ML Classifier Fusion for Three Data Streams with Quality Inversely Proportional to Time Resolution

Nageswara S. V. Rao
Oak Ridge National Laboratory
Oak Ridge, TN 37831, USA
raons@ornl.gov

Chris Y. T. Ma
The Hang Seng University of Hong Kong
chrisma@hsu.edu.hk

Fei He
Texas A&M University-Kingsville
Kingsville, TX, USA
fei.he@tamuk.edu

***Abstract*—We consider a monitoring scenario of phenomenon using three different streams of measurements whose quality is proportional to their constant inter-arrival times. Each measurement of a stream needs to be binary-classified to reflect the state of interest of the phenomenon. A set of classifiers is separately trained and fused for each stream at its time resolution using measurements collected under known states. We present a machine learning method to fuse the outputs of these fusers to provide a final classification at the finest time resolution. We show that this fused-fusers method provides decisions with likely superior classification probability compared to the best individual classifiers and fused-classifiers. We derive generalization equations that guarantee a superior classification probability of fused-fusers with a confidence probability specified by the classifiers' generalization equations. We apply these results to study a practical problem of classifying Pu/Np target dissolution events at a radiochemical processing facility using gamma spectral measurements of effluent flows.**

***Index Terms*—classifier, fuser, fused-fusers, fused-classifiers, switched-fusers, generalization equation, ROC, time resolution, machine learning**

I. Introduction

We consider a monitoring scenario with three distinct data streams, each providing measurements of a phenomenon at different time resolutions with constant inter-arrival times (or rates) with different quality. A binary classification decision is required upon the arrival of each measurement of a stream to reflect the state of interest of the phenomenon, for example, a specific dissolution operation at a radiochemical facility, or the presence of a target in a monitored area, or an anomaly in computer network traffic. The measurements at the coarsest time resolution are of the highest quality, and are classified by set $\mathbf{C}_H$ of binary classifiers. At the finest time resolution, measurements are of the lowest quality and are classified by set $\mathbf{C}_L$ of binary classifiers. The measurements of the third stream have the time resolution and quality in between the two, and are classified by set $\mathbf{C}_M$ of binary classifiers. The receiver operating characteristic (ROC) curves of classifiers

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy (DOE). This research is supported in part by RAMSES project, Advanced Scientific Computing Research, Office of Science, DOE . The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

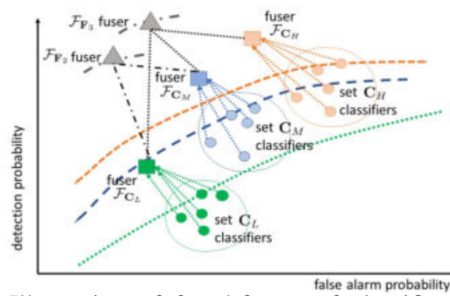


Fig. 1: Illustration of fused-fusers of classifier sets for improved classification probability with quantified confidence.

of higher quality are overall higher than those of others, but their classification outputs are provided at a coarser time resolution or less frequently. We address the problem of classifying the measurements of these streams as they arrive by "combining" the outputs of individual classifiers and fusers. This formulation is a specific abstraction of the classification [1], [2] and classifier fusion [3], [4], and is an extension of a two-stream formulation in [5]. It is motivated by the nuclear dissolutions application [6], and is a special case of the well-studied information fusion problem [7] applied to multi-rate [8] and multi-resolution [9] sensors or data sources.

A straight-forward method is to choose the "best" classifier from each set, and use it to classify the corresponding measurements as they arrive. We present a superior classification approach with generalization performance better than the best classifier or fused-classifiers from each stream, with quantified confidence. We employ a method that first fuses the classifiers within sets $\mathbf{C}_H$, $\mathbf{C}_M$ and $\mathbf{C}_L$, and then fuses the fused-classifiers as illustrated in Fig. 1. We show that this fused-fusers method provides superior classification probability compared to the best classifiers of $\mathbf{C}_H$, $\mathbf{C}_M$ and $\mathbf{C}_L$, and their fused-classifiers, with a confidence probability specified by its generalization equation.

We consider that machine learning (ML) methods [10] are used to train the classifiers using finite training sets consisting of measurements collected under known states of the phenomenon. Consequently, the best classification method does not exist [11], and the classification probability of a classifier can only be specified with a confidence bound provided by its generalization equation [12], [13]. We derive sufficiency conditions for the superior performance of our method under the isolation property [14] of the fuser classes, and derive the generalization equations [15]. Under this formulation, solutions using the existing multi-rate, multi-resolution methods require the knowledge of the underlying system or distribution models [7]–[9]. In addition, ML methods without such knowledge requirements [10] do not provide the needed generalization equations in most cases. The previous switched-fuser method [5] provides the generalization equations for the special case of two streams, which we generalize here to three streams and broader classes of ML fusers.

This problem formulation is motivated by a practical task of detecting events associated with the Plutonium/Neptunium (Pu/Np) target dissolutions at a radiochemical facility, by using gamma spectral measurements of effluent streams [6], [16], [17]. Previously, measurements at two different time resolutions are considered [5], wherein multiple classifiers are fused for each stream, and the fusers are switched as measurements arrive. The classification performance of this method is superior to using a single (best) classifier for each stream. In this paper, we consider measurements at three time resolutions, namely, 1, 3 and 6 hours, and present a solution based on fused-fusers of three time resolutions. This approach improves the classification performance over the best classifiers and switched-fusers in [5]. In addition, we compare various solutions based on switching the classifiers, fused-classifiers, and fused-fusers.

The organization of the paper is as follows. The problem formulation is provided in Section II, and our overall approach is described in Section III. Generalization equations of fusers, fused-classifiers, fused-fusers, switched-classifiers, and switched-fusers are presented in Section IV. The experimental results of dissolution classification problem are related to the analytical results in Section V. Conclusions and directions for future work are presented in Section VI.

II. Problem Formulation

A binary classifier $C : \Re^d \mapsto \{0,1\}$ is characterized by the measurement input $X \in \Re^d$ mapped to binary or Boolean output $Y = C(X) \in \{0,1\}$. We consider three separate sets of classifiers used for the three measurement streams. The ordered set of classifiers $\mathbf{C}_H = \{C_{H_1}, C_{H_2}, \ldots, C_{H_{n_H}}\}$ handles measurements at the coarsest time resolution with time interval t_H, another ordered set of classifiers $\mathbf{C}_M = \{C_{M_1}, C_{M_2}, \ldots, C_{M_{n_M}}\}$ handles those at the medium time resolution with time interval t_M, and a third ordered set of classifiers $\mathbf{C}_L = \{C_{L_1}, C_{L_2}, \ldots, C_{L_{n_L}}\}$ handles at the finest resolution, with the corresponding fixed time intervals $t_L < t_M < t_H$. The rates of classification output from classifiers $C_H \in \mathbf{C}_H$, $C_M \in \mathbf{C}_M$, and $C_L \in \mathbf{C}_L$ are $1/t_H$, $1/t_M$, and $1/t_L$, respectively. The *rate fraction* of $\mathbf{C}_R$, for $R = H, M, L$, is $\rho_R = \frac{1/t_R}{(1/t_H + 1/t_M + 1/t_L)} = t_H t_M t_L / [t_R(t_H t_M + t_M t_L + t_H t_L)]$ such that a larger value represents more frequent but lower quality measurements and vice versa. A classification decision is required at the arrival of every measurement, namely, at an average rate of $1/t_H + 1/t_M + 1/t_L$.

To represent multiple resolutions, we use a generic ordered set of classifiers $\mathbf{C} = \{C_1, C_2, \ldots, C_{n_C}\}$. Let $D_C(F)$ of classifier C denote its *detection probability* at the *false alarm probability* F. The *operating point* (OP) of C is given by $P_C = (F_C, D_C(F_C))$, simply denoted by (F_C, D_C). We consider that ROC, $D_C(F)$, as a function of F is non-decreasing. The classifier C_i is *superior* to classifier C_j in terms of ROC if for all F, we have $D_{C_i}(F) \geq D_{C_j}(F)$. This condition implies that at any false alarm probability, the detection probability of C_i is at least as high as that of C_j, and combined with the non-decreasing property of $D_C(F)$ implies that at any detection probability, the false alarm probability of C_i is no higher than that of C_j. We consider that the classifier set $\mathbf{C}_A$ to be superior to $\mathbf{C}_B$ if C_{A_i} is superior to C_{B_j} for all $i = 1, 2, \ldots, n_{C_A}$, and $j = 1, 2, \ldots, n_{C_B}$.

