-Vitro and APECS-Vivo GUI software.