

Reviewer 3

Improving the welfare and QoL of dogs in shelters and laboratories is indeed important. Thus the paper is worth publishing

However the design in my opinion has some shortcomings which at least need to be discussed insofar as cannot be improved, or in cases where improvement is possible should be done:

Reviewer comments	Author responses
a) the behavioural categories are very coarse: "bark": which type of barking?? see Taylor/Reby, Morton etc regarding motivation behind vocalizations in mammals; locomotion : walk, trot, or gallop; sniffing : which nostril (Siniscalchi re motivation); grooming: autochthonous or displacement eg in scratching...	The development of a suitable random forest machine learning algorithm is a balance between the number of behavioural classification and model accuracy. For example when locomotion was classified as walking, trotting and galloping, the accuracy of the model declined see Redmond et al 2024. While interesting, these behavioural distinctions are too fine to be modelled accurately.
b) saliva cortisol would have been another additional measurementparameter	Yes it would have been interesting however the budget did not allow for this.
statistics: please use boxplots, medians and quartiles instead; why do you use ordinal statistics ((K-W, Wilcoxon) for metric data?? randomization test, especially for small samples, would be stronger; Why only FIVE instead of at least SIX days iof test duration (for nonparametric statistics n => 6 is considered minimum; you use data sets for repeated testuing so you need a correction e.g. Bonferroni?!; please give exact p-values instead of < 0,.005 etc wherever possible	The Bonferroni correction was embedded in the code for repeated measures ANOVA. This has been added to the methods. Box plots were considered but we felt these would make the effects harder to see. The Kruskal-Wallis test was used as the metric data was not normally distributed and therefore violated the assumptions of the parametric test such as ANOVA and GLM. Furthermore, the data could not be normalised using the transformTukey function in R.
c) some improvements for content: Please give at least some data on statistics (tests, p-values) in Abstract	The requirement of the journal is that the abstract is a maximum of 200 words which makes including results problematic. We chose not to include p-values to ensure that the abstract was as informative as possible.
please explain what the "forest test" is - this paper may be very interesting to practinioners who are not so familiar with modern statistics	Assuming you are referring to the random forest algorithm, we referred to this information in our model validation paper. This paper is built on previous work that has already been described.
did you change the ropes everyday, or only for each new pair of dogs??	Rope toys were replaced for each new pair of dogs. This has been clarified in the methods.
I 259: you cannot determine whether "hour of day" has an influence if you start same hour each day. Just that a decrease over time occurs	This has been changed to change over time rather than hour.
I 312: the verb is missing	It is not clear where this error is
I 507 ff: no content given	Oops now fixed