Deep Neural Network Hard Parameter Multi-Task Learning for Condition Monitoring of an Offshore Wind Turbine

Abstract

Breaking the curse of small datasets in machine learning is but one of the major challenges that cause several real-life prediction problems. In offshore wind application, for instance, this issue presents when monitoring an asset in an attempt to reduce its infant mortality failures. Another challenge could emerge when reducing the number of sensors installed in order to limit the investment in monitoring systems. To tackle these issues, the aim of this article is to investigate the impact of small data-set on conventional machine learning methods, and to outline the improvement achievable by the implementation of transfer learning approach. It provides a solution to mitigate this issue by applying a hard parameter multi-task learning approach to the supervisory control and data acquisition data from an operational wind turbine, allowing for smaller datasets to efficiently predict the status of the gearbox's vibration data. Two experiments are carried out in this paper. The first is to envisage the possibility of using hard parameter transfer on the operational data from two wind turbines. The second is to compare the results of this model to the findings from a conventional deep neural network model trained on the data from a single turbine.

Innes Murdo Black[1],Debora Cevasco[1,2], Athanasios Kolios[1].

[1]University of Strathclyde, 16 Richmond St Glasgow G1 1XQ, Glasgow, Scotland

[2]Ramboll, Jürgen-Töpfer-Straße 48, 22763, Hamburg, Germany.

Author contact email: innes.black@strath.ac.uk

Keywords: Deep Neural Network, Hard Parameter Transfer, Multi-Task Learning, Condition Monitoring, Offshore Wind Turbine, Supervisory Control and Data Acquisition (SCADA),Condition Monitoring Systems (CMS), Machiene Learning, Regression, Classification

1 Introduction

Consistency is a great way of de-risking offshore wind, but unfortunately, some failures are inherent. As the popularity of wind turbines increases globally so do the varying effects which make the failures of the structure more difficult to predict. Furthermore, new technologies and optimised designs can suddenly fail due to quality or stress related failures, respectively. This type of failure is termed infant mortality failure, causing a potential significant loss in revenue, especially if employed in the offshore wind sector.

Anomalies detection and failure modes diagnosis can be utilised to identify the structural health of offshore wind turbine components, by using intelligent computing such as artificial neural networks. Richmond et. al [10] performs a stochastic assessment of the aerodynamics of an offshore wind turbine using an artificial neural network among other machine methods to determine wind speeds and directions. By extending this work, one could make assumptions on the fatigue life of an offshore wind turbine. Bao et al. [1] utilise a one-dimensional convolutional neural network to determine the occurrence of damage to the support structure of an offshore wind turbine. In this example, the examination looks at localised damage to a jacket support structure under regular waves with an incredibly high accuracy of 98.4%.

Artificial neural networks have been used to examine the 'life percentage' of an offshore wind turbine based on the failure time distributions. Yang et al. [15] apply a two-level failure probability procedure for the gearbox, the rotor, the pitch mechanism and the generator. This simplified method has shown that condition-based maintenance schemes can reduce the cost of preservation compared to classic time-based maintenance, where regular intervals are used to assess the asset. A breakdown of convectional models suited to wind turbine maintenance is discussed in [2] . The components of the drive-train have the potential highest impact on the maintenance cost of the next generation of offshore wind farms [3]. These are among the most expensive components for an offshore wind turbine, and are continuously undergoing remodelling where innovations are to accommodate bigger power outputs [4] with larger loads. Condition Monitoring (CM) signals, in combination with high-frequency Supervisory Control and Data Acquisition (SCADA) data have been extensively used to train machine learning (ML) models to predict failures in the drive-train components [11] [12]. In [11], Stetco et al. document the state-of-the-art ML methods and processes for the wind turbine condition monitoring. Tautz-Weinert and Watson [12] examine and discuss the effectiveness of the numerous ML methods based on the type and amount of data available. However, the information available from the literature mainly focuses on training the algorithms based on the availability of relatively big sets of data – generally for at least more than three years. The authors, thus, identified a knowledge gap in the field of offshore wind applications regarding the investigation of a detection algorithm suitable for small datasets.

