Revealing gait as a murine biomarker of injury, disease, and age with multivariate statistics and machine learning

Bilal A. Naved[1, 2, 3], Shuling Han[1, 2], Kyle M. Koss[1, 2], Mary J. Kando[4], Jiao-Jing Wang[1,2], Jill Chang[5], Craig Weiss[4], Maya G. Passman[5], Jason A. Wertheim[1,2], Yuan Luo[7], Zheng J. Zhang[1,2]

[1]Dept. of Surgery, [2]Comprehensive Transplant Center, Feinberg School of Medicine, [3]Dept. of Biomedical Engineering, [4]Behavioral Phenotyping Core, [5]University of Colorado School of Medicine, [6]Barnard College, [7]Division of Health and Biomedical Informatics, Dept. of Preventive Medicine.

Background

- 15% of people experience a gait abnormality by the age of 60 (1). That number increases to over 80% over the age of 85.
- Rodent models are the most prevalent pre-clinical model for evaluating gait.
- Hundreds of rodent gait studies rely on univariate analyses of treadmill data.
- Systems like DigiGait generate multidimensional data (30+ measures).
- Traditional approaches compare individual gait parameters in isolation, offering limited insight into how features interrelate in distinct gait phenotypes.
- Our lab recently developed a multivariate, machine learning-based statistical analysis pipeline that can characterize, quantify, and distinguish gaits.
- The question stands if the developed pipeline may apply more generally to other etiologies of gait deficit including central nerve injury or normal aging?

Hypotheses

- Multivariate gait analysis will reveal biologically consistent relationships.
- There are latent factors that help intuitively understand gait.
- Not all measurable features are relevant for characterizing gait, there is a subset
- Using this subset to train models will be more accurate than using all features
- These models will be able to distinguish between different gait phenotypes, like a human eye can just by watching.
- This can all be done across various etiologies in rodent models (central, peripheral, aging) (**Fig. 1**) (**Fig. 2**).

Experimental Design

- We modeled three different etiologies of gait deficit (peripheral nerve injury, central nerve injury, and natural aging).
- Groups shown in **Table 1** below
- 14 animals in each group.
- After a 2-week recovery period animals were video taped walking on DigiGait **(Fig 1).**

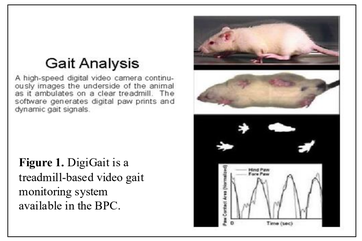

Figure 1. DigiGait is a treadmill-based video gait monitoring system available in the BPC.

Experimental Groups				
Etiology	Group 1	Group 2	Group 3	Group 4
Peripheral nerve injury	Control	Nerve transection	Limb transplant	
Central nerve injury	Control	Hyperoxia	IUGR	Hyperoxia + IUGR
Aging	Younger mice	Older mice		

Methods

- Collected DigiGait data of three different gait deviations **(Fig 1).**
- Multivariate factor analysis (using MATLAB's factoran) and forward feature selection (with ten-fold cross-validation) was conducted to identify those features and factors most descriptive of each gait state.
- Five machine learning classifier models were trained with ten-fold cross-validation and evaluated (e.g. random forest, regression, discriminant analysis, support vector machine, and ensemble) in a 70-30 training-testing split for their accuracy, precision, recall, and F-score.
- The highest performing model was used to score each type of gait.
- The score distributions were plotted on a histogram for direct comparisons of score populations between various gait states**.**

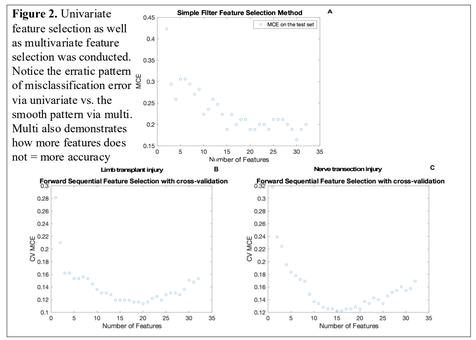

Figure 2. Univariate feature selection as well as multivariate feature selection was conducted. Notice the erratic pattern of misclassification error via univariate vs. the smooth pattern via multi. Multi also demonstrates how more features does not = more accuracy