The outputs of an ordered set of classifiers $\mathbf{C}$ with corresponding input vectors $\mathbf{X} = \{X_1, X_2, \ldots, X_{n_C}\}$, $X_i \in \Re^d$, is combined by a *fuser* $\mathcal{F}_{\mathbf{C}}$, which itself is a classifier with input vector $(C_1(X_1), C_2(X_2), \ldots, C_{n_C}(X_{n_C}))$ of binary components and binary output. The *fused-classifier*, $\mathcal{F}_{\mathbf{C}} \circ \mathbf{C}$, is characterized by the Boolean valued function

$$(\mathcal{F}_{\mathbf{C}} \circ \mathbf{C})(\mathbf{X}) = \mathcal{F}_{\mathbf{C}}\left(C_1(X_1), C_2(X_2), \ldots, C_{n_C}(X_{n_C})\right)$$

defined on measurements space $\Re^{d \times n_C}$. When all classifiers use the same input, that is, $X_1 = X_2 = \cdots = X_{n_C} = X$, we denote this function's output simply by $(\mathcal{F}_{\mathbf{C}} \circ \mathbf{C})(X)$, which is defined on the measurements space $\Re^d$.

The *switched-classifier* $\mathcal{S}_{\{C_{H_i}, C_{M_j}, C_{L_k}\}}$ outputs that of the classifier C_{R_l}, for $R = H, M, L$ that reflects its time resolution, at the corresponding $l = i, j, k$ indices. When C_{R_l} is a fused-classifier, it is simply referred to as a switched-fuser.

III. Classifiers: Enhancement and Fusion

Our method consists of fusing the outputs of classifier sets $\mathbf{C}_H$, $\mathbf{C}_M$, and $\mathbf{C}_L$ using fusers $\mathcal{F}_{\mathbf{C}_H}$, $\mathcal{F}_{\mathbf{C}_M}$, and $\mathcal{F}_{\mathbf{C}_L}$, respectively, and reporting the output at the respective time resolution. Then, a classification output is produced at the finest time resolution by a fused-fuser $\mathcal{F}_{\mathbf{F}_3}$ for $\mathbf{F}_3 = \{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$, that combines fusers from each time resolution with the latest inputs at their individual resolutions. The overall output is produced by the *switched-fused-fuser* $\mathcal{S}_{\{\mathcal{F}_{\mathbf{F}_3}, \mathcal{F}_{\mathbf{F}_2}, \mathcal{F}_{\mathbf{C}_L}\}}$, for $\mathbf{F}_2 = \{\mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$, which outputs that of a fuser as its arrival rate, and has the output rate of $1/t_H + 1/t_M + 1/t_L$.

A. Overall Approach

Our method is based on developing fusers $\mathcal{F}_{\mathbf{C}_H}$, $\mathcal{F}_{\mathbf{C}_M}$, and $\mathcal{F}_{\mathbf{C}_L}$ that are superior to the corresponding individual classifiers, and then utilizing their fusers $\mathcal{F}_{\mathbf{F}_3}$ and $\mathcal{F}_{\mathbf{F}_2}$, as illustrated in Fig. 1. Let $X_H, X_M, X_L \in \Re^d$ denote the latest measurements from the three streams. Then, the fused-classifier function for data stream of resolution $R = H, M, L$ is $\mathcal{F}_{\mathbf{C}_R} \circ \mathbf{C}_R$ with all classifiers of $\mathbf{C}_R = \{C_{R_1}, C_{R_2}, \ldots, C_{R_{n_R}}\}$ using the same input X_R. Thus, we have

$$\begin{aligned}
&(\mathcal{F}_{\mathbf{C_R}} \circ \mathbf{C}_R)(X_R) \\
&\quad = \mathcal{F}_{\mathbf{C_R}}\left(C_{R_1}(X_R), C_{R_2}(X_R), \ldots, C_{R_{n_R}}(X_R)\right)
\end{aligned}$$

for resolution $R = H, M, L$. Then, their fuser $\mathcal{F}_{\mathbf{F}_3}$, for $\mathbf{F}_3 = \{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$, corresponds to the fused-classifier function $\mathcal{F}_{\mathbf{F}_3} \circ \mathbf{F}_3$. Using the composite input $\mathbf{X} = \{X_H, X_M, X_L\}$, we denote the output of the individual stream fusers as

$$(\mathbf{F}_3 \circ \mathbf{C})(\mathbf{X}) \triangleq ((\mathcal{F}_{C_H} \circ \mathbf{C}_H)(X_H), (\mathcal{F}_{C_M} \circ \mathbf{C}_M)(X_M), (\mathcal{F}_{C_L} \circ \mathbf{C}_L)(X_L)),$$

which is the input to the fuser function $\mathcal{F}_{\mathbf{F}_3}$. We denote the fused-fusers process as a composition of outputs of classifiers and fusers using their respective functions as follows:

$$\begin{aligned}(\mathcal{F}_{\mathbf{F}_3} \circ \mathbf{F}_3 \circ \mathbf{C})(\mathbf{X}) &= (\mathcal{F}_{\mathbf{F}_3} \circ (\mathbf{F}_3 \circ \mathbf{C}))(\mathbf{X}) \\ &\triangleq \mathcal{F}_{\mathbf{F}_3}((\mathcal{F}_{C_H} \circ \mathbf{C}_H)(X_H), (\mathcal{F}_{C_M} \circ \mathbf{C}_M)(X_M), \\ &\qquad (\mathcal{F}_{C_L} \circ \mathbf{C}_L)(X_L)) \\ &\triangleq \mathcal{F}_{\mathbf{F}_3}(\mathcal{F}_{\mathbf{C}_H}(\mathcal{C}_H(X_H)), \mathcal{F}_{\mathbf{C}_M}(\mathcal{C}_M(X_M)), \\ &\qquad \mathcal{F}_{\mathbf{C}_L}(\mathcal{C}_L(X_L))),\end{aligned}$$

where $\mathcal{C}_R(X_R) = (C_{R_1}(X_R), C_{R_2}(X_R), \ldots, C_{R_{n_R}}(X_R))$, for $R = H, M, L$.

We similarly consider $\mathcal{F}_{\mathbf{F}_2}$, for $\mathbf{F}_2 = \{\mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$ that corresponds to the fused-classifier function $\mathcal{F}_{\mathbf{F}_2} \circ \mathbf{F}_2$ with composite input $\{X_M, X_L\}$ and output $\mathcal{F}_{\mathbf{F}_2}(\mathcal{F}_{\mathbf{C}_M}(\mathcal{C}_M(X_M)), \mathcal{F}_{\mathbf{C}_L}(\mathcal{C}_L(X_L)))$.

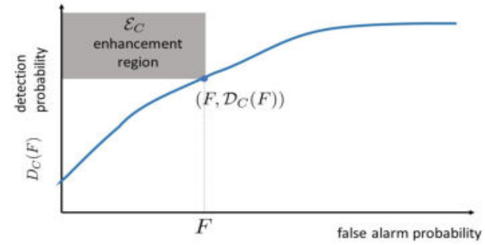
The *enhancement region* $\mathcal{E}_C$ of a classifier C is the rectangular region with lower false alarm and higher detection probabilities than those of C as shown in Fig. 2(a), and that of a set of classifiers is the intersection of their enhancement regions as shown in Fig. 2(b). Our approach is to utilize combinations of fusers and switching to improve the performance within each set such that the resultant operating point is within one or both coordinate ranges of the intersection of enhancement regions of $\mathbf{C}_H$, $\mathbf{C}_M$, and $\mathbf{C}_L$, as illustrated in Fig. 2(b). For instance, outputs of three fusers are switched by $\mathcal{S}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}$ so that their detection and false alarm probabilities are linear combinations with coefficients that reflect the rate fraction ρ_R. For a suitably high ρ_H and/or ρ_M, the operating point of $\mathcal{S}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}$ lies within the intersection of enhancement regions of the three fusers with certain probability at the expense of requiring more high quality data.

