

**TITLE**  
**May 6, 2020**

ABSTRACT.

CONTENTS

1. Introduction	2
2. Methods	3
3. Regression	5
4. Classification	6
References	8

## 1. INTRODUCTION

On average, mortality rates in humans tend to increase exponentially with age. However, this mortality rate does not apply uniformly to all humans in an age group; there is an observed heterogeneity in mortality and health status among individuals in an age cohort. One 80-year-old may have a long list of health issues, limited mobility, and cognitive deficits, while another may seem a decade younger than their true age. The difference in the rate of aging between individuals creates a difference between chronological and biological age. The concept of frailty is used to capture this phenomenon.

Frailty is defined in the literature as the state of increased vulnerability to adverse health outcomes. Identifying frailty is an important clinical challenge as frail individuals tend to have both increased risk of diseases and disorders, worse health outcomes from the same disease, and even different symptoms of the same disease (Whitehead et al., 2014). Their all-cause mortality risk is also higher than non-frail individuals of a similar age.

Though how to best quantify frailty is still controversial, one popular method is to use a frailty index (FI). In this method, an individual is scored on a set of age-related deficits as either 0 (no deficit), 0.5 (slight deficit), or 1 (deficit present). The scores are added up and divided by the total number of deficits in the frailty index, giving a final FI score. Studies have shown that the specific items included in a FI, provided they meet certain criteria, are not important (Rockwood et al., 2017). The criteria for items to be included in a FI are: 1) that the deficits are health-related; 2) that the deficits generally increase with age; 3) that the FI is composed of more than 30 different items; and 4) that the items in the FI do not saturate (e.g. become apparent in all individuals) at too early an age (Searle et al., 2008). Frailty indexes constructed which fulfill the above requirements have consistent characteristics shared between them: average FI scores increase exponentially with age, variability in FI scores also increases with age, the score strongly predicts risk of adverse outcomes and mortality, and there is a submaximal limit of about  $\frac{2}{3}$  of deficits accumulated after which the individual dies (Searle et al., 2008). Although human frailty indexes have been studied extensively, it is only recently that researchers have recreated one for mice.

In mice, the frailty index consists of 30 set features which can be graded without the use of specialized equipment. In a 2017 study by Rockwood et al. (2017) comparing the mouse frailty index across lifespan of two hundred fifty-one male C57BL/6J to human FI data, the mouse CFI showed several essential features seen with human aging. The mouse CFI had a similar distribution to humans where the mean FI score increased exponentially with age with comparable values at comparable life stages. The FI scores also showed a dose-response relationship with mortality and a submaximal limit. Although individual deficit accumulation patterns varied, the distribution of the frailty index broadened with age and mean deficit accumulation increased. At all ages, greater deficit accumulation was associated with reduced survival in age-adjusted models. These results point to the potential usefulness of the FI concept for a preclinical model of frailty and aging. Indeed, the mouse CFI has since been used to evaluate interventions on

frailty such as administering resveratrol and caloric restrictions, and to study the changes in aging of the microbiome, inflammation, genetics, and heart health (Kane et al., 2017).

Unlike the human frailty index, most mouse frailty indexing has been done with the same set of 30 features. It is likely though, that just as the human frailty index is robust to different sets of deficits, a new mouse CFI can be constructed. Indeed, some studies have already substituted a small number of mouse CFI features for ones that fit their specific strain or experiment (Kane et al., 2017). We propose that we can use this characteristic of frailty indexes to construct our own. We hypothesize that biological aging produces changes in behavior and physiology which can be captured with our video assay. Our goal is to develop an efficient and scalable visual frailty index from video metrics using modeling. Such an FI could be used to study mechanisms, exacerbators, and interventions for biological aging on a large scale.

## 2. METHODS

Two hundred fifty-three C57BL/6J mice of varying ages were put in the one-hour open field test with top-down video taken. They were then given a frailty index score by trained scorers. The ages of the mice ranged from 20 to 134 weeks. There were 161 males and 92 females. Below is a plot showing CFI scores by age, colored for sex (males in blue, females in red):

Leinani's figure

Though the overall trend for the mean is to increase with age group, there is a slight dip between 80 to 100 weeks. Further, although standard deviation of the group increased with age group, after 40 weeks the increase is only slight.