1.1 Conventional Machine Learning Versus Multi-Task Learning

Conventional ML typically involves optimising for a particular task $T = (y, f(x))$, where y is the output feature domain and $f(x)$ is the predicative function made up of X feature data. The model is trained for a single task, this generally may achieve an acceptable performance for a single domain $D = (x, p(x))$ of marginal probability distrabution $p(x)$, but by focusing on one signal task we ignore information that may help us do better on other metrics relating to that task. By sharing representations of a global task trained on the source domain D_s and target domain D_t, with a similar probability distribution, we may be-able to better represent our general task. This is Multi-Task Learning.

A benefit of multi-task learning is that knowledge can be transferred. Inductive transfer learning is a sub category of transfer learning. Where an accurate model is usually trained on the source domain D_s to determine the hypothesis space, this article implements the hard-parameter transfer technique of inductive transfer. The hypothesis space generated by hard parameter transfer can help improve the target task results. Particularly, if there are small amounts of data or class labels for the target task. Performing inductive transfer learning using source domain data to train the general model and applying the target task data for fine tuning can lead to a more accurate model[9].

The multi-task learning (MTL) theory has been employed across many fields of application requiring a supervised prediction of one or more classes; two-stage facial recognition [14], quality assessment of fetal head ultrasonic assessment [8], or for bandwidth allocation for multiple mobile users [6]. The MTL models come in a plethora of forms: joint learning, learning to learn, and learning with an auxiliary task, are among some of the names that have been allocated to its predictive assignment. To generalise the need for its application, it can be stated that a MTL approach is worth being investigated as soon as the problem requires optimising for more than one task.

1.2 Problem statement and scope of the analysis

The availability data from offshore wind turbines is limited, with CMS data on the generators is less accessible. When wind turbines are installed this restricts the amount of data available. This is a sensitive period where novel failure events are more likely to happen in the early stages of the structure (infant mortality failures). Additionally, low-cost monitoring campaigns might be preferred to reduce the cost related to the hardware and the storage of data, limiting the amount of assets with SCADA or CMS systems. To address this issue, this work aims to apply a data-driven multi-task learning approach to monitoring the health status of an offshore wind turbine gearbox.

The purpose is to demonstrate how a hard parameter transfer model can achieve greater results than a conventional machine learning model when applied to a limited amount of training data. Therefore, the scope of this paper is to make a step towards understanding of the setup of a suitable monitoring algorithm based on CMS with a small set of data of an offshore wind turbine gearbox.

The remainder of this article provides a discussion of the literature and the theoretical basis the of MTL method in Section 1. In Section 2, the methodology of this study is introduced, with details on the methods applied for the data pre-processing, the model training, and the evaluation metrics used to determine the effectiveness of the detection. Section 3 includes the results highlighting the main findings and showing the clear context for which MTL is effective, to finally closes with a discussion and conclusion.

2 Methodology

This section introduces the methodology for training and comparing models predicting, in binary form, the status of an offshore wind turbine gearbox. The data collected, pre-processing, and their division into data-sets for the training of a conventional and a MTL algorithm are introduced. The overall workflow for the training of the models is described in detail. Finally, the metrics used for the evaluation of the models are defined.

2.1 Data Collection

The analysis presented in this paper is built on time-series data from one MW offshore wind turbines, in normal operation. The signals from the SCADA and condition monitoring (CM) systems consist of eight monitoring channels, recorded with a ten-minute resolution. These channels include meteorological information, the operational data of the wind turbine, and the vibration data from the gearbox, with the associated flag warning raised in case of anomalies. This latter provides the (binary) label targeted in the training of the classification models.

There are two data-sets, the source domain data contains 31804 time-steps (220 operational days) is collected from a turbine. From another turbine of the same population, considered homogeneous. The target domain data, has a reduced length of 8141 time-steps These two data-sets are, thus, representative of an existing wind turbine and a newly installed turbine. Both wind turbines have the same form and hence similar distributions in each of the identical features. The source domain data has alarm signatures for total of 19% of the data set and the target domain data has a similar a portion with 18%.