B. Fused-Classifiers

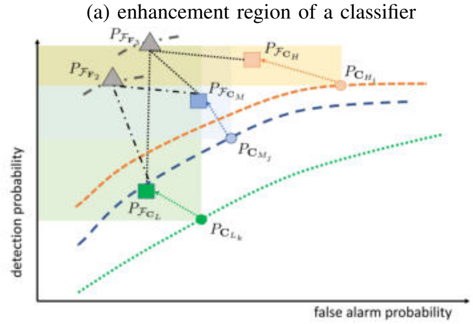
The fuser $\mathcal{F}_{C_R}$, $R = H, M, L$, is considered to be $\mathbf{I}$-superior to the corresponding individual classifier subset if

$$D_{\mathcal{F}_{C_R}} \geq D_{C_{R_i}} \quad \text{and} \quad F_{\mathcal{F}_{C_R}} \leq F_{C_{R_i}}$$

with certain probability for $i \in \mathbf{I}$; it is simply called superior if the subset is the entire classifier set. These conditions are satisfied with a confidence probability if a fuser $\mathcal{F} : [0,1]^n \mapsto [0,1]$ is chosen from a class $\Phi_{\mathcal{F}}$, as in the case of ML methods, under the isolation property [14] such that it contains $\mathcal{F}_i$, $i \in \{1, 2, \ldots, n\}$, where $\mathcal{F}_i(X) = x_i$ for $X = (x_1, x_2, \cdots, x_n)$. Analytical bounds are derived for the confidence probability for $\mathcal{F}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}$ using the generalization equations of the classifiers in Theorem 4.1.



(a) enhancement region of a classifier



(b) ROC curves and OPs of individual classifiers, fusers, and fused-fusers

Fig. 2: Enhancement regions and comparison of operating points and ROC curves of classifiers and fusers.

C. Switched-Classifiers

The rate of classification output from switched-classifier $\mathcal{S}_{\{C_{H_i}, C_{M_j}, C_{L_k}\}}$ is $1/t_H + 1/t_M + 1/t_L$, since it outputs every time any of its component classifiers outputs. The detection and false alarm probabilities of $\mathcal{S}_{\{C_{H_i}, C_{M_j}, C_{L_k}\}}$ are

$$D_{\mathcal{S}_{\{C_{H_i}, C_{M_j}, C_{L_k}\}}} = \sum_{R = H_i, M_j, L_k} \rho_R D_{C_R},$$

$$F_{\mathcal{S}_{\{C_{H_i}, C_{M_j}, C_{L_k}\}}} = \sum_{R = H_i, M_j, L_k} \rho_R F_{C_R},$$

respectively, for $\rho_R \in [0,1]$, $R = H_i, M_j, L_k$.

The conditions for $\mathcal{S}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}$ to be in the intersection of enhancement regions of best individual classifiers are

$$D_{\mathcal{S}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}} \geq \max\{D_{\mathbf{C}^*_H}, D_{\mathbf{C}^*_M}, D_{\mathbf{C}^*_L}\} \quad \text{and}$$

$$F_{\mathcal{S}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}} \leq \min\{F_{\mathbf{C}^*_H}, F_{\mathbf{C}^*_M}, F_{\mathbf{C}^*_L}\},$$

where $D_{\mathbf{C}^*_H} = \max_i D_{\mathbf{C}_{H_i}}$, $D_{\mathbf{C}^*_M} = \max_j D_{\mathbf{C}_{M_j}}$, $D_{\mathbf{C}^*_L} = \max_k D_{\mathbf{C}_{L_k}}$, $F_{\mathbf{C}^*_H} = \min_i F_{\mathbf{C}_{H_i}}$, $F_{\mathbf{C}^*_M} = \min_j F_{\mathbf{C}_{M_j}}$, and $F_{\mathbf{C}^*_L} = \min_k F_{\mathbf{C}_{L_k}}$.

This detection probability condition depends on the rate fraction ρ_R, $R = H, M, L$, and is given by

$$\rho_H D_{\mathcal{F}_{\mathbf{C}_H}} + \rho_M D_{\mathcal{F}_{\mathbf{C}_M}} + \rho_L D_{\mathcal{F}_{\mathbf{C}_L}} \geq \max\left\{D_{\mathbf{C}^*_H}, D_{\mathbf{C}^*_M}, D_{\mathbf{C}^*_L}\right\}.$$

The corresponding false alarm probability condition is

$$\rho_H F_{\mathcal{F}_{\mathbf{C}_H}} + \rho_M F_{\mathcal{F}_{\mathbf{C}_M}} + \rho_L F_{\mathcal{F}_{\mathbf{C}_L}} \leq \min\left\{F_{\mathbf{C}^*_H}, F_{\mathbf{C}^*_M}, F_{\mathbf{C}^*_L}\right\}.$$

In general, these two conditions may not be simultaneously satisfied, but they are under some statistical independence conditions [5].

IV. Machine Learned Classifiers

For a classifier based on ML methods, the detection and false alarm probabilities are jointly characterized by its generalization equation that provides a confidence probability bound $1-\delta(\epsilon, l)$ which ensures that the generalization error is within a precision parameter ϵ of the optimal, based on l training measurements [15]. We now derive the generalization equations for a computed generic-form fuser $\tilde{\mathcal{F}}_{\mathbf{A}}$ for $\mathbf{A} = \mathbf{C}_H, \mathbf{C}_M, \mathbf{C}_L, \mathbf{F}_2, \mathbf{F}_3$, based on $l_{\mathbf{C}_H}$, $l_{\mathbf{C}_M}$, and $l_{\mathbf{C}_L}$ training examples for the classifier sets $\mathbf{C}_H$, $\mathbf{C}_M$, and $\mathbf{C}_L$, respectively, and $l_{\mathcal{F}_{\mathbf{A}}}$ training samples for the fuser $\mathbf{A} = \mathbf{F}_2, \mathbf{F}_3$.

A. Classifier Generalization Equations

Let $C(X)$ be the output of a classifier C based on input X that corresponds to output Y distributed according to an unknown distribution $\mathbb{P}_{X,Y}$. A fused-classifier, namely, $\mathcal{F}_{\mathbf{C}_R} \circ \mathbf{C}_R$ or $\mathcal{F}_{\mathbf{F}} \circ \mathcal{F}_{\mathbf{C}_R} \circ \mathbf{C}_R$, for $R = H, M, L$, is a composition of classifiers and one or two fusers, and hence itself is a classifier with $X \in \Re^d$ and $Y \in \{0,1\}$. The classifier C is chosen from a function class Φ_C according to a cost criterion. Its *expected error* is defined as

$$I_C = \int \left(C(X) \oplus Y\right) d\mathbb{P}_{X,Y},$$

where $\oplus$ is the exclusive-OR operation. The false alarm probability is $F_C = I_C$ under $Y = 0$, and the missed detection probability is $1 - D_C = I_C$ under $Y = 1$; thus, we have $I_C = F_C + 1 - D_C$, that is, the expected error is a linear combination of false alarm and detection probabilities. For two classifiers C_i and C_j, the condition $I_{C_i} \leq I_{C_j}$ implies $F_{C_i} \leq F_{C_j}$ or $D_{C_i} \geq D_{C_j}$, that is, either F_{C_i} or D_{C_i}, or both, lie in the enhancement-region $\mathcal{E}_{C_j}$ of C_j.

Given the training set $(X_1, Y_1), (X_2, Y_2), \ldots, (X_{l_C}, Y_{l_C})$, the *empirical error* is

$$\hat{I}_C = \sum_{i=1}^{l_C} \left(C(X_i) \oplus Y_i\right),$$

where the measurement X_i is collected under the known state Y_i of phenomenon. The best classifiers that minimize the expected and empirical error are denoted by C^* and $\hat{C}$, respectively. Thus, for any $C \in \Phi_C$, $I_C \geq I_{C^*}$ and $\hat{I}_C \geq \hat{I}_{\hat{C}}$. In general, C^* is unknown since the underlying distribution $\mathbb{P}_{X,Y}$ is unknown, and $\hat{C}$ is not precisely computable due to its complexity (e.g. NP-hard) or limitations of the computing systems. In practice, we consider a *computed classifier* $\tilde{C}$ that achieves the minimum empirical error within $\hat{\epsilon}_{\tilde{C}}$ such that $\hat{I}_{\tilde{C}} = \hat{I}_{\hat{C}} + \hat{\epsilon}_{\tilde{C}}$, and its expected error is given by $I_{\tilde{C}} = I_{C^*} + \epsilon_{\tilde{C}}$, where $\epsilon_{\tilde{C}} \geq 0$.