Table 2. Factor analysis results of various etiologies of gait deficit

Gait Deficit Due To Limb Transplantation					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Swing	SL Var	StrideLength	Absolute Paw Angle	Swing	Brake
Propel	Stride Length CV	Belt Speed	Midline Distance	%SwingStride	Propel
Stride Frequency					%PropelStance
Gait Deficit Due To Nerve Transection					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Propel	Brake	SL Var	Stride Length	Swing	Paw Area at Peak Stance in sq. cm
Stride Frequency	%Propel Stance	Stride Length	Belt Speed	%SwingStride	Paw Area Variability at Peak Stance in sq. cm
					Min ΔA/dT
Latent factors that describe gait in the setting of peripheral nerve injury					
Relationship of stride length and speed	Stride length and its variation		Measures of the paw	Measures of symmetry	Phases of the stride
Gait Deficit Due To IUGR					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Propel	Stride Length	Paw Area at Peak Stance in sq. cm	SL Var	Brake	Swing
%Propel Stance	Belt Speed	Min ΔA/dT	Stride Length CV		%SwingStride
Gait Deficit Due To Hyperoxia					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Propel	Paw Area at Peak Stance in sq. cm	%Propel Stance	SL Var	Swing	Stride Length
Stride Frequency	Paw Area Variability at Peak Stance in sq. cm	Absolute Paw Angle	Stride Length CV	%SwingStride	Belt Speed
Belt Speed					
Gait Deficit Due To Both IUGR + Hyperoxia					
Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Propel	Paw Area at Peak Stance in sq. cm	Brake	Stride Length	Swing	Swing
%Propel Stance	Paw Area Variability at Peak Stance in sq. cm	%Propel Stance	Belt Speed	Stride Length CV	%SwingStride
Stride Frequency	Min ΔA/dT				
Latent factors that describe gait in the setting of central nerve injury					
Relationship of stride length and speed	Stride length and its variation		Measures of the paw		Phases of the stride

Notice key similarities (underlined) and differences (**bolded**) between the two datasets. The bottom two rows describe the latent or implied factors that make sense of the various groups of observable measures shown above. These latent factors are an interpretation of the factor groupings shown above. The factor groupings were identified via factor analysis, which quantifies the relationship between individual, measurable features and can be used to identify the most highly grouped features. Table reproduced from the primary and corresponding authors' prior published study and with their permission (33).

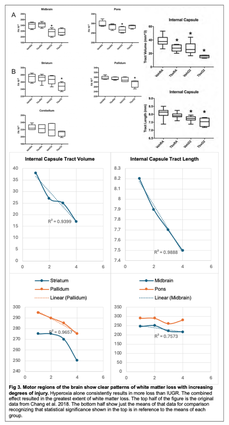

Fig 3. Motor regions of the brain show clear patterns of white matter loss with increasing degrees of injury. Hyperoxia alone consistently results in more loss than IUGR. The combined effect resulted in the greatest extent of white matter loss. The top half of the figure is the original data from Chang et al. 2018. The bottom half show just the means of that data for comparison recognizing that statistical significance shown in the top is in reference to the means of each group.

Table 3. Evaluating the performance of 4 different model architectures in distinguishing between healthy and pathological phenotypes of gait

