We thank all 3 reviewers for their thoughtful comments.

Reviewer 1: "*nearest neighbor theory papers have largely not worried too much about constants......This analysis is neat albeit limited in scope. I suspect that this work will be of interest mostly to researchers working on theory for nonparametric methods.*" In the evolution of the study of nearest neighbor, early work focused on consistency, and later work goes beyond consistency and focuses on rate of convergence. The logical next step of theoretical interest would be on the constant. You are absolutely correct that very few work studies the constant. We argue that this is "a feature, not a bug". The seemingly relative unpopularity of this type of analysis may be due to its technical challenge and depth of the analysis.

"*The scope of the analysis is very limited to distributed nearest neighbor classification (along with some distributional assumptions that the authors point out in the discussion section to be a bit restrictive), and it's a bit unclear whether insights here generalize to other methods.*" Our analysis generalizes to other ensemble methods as well as data-interpolation weighted methods. The latter is a fairly interesting direction, due to its connection with deep learning. We leave these explorations as future works.

"*Currently the paper has lots of small typos. Please proofread carefully and revise..*" Thanks for pointing out, and we will fix the typos in the final version, if accepted.

"*Also, I find Table 1 ... How is the risk percentage defined in comparison to the oracle KNN/OWNN? Additive?*" The risk in the table are empirical classification error; the risks of our proposed methods and the oracle KNN/OWNN are calculated separately so that one can compare the numerical values in the table directly.

"*I'd suggest adding error bars to Table 1 (for example, to denote standard deviations across experimental repeats). Also, the multi-cohort medical study example could be good to mention in broader impacts...* " Thank you for the suggestions. We will add them in the final version, if accepted.

Reviewer 2: "*The derivations are done for only two classes (for binary classification) which significantly degrades the importance of these findings and limits its usage in real-world problems.*" The proofs to our main results are very convoluted even for the binary case. In order not to distract from the main message, we choose to state the results in terms of the binary classification setting. All the results can be naturally generalized to the multi-class setting. We stress that we are not alone in this choice. Almost all the major theoretical work on nearest neighbor in the last 10 years focus on binary classification. For example, [46], [11], and [25] cited in the paper.

"*nearest neighbor classification is no longer popular as before as there are numerous good alternatives.... In one year, it ([45]) is cited only once. This is quite natural in my opinion since deep neural networks dominated classification....*" Due to the good interpretability and relative low time complexity, nearest neighbor methods are still of high interest among the practitioners and the nonparametric community. It is true that deep neural networks dominated the classification literature but it does not mean that there is no room for other important and widely used methods. In particular, the sheer volume of works on deep neural networks may be due to the fact that they are relatively new and not very well understood, especially their statistical guarantees. In contrast, many aspects of nearest neighbor have been studied and the rest are really difficult to analyze. The latter is the gap we try to fill. Lastly, we stress that deep and insightful theoretical work in nearest neighbor, such as [46] and [11], are still highly cited (for 234 times and 85 times respectively.)

"*there are good alternatives to the nearest neighbor classification for large-scale data as hashing, approximate nearest neighbor classification methods, etc.*" All these acceleration approaches provide approximate (not exact) nearest neighbor classification. However, there is no statistical guarantees in terms of their learning performance. In practice, distributed learning can be combined with such approaches to further speed up the learning process.

"*Lastly, advantages of using weighted voting scheme instead of majority voting must be clear since it is already well studied in general classification combination. It is also very intuitive ...*" These may be intuitive, but to the best of our knowledge, we are the first to rigorously prove them and quantify the multiplicative constant. We also give the exact loss on accuracy due to the choice of the weighting scheme. This finding is subtle, but important.

"*the authors mention divergence of s, which is the number of subsamples. It is not a sequence, thus I do not understand what the authors meant with this?.* " s increases as N (size of the whole dataset) increases. One of our main contributions includes proving that the sharp upper bounds of s (number of subsamples) are $s \asymp N^{2/(d+4)}$ and $s \asymp N^{4/(d+4)}$ for M-DiNN and W-DiNN, respectively. In this sense, the s can be considered as a sequence $s(N) \to \infty$ as $N \to \infty$.

"*But, it will be nice to write a paragraph summarizing the basic differences with respect to the papers given in [45] and [46].* " In the introduction section, the 3 paragraphs started from line 52 have summarized our contribution compared to [45] and [46]. We will re-summarize them in a separate paragraph, in the final version, if accepted.

Reviewer 3: We thank you for your very positive comments.