In order to see how the scorer had an effect on ratings, a plot graphing FI by age colored by the scorer was made:

Leinani's figure

The plot reveals some obvious scorer effects. The red scorer tended to have slightly higher CFI scores while the purple scorer tended to have slightly lower CFI scores. In the 70 to 90 week age range, the turquoise scorer was dominant, combined with the purple scorer, and they tended towards lower CFI scores. This may explain the slight dip seen in group means.

The videos of each mouse were run through our open field pipeline which consists of the standard gait metrics and the open field metrics. For a review of these metrics refer to (\*\*\*\*\*)

In addition to our standard pipeline features, we engineered features from the video data. The first feature we looked at was flexibility. Spinal mobility in humans is a predictor for quality of life in aged populations (Imagama et al., 2011) and the mouse is used as a model for the aging human spine (Holguin et al., 2014). Thus, we hypothesized that aged mice may bend their spine to a lesser degree, or less often due to reduced flexibility or spine degeneration. This change in flexibility can be captured by the pose estimation. In this approach, we take the coordinates of three points on the mouse at each video frame:

the back of the head (A), the middle of the back (B), and the base of the tail (C). A line is drawn through the points. Tracking the length of the line AC normalized for mouse length (dAC), the orthogonal distance of the middle of the back B from the line (dB), and the angle of the three points (aABC) is used as a measure of the bend in the spine. For each of the three per frame measures (dAC, dB, and aABC) a mean, median, standard deviation, minimum, and maximum are calculated per video. These become the features to be used in the model.

#### Leinani’s figure

The next engineered features were body composition features. In humans, anthropometric measures such as BMI and waist-to-hip ratio are predictors for certain health conditions and mortality risk. Aging is associated with changes in body composition, increased visceral fat and decreased subcutaneous fat, progressive dysfunction of adipose tissue, and changes in ratio of brown and white adipose tissue ([Huffman and Barzilai, 2009](#)). These changes are also seen in obese patients, making them present as prematurely aged. Though obesity, as defined by BMI, is a predictor for various metabolic disorders, cardiovascular disorders, cognitive disorders, and overall mortality risk, it seems that visceral fat is a strong causal factor ([Huffman and Barzilai, 2009](#)). Visceral fat creates a systemic inflammatory environment via secretion of inflammasomes. This ties visceral fat to the popular theory of aging: “Inflammaging”, where aging is caused or exacerbated by chronic low grade inflammation ([Pérez et al., 2016](#)). Visceral fat is a better predictor for various health issues and mortality than BMI ([Scafoglieri et al., 2014](#)). This is thought to be because BMI can sometimes be a poor predictor of visceral fat. Further, anthropometric measures tend to vary based on measurement points and are also unreliable measures of visceral fat, though they may still be good predictors of risk of disorders and mortality.

In mice, changes in body composition are observed as well ([Pappas and Nagy, 2019](#)). So far, no body measurements have been shown to solidly predict visceral fat, and specialized equipment must be used to predict fat composition. We observed that body weight was very highly correlated with the gait metrics with the highest correlations with FI score. Thus, we hypothesize that body composition metrics may show some signal for aging and frailty. On each frame, an ellipse was fitted to the mouse, giving a width and a length. The means and medians of the mouse width were used as features. Further, we took the coordinates of the rear paws of the mice per frame as a estimation of hip width. The mean and median rear paw width were also used as features for the model.

We also looked into features related to intra-mouse variability of gait. A known part of aging is the heterogeneity in the presence of aging related metrics across an age group—heterogeneity in the age group increases with age group. This is a characteristic of the frailty index. There is also interest in heterogeneity at the individual level. In addition to looking at inter-group heterogeneity in gait, we looked at intramouse heterogeneity in gait using standard deviation and interquartile range of the mouse for each gait metric. The standard deviations and interquartile ranges were used as features for the model.

### 3. REGRESSION

The association between sex and frailty score variable is provided in [Table 1](#). For these animals, the association is statistically significant based on a chi-squared test of association ( $p$ -value = 0.004).

Sex	Score < 6	Score $\geq$ 6
Male	102	59
Female	74	17

TABLE 1. Association between sex and frailty score.