2.2 Data Pre-processing

The data-sets used for the training and testing of a machine learning algorithm generally require some pre-processing to ensure satisfactory performances of the prediction model. The application in this paper consists of a two-step procedure. First, a data cleaning process is performed for removing outliers and missing values. For the treatment of missing values, time instances of the database are removed from the analysis if over 50% of the data is missing. The outlier removal is processed by removing vector instances where the values should be scalar. For the remainder of the data, a K-nearest neighbours (K-NN) imputation method is applied [13].

The last step is to split the data into training, and testing sets. In the specific of this experiment, 80% of the data is employed for the training of the models, while the remaining 20% is used for the final testing.

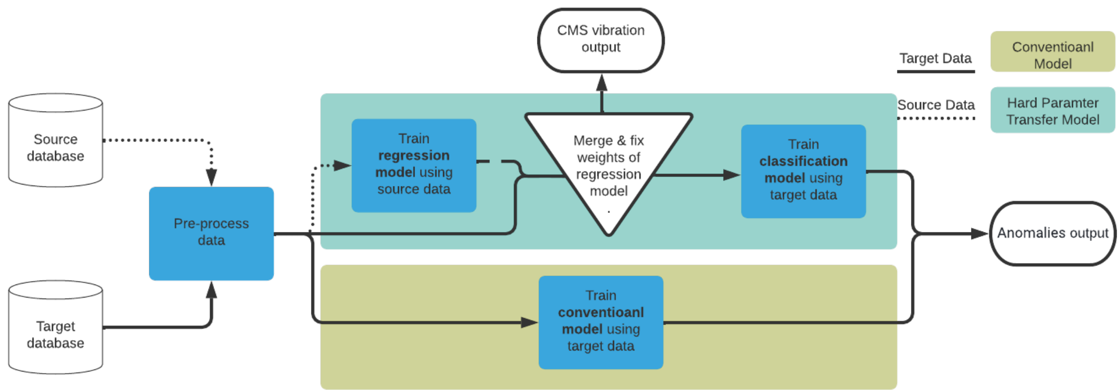


Figure 1: Flowchart detailing the data flow and construction of the conventional model on the left and the multi-task learning model on the right.

2.3 Models

The MTL model is built up in two training stages, with information being leveraged from two data-sets. The knowledge acquired from the source domain data D_s with a greater number of samples is used to train, a feature extractor (regression model). Taking the SCADA data, and making predictions on the vibration data from the gearbox. A classifier is connected to this first model, and uses a hard parameter transfer to merge the last neurons from the artificial neural network to the first layer of the convolutional neural network (classification model). The weights of the neurons of the regression model are fixed, and the remaining weights of the model are trained using the small target domain D_t data-set.

The conventional model takes the same architectural form as the regression and the convolutional neural network classifier together, but it is just trained by only using the target domain data D_t, smaller data-set. Therefore, the model receives the SCADA data as the input, and it outputs the anomalous behaviour of the gearbox. Both models have the same desired task but one model uses hard-parameter transfer to transfer knowledge and increase the amount of knowledge representing the anomalous behaviour, since it has two outputs the CMS vibration predictions and the binary error message.

2.3.1 Training of the models

Three distinct models are trained for the purpose of this paper. One is the feature extraction model, another is the classification model used for the MTL procedure, and the last is a conventional classification model.

The transfer learning model is built up of three blocks, as it can be observed in Figure 1. The feature extractor; this takes in input from the meteorological data, the wind turbine operational data, and the gear oil temperature, and then outputs the gearbox vibration features. This model is built up with a deep neural netowrk (DNN) architecture. In particular, it consists of 14 sequential layers all implementing a rectified linear unit (RELu) activation function. Utilising a uniform variance scaling [5] allows the neural network to train extremely deep rectified models directly from scratch. The optimiser for the regression feature extraction is called 'Adam', which is derived from the adaptive moment estimation [7]. This is effective for noisy, nonlinear data.

Hard parameter MTL is carried out using the classification model highlighted in Figure 1. This convolutional neural network (CNN) architecture is exactly the same as the classifier used for the conventional model. Both classifiers feature one convolutional layers of width 64, and similarly utilise the RELu activation function. This is followed by a drop out of 0.5 to further four layers of a one-dimensional convolution, which implements the sigmoid function - commonly used in classification models. The kernel initialisation of the weights for the CNN uses uniform variance scaling. A standard gradient descent with Nesterov Momentum is employed to improve the accuracy while dealing with noisy data from the vibration signals, with a learning rate of 0.1 and a momentum of 0.9. Lastly, the cross binary entropy loss function is implemented to distinguish the gearbox status class.

