I. Appendix A

The schematic of the formulated open-world fault diagnosis scheme based on uncertainty and prototypes (OWFD-UP) is shown in Fig. 1, where (h) and (i) are flowcharts of the proposed solution and (a)-(g) are descriptions of each step.

(h) is a completely offline process. Firstly, the available manually labeled data are split into training set, validation set and test set by 6:2:2. Secondly, data standardization is implemented by Eq. 1 of the main article, which is based on statistical indicators of fault-free training data and alarm mechanism. Then, the prototypes are selected from the training data of each class by the dissimilarity-based sparse subset selection (DS3) algorithm, i.e., Eq. 5 of the main article, and stored evenly in the buffer. Next, the classifier is fitted on the pre-stored prototypes. Finally, the optimal regulatory factor α is determined by the open-set grid search protocol.

Once the above procedures are completed, the online program in (i) can be switched on. Firstly, new raw samples, either individually or in batches, are standardized by Eq. 1 of the main article, where the parameters are prepared at (h). Then, the standardized data are fed into the classifier fitted at (h) to obtain the closed-set probability prediction. The uncertainty of the closed-set decision is evaluated by Eq. 6 of the main article, and then multiplied by α determined at (h) to obtain the unknown probability. After normalization by Eq. 8 of the main article, the open-set probability prediction containing both known and unknown is obtained. When the number of unknowns discriminated by Eq. 9 of the main article reaches the expected number, the incremental learning (IL) phase can be entered. Otherwise, the online detection continues.

In the offline IL phase, the new samples are first manually labeled. Then, they are split by 6:2:2 and the training data from them are merged with the prototypes pre-stored at (h). Afterwards, the prototypes are reselected on the merged data by Eq. 5 of the main article and deposited in the buffer. The classifier is refitted on the selected new prototypes. Next, a new regulatory factor α is redetermined by the open-set grid search protocol. At this point, IL has been completed. Finally, the corresponding components at (h) are replaced with the latest prototypes, classifier and regulatory factor α, and the online detection is restarted.