There are several different sets of features (or predictors) available for modeling the overall frailty score. To assess each set’s predictive ability, we will train models using animal features (weight, sex), gait features ( $p = 11$ ), engineered features ( $p = 3$ ), and different combinations of the three.

Before building the models, we will split the data into training and test sets. The training set will be used to explore relationships and build the model. The test set will be used to assess model performance. To partition the data, we split the original data ( $n = 252$ ) in a *stratified* manner to preserve the class distribution of the score. This will keep the distribution of animal scores approximately the same ([Figure 1](#)).



FIGURE 1. Distribution of scores

We fit a variety of models to the frailty index aging dataset. Figures [3](#) and [4](#) show performance metrics calculated using cross-validation and the test data. Linear Model (LM) was the weakest performer closely followed by the robust linear model (RLM). While both penalizing and boosting the linear model did help, it did not make the models (LMnet, LMB) very competitive.

Another group of models included support vector machines (SVM), random forest (RF) and gradient boosting machines (GBM) had a better performance than others on the training set but performed poorly on the test set. There was a pack of two-stage (first dimension reduction, then regression) models (PLS, PCR) that showed better results but would not be considered interpretable given the dimension reduction approach applied to the original features in the first step.

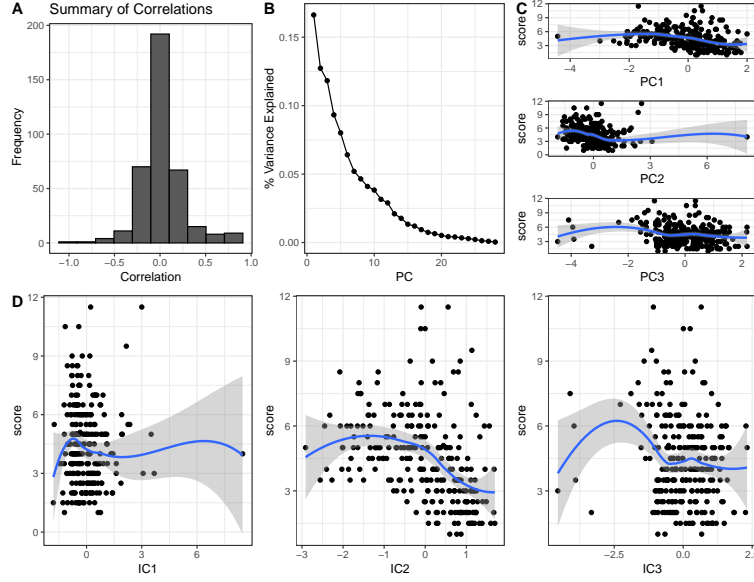


FIGURE 2. (A) We see a few high correlations in a histogram plot of inter-correlations between all the features. (B) Proportion of variance explained by principal components (PCs) showing that we need quite a few of them to explain the variability in the original predictor space. (C) A scatterplot between score and first three PCs. (D) A scatterplot between score and first three Independent Componentnets (ICs).

Given this, if one wants interpretability then using either a penalized linear model (LMnet) or random forest (RF) would be a reasonable choice but if one wants a higher predictive power, then two-stage models (PLS, PCR) may be better.

#### 4. CLASSIFICATION

While frailty index is continuous in nature, we can represent it qualitatively and define

$$Y = \begin{cases} \text{Class 1} & \text{score} \leq 4 \\ \text{Class 2} & \text{score} > 4 \end{cases}$$

The question of interest is whether we can predict the frailty class of an animal using gait, open field, animal and our own engineered features.

As earlier, there are 252 samples available; 80% will be used for training the algorithms while the remainder will be used to evaluate the final candidate models. The data were split using stratified random sampling to preserve the class distribution of the outcome. 10 repeats of 10-fold cross-validation were used to tune the models.