For the consistency of the comparison, the conventional model takes the overall same architecture of both the feature extraction and the classification model combined. The main differences from the conventional model to the hard parameter transfer model is:

- The binary cross entropy optimiser is applied to the whole model.
- The entire model is trained in one process and one data-set.
- The model only has one output stream of information representing the gearbox status.

2.4 Evaluation metrics

The regression results are evaluated using the mean absolute percentage error $MAPE$, which represents the average absolute percentage error for each time period minus the actual values divided by actual values.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{\hat{y}_i - y_i}{y_i}| \tag{1}$$

where $\hat{y}_i$ is the forecasted value, y_i is the actual value and n is the number of samples. The metrics of the classification models are calculated using $F1_{score}$ and the $Accuracy$. The $F1_{score}$ metric conveys the balance between the precision, the true positive predictive value, the recall, and the true positive rate, by calculating their harmonic mean. The accuracy represents the percentage the model correctly calculates.

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \tag{2}$$

$$F1_{score} = \frac{t_p}{t_p + 0.5(f_p + f_n)} \tag{3}$$

The true positive t_p, the false positive f_p and the false negative f_n are used to calculate the F1-score and are going to be explicitly reported to judge the quality of the classification. W Yang et. al. [15] conducted as study on wind turbine monitoring and indicate that predicting more than 60% of a wind turbine faults will reduce the cost of operations and maintenance. This has been implemented as a threshold for our models.

3 Results

The correlation plot in Figure 2 highlights the Pearson's correlation of the variables to one another. The set of features proves the potential advantages of the hard perimeter transfer learning. In machine learning, the higher the correlation to the data this increases the chances of the the predicative function describing the task. It can be observed that from the "Power Bin" to "Wind Speed", the features have no relation to the gearbox "Error". On the other hand, the correlation between the vibration data and the gearbox status is higher. This increased correlation help improve the predictive function.

3.1 Regression model

This model takes the SCADA data from the "larger" source domain data set and makes predictions on the CMS vibration data after pre-processing. The training stage is carried out over 1000 Epochs having a total of 388,803 trainable parameters. To validate the accuracy of the model new, unseen data from the wind turbine is fed into the model producing a: MAPE = 27.00%, an accuracy = 99.92% and R2 = 68.61%. This model is the highest performing model in the process. The predictions from the model are highlighted in Figure 3. With an R2 score of 68.61%, this is a more than acceptable indicator of anomalous behaviour. The plot highlights the feature extractor's ability to do so.

Implementing a MTL model, this information would normally be fed into the model or in the worst case be ignored. This procedure utilises the information to make strong estimations on the erroneous behaviour and it could be used in another maintenance methods in a multi-agent manor for predictive trending methods. Where one could monitor the trends of the vibrations vibrations more closely to determine long term effects.

3.2 Multi-Task Learning Model

Similarly, the MTL classification model participates in training over 1000 Epochs of 83, 329 trainable parameters. This model uses significantly fewer data points compared to the feature extractor of 8441 producing an accuracy of 91.29%, and an F1-score of 69.54%. The classification for the anomaly detection utilising MTL is more than acceptable, Yang et al [15] indicate that predicting more than 60% of wind turbine faults will reduce the cost of operations and