A. Performance in Distinguishing Peripheral Gait Deficit from Healthy Gait	Accuracy	Precision	Recall	F-Score
Random Forest	0.7294	0.7560	0.7521	0.7492
Discriminant Analysis	0.7477	0.8022	0.7634	0.7742
Support Vector Machine	0.7744	0.8108	0.7915	0.7948
Regression	0.7868	0.8570	0.7826	0.8130
Ensemble	0.9099	0.9283	0.9086	0.9165
B. Performance in Distinguishing Between Two Phenotypes of Peripheral Gait Deficit: Limb Transplant from only Nerve Transection	**Accuracy**	**Precision**	**Recall**	**F-Score**
Random Forest	0.6435	0.7072	0.6832	0.6878
Discriminant Analysis	0.7165	0.7827	0.7188	0.7388
Support Vector Machine	0.6987	0.7806	0.7341	0.7457
Regression	0.7237	0.7882	0.7270	0.7456
Ensemble	0.8780	0.9263	0.8781	0.8984
C. Performance of an Ensemble Model in Distinguishing Between Various Pairwise Comparisons of Central Gait Deficit	**Accuracy**	**Precision**	**Recall**	**F-Score**
All Central vs. Control	1.0000	1.0000	1.0000	1.0000
IUGR vs. Control	0.9848	0.9000	1.0000	0.9474
Hyperoxia vs. Control	0.9783	0.9444	1.0000	0.9714
IUGR + Hyperoxia vs. Control	0.9726	0.9824	1.0000	0.9375
Hyperoxia vs. IUGR + Hyperoxia	0.7736	0.5882	0.6667	0.6250
IUGR vs. IUGR + Hyperoxia	0.9615	0.9412	1.0000	0.9697
Hyperoxia vs. IUGR	0.8667	0.9722	0.8750	0.9211
D. Performance of an Ensemble Model in Distinguishing Between Additional Pairwise Comparisons	**Accuracy**	**Precision**	**Recall**	**F-Score**
Younger vs. Older	0.7143	0.7273	0.6667	0.6957
Random Control vs. Random Control	0.5098	0.4643	0.5652	0.5098

Using the identified features from feature selection + factor analysis 4 different model architectures were evaluated for their accuracy, precision, recall, and F-score in their ability to distinguish (**A**) healthy gait from gait deficit due to peripheral nerve injury and (**B**) gait deficit due limb transplantation from gait deficit due to total nerve transection alone. Then using the Ensemble model, (**C**) various additional pairwise comparisons of Central Gait Deficit were evaluated and (**D**) additional comparisons of gait deviation due to age or between randomly selected controls. Notice how a more moderate deviation due to aging is observed in **D** as compared to the performance of distinguishing between disease and injury states (**A-C**). Table partially reproduced from the primary and corresponding authors' prior published study and with their permission (33).

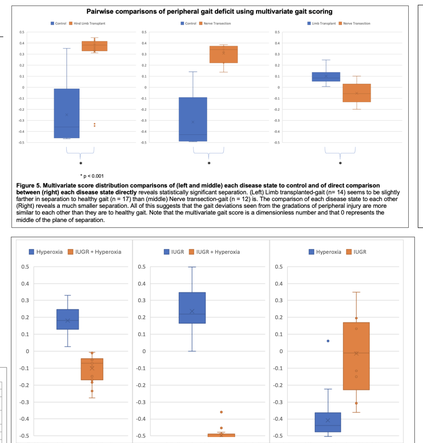

Figure 5. Multivariate score distribution comparisons of (left and middle) each disease state to control and of direct comparison between (right) each disease state directly reveals statistically significant separation. (Left) Limb transplant-gait (n= 14) seems to be slightly farther in separation to healthy gait (n = 17) than (middle) Nerve transection-gait (n = 12) is. The comparison of each disease state to each other (Right) reveals a much smaller separation. All of this suggests that the gait deviations seen from the gradations of peripheral injury are more similar to each other than they are to healthy gait. Note that the multivariate gait score is a dimensionless number and that 0 represents the middle of the plane of separation.

Fig 7. Multivariate score distribution comparisons of (left and middle) each central disease state and the combined and (right) each disease state directly reveals statistically significant separation. (Left) Hyperoxia-gait seems to be closer in separation to the combined score than (middle) IUGR-gait is to the combined. The comparison of each disease state to each other (Right) reveals a shift of the plane of separation strongly towards hyperoxia-gait suggesting its gait had a stronger "pull" on the multivariate scores. All of this suggests that the gait deviations seen in the combined, IUGR + hyperoxia-gait, state are influenced more heavily by hyperoxia than by IUGR.

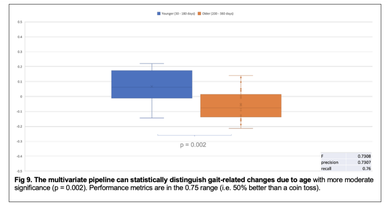

Fig 9. The multivariate pipeline can statistically distinguish gait-related changes due to age with more moderate significance (p = 0.002). Performance metrics are in the 0.75 range (i.e. 50% better than a coin toss).