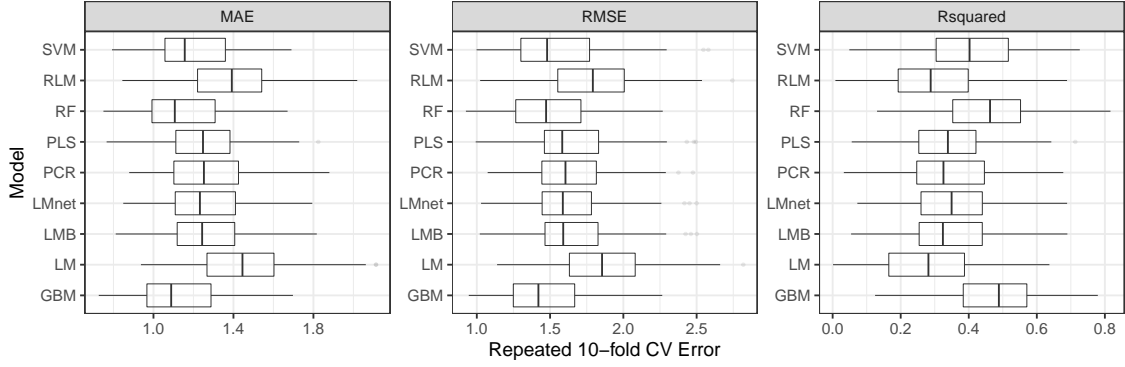


FIGURE 3. Performance of different models on training data plotted in terms of MAE (median absolute error), RMSE (root mean squared error), and Rsquared ( $R^2$ ). A lower MAE, RMSE and higher  $R^2$  are preferred.

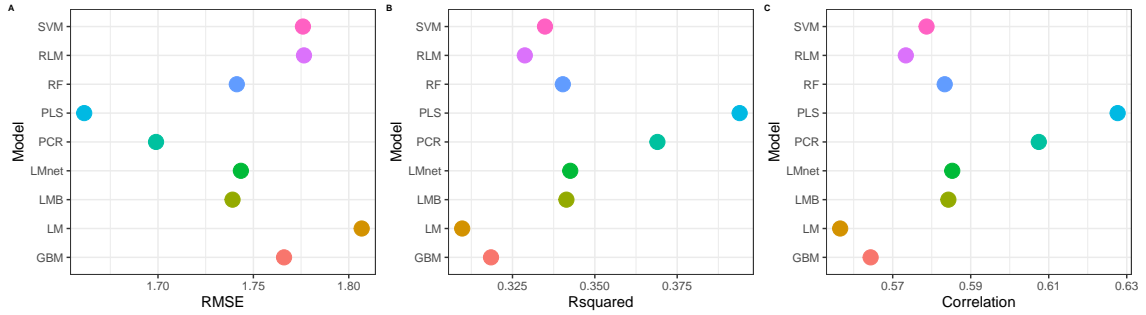


FIGURE 4. Performance of different models on test data plotted in terms of RMSE (root mean squared error), and Rsquared ( $R^2$ ) and correlation ( $\text{cor}(Y_{\text{obs}}, Y_{\text{pred}})$ ). A higher  $R^2$ , correlation and lower RMSE are preferred.

A series of models were fit to the training set. The hyperparameter combination associated with the smallest (which error?) was chosen for the final model and used in conjunction with the entire training set. The following models were investigated,

- Linear Discriminant Analysis and its variants
- Logistic Regression with Elastic Net
- Support Vector Machine
- Random Forest
- Gradient Boosting Machine

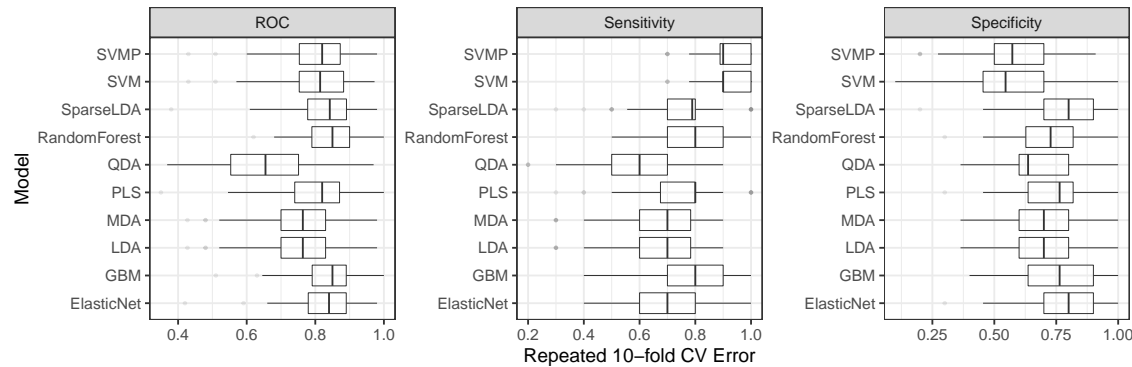


FIGURE 5. Performance of different models on training data plotted in terms of ROC (area under the Receiver operating characteristic curve), sensitivity and specificity.

