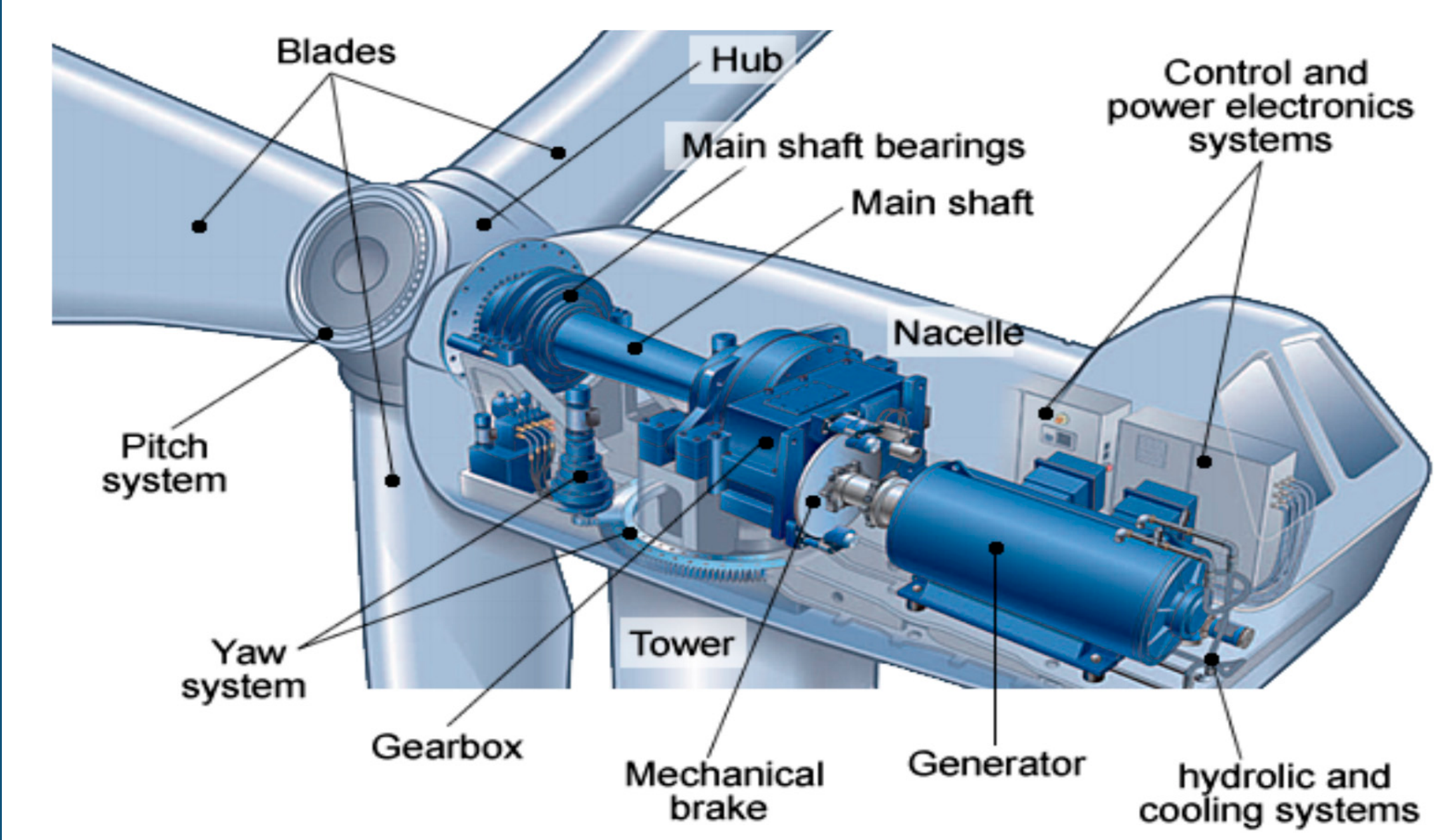# Machine Learning Techniques Used for Wind Turbine Monitoring

Yohann Jacob Sandvik

Institute of Electronic Systems - NTNU

The Wind Turbine

Objectives

Simplified a wind turbine works by wind pushing the blades, generating torque that makes the hub rotate. The hub is connected to the gearbox through the main shaft. The gearbox then gears down the torque and gears up the rotational speed to a level that the generator can use to induce current, that goes to a station that transforms the voltage to a level that can be used in the electrical grid.

Information about a wind turbine can come from many sources, it can come from external sources such as images from a camera, or from internal sensors measuring operational data. The collective term for systems measuring operational data is supervisory control and data acquisition (SCADA) systems.

Table

Machine learning model	Articles
ANNs	[31, 33, 35–40]
GP regression	[28, 29]
SVM	[27, 30, 32]
PCA	[20]
KNN	[27]
RF regression	[27]
Network of RBMs	[4, 34]

Regression Based Models

- Abbreviations: Artificial Neural Networks (ANN), Support Vector Machine (SVM), Gaussian Process (GP), K-Nearest Neighbours (KNN), Principle Component Analysis (PCA), Random Forest (RF) and Restricted Boltzman Machine (RBM).
- A frequent approach is to use a supervised learning model to capture the normal behaviour of a wind turbine, or wind turbine component, by predicting the value of one time dependent variable, or by recreating the input variables.
- Then the deviation between the prediction and the actual value is used to detect anomalous behaviour.
- The authors then set a fixed threshold, or a variable threshold to which behaviour is considered anomolous.

Table

Machine learning model	Articles
ANN based	[7, 8, 10, 41–43]
Tests several individual classifiers	[6, 13, 15, 16, 44]
Uses fusion / ensamble of classifiers	[9, 14, 45]
SVM	[17, 18, 22, 23, 46, 47]
DBN	[48]
Hidden Markov model (HMM)	[49]

Supervised Classification Based Models

- Abbreviations: Deep Belief Network (DBN).
- For supervised classifiers the models required labeled datasets to train.
- Many of the supervised models use images to as input to detect damage in the blades.
- Many of the remaining models use vibrational sensors to detect structural damage in the blades as well.

Table

Machine learning model	Articles
RBMs	[51]
K-means clustering	[21]
One-class SVM (OCSVM)	[11]
Multiway PCA	[19]

Unsupervised Classification Based Models

- As evident from the tables, the unsupervised classification models are fairly rare compared to the supervised classification models.
- There was found one implementation of clustering for condition monitoring of wind turbines, by Zhang and Ma [21]. They use a generalization of PCA called parallell factor analysis (PARAFAC) for dimensionality reduction and K-means clustering on the reduced feature space. They are able to identify distinct operation modes of the wind turbines, and reduce data redundancy by PARAFAC.
- However, they mention that there is still further work that can be done, especially in testing their model with data from turbines operating at different wind speeds.
- Pozo, Vidal, and Salgado [19] also have an interesting approach. They use multiway PCA to set up a baseline model, and then uses hypothesis testing to determine whether the power curve of a wind turbine is an anomaly or not.

Discussion

- As can be expected to acheive high accuracy the amount of training data required, and the complexity of the model increases. Additionally, there is a trade-off between complexity and interpretability.
- The supervised classification based models are the most popular choice among the articles included, and show great promise in terms of accuracy, however they are quite hard to interpret. In addition, they require labelled datasets and the dataset available is not labelled so they are not considered relevant.
- The regression based models do not require explicit labeling to model normal behaviour, so they could be used to detect some anomalous behaviour. What is problematic with using complex regression based models is that without labelled data it might be hard to validate their performance.
- It is promising to see that K-means clustering paired with PARAFAC showed such good results.