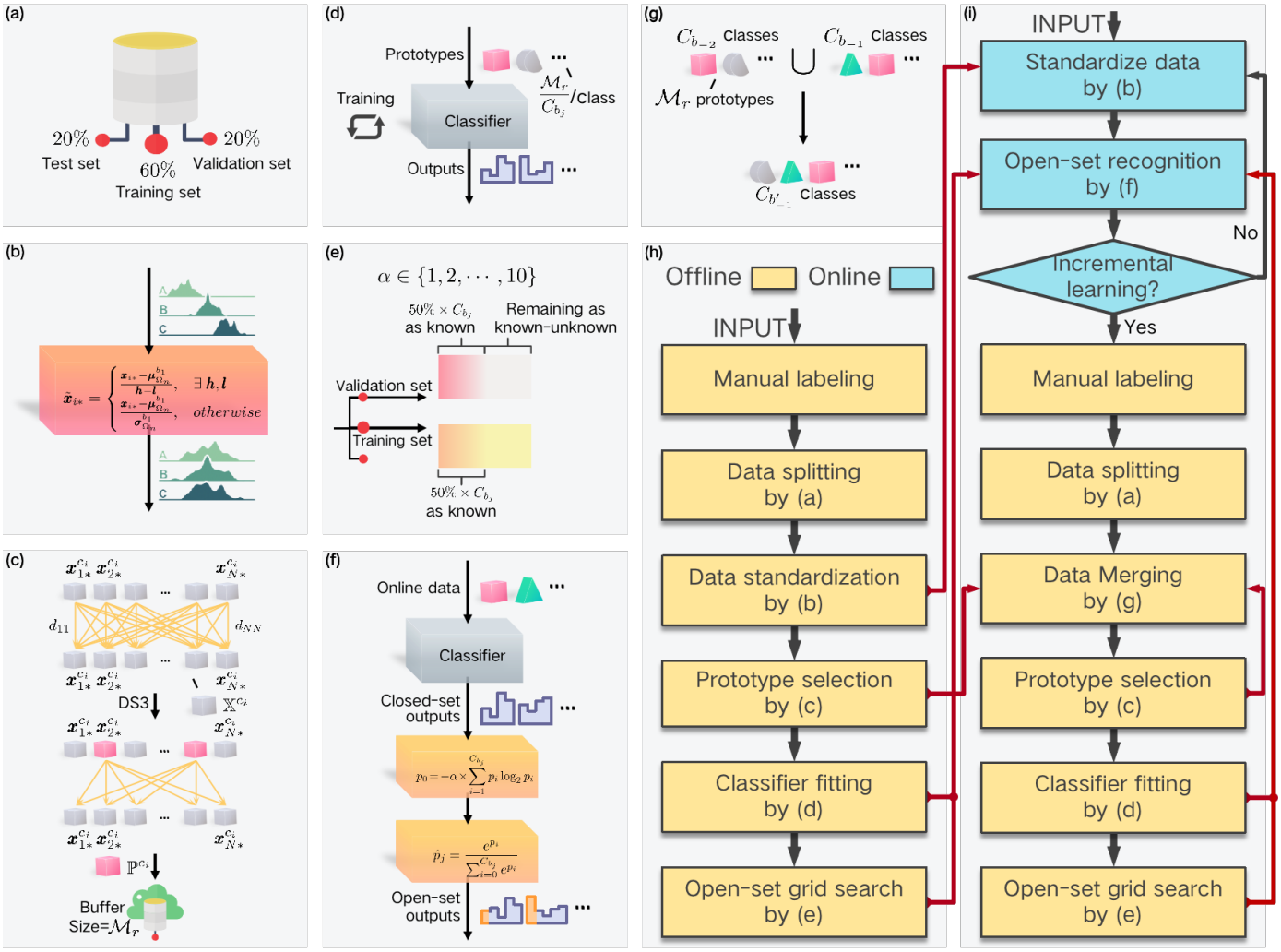

Fig. 1. Schematic of OWFD-UP. (a) Data splitting criterion. The training set, validation set, and test set are split by 6:2:2 for fitting the closed-set classifier, parameter search, and closed-set performance evaluation, respectively. (b) Data standardization process based on normal data and alarm mechanism. A, B, and C are 3 different variables and the curves on the axes are their kernel density estimates. (c) DS3-based prototype selection process. The DS3 algorithm searches for prototypes $\mathbb{P}^{c_i}$ in $\mathbb{X}^{c_i}$ that represent itself well based on dissimilarity. (d) Classifier fitting process. The classifier is fed with $\mathcal{M}_r$ selected prototypes, $\mathcal{M}_r/C_{b_j}$ per class, and outputs closed-set class probability estimates. (e) Open-set grid search protocol. Set 50% of the classes in the validation set as known and the rest as known-unknown. The classifier is fitted on 50% of the known classes in the training set. The validation during the search for regulatory factor α is performed on all C_{b_j} classes. (f) Uncertainty-based open-set fault diagnosis process. For an online sample, the uncertainty in its closed-set output serves as an estimate for the unknown. (g) Data merging process. The pre-stored $\mathcal{M}_r$ prototypes are merged with the training data from the latest batch. (h) Offline modeling procedures. Execute after the first batch of data has been fetched. Prepare the mean $\boldsymbol{\mu}_{\Omega_n}^{b_1}$ and standard deviation $\boldsymbol{\sigma}_{\Omega_n}^{b_1}$ of the normal sample set, the pre-stored prototypes, the fitted classifier, and the regulatory factor α for the subsequent procedures. (i) Open-world fault diagnosis procedures. Online data can be input separately or in batches. Once a batch size open-set fault diagnosis is completed, the IL can be turned on. IL aims to update pre-stored prototypes, fitted classifier, and the regulatory factor α, but it has no effect on data standardization.