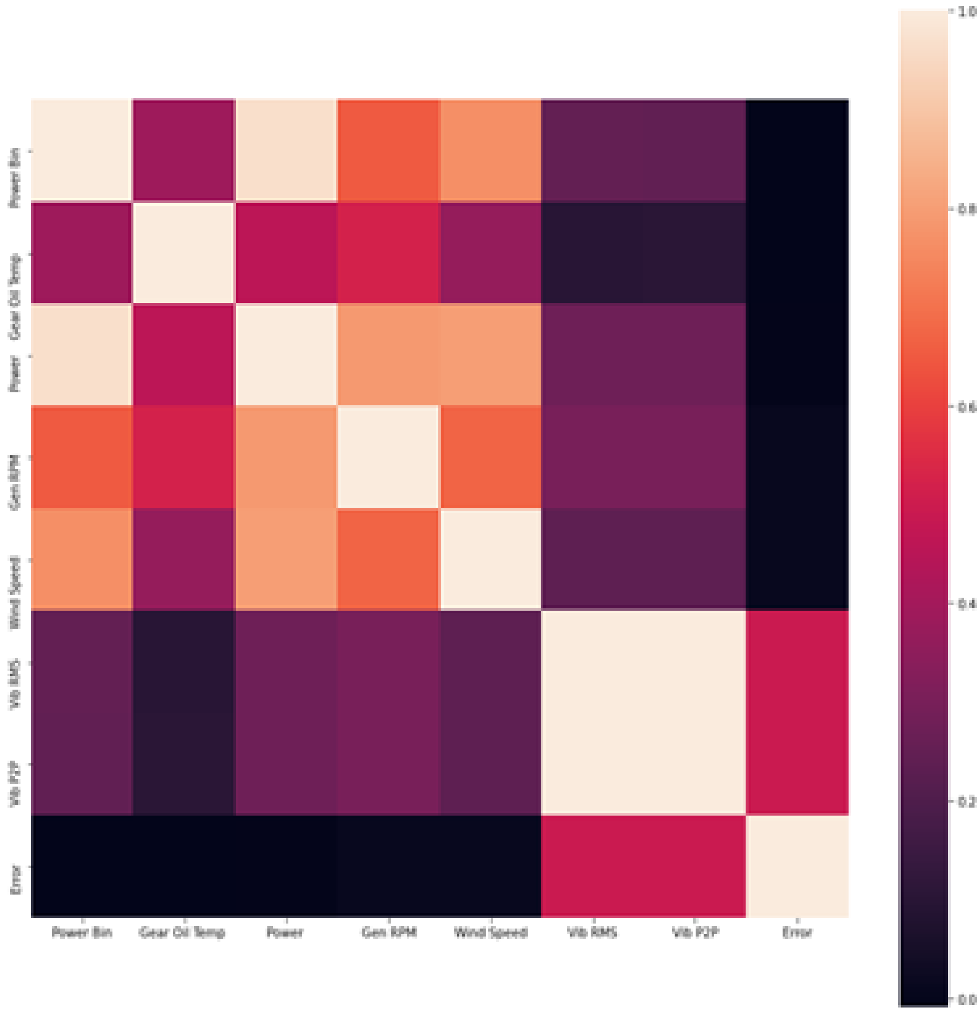
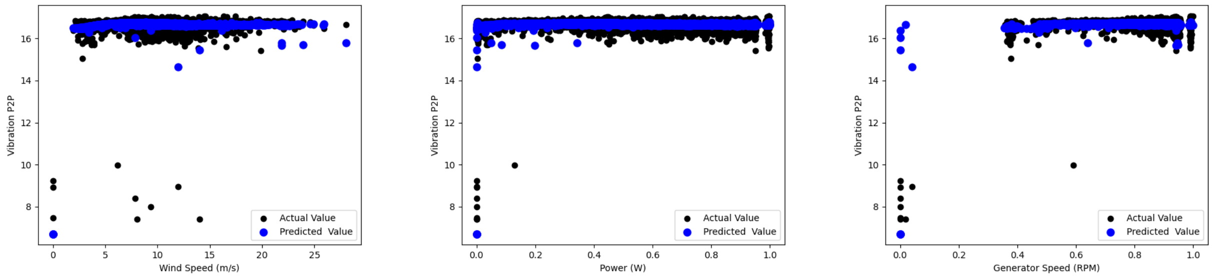


Figure 2: Linear correlation for all the features, the first five features are the inputs for both the regression model and the traditional model. The vibration signals are the outputs of the regression aspect hard parameter transfer model, and the Error is the output of the full hard parameter transfer model and the conventional model.