The performance of a computed classifier $\tilde{C}$ is characterized by its generalization equation

$$\mathbb{P}^l_{X,Y}\left\{I_{\tilde{C}} - I_{C^*} > \epsilon + \hat{\epsilon}_{\tilde{C}}\right\} < \delta_{\Phi_C}(\epsilon, l_C), \tag{4.1}$$

where l_C is the number of training measurements, and ϵ and δ are precision and uncertainty parameters, respectively. In the special case where the computed $\tilde{C}$ minimizes the empirical error, that is, $\tilde{C} = \hat{C}$, the above equation simplifies to

$$\mathbb{P}^l_{X,Y}\left\{I_{\hat{C}} - I_{C^*} > \epsilon\right\} < \delta_{\Phi_C}(\epsilon, l_C).$$

This result is derived from the uniform convergence of means and expectations given by [13]

$$\mathbb{P}^l_{X,Y}\left\{\sup_{C \in \Phi_C} \left|\hat{I}_C - I_C\right| > \epsilon\right\} < \delta_{\Phi_C}(\epsilon/2, l_C),$$

or $\mathbb{P}^l_{X,Y}\left\{\sup_{C \in \Phi_C} \left|\hat{I}_C - I_C\right| < \epsilon\right\} > 1 - \delta_{\Phi_C}(\epsilon/2, l_C).$

B. Fuser Improvement

In a set of classifiers $\mathbf{C}_R$, for $R = H, M, L$, a classifier C_{R_i} is chosen from a function class $\Phi_{C_{R_i}}$, and the corresponding expected best, empirical best, and computed versions are denoted respectively by

$$\mathbf{C}^*_R = \left\{C^*_{R_1}, C^*_{R_2}, \cdots, C^*_{R_{n_R}}\right\},$$
$$\hat{\mathbf{C}}_R = \left\{\hat{C}_{R_1}, \hat{C}_{R_2}, \cdots, \hat{C}_{R_{n_R}}\right\},$$
$$\tilde{\mathbf{C}}_R = \left\{\tilde{C}_{R_1}, \tilde{C}_{R_2}, \cdots, \tilde{C}_{R_{n_R}}\right\}.$$

For each R_i, the expected best $C^*_{R_i}$, empirical best $\hat{C}_{R_i}$, and computed $\tilde{C}_{R_i}$ are all chosen from the same function class $\Phi_{C_{R_i}}$ such that in terms of notation we have

$$\Phi_{C_{R_i}} = \Phi_{C^*_{R_i}} = \Phi_{\hat{C}_{R_i}} = \Phi_{\tilde{C}_{R_i}}.$$

Then, the best among these sets of classifiers are given by $C^*_{R_{\min}} = \operatorname*{argmin}_{R_i=1}^{n_R} I(C^*_{R_i})$, $\hat{C}_{R_{\min}} = \operatorname*{argmin}_{R_i=1}^{n_R} \hat{I}(\hat{C}_{R_i})$, and $\tilde{C}_{R_{\min}} = \operatorname*{argmin}_{R_i=1}^{n_R} \hat{I}(\tilde{C}_{R_i})$, respectively. Each set of classifiers are fused by a fuser $\mathcal{F}_{\mathbf{C}_R} : \{0,1\}^{n_R} \mapsto \{0,1\}$ chosen from class $\Phi_{\mathcal{F}_{\mathbf{C}_R}}$, and these fusers are fused by $\mathcal{F}_{\mathbf{F}_3} : \{0,1\}^3 \mapsto \{0,1\}$ and $\mathcal{F}_{\mathbf{F}_2} : \{0,1\}^2 \mapsto \{0,1\}$ chosen from class $\Phi_{\mathcal{F}_{\mathbf{F}_3}}$ and $\Phi_{\mathcal{F}_{\mathbf{F}_2}}$, respectively. Let $\mathcal{F}^*_{\mathbf{A}^*}$, $\hat{\mathcal{F}}_{\hat{\mathbf{A}}}$, and $\tilde{\mathcal{F}}_{\tilde{\mathbf{A}}}$ represent the expected best, empirical best, and computed fusers, respectively, wherein the index $\mathbf{A} = \mathbf{C} \in \{\mathbf{C}_H, \mathbf{C}_M, \mathbf{C}_L\}$ corresponds to an individual stream. We use $\mathbf{A} = \mathbf{F}_2, \mathbf{F}_3$ to denote the fusers with composite inputs. The *expected fuser improvement* is defined for $\mathbf{A} = \mathbf{C}_H, \mathbf{C}_M, \mathbf{C}_L, \mathbf{F}_2, \mathbf{F}_3$ and $\mathbf{A} = \{A_1, A_2, \ldots, A_{n_A}\}$ as

$$\Delta_{\mathcal{F}_{\mathbf{A}}} = \min_{i=1}^{n_A} I_{A_i} - I_{\mathcal{F}_{\mathbf{A}}},$$

which represents the difference between the expected error of the best component classifier or fuser and the fuser. The corresponding *empirical fuser improvement* is defined as

$$\hat{\Delta}_{\mathcal{F}_{\mathbf{A}}} = \min_{i=1}^{n_A} \hat{I}_{A_i} - \hat{I}_{\mathcal{F}_{\mathbf{A}}}.$$

The expected best, empirical best, and computed versions of the fuser chosen from fuser class $\Phi_{\mathcal{F}_{\mathbf{A}}}$ are denoted by $\mathcal{F}^*_{\mathbf{A}^*}, \hat{\mathcal{F}}_{\hat{\mathbf{A}}}$, and $\tilde{\mathcal{F}}_{\tilde{\mathbf{A}}}$, respectively.

We consider the *expected best fuser improvement* as

$$\Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}} = \min_{i=1}^{n_A} I_{A^*_i} - I_{\mathcal{F}^*_{\mathbf{A}^*}}, \tag{4.2}$$

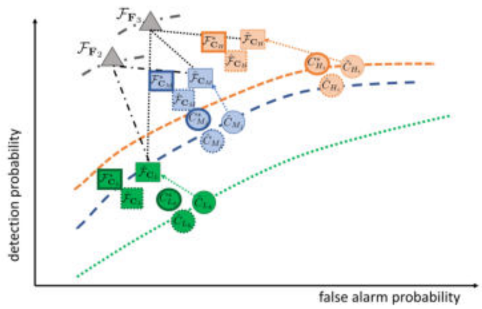


Fig. 3: Optimal expected C^*, optimal empirical $\hat{C}$, and computed $\tilde{C}$ versions for classifier/fuser $C \in \{C_{H_i}, C_{M_j}, C_{L_k}, \mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$.

which represents the difference between the expected error of the expected best classifiers or fusers and their expected best fuser. Under the isolation property of the class $\Phi_{\mathcal{F}_\mathbf{A}}$, it consists of a function that makes it identical to A_i^* for each i. Consequently, $\Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}}$ is non-negative since the expected error is minimized by the expected best classifiers and fusers.

C. Fusers Generalization Equations

We now derive the generalization equations of the computed fuser $\tilde{\mathcal{F}}_{\tilde{\mathbf{A}}}$, for $\mathbf{A} = \mathbf{C}_H, \mathbf{C}_M, \mathbf{C}_L, \mathbf{F}_2, \mathbf{F}_3$, trained with $l_{\mathcal{F}_\mathbf{A}}$ samples, of computed classifiers or fusers corresponding to set $\tilde{\mathbf{A}}$ each trained with $l_\mathbf{A}$ samples. The individual fuser $\tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_R}$, for $R = H, M, L$, is an ML-estimate of the fuser based on the ML-estimates $\tilde{\mathbf{C}}_R = \left\{\tilde{\mathbf{C}}_{R_1}, \tilde{\mathbf{C}}_{R_2}, \cdots, \tilde{\mathbf{C}}_{R_{n_R}}\right\}$ of the classifiers of $\mathbf{C}_R$. For example, $\tilde{\mathcal{F}}_{\{\tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_H}, \tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_M}, \tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_L}\}}$ is an ML-estimate of $\mathcal{F}_{\mathbf{F}_3}$ based on the training samples.

The closeness of the empirical best classifiers and fusers in terms of the best expected fuser improvement $\Delta^*_{\mathcal{F}^*_{\mathbf{C}^*}}$ is probabilistically guaranteed by the following result, which is a generalization of a result in [5].

