# Supplementary Materials

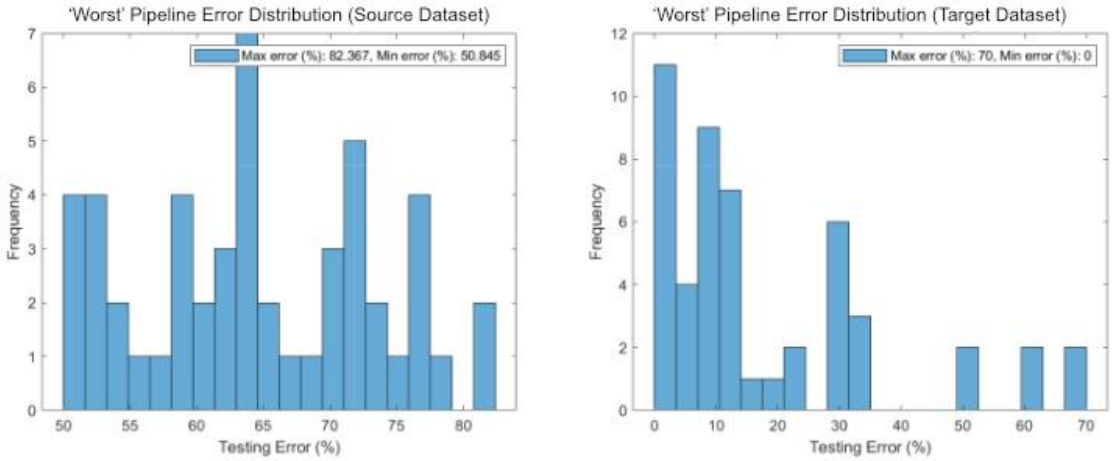


Figure C1: (left) Histogram of classification errors for the 'worst' design choices for the source dataset (Case Western reserve dataset). (right) Histogram of the classification errors for the 'worst' design choices identified for the target dataset (MFPT dataset). The 'worst' choices across each stage for each dataset were summarised in Table 3 of the paper.

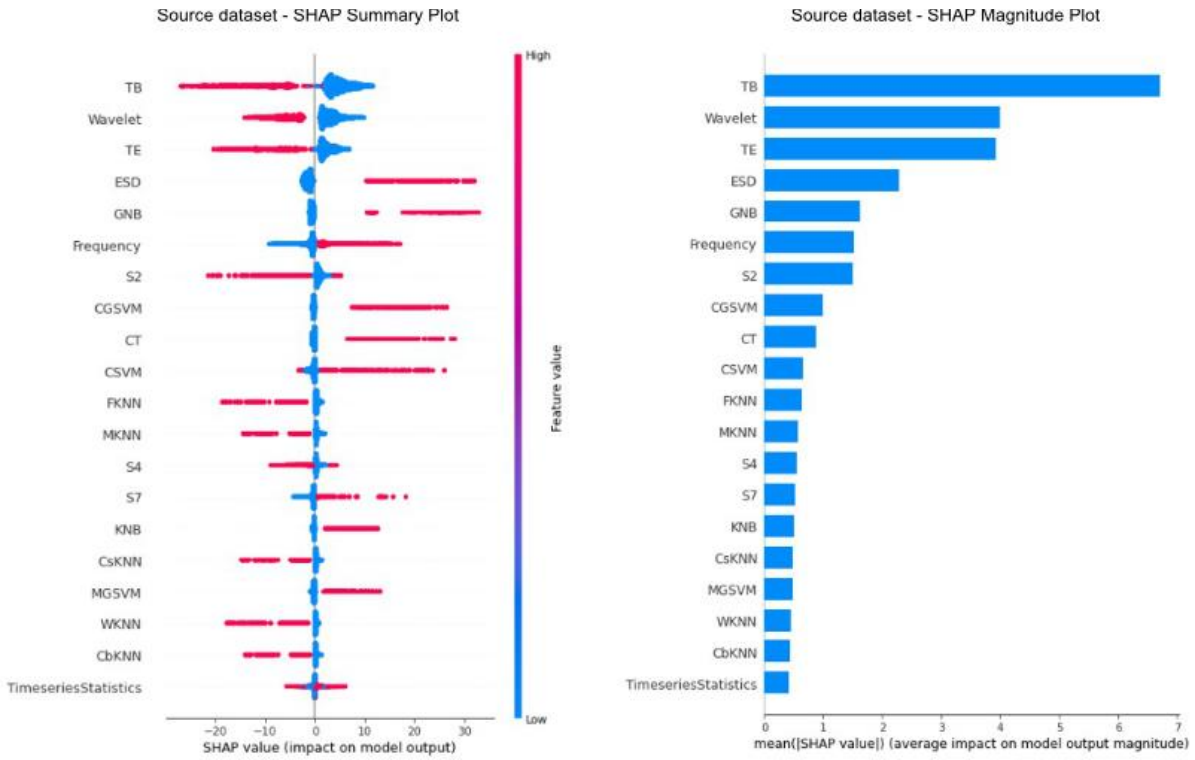


Figure C2: (Left) Summary plot showing the pipeline choice ranking and direction of impact on the pipeline classification error. When values of the choice is high, the choice is present in the design, while a low value denotes the absence of the choice in a pipeline. Some values, like the timeseries statistics choice, have generally low impact regardless of its presence or absence in

