Machine Learning meets Animal Physiology: Understanding Disease Progression and Predicting Impending Death

Jie Mei1, Stefanie Banneke, Matthias Endres1, Christoph Harms1,2 and Julius V Emmrich1

1. Department of Neurology and Department of Experimental Neurology, Charité – Universitätsmedizin Berlin; 2. German Federal Institute for Risk Assessment, German Center for the Protection of Laboratory Animals (BfR); QUEST – Center for Transforming Biomedical Research, Berlin Institute of Health (BHI). Berlin, Germany.

1. Introduction

Humane endpoints help to prevent severe suffering in experimental animals. However, properly determining humane endpoints remains challenging due to:

- Wide variety of animal models of disease;
- Lack of standardized protocols for evaluating the physiological state of animals;
- Heterogeneity of humane endpoints in published studies.

To address this issue, we:

- Propose a novel method using machine learning models to generalize the process of endpoint refinement;
- Validated the applicability of the machine learning models in death prediction across different animal models.

The study showed that:

- Machine learning models could detect animals at increased risk of death with high accuracy in mouse models of stroke and sepsis.

2. Animal models and timeline

Mouse model of stroke – 147 mice of 3 strains: C57BL/6J, C3H/HeJ and 129S6/SvEvTac. Animals were treated with temporary filamentous middle cerebral artery occlusion (MCAo) or sham procedure (control).

Mouse model of sepsis – 435 mice of 4 strains: C57BL/6J, Mekj (Jackson Lab 29-Merk tm1(Grr)); C3H/HeJ, 129S6/SvEvTac and C57BL/6J. Animals were treated with intraperitoneal injection of lipopolysaccharide (LPS) or physiological saline solution (control).

Parameters obtained in the stroke model:

- Body weight and core body temperature: daily;
- Sickness score: on the day of treatment and on the 1st, 2nd, 7th, 14th and 28th day post-treatment.

Parameters obtained in the sepsis model:

- Surface body temperature and sickness score: 8 times/day on the day of treatment; 3 times/day for 2 days after the 2nd treatment; daily from the 3rd day post-treatment onwards;
- Body weight: 3 times/day from the 1st day of treatment to the 2nd day post-treatment; daily from the 3rd day post-treatment onwards.

Humane endpoints: Severity of disease was evaluated on a scale of 0-5 (0: normal; 5: maximum severity). Animals were sacrificed upon reaching a sickness score greater than 4 once or a score of 4 twice within 2 hours.

3. Models for predicting death

Body temperature, body weight and sickness score were used separately or in combination to train machine learning models including:

- Logistic regression;
- Decision tree (of max_depth = 1, 2, 3 or 4);
- Support vector machine (with linear or radial basis function (RBF) kernels; C = 100 or 10; gamma = 0.01, 0.001 or 0.0001);
- Random forest classifier (with n_estimators = 2, 4 or 8).

Stratified k-fold cross-validated grid search was applied (stroke model, k = 2; sepsis model, k = 3) to assess and identify:

- General performance of machine learning models;
- Usability of physiological parameters obtained at different time points for death prediction;
- Hyperparameters of the model.

Available data from all time points before the averaged time of death was included in the analysis.

4. Results

Death as an outcome could be predicted with high accuracy in mouse models of stroke and sepsis:

- Stroke model: 97.5% accuracy with weight measurements obtained on the 3rd day post-treatment;
- Sepsis model: 96.2% accuracy, with body temperature and sickness scores obtained 24 hours post-treatment.

For the stroke model, adding an additional physiological parameter in model training did not lead to higher performance in death prediction. In the sepsis model, the use of additional physiological parameters in model training led to higher model performance.

However, this does not apply to all time points (Table 1):

- When training machine learning models with data from the 24th hour post-treatment, adding additional parameters (i.e., sickness score) in model training led to an increase of 21.5%, 10.9% and 1.1% in sensitivity, F1 score and accuracy, respectively.
- When training machine learning models with data from the 36th hour post-treatment, improvement in model performance was not observed.

By applying endpoints determined with machine learning models, non-survivors could be euthanized at an earlier time point (Figure 2):

- In the stroke model, 2 out of 3 (67%) non-survivors could have been euthanized on the 3rd day post-treatment (average time of death of the 2 animals = 4.5 (0.7) days post-treatment);
- In the sepsis model, 25 out of 28 (89%) non-survivors could have been euthanized 24 hours post-treatment (average time of death of the 25 animals = 58.7 (35.0) hours post-treatment).

The false positive rate was low. None and 6 out of 254 (2.3%) animals that survived until the end of the experiment, were falsely predicted to die at an earlier time point during the experiment for stroke and sepsis groups, respectively.

Results from the present study may also suffer from small size of datasets and missing data points, which are the common features of animal experiments.

5. Conclusions

- Machine learning can be used to predict impending death in mouse models of acute disease and to refine humane endpoints;
- Automated parameter search minimizes manual effort and allows for objective identification of physiological parameters and time points with high relevance for death prediction in mouse models of acute disease;
- Combining more than one physiological parameter during model training increases prediction accuracy;
- Application of machine learning-based humane endpoints allows for earlier termination of experiments than conventional manual scoring methods, thus reducing an animal’s unnecessary suffering and distress.

Acknowledgements

We thank all animal technicians from the Center for the Protection of Laboratory Animals at the German Federal Institute for Risk Assessment (BfR), especially Paolo Rosellini Tognetti and Lisa Gordingken, for their technical assistance. This study was funded by a grant from Einstein Stiftung Berlin.