Theorem 4.1: Generalization Equations of Fusers: Consider the computed fuser $\tilde{\mathcal{F}}_{\tilde{\mathbf{A}}}$, for $\mathbf{A} = \mathbf{C}_H, \mathbf{C}_M, \mathbf{C}_L, \mathbf{F}_2, \mathbf{F}_3$, trained with $l_{\mathcal{F}_\mathbf{A}}$ samples, of computed classifiers or fusers corresponding to set $\tilde{\mathbf{A}}$ each trained with $l_\mathbf{A}$ samples with empirical error $\tilde{\epsilon}$. Under the isolation properties of all fuser classes, we have, for $\mathbf{A} = \{A_1, A_2, \ldots, A_{n_A}\}$,

$$\mathbb{P}\left\{I_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}} - \min_{i=1}^{n_A} I_{A_i^*} + \Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}} > \epsilon + \tilde{\epsilon}\right\} < \delta_{\Delta_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}}(\epsilon, l_\mathbf{A}, l_{\mathcal{F}_\mathbf{A}}),$$

where $\tilde{\epsilon} = \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}$ such that $I_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}} = I_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} + \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}$ and

$$\delta_{\Delta_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}}(\epsilon, l_\mathbf{A}, l_{\mathcal{F}_\mathbf{A}}) = \delta_{\Phi_{\mathcal{F}_\mathbf{A}}}(\epsilon/2, l_{\mathcal{F}_\mathbf{A}}) + \sum_{i=1}^{n_A} \delta_{\Phi_{A_i}}(\epsilon/(2n_A), l_\mathbf{A}).$$

Proof: The condition $\left\{I_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}} - \min_{i=1}^{n_A} I_{A_i^*} + \Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}} > \epsilon + \tilde{\epsilon}\right\}$ is equivalent to the condition $\left\{I_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} - \min_{i=1}^{n_A} I_{A_i^*} + \Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}} > \epsilon\right\}$ since $I_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}} = I_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} + \tilde{\epsilon}$, where we suppressed the subscript of the empirical error term $\tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}$ to simplify the notation.

Then, the condition $\left\{I_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} - \min_{i=1}^{n_A} I_{A_i^*} + \Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}} > \epsilon\right\}$ is the same as $\left\{I_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} - I_{\mathcal{F}^*_{\mathbf{A}^*}} > \epsilon\right\}$, and this condition implies the condition on fuser and fused parts: $I_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} - I_{\mathcal{F}^*_{\mathbf{A}^*}} > \epsilon/2$ or, for some i, $I_{\hat{A}_i} - I_{A_i^*} > \epsilon/(2n_A)$. Thus, the probability of the latter condition upper bounds that of the former. We now obtain an upper bound for the probability of this condition by applying Eq. (4.1) for fuser class $\Phi_{\mathcal{F}_\mathbf{A}}$ and component class Φ_{A_i} with sample sizes $l_{\mathcal{F}_\mathbf{A}}$ and $l_\mathbf{A}$, respectively. This probability is upper bounded by the sum of probabilities of the two conditions, which are upper-bounded by $\delta_{\Phi_{\mathcal{F}_\mathbf{A}}}(\epsilon/2, l_{\mathcal{F}_\mathbf{A}})$ and $\delta_{\Phi_{A_i}}(\epsilon/(2n_A), l_\mathbf{A})$ for each i, respectively. $\square$

This theorem ensures that with confidence $1 - \delta_{\Delta_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}}(\epsilon, l_\mathbf{A}, l_{\mathcal{F}_\mathbf{A}})$, the expected error of the computed fuser $\tilde{\mathcal{F}}_{\tilde{\mathbf{A}}}$ is within the precision parameter $\epsilon + \tilde{\epsilon}$ of the best expected error $\min_{i=1}^{n_A} I_{A_i^*}$ reduced by fuser improvement $\Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}}$. This guarantee is distribution-free in that it is independent of the underlying distribution $\mathbb{P}_{X,Y}$, which could be quite complex. The lower the error $\tilde{\epsilon}$ due to computed fuser, the tighter will be the precision parameter. Also, the confidence improves in general with the increasing training sample sizes.

1) Fused-Fusers Generalization Equations: We obtain complete generalization equations of our fused-fusers by applying Theorem 4.1 in two steps. Using $\tilde{\mathcal{F}}_{\{\tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_H}, \tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_M}, \tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_L}\}} = \tilde{\mathcal{F}}_{\tilde{\mathbf{F}}_3}$, we first obtain

$$\mathbb{P}\left\{I_{\tilde{\mathcal{F}}_{\hat{\mathbf{F}}_3}} - \min_{R=H,M,L} I_{\mathbf{F}^*_{\mathbf{C}_R}} + \Delta^*_{\mathcal{F}^*_{\mathbf{F}_3^*}} > \epsilon + \tilde{\epsilon}\right\} < \delta_{\Delta_{\hat{\mathcal{F}}_{\hat{\mathbf{F}}_3}}}(\epsilon, l_{\mathbf{F}_3}, l_{\mathcal{F}_{\mathbf{F}_3}}),$$

where $\tilde{\epsilon} = \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{\mathbf{F}}_3}}$ such that $I_{\tilde{\mathcal{F}}_{\hat{\mathbf{F}}_3}} = I_{\hat{\mathcal{F}}_{\hat{\mathbf{F}}_3}} + \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{\mathbf{F}}_3}}$ and

$$\delta_{\Delta_{\hat{\mathcal{F}}_{\hat{\mathbf{F}}_3}}}(\epsilon, l_{\mathbf{F}_3}, l_{\mathcal{F}_{\mathbf{F}_3}}) = \delta_{\Phi_{\mathcal{F}_{\mathbf{F}_3}}}(\epsilon/2, l_{\mathcal{F}_{\mathbf{F}_3}}) + \sum_{R=H,M,L} \delta_{\Phi_{\mathcal{F}^*_{\mathbf{C}_R}}}\left(\epsilon/6, l_{\mathcal{F}_{\mathbf{C}_R}}\right).$$

Then, the second application gives us, for $R = H, M, L$,

$$\delta_{\Phi_{\mathcal{F}^*_{\mathbf{C}_R}}}\left(\epsilon/6, l_{\mathcal{F}_{\mathbf{C}_R}}\right) = \delta_{\Phi_{\mathcal{F}_{\mathbf{C}_R}}}(\epsilon/12, l_{\mathcal{F}_{\mathbf{C}_R}}) + \sum_{i=1}^{n_R} \delta_{\Phi_{\mathbf{C}^*_{R_i}}}\left(\epsilon/(12n_R), l_{\mathbf{C}_{R_i}}\right).$$

2) Fuser Improvement Estimates: The *best empirical fuser improvement* is given by

$$\hat{\Delta}_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}} = \min_{i=1}^{n_A} \hat{I}_{\hat{\mathbf{A}}_i} - \hat{I}_{\hat{\mathcal{F}}_\mathbf{A}},$$

for $\mathbf{A} = \mathbf{C}_H, \mathbf{C}_M, \mathbf{C}_L, \mathbf{F}_2, \mathbf{F}_3$, which represents the difference between the empirical error of the best of empirical best classifiers and their empirical best fuser. Under the isolation property of the class $\Phi_{\mathcal{F}_\mathbf{A}}$, both $\Delta^*_{\mathcal{F}^*_{\mathbf{C}^*}}$ and $\hat{\Delta}_{\hat{\mathcal{F}}_{\hat{\mathbf{C}}}}$ are non-negative, since the expected and empirical errors are minimized by the expected and empirical best fusers, respectively. These two quantities are not computable using only the computed classifiers $\tilde{A}_i$ and fusers for two different reasons: the former due to the approximate minimization of the empirical error by $\tilde{\mathcal{F}}$ and latter due to the lack of knowledge about $\mathbb{P}_{X,Y}$. Instead, we compute their computable version

$$\tilde{\Delta}_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}} = \min_{i=1}^{n_A} \hat{I}_{\tilde{A}_i} - \hat{I}_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}},$$

which is not guaranteed to be non-negative even under isolation property of the class $\mathbf{F}_{\mathcal{F}_{\mathbf{A}}}$, unlike $\Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}}$ and $\hat{\Delta}_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}$. It is shown to be closer to $\Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}}$ by using the empirically best estimates in [18], which provides the following result

$$\mathbb{P}\left\{\left|\Delta^*_{\mathcal{F}^*_{\mathbf{A}^*}} - \tilde{\Delta}_{\tilde{\mathcal{F}}_{\hat{\mathbf{A}}}}\right| > \epsilon + \tilde{\epsilon}\right\} < \delta_{\Delta_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}}(\epsilon, l_{\mathbf{A}}, l_{\mathcal{F}_{\mathbf{A}}}),$$

where $\tilde{\epsilon} = \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{\mathbf{A}}}}$.