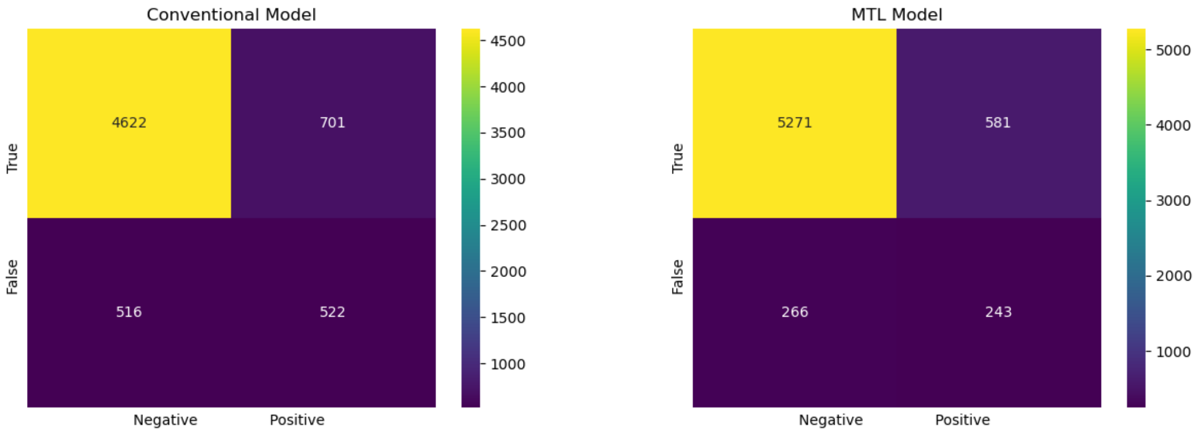
maintenance. The classification results are displayed in Figure 4. With anomaly detection, it is expected that the true negative is the median. The most detrimental prediction is a false positive, and this presents one significant portion of this model that would need improvements.

For condition monitoring, a false positive reading would highlight to the operator that the machine is currently running 'normally' but it is in a state of potential error. In the moment of a failure event, a chain reaction of issues can lead to catastrophic failure. To prevent this one might call the machine to turn off. However, this is not possible with a false positive, the machine running in a state where there is an error but the observation from the model is contradicting reality. Alternatively, if the anomaly signature is not so serious the machine could be asked to operate at a reduced rate. False negative does not present as much of a risk as it will not lead to catastrophic failure .Both scenarios are not ideal, both cause a loss in earnings and the whole objective with condition-based maintenance is to optimise the up-time of the wind turbine.



(a) Vibration vs Wind (b) Vibration vs Power (c) Vibration to Generator Speed

Figure 3: Regression model, trained on the large data-set, used as the basis for the MTL. This figure details the ability to determine the vibration signals from some of the input features.



(a) Conventional classification test results (b) MTL classification test results

Figure 4: Classification results from the MT model, highlighting the true negative, false positive, false negative, and true positive rates of the test data-set.

3.3 Model Comparison

Table 1: Model comparison of the conventionally trained model vs the hard-parameter MTL model

	Accuracy	F1 Score
Conventional Model	83.76%	57.56%
Hard Parameter MTL Model	91.29%	69.54%

The conventional model undertook the same treatment as the hard parameter transfer model, with 1000 Epochs. However, the model has suffered from an over-fitting of the result. It is unable to predict many of the errors leading to failure. A comparison of the results from the conventional compared to the MTL model is made in table 1. This highlights that with a reduced data-set, the results can be improved implementing hard parameter transfer with homogeneous dat. The conventional model training is hindered because there is simply not enough data-points and the correlation of the data is not as great to fully converge the model to surpass the MTL model accuracy.

One significant aspect that is not explored in this paper, is varying the sample size of the data. The current understanding of machine learning is both the larger the data and the network the greater the accuracy. With this current understanding, and the results displayed here, there must be a crossover point where the benefit of transfer learning will diminish, and no longer be a suitable candidate to determine the anomaly detection of an offshore wind turbine. An examining of this crossover point where conventional methods are more accurate is would provide a time-frame where this technique is recommended.

There are two ways in which time-series data varies, firstly the frequency, how many samples are taken over a specific period. The second is the length. Both are contributing factors to the size of the time series data used to train the model. In the case of this report, there are 8451 points over a two-year period. This limiting factor will determine the level of improvement from conventional model training to hard parameter transfer.