the design, while others such as the 'ESD' (Ensemble subspace discriminant) model have little impact when absent, but a large negative impact when present in the design. (Right) Magnitude plot showing the pipeline choice ranking and magnitude of impact on the pipeline classification error. The top 5 most impactful choices in descending order are 'TB' (Binary classification task), 'Wavelet' (data processing domain), 'TE' (motor end classification task), 'ESD' (ensemble subspace discriminant model) and 'GNB' (Gaussian Naive Bayes model). As shown, much of the impact has reduced by the 'S2' (Fan end sensor, only) choice, showing much of the impact is concentrated in a relatively small fraction of the decisions.

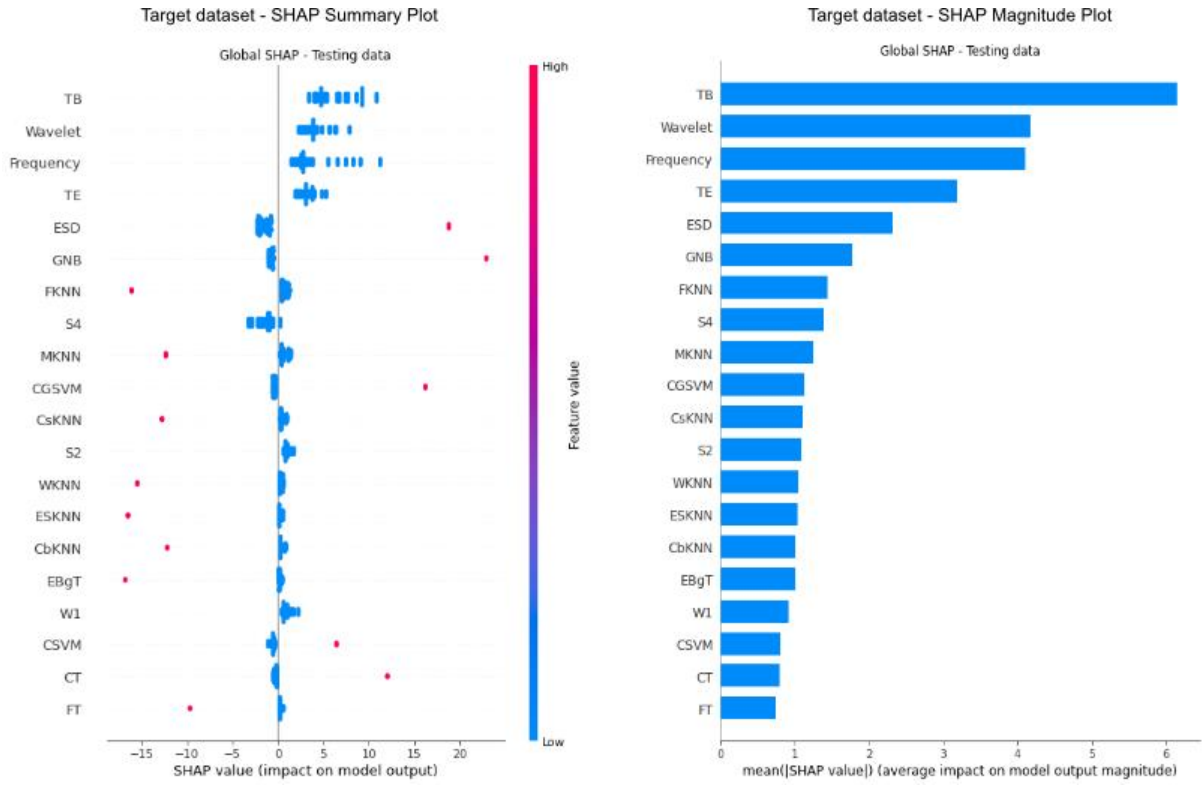


Figure C3: (Left) Summary plot showing the pipeline choice ranking and direction of impact on the pipeline classification error for the target design subset. For most variables, their absence (low value) are concentrated nearer a SHAP value of 0, while their inclusion (high value) tends to generate a more extreme SHAP value response, showing the model is more sensitive to a design inclusion than absence in this case. (Right) Magnitude plot showing the pipeline choice ranking and magnitude of impact on the pipeline classification error for the target design subset. The top 5 most impactful choices have 4 common choices with source case, with the addition being the frequency data processing domain taking third place. Compared to the SHAP plots for the source dataset, the impact of all variables is higher, however the most impact is still concentrated within the top few variables ('TB'-'ESD').