D. Switched-Fuser Generalization Equations

We consider a switch $\mathcal{S}_{\{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}}$ whose output is that of fused-classifier $\mathcal{F}_{\mathbf{C}_R}$ at resolution $R = H, M, L$, a fraction of ρ_R times within a time window such that $\sum_{R=H,M,L} \rho_R = 1$.

Theorem 4.2: Generalization equations of switched-fuser: The computed switched-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{C}_H}, \tilde{\mathcal{F}}_{\mathbf{C}_M}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}$ uses the computed classifiers trained with $l_{\mathcal{F}_{\mathbf{C}_H}}$, $l_{\mathcal{F}_{\mathbf{C}_M}}$, and $l_{\mathcal{F}_{\mathbf{C}_L}}$ samples, respectively. The computed basic classifiers used by these fusers are from $\mathbf{C}_H$, $\mathbf{C}_M$, and $\mathbf{C}_L$, which are trained with $l_{\mathbf{C}_H}$, $l_{\mathbf{C}_M}$, and $l_{\mathbf{C}_L}$ samples, respectively. Under the isolation properties of fuser classes, we have

$$\mathbb{P}\left\{I_{\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{C}_H}, \tilde{\mathcal{F}}_{\mathbf{C}_M}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}} - I_{\mathcal{S}_{\{\mathcal{F}^*_{\mathbf{C}_H}, \mathcal{F}^*_{\mathbf{C}_M}, \mathcal{F}^*_{\mathbf{C}_L}\}}} + \Delta^*_{\mathcal{S}} > \epsilon + \tilde{\epsilon}_{\mathcal{S}}\right\} < \sum_{R=H,M,L} \delta_{\Delta_{\tilde{\mathcal{F}}_{\mathbf{C}_R}}}(\rho_R \epsilon, l_{\mathbf{C}_R}, l_{\mathcal{F}_{\mathbf{C}_R}})$$

where $\mathcal{S}_{\{\mathcal{F}^*_{\mathbf{C}_H}, \mathcal{F}^*_{\mathbf{C}_M}, \mathcal{F}^*_{\mathbf{C}_L}\}}$ is the switched-fuser classifier based on the expected best fusers of expected best basic classifiers, $\Delta^*_{\mathcal{S}} = \sum_{R=H,M,L} \rho_R \Delta^*_{\mathcal{F}^*_{\mathbf{C}_R}} \geq 0$, $\tilde{\epsilon}_{\mathcal{S}} = \sum_{R=H,M,L} \rho_R \tilde{\epsilon}_{\tilde{\mathcal{F}}_{\mathbf{C}_R}}$, and $\sum_{R=H,M,L} \rho_R = 1$.

Proof: We consider the decomposition

$$\epsilon + \tilde{\epsilon}_{\mathcal{S}} = \sum_{R=H,M,L} \rho_R \epsilon + \sum_{R=H,M,L} \rho_R \tilde{\epsilon}_{\tilde{\mathcal{F}}_{\mathbf{C}_R}}.$$

The condition

$$\left\{I_{\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{C}_H}, \tilde{\mathcal{F}}_{\mathbf{C}_M}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}} - I_{\mathcal{S}_{\{\mathcal{F}^*_{\mathbf{C}_H}, \mathcal{F}^*_{\mathbf{C}_M}, \mathcal{F}^*_{\mathbf{C}_L}\}}} + \Delta^*_{\mathcal{S}} > \epsilon + \tilde{\epsilon}_{\mathcal{S}}\right\}$$

implies

$$I_{\hat{\mathcal{F}}_{\mathbf{C}_R}} - I_{\mathcal{F}^*_{\mathbf{C}_R}} + \Delta^*_{\mathcal{F}^*_{\mathbf{C}_R}} > \rho_R \left(\epsilon + \tilde{\epsilon}_{\hat{\mathcal{F}}_{\mathbf{C}_R}}\right)$$

for $R = H, M, L$. Thus, the probability in the theorem is upper bounded by the sum of the probabilities of the above term, which is upper-bounded by $\delta_{\Delta_{\tilde{\mathcal{F}}_{\mathbf{C}_R}}}(\rho_R \epsilon, l_{\mathbf{C}_R}, l_{\mathcal{F}_{\mathbf{C}_R}})$ for $R = H, M, L$, obtained by applying Theorem 4.1 to classifier sets $\mathbf{C}_H$, $\mathbf{C}_M$, and $\mathbf{C}_L$. □

Consider the computed switched-fuser improvement $\tilde{\Delta}_{\mathcal{S}} = \sum_{R=H,M,L} \rho_R \tilde{\Delta}_{\tilde{\mathcal{F}}_{\mathbf{C}_R}}$. By following the approach used in Theorem 4.1, we obtain

$$\mathbb{P}\left\{\left|\Delta^*_{\mathcal{S}} - \tilde{\Delta}_{\mathcal{S}}\right| > \epsilon + \tilde{\epsilon}_{\mathcal{S}}\right\} < \sum_{R=H,M,L} \delta_{\Delta_{\tilde{\mathcal{F}}_{\mathbf{C}_R}}}(\rho_R \epsilon, l_{\mathbf{C}_R}, l_{\mathcal{F}_{\mathbf{C}_R}}),$$

which characterizes the closeness between the computed and expected best fuser improvements.

We consider two special cases of this theorem:
(a) The switched-classifier $\mathcal{S}_{\{C_H, C_M, C_L\}}$ corresponds to choosing the fuser class with an identity function and the classifier set with a single classifier, $\Phi_{\mathcal{F}_{\mathbf{C}_R}} = \{I\}$, $\mathbf{C}_R = \{C_R\}$, for $R = H, M, L$. Consequently, $\Delta^*_{\mathcal{F}^*_{\mathbf{C}_R}} = \tilde{\Delta}_{\tilde{\mathcal{F}}_{\mathbf{C}_R}} = 0$, for $R = H, M, L$, and $\Delta^*_{\mathcal{S}} = \tilde{\Delta}_{\mathcal{S}} = 0$.
(b) The switched-fused-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{F}_3}, \tilde{\mathcal{F}}_{\mathbf{F}_2}, \tilde{\mathcal{F}}_{\tilde{\mathbf{C}}_L}\}}$ corresponds to choosing classifier sets as fused-classifiers such that $\mathbf{C}_H = \mathbf{F}_3$, and $\mathbf{C}_M = \mathbf{F}_2$, since fused-classifiers are also classifiers.

E. Comparison of Switched-Classifiers and Switched-Fusers

As measurements of different streams arrive, a classifier and a fuser are used to produce output by switched-classifier and switched-fuser, respectively. Using only classifiers, switched-classifier $\mathcal{S}_{\{\tilde{C}_{H_{\min}}, \tilde{C}_{M_{\min}}, \tilde{C}_{L_{\min}}\}}$ outputs those of classifiers with the lowest empirical error, which approximates the best performance provided by $\mathcal{S}_{\{C^*_{H_{\min}}, C^*_{M_{\min}}, C^*_{L_{\min}}\}}$. The switched-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{C}_H}, \tilde{\mathcal{F}}_{\mathbf{C}_M}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}$ outputs those of the empirically best fusers of individual streams, which approximates the best performance provided by $\mathcal{S}_{\{\mathcal{F}^*_{\mathbf{C}_H}, \mathcal{F}^*_{\mathbf{C}_M}, \mathcal{F}^*_{\mathbf{C}_L}\}}$. Furthermore, the switched-fused-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{F}_3}, \tilde{\mathcal{F}}_{\mathbf{F}_2}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}$ outputs those of the empirically best of fused-fusers, for $\mathbf{F}_3 = \{\mathcal{F}_{\mathbf{C}_H}, \mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$ and $\mathbf{F}_2 = \{\mathcal{F}_{\mathbf{C}_M}, \mathcal{F}_{\mathbf{C}_L}\}$, which approximates the best performance provided by $\mathcal{S}_{\{\mathcal{F}^*_{\mathbf{F}_3}, \mathcal{F}^*_{\mathbf{F}_2}, \mathcal{F}^*_{\mathbf{C}_L}\}}$. Under the isolation property of all component fuser classes, we have the following conditions on the expected errors

$$I_{\mathcal{F}^*_{\mathbf{F}_3}} \leq \min\{I_{\mathcal{F}^*_{\mathbf{C}_H}}, I_{\mathcal{F}^*_{\mathbf{C}_M}}, I_{\mathcal{F}^*_{\mathbf{C}_L}}\}$$

