

Dear Reviewers,

We graciously thank you for your feedback and criticisms. We have carefully taken your comments into consideration and would like to respond to some of the major concerns.

**Reviewer’s Comment:** “The method is explained clearly and tested on both synthetic data and a large number of data sets. But it seems to me that the novelty is too small.”

**Response:** A major point of this work is to emphasize that it is possible to construct a decision forest with the same ease-of-use and time/space complexity as random forest that simultaneously empirically outperforms Breiman’s random forest (RF) on a variety of synthetic and real-world datasets. While various oblique decision forests have been proposed over the past decade and a half, RF still remains the most widely used. We speculate that part of this is due to its relative simplicity compared with other learning algorithms. This is supported by the fact that while even Breiman suggests that his oblique Forest-RC demonstrates superior performance to axis-aligned RF when tuned properly, it remains relatively unused as compared to its axis-aligned counterpart. A second reason RF is still the most widely used decision forest algorithm is due to its superior empirical performance, as demonstrated by two recent benchmark studies (Delgado2014; Caruana2008). In both studies, which evaluated a variety of axis-aligned and oblique decision forest methods, RF was found to be the best overall classifier. In our work, we used the same benchmark datasets as in (Delgado2008), even preprocessed in the same exact way, and demonstrate that RerF outperforms RF. Although our proposed method is relatively simple, we believe our findings to be significant to the scientific community. Furthermore, we believe that the simplicity in our oblique method makes it amenable to theoretical analysis, as has been done recently for RF (Scornet2015).

**Reviewer’s Comment:** “So, as admitted by the authors in several parts of the paper, the current proposal is highly similar to Breimans Forest-RC. More specifically, the only difference is that Forest-RC sets the number of nonzero in each column of  $A$  to be fixed and set by a parameter, while the current proposal randomly sets this number. Now this seems a very narrow difference, i.e. it renders the current proposal as something very incremental compared to Forest-RC. To further aggravate the problem of this incrementality Forest-RC is not included in the experimental comparisons. This means in the end we have a new system that is marginally different from Forest-RC and we do not know if the (small) differences bring any advantage at all This seems too short to me for a conference like ICML.”

**Response:** We are in the process of comparing our proposed method to Forest-RC. Preliminary results suggest that they achieve comparable classification performance, with RerF requiring less training time due to one less hyperparameter

that has to be tuned. We will provide a thorough and fair comparison of the two in the final version.

**Reviewer’s Comment:** “The paper is fairly lacking in a proper introduction of random projection. Merely leaving a pointer to Li et al., 2006 without even briefly reviewing its basic ideas and properties leaves the paper far from being self-contained.”

**Response:** We agree. We will provide a more thorough review of the principle and properties of random projections, including the Johnson-Lindenstrauss lemma and applications to dimensionality reduction.

**Reviewer’s Comment:** “The way random projection is used in this paper is fairly straightforward. And without any analysis, whether theoretical or empirical, about its effect on the bias/variance of the learnt random forest, the blending of two seems almost ad-hoc and not well grounded.”

**Response:** We are currently exploring the effect of random projections on the bias-variance tradeoff of the decision forest. Intuitively, random projections, as opposed to variable subsampling, likely reduces the bias by relaxing the constraint of axis-aligned splits. Furthermore, theoretical results by (Louppe2015) suggest that random projections decrease variance of the ensemble by reducing correlation among trees. A more rigorous treatment will be included in the final version.