4 Conclusions

This paper has successfully highlighted how MTL accelerates the accuracy of data-driven condition monitoring of a NN with limited data. This is a novel approach to offshore wind energy but is consistent with other areas where this methodology is implemented. The main observation is that the infant mortality failure can be quickly detected, and scheduled maintenance can be planned.

By implementing the two different cost functions the model is better suited at extracting the features and classifying through reduced noise and overfitting compared to the traditional method. One observation is that this model is only suitable for limited period. The longer the wind turbine is in operation, there is less probability that the components will fail, coupled with an increased amount of data.

One of the most significant properties of the multi-task learning method is the increased amount of useable information. You have the vibration and error information both of which can be used in tandem to authenticate the diction process. One detecting patterns for anomaly detection and another making time series predictions. This is appose to the classical method which has only one output, the error evaluation, which is useful but ideally more information is advisable in maintenance.

A progression from this concept would be to both: investigate the time-frame in which MTL is most advisable for maximum accuracy of a condition monitoring system during the infant mortality period. Secondly, comparing varying types of MTL models to determine a baseline model for continuity in the industry.

References

[1] Xingxian Bao, Tongxuan Fan, Chen Shi, and Guanlan Yang. One-dimensional convolutional neural network for damage detection of jacket-type offshore platforms. *Ocean Engineering*, 219:108293, 2021.

[2] Innes Murdo Black, Mark Richmond, and Athanasios Kolios. Condition monitoring systems: a systematic literature review on machine-learning methods improving offshore-wind turbine operational management. *International Journal of Sustainable Energy*, 40:923–946, 2021.

[3] D Cevasco, M Collu, and Z Lin. Om cost-based fmeca: Identification and ranking of the most critical components for 2-4 mw geared offshore wind turbines. volume 1102, 2018. cited By 6.

[4] D Cevasco, S Koukoura, and A J Kolios. Reliability, availability, maintainability data review for the identification of trends in offshore wind energy applications. *Renewable and Sustainable Energy Reviews*, 136:110414, 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. pages 1026–1034, 2015.

[6] Liang Huang, Xu Feng, Cheng Zhang, Liping Qian, and Yuan Wu. Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing. *Digital Communications and Networks*, 5:10–17, 2019. Artificial Intelligence for Future Wireless Communications and Networking.

[7] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[8] Zehui Lin, Shengli Li, Dong Ni, Yimei Liao, Huaxuan Wen, Jie Du, Siping Chen, Tianfu Wang, and Baiying Lei. Multi-task learning for quality assessment of fetal head ultrasound images. *Medical Image Analysis*, 58:101548, 2019.

[9] Emilio Soria Olivas, José David Martín Guerrero, Marcelino Martinez-Sober, Jose Rafael Magdalena-Benedito, and Antonio José Serrano López, editors. *Handbook of Research on Machine Learning Applications and Trends*. IGI Global, 2010.

[10] M Richmond, A Sobey, R Pandit, and A Kolios. Stochastic assessment of aerodynamics within offshore wind farms based on machine-learning. *Renewable Energy*, 161:650–661, 2020.

[11] Adrian Stetco, Fateme Dinmohammadi, Xingyu Zhao, Valentin Robu, David Flynn, Mike Barnes, John Keane, and Goran Nenadic. Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy*, 133:620–635, 2019.

[12] Jannis Tautz-Weinert and Simon J Watson. Using scada data for wind turbine condition monitoring - a review. *IET Renewable Power Generation*, 11:382–394, 2016.

[13] O Troyanskaya, M Cantor, G Sherlock, P Brown, T Hastie, R Tibshirani, D Botstein, and R B Altman. Missing value estimation methods for dna microarrays. *Bioinformatics*, 17:520–525, 2001. cited By 2271.

[14] Wang Xiaohua, Peng Muzi, Pan Lijuan, Hu Min, Jin Chunhua, and Ren Fuji. Two-level attention with two-stage multi-task learning for facial emotion recognition. *Journal of Visual Communication and Image Representation*, 62:217–225, 2019.

[15] Wenxian Yang, Peter J Tavner, Christopher J Crabtree, Y Feng, and Y Qiu. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, 17:673–693, 2014.