$$I_{\mathcal{F}^*_{\mathbf{F}_3}} \leq I_{\mathcal{F}^*_{\mathbf{F}_2}} \leq \min\{I_{\mathcal{F}^*_{\mathbf{C}_M}}, I_{\mathcal{F}^*_{\mathbf{C}_L}}\}$$

$$I_{\mathcal{F}^*_{\mathbf{C}^*_R}} \leq I_{C^*_{R_{\min}}} = \min\{I_{C^*_{R_1}}, I_{C^*_{R_2}}, \ldots, I_{C^*_{R_{n_R}}}\},$$

for $R = H, M, L$. These conditions are satisfied by the corresponding empirical estimates with the confidences specified by Theorem 4.1. By using $\leq_\delta$ to denote the inequality satisfied under the uncertainty function δ, the above inequalities are satisfied using their counterpart empirical estimates. Thus, by using Theorem 4.2, we have the following conditions on the expected errors

$$I_{\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{F}_3}, \tilde{\mathcal{F}}_{\mathbf{F}_2}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}} \leq_\delta I_{\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{C}_H}, \tilde{\mathcal{F}}_{\mathbf{C}_M}, \tilde{\mathcal{F}}_{\mathbf{C}_L}\}}} \leq_\delta I_{\mathcal{S}_{\{\tilde{C}_{H_{\min}}, \tilde{C}_{M_{\min}}, \tilde{C}_{L_{\min}}\}}},$$

wherein confidence functions are specified by Theorems 4.1 and 4.2. We follow this approach to obtain the generalization equation for the switched-classifier $\mathcal{S}_{\{\tilde{C}_{H_i}, \tilde{C}_{M_j}, \tilde{C}_{L_k}\}}$ by using decomposition $\tilde{\epsilon}_{\mathcal{S}_{\{\tilde{C}_{H_i}, \tilde{C}_{M_j}, \tilde{C}_{L_k}\}}} = \rho_i \tilde{\epsilon}_{\tilde{C}_{H_i}} + \rho_j \tilde{\epsilon}_{\tilde{C}_{M_j}} + \rho_k \tilde{\epsilon}_{\tilde{C}_{L_k}}$ as

$$\mathbb{P}\left\{I_{\mathcal{S}_{\{\tilde{C}_{H_i}, \tilde{C}_{M_j}, \tilde{C}_{L_k}\}}} - I_{\mathcal{S}_{\{C^*_{H_i}, C^*_{M_j}, C^*_{L_k}\}}} > \epsilon + \tilde{\epsilon}_{\mathcal{S}_{\{\tilde{C}_{H_i}, \tilde{C}_{M_j}, \tilde{C}_{L_k}\}}}\right\} < \sum_{R=H,M,L} \delta_{\Phi_{C_R}}(\rho_R \epsilon, l_{\mathbf{C}_R}),$$

which shows the improvement $\Delta^*_{\mathcal{S}}$ due to the fusers of the switched-fuser in Theorem 4.2.

V. Radiochemical Dissolutions Classification

At a radiochemical processing facility, radioactive Pu-238 is produced by chemically dissolving irradiated Np-237 targets. We consider a practical task of detecting these dissolution events using gamma spectral measurements (detailed descriptions in [6], [16]), which is an important part of facility analytics needed for operational and non-proliferation purposes. The gamma spectra of the effluents are collected by a High Purity Germanium (HPGe) detector located at the facility's off-gas stack, which contain spectral signatures of the materials produced by dissolutions.

The measurements are aggregated every 1, 3, and 6 hours, and are processed and analyzed to estimate 15 isotope counts in their gamma spectral regions, which are used as features of classifiers to detect Pu-238 material. The 15 isotopes are five iodine isotopes, I-131, I-132, I-133, I-134 and I-135; four krypton isotopes, Kr-85, Kr-87, Kr-88 and Kr-89; four xenon isotopes, Xe-135, Xe-135m, Xe-137 and Xe-138; one barium isotope Ba-138; and one cesium isotope Cs-138. These gamma counts are Poisson distributed with means that depend on the target being dissolved. Previous works used features based on single aggregation periods with subsets of isotope counts in [6], [16] and their ratios in [17], and two periods in [5].

A. Classifiers' Performance

Different classifiers (as in [16]) for Pu/Np target signatures are trained, with the lowest quality measurements every 1 hour, medium quality measurements every 3 hours, and the highest quality measurements every 6 hours. Eight ML classifiers are trained, namely, Classification Trees (CT), Discriminant Analysis Classifier (DAC), Error Correcting Output Codes (ECOC), Ensemble of Trees (EOT), Kernel Method (KM), k Nearest Neighbors (kNN), Naive Bayes (NB), and Support Vector Machine (SVM). Thus, we have $\mathbf{C} = \{$ CT, DAC, ECOC, EOT, KM, kNN, NB, SVM $\}$. These are chosen to represent the design diversity, namely, smooth (KM and SVM) and non-smooth (CT, EOT and kNN), and statistical (DAC, ECOC, NB) and structural (kNN), since there is no single best classification method under finite samples [11].

B. Fused-Classifiers and Fused-Fusers' Performance

As the non-smooth classifiers CT, EOT and kNN perform significantly better than others, they are fused using the ensemble fuser (EOT-F) which satisfies the isolation property. The ROC plots and OPs of classifiers and fusers trained using measurements at three time resolutions are shown in Fig. 4. Overall, measurements at coarser time resolution have higher quality as indicated by the performance of the corresponding classifiers as well as their fusers, indicated left to right in Fig. 4. The OPs of classifiers, fused-classifiers and fused-fusers at three time resolutions are shown in Fig. 5. For all three streams, fused-classifier's OP is inside the enhancement regions of the component classifiers. The classification errors of classifiers of individual streams with the lowest computed errors, as shown in Fig. 5(a), are listed in Table I. The classification error is lowered or the same by fused-classifiers of individual streams, and further lowered by their fused-fusers as shown in Fig. 5(b) and Table I, in agreement with Theorem 4.1.

time	classifier	fused-classifiers	fused-fusers
1 hour	27.27%	25.57%	25.57%
3 hours	10.79%	10.79%	0%
6 hours	9.03%	7.43%	0%

TABLE I: Classification errors of best classifiers, fused-classifiers, and fused-fusers.

switched components	classification error
best classifiers	7.19%
fused-classifiers	6.76%
fused-fusers	5.65%

TABLE II: Switched classifiers and fusers.

C. Switched-Classifier and Switched-Fuser's Performance

The switched-classifiers and switched-fusers provide output at three different resolutions by appropriately switching the output of a component classifier or fuser. The classifiers with the lowest computed error $\tilde{C}_{R_{\min}}$ for stream $R = H, M, L$ are utilized by the switched-classifier $\mathcal{S}_{\{C_{H_{\min}}, C_{M_{\min}}, C_{L_{\min}}\}}$. The top three classifiers of each resolution are fused and then switched by switched-fusers $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\tilde{C}_H}, \tilde{\mathcal{F}}_{\tilde{C}_M}, \tilde{\mathcal{F}}_{\tilde{C}_L}\}}$, whose performance is superior to the above switched-classifier, as shown in Fig. 6. Then, the fused-fusers at resolutions H and M are used by switched-fused-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{F}_3}, \tilde{\mathcal{F}}_{\mathbf{F}_2}, \tilde{\mathcal{F}}_{\tilde{C}_L}\}}$, which has a lower classification error, as indicated by Eq (4.1)-(4.2). This analytical result is confirmed by OPs shown in Fig. 6, and their computed errors shown in Table II.

Theorems 4.1 and 4.2 analytically characterize the performance of various fusers and switched versions, and also provide practical guidance. Among classifiers, those with lower training errors are associated with smaller $\tilde{\epsilon}$ values in Theorem 4.1, and hence have stronger performance guarantees. As a result, they are used as inputs to fusers, subsequently in fused-fusers, and finally in switched-fused-fuser, which provides performance superior to all classifiers, switched-classifiers, fused-classifiers, and fused-fusers at the finer time resolution.

VI. Conclusion and Future Works

In practical multiple-sensor systems, the quality and rates of measurements may vary significantly [7]. We formulated an abstract sensor fusion problem using three measurement streams with their quality inversely correlated to time resolution. We proposed a fused-fusers method that utilizes classifiers and fusers for each type, and fuses and switches the fusers in time dimension to provide classification decisions at three time resolutions. We analytically showed superior classification probabilities for fused and switched ML classifiers, with a confidence probability specified by their generalization equations. These results provide analytical foundations to solutions for a classification task associated with target dissolution events at a radiochemical processing facility.

