Response to Reviewer 1 Comments

Point 1: Please add one section to discuss relevant studies as only a few studies were highlighted in the introduction.

Response 1: In the revision, some relevant studies have been added in Section of Introduction. (in red)

Point 2: The caption for Figs 2 and 3 requires additional information to provide a more comprehensive description of the figure.

Response 2: A more comprehensive description of Figs. 2-3 has been added in the revision. (in red)

Point 3: Please cite references for LSTM and BI-LASTM models.

Response 3: Two references for LSTM and BI-LASTM models have been cited in the revision. (in red)

[1] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735-1780.

[2] Wang S X, Wang X, Wang S M, Wang D. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. Electrical Power and Energy Systems, 2019, (19): 470-479.

Point 4: Line 173 'are all the parameters' what do you mean? Please provide more details.

Response 4: Line 173 has been revised, that is, u_w, W_s, b_s are the parameters in the attention gate. (in red)

Point 5: Please provide a link/reference for the ENET dataset.

Response 5: A link for the ENET dataset has been provided in the revision. (in red)

Point 6: Please make the description much clearer between metadata and signal data for the ENET dataset.

Response 6: We have made the description much clearer between metadata and signal data in the revision. (in red)

Point 7: Line 230, 'which are smoother than the original signals.' Please refer to Figures 9 and 10.

Response 7: Line 230 has been revised, that is, "which are smoother than the original signals in Figs. 9-10." (in red)

Point 8: Lines 248-249 'One-phase PD signal has 19 statistical features, so three-phase PD signals 248 have 57 dimensional features, which are shown in Fig.13.' How do these extracted features become the signals shown in Fig13?

Response 8: Lines 248-249 have been revised, that is, One-phase PD signal has 19 statistical features, so three-phase PD signals have 57 dimensional features in all. The 57 dimensional features form a vector, which can be depicted in Fig.13. (in red)

Point 9: Please add parameter setting values used with the proposed and compared models.

Response 9: Parameter setting values used with proposed model are shown in Table 2 , and those of compared models are same with the proposed model. (in red)

Point 10: Regarding the comparison between different algorithms, did the authors conduct simulations for the compare models, or it is based on some published results? Please highlight that in the manuscript. If it is based on some published results, please cite those works.

Response 10: The comparison simulations between different algorithms are conducted by us, and we have highlight it in the revision. (in red)

Point 11: Please add results as numbers as well to Figure 21.

Response 11: Some results about Fig.21 have been added in the revision. (in red)