# Referee report

## General quality

**Scientific significance: Excellent**

- The goal of the paper (to see whether it is possible to predict problems with the AFlaps using readily available SCADA data and simple ML models) seems to me to be relevant. SCADA data is readily available, which means it can be used at no additional cost. This makes it especially relevant for the industry. Furthermore, the fact that this research focusses on simple ML models first is useful because it gives an idea of what the lower-bound is of the modelling complexity.

**Scientific quality: Good**

- The research quality is good. The methodology seems to be sound. I could not find any clear methodological flaws or shortcomings. The methodology is well described in the manuscript. However, the manuscript might benefit from a schematic overview of the methodology. The results are well presented. Both the positives and negatives of the results are pointed out by the authors, which is a good thing because it gives a better understanding of the performance of the methodology. The literature study is however limited. Only a small number of papers are mentioned even though condition monitoring of wind turbines is currently a hot research topic. More recent papers that give an overview of techniques that are used for condition monitoring of wind turbines exist.

**Presentation quality: Good**

- The figures are clear. The result tables can, due to the usage of abbreviations for the different cases or scenarios, be somewhat difficult to understand at a glance. This is to a certain extent solved by table 2. The text however does contain multiple typo's and sentence errors.

## Suggestions for revision

1. The addition of a schematic overview of the methodology to the manuscript. This will make it more clear for the reader what the different steps of the methodology are.
2. Expansion of the literature study on page 2 with more recent (overview) papers on condition monitoring of wind turbines.
3. Check the text to make sure that all abbreviations are at their first appearance preceded by their meaning. See for example the abstract.
4. Check the text on typo's and sentence errors.
5. Line 250: … k the ridge coefficients to be minimized. Where in Equation 7 is k? Please clarify in the manuscript.
6. Give a more thorough/in-depth explanation why some methodological decisions were taken: feature generation techniques, feature selection techniques, used models. For example why was MiniRocket selected and not a different feature engineering method? Why was the random forest selected instead of other simple models like SVM, ...? Please clarify in the manuscript.
7. Typo in line 367: …, and active with fault (AF_Off_Fault) -> should it not be AF_On_Fault? Please clarify in the manuscript.
8. It might be useful to add a table with abbreviations to the paper so that it is easier to look up the meaning.