EasyChair Preprint

№ 8172

Classification and Fusion of Two Disparate Data Streams and Nuclear Dissolutions Application

Nageswara Rao, Yu Tak Ma and Fei He

June 1, 2022

Classification and Fusion of Two Disparate Data Streams and Nuclear Dissolutions Application

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 two streams of data or measurements with disparate qualities and time resolutions that need to be classified. The first stream consists of higher quality data at a coarser time resolution, and the other consists of lower quality data at a finer time resolution. We present a fuser-switch method that fuses the set of classifiers of each stream separately and switches between them. We show that this method provides classification decisions at a finer time resolution with superior detection and false alarm probabilities compared to individual classifiers, under the statistical independence and time resolution ratio conditions. When classifiers are trained using machine learning methods, we show that this superior performance is guaranteed with a confidence probability specified by the classifiers' generalization equations. We use these results to provide analytical foundations for previous practical results that achieved significant performance improvements in classifying Pu/Np target dissolution events at a radiochemical processing facility.**

***Index Terms*—classifier, fuser, generalization equation, ROC, time resolution, statistical independence**

I. Introduction

We consider a scenario of two separate streams of data or measurements with different time resolutions (or rates), wherein a classification decision is required upon arrival of each data or measurement. The data at the coarser time resolution (or lower rate) are of higher quality, and are classified by the set $\mathcal{A}$ of binary classifiers. At the finer time resolution (or higher rate), data are of lower quality and are classified by the set $\mathcal{B}$ of binary classifiers. The receiver operating characteristic (ROC) curves of $\mathcal{A}$-classifiers are overall higher than those of $\mathcal{B}$-classifiers, but their classification outputs are provided at the coarser time resolution or less frequently. We address the problem of classifying the data of either stream as they arrive by "combining" the outputs of classifiers. This formulation is a specific abstraction of the classification [1], [2] and classifier fusion problem [3], [4]. It is motivated by the nuclear dissolutions application [5], and is a special case of the well-studied information fusion problem [6] applied to multi-rate [7] and multi-resolution [8] sensors or data sources.

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).

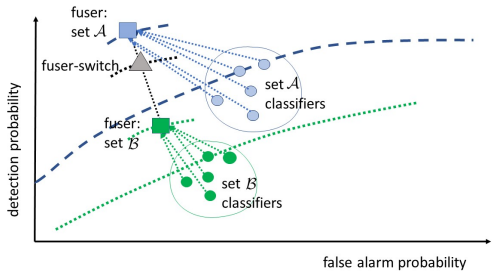


Fig. 1: Fuser-switch utilizes fusers of individual classifier sets and switches between them for improved ROC performance.

A straight-forward method is to choose the "best" classifier from each set and use them to classify the corresponding data items as they arrive. We present a superior approach that provides the classification performance better than the best classifier from each set with quantified confidence. We employ a *fuser-switch* method that first fuses the classifiers within sets $\mathcal{A}$ and $\mathcal{B}$, and uses the fused classifiers to classify the corresponding items of coarse and fine time resolution, as illustrated in Fig. 1. We show that this method provides superior detection and false alarm probabilities compared to the best classifiers of $\mathcal{A}$ and $\mathcal{B}$ with confidence probabilities.

We analyze the performance of this method in two steps. First, under (ideal) statistical independence conditions, we utilize the classical fusion results (Condorcet Jury theorem [9] and Chow's fuser [10]) to establish the superior performance under sufficiently high percentage of the higher quality, coarser time resolution data. The solutions in this case require a complete knowledge of the detection and false alarm probabilities. Second, under more general conditions, we consider that the classifiers and fusers are selected using machine learning (ML) methods [11] that utilize training sets. Consequently, the statistical independence property is not guaranteed and the underlying detection and false alarm probabilities are known within confidence bounds based on their generalization equations [12], [13]; in particular, the best classifier can only be approximately identified [14]. We derive the conditions for the superior performance of our fuser-switch method using fusers that satisfy the isolation property [15], and derive the confidence probabilities using distribution-free generalization equations of the classifiers that characterize their performance [16]. For this problem, the existing multi-rate, multi-resolution methods require the knowledge of the underlying system or distributions models [6]–[8], and ML

methods without those requirements [11] do not provide the needed generalization equations.