Method	Tuning Parameters
Linear Discriminant Analysis (LDA)	None
Mixture Discriminant Analysis (MDA)	Number of Subclasses

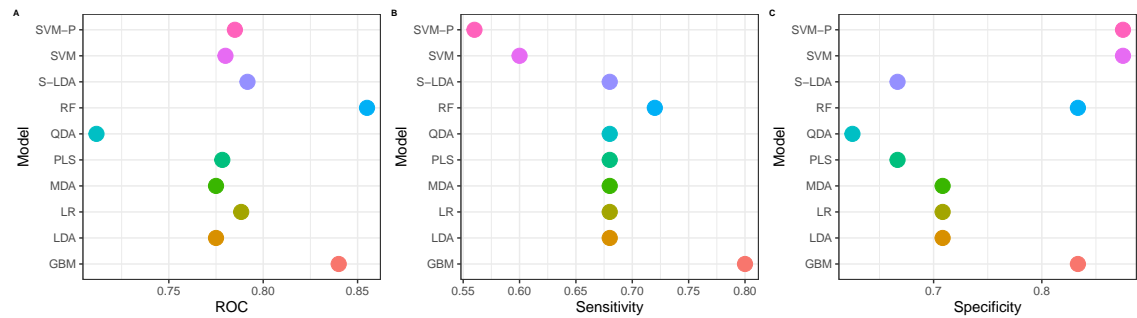


FIGURE 6. Performance of different models on test data plotted in terms of ROC (area under the Receiver operating characteristic curve), sensitivity and specificity.

# REFERENCES

HUFFMAN, D. M. and BARZILAI, N. (2009). Role of visceral adipose tissue in aging. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1790** 1117–1123.

IMAGAMA, S., MATSUYAMA, Y., HASEGAWA, Y., SAKAI, Y., ITO, Z., ISHIGURO, N. and HAMAJIMA, N. (2011). Back muscle strength and spinal mobility are predictors of quality of life in middle-aged and elderly males. *European Spine Journal* **20** 954–961.

KANE, A. E., AYAZ, O., GHIMIRE, A., FERIDOONI, H. A. and HOWLETT, S. E. (2017). Implementation of the mouse frailty index. *Canadian journal of physiology and pharmacology* **95** 1149–1155.



- PAPPAS, L. E. and NAGY, T. R. (2019). The translation of age-related body composition findings from rodents to humans. *European journal of clinical nutrition* **73** 172–178.
- PÉREZ, L. M., PAREJA-GALEANO, H., SANCHIS-GOMAR, F., EMANUELE, E., LUCIA, A. and GÁLVEZ, B. G. (2016). 'adipaging': ageing and obesity share biological hallmarks related to a dysfunctional adipose tissue. *The Journal of physiology* **594** 3187–3207.
- ROCKWOOD, K., BLODGETT, J., THEOU, O., SUN, M., FERIDOONI, H., MITNITSKI, A., ROSE, R., GODIN, J., GREGSON, E. and HOWLETT, S. (2017). A frailty index based on deficit accumulation quantifies mortality risk in humans and in mice. *Scientific reports* **7** 43068.
- SCAFOGLIERI, A., CLARYS, J. P., CATTRYSE, E. and BAUTMANS, I. (2014). Use of anthropometry for the prediction of regional body tissue distribution in adults: benefits and limitations in clinical practice. *Aging and disease* **5** 373.
- SEARLE, S. D., MITNITSKI, A., GAHBAUER, E. A., GILL, T. M. and ROCKWOOD, K. (2008). A standard procedure for creating a frailty index. *BMC geriatrics* **8** 24.
- WHITEHEAD, J. C., HILDEBRAND, B. A., SUN, M., ROCKWOOD, M. R., ROSE, R. A., ROCKWOOD, K. and HOWLETT, S. E. (2014). A clinical frailty index in aging mice: comparisons with frailty index data in humans. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **69** 621–632.