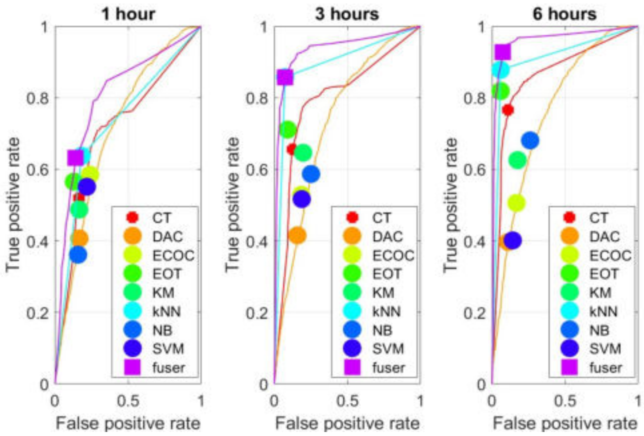
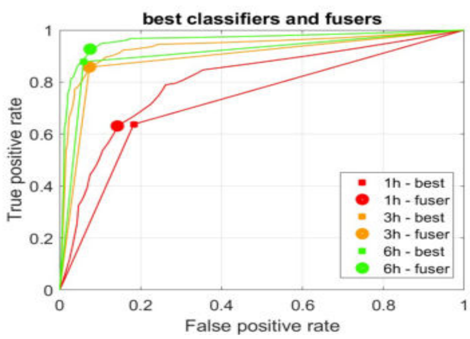
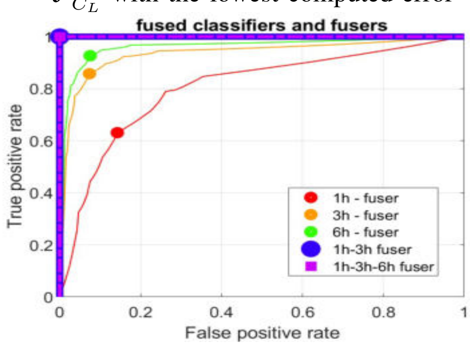


Fig. 4: Operating points of classifiers including $\tilde{C}_{H_{\min}}$, $\tilde{C}_{M_{\min}}$, $\tilde{C}_{L_{\min}}$ with the lowest computed errors, and fusers $\tilde{\mathcal{F}}_{\tilde{C}_H}$, $\tilde{\mathcal{F}}_{\tilde{C}_M}$, $\tilde{\mathcal{F}}_{\tilde{C}_L}$ at 1, 3, and 6 hours time resolutions. ROC plots are shown for the fusers, the best and two other classifiers.



(a) classifiers $\tilde{C}_{H_{\min}}$, $\tilde{C}_{M_{\min}}$, $\tilde{C}_{L_{\min}}$, and fusers $\tilde{\mathcal{F}}_{\tilde{C}_H}$, $\tilde{\mathcal{F}}_{\tilde{C}_M}$, $\tilde{\mathcal{F}}_{\tilde{C}_L}$ with the lowest computed error



(b) fused-classifiers $\tilde{\mathcal{F}}_{\tilde{C}_H}$, $\tilde{\mathcal{F}}_{\tilde{C}_M}$, $\tilde{\mathcal{F}}_{\tilde{C}_L}$, and fused-fusers $\tilde{\mathcal{F}}_{\{\tilde{\mathcal{F}}_{\tilde{C}_M},\tilde{\mathcal{F}}_{\tilde{C}_L}\}}$ and $\tilde{\mathcal{F}}_{\{\tilde{\mathcal{F}}_{\tilde{C}_H},\tilde{\mathcal{F}}_{\tilde{C}_M},\tilde{\mathcal{F}}_{\tilde{C}_L}\}}$ with the lowest computed error

Fig. 5: Operating points of the best classifiers, fused-classifiers and fused-fusers.

Future directions include generalizations of this method to $N > 3$ measurement streams, and their application to other practical scenarios. Also, hybrid scenarios with a combination of sample-trained ML classifiers and Bayesian classifiers derived under known distributions would be of future interest.

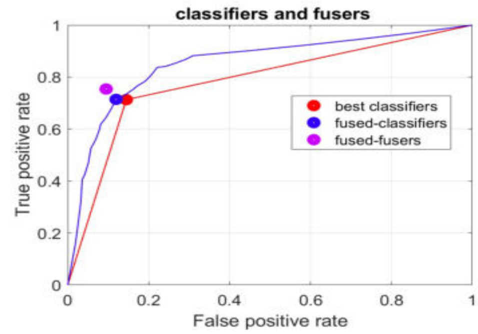


Fig. 6: Operating points of switched-classifier $\mathcal{S}_{\{\tilde{C}_{H_{\min}},\tilde{C}_{M_{\min}},\tilde{C}_{L_{\min}}\}}$, switched-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\tilde{C}_H},\tilde{\mathcal{F}}_{\tilde{C}_M},\tilde{\mathcal{F}}_{\tilde{C}_L}\}}$, and switched-fused-fuser $\mathcal{S}_{\{\tilde{\mathcal{F}}_{\mathbf{F}_3},\tilde{\mathcal{F}}_{\mathbf{F}_2},\tilde{\mathcal{F}}_{\tilde{C}_L}\}}$.

References

[1] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2020, fourth edition.

[2] K. P. Murphy, *Probabilistic Machine Learning: An introduction*. MIT Press, 2022.

[3] K. Woods, W. P. Kegelmeyer, and K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 405–410, 1997.

[4] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, pp. 66–75, 1994.

[5] N. S. V. Rao, C. Y. T. Ma, and F. He, "Classification and fusion of two disparate data streams and nuclear dissolutions application," in *International Conference on Information Fusion*, 2022.

[6] N. S. V. Rao, C. Greulich, S. Sen, K. Dayman, A. Nicholson, M. R. Chatin, K. M. Buckley, R. D. Hunley, J. Johnson, H. H. Hesse, and R. Hale, "Classifiers for dissolution events in processing facility using effluents measurements," in *Institute of Nuclear Materials Management Annual Meeting*, 2019.

[7] B. Khaleghi, A. Khamis, and F. O. Karray, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, 2011.

[8] L. Yan, L. Jiang, Y. Xia, and M. Fu, "State estimation and data fusion for multirate sensor networks," *International Journal of Adpative Control and Signal Processing*, vol. 30, pp. 3–15, 2016.

[9] X. Du and A. Zare, "Multiresolution multimodal sensor fusion for remote sensing data with label uncertainty," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2755–2769, 2020.

[10] R. F. Brena, A. A. Aguileta, L. A. Trejo, E. Molino-Minero-Re, and O. Mayora, "Choosing the best sensor fusion method: A machine-learning approach," *Sensors*, vol. 20, no. 8, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/8/2350

[11] L. Devroye, L. Gyorfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.

[12] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2018, second edition.

[13] V. N. Vapnik, *Statistical Learning Theory*. New York: John-Wiley and Sons, 1998.

[14] N. S. V. Rao, "On fusers that perform better than best sensor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 904–909, 2001.

[15] ——, "A generic sensor fusion problem: classification and function estimation," in *Multiple Classifier Systems*, F. Roli, Ed., 2004.

[16] N. S. V. Rao, C. Greulich, S. Sen, J. Hite, K. J. Dayman, A. D. Nicholson, D. E. Archer, M. J. Willis, I. G. R. D. Hunley, J. Johnson, A. J. Rowe, I. R. Stewart, and J. M. Ghawaly, "Classification of dissolution events using fusion of effluents measurements and classifiers," in *Institute of Nuclear Materials Management Annual Meeting*, 2020.

[17] N. S. V. Rao, C. Greulich, M. P. Dion, J. Hite, K. J. Dayman, A. D. Nicholson, D. E. Archer, M. J. Willis, J. M. Ghawaly, I. G. R. D. Hunley, and J. Johnson, "Isotope ratio features for classification of dissolution events using effluents measurements," in *Institute of Nuclear Materials Management Annual Meeting*, 2021.

[18] N. S. V. Rao, S. Sen, Z. Liu, R. Kettimuthu, and I. Foster, "Learning concave-convex profiles of data transport over dedicated connections," in *Machine Learning for Networking*, E. Renault, P. Muhlethaler, and S. Bourmerdassi, Eds. Lecture Notes in Computer Science 11407, Springer, 2019.