We apply these results to provide analytical insights and justification for a solution that was developed for the classification of events associated with the Plutonium/Neptunium (Pu/Np) targets being dissolved at a radiochemical facility using gamma spectral measurements of effluent streams [5], [17], [18]. In this scenario, previous experiments showed very promising results that when a diverse set of 8 classifiers are fused and switched based on measurement time-windows, the classification performance is significantly improved, from 70% detection rate at 20% false alarm rate in [5] to 95% detection rate at 5% false alarm rate in [17]. These practical results motivated our two-step analytical study of this approach: ideal cases provide basic insights into the performance improvements using classical results, and their generalizations to correlated, finite sample conditions make them applicable to more recent multiple (smooth, non-smooth, statistical and structural) ML classifiers that are being increasingly used in several applications [1], [12].

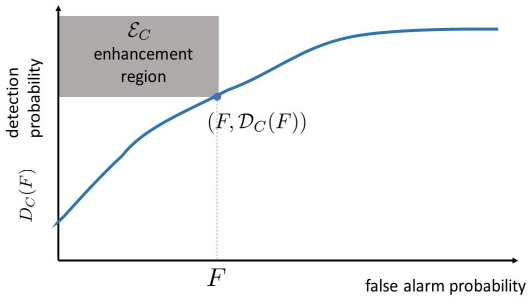
The organization of the paper is as follows. The problem formulation is described in Section II, and our overall approach is described in Section III. The performance equations are derived under the statistical independence conditions in Section IV. Performance equations of the fuser-switch classifiers are presented in Section V. The experimental results of dissolution classification problem are related to the analytical results in Section VI. Conclusions and directions for future work are presented in Section VII.

II. Problem Formulation

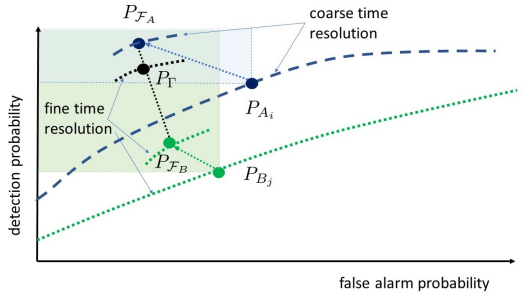
A binary classifier $C : \Re^d \mapsto \{0, 1\}$ maps the input $X \in \Re^d$ to Boolean output $Y = C(X) \in \{0, 1\}$. We consider that two separate sets of classifiers are used for two data streams: a set of classifiers $\mathcal{A} = \{A_1, A_2, \ldots, A_{n_A}\}$ handles data items at a coarser time resolution with time interval t_A, and a complementary set of classifiers $\mathcal{B} = \{B_1, B_2, \ldots, B_{n_B}\}$ handles data items at a finer time resolution with time interval $t_B < t_A$. The rates of classification output from classifiers $A \in \mathcal{A}$ and $B \in \mathcal{B}$ and $\Gamma_{A,B}$ are $1/t_A$, $1/t_B$, respectively. The *rate fraction* of $\mathcal{A}$ is $\rho = \frac{1/t_A}{(1/t_A+1/t_B)} = t_B/(t_A + t_B)$ such that a larger value represents more frequent higher quality data and vice versa. A classification decision is required at the arrival of every data or measurement, namely, at a minimum time interval $t_A t_B/(t_A + t_B)$. For discussions that apply to both sets of classifiers, we use a generic set of classifiers $\mathcal{C} = \{C_1, C_2, \ldots, C_{n_C}\}$.

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

$$D_{C_i}(F) \geq D_{C_j}(F).$$



(a) improvement region of a classifier



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

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

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 the classifier set $\mathcal{A}$ to be superior to $\mathcal{B}$ if A_i is superior to B_j for all $i = 1, 2, \ldots, n_A$, and $j = 1, 2, \ldots, n_B$. Let Γ_{C_i,C_j} denote a *switch* classifier that outputs those of two classifiers C_i and C_j corresponding to their individual data as their arrive. $1/t_A$, $1/t_B$ and $1/t_A + 1/t_B$, respectively. The rate of classification output from $\Gamma_{A,B}$ is $1/t_A + 1/t_B$, since it outputs every time its either constituent classifier outputs. The detection and false alarm probabilities of $\Gamma_{A,B}$ are

$$D_{\Gamma_{A,B}} = \rho D_A + (1-\rho) D_B = (t_B D_A + t_A D_B)/(t_A + t_B),$$

$$F_{\Gamma_{A,B}} = \rho F_A + (1-\rho) F_B = (t_B F_A + t_A F_B)/(t_A + t_B),$$

respectively, for $\rho \in [0, 1]$.

III. Classifiers: Enhancement and Fusion

Our method consists of fusing the outputs of classifier sets $\mathcal{A}$ and $\mathcal{B}$ using fusers $\mathcal{F}_A$ and