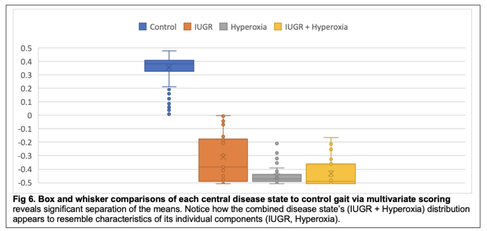

Fig 6. Box and whisker comparisons of each central disease state to control gait via multivariate scoring reveals significant separation of the means. Notice how the combined disease state's (IUGR + Hyperoxia) distribution appears to resemble characteristics of its individual components (IUGR, Hyperoxia).

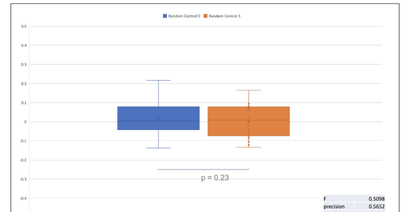

Fig 8. Randomly chosen control animals have statistically indistinguishable gait with performance metrics in the 0.50 range (a coin toss).

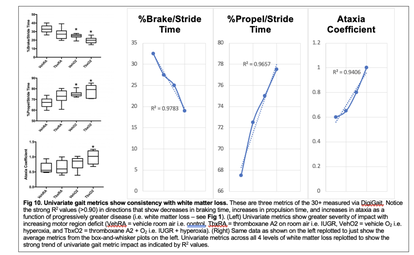

Fig 10. Univariate gait metrics show consistency with white matter loss. These are three metrics of the 30+ measured via DigiGait. Notice the strong R^2 values (>0.90) in directions that show decreases in braking time, increases in propulsion time, and increases in ataxia as a function of progressively greater disease (i.e. white matter loss – see **Fig 1**). (Left) Univariate metrics show greater severity of impact with increasing motor region deficit (VehRA = vehicle room air i.e. control, ThaRA = thromboxane A2 on room air i.e. IUGR, VehO2 = vehicle O_2 i.e. hyperoxia, and ThaO2 = thromboxane A2 + O_2 i.e. IUGR + hyperoxia). (Right) Same data as shown on the left replotted to just show the average metrics from the box-and-whisker plots on the left. Univariate metrics across all 4 levels of white matter loss replotted to show the strong trend of univariate gait metric impact as indicated by R^2 values.

Results, Discussion, and Conclusion

- 16 features maximized predictive ability vs. 32 total **(Fig 2).**
- Univariate selection performed 6% less accurately than multivariate selection (**Fig 2**).
- Factor analysis allowed us to understand exact relationships and tie them to consistent biological phenomena **(Table 2).**
- Using the identified features, various models were trained. Ensemble-based classifiers achieved >90% classification accuracy with similarly high precision, recall, and F-score **(Table 3).**
- Moreover, these classifiers were able to distinguish between the varying etiologies of gait with almost 90% accuracy and statistical significance **(Table 3, Fig 5-9).**
- Plotting multivariate gait score distributions revealed statistically significant score separation between types of peripheral nerve injury, central nerve injury, and even in the course of normal aging **(Fig 5-9).**
- Changes in uni- and multivariate gait scores tracked with degree of white matter loss **(Fig 3, 10, and 11).**

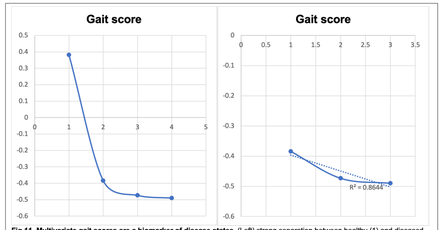

Fig 11. Multivariate gait scores are a biomarker of disease states. (Left) strong separation between healthy (1) and diseased (2-4) states. (Right) physiologically consistent separation among disease states. Hyperoxia (2) and the combined state (3) are closer to each other. IUGR on its own has less severe impact. This same trend was seen in the white matter volume data as well as in the univariate gait metrics published by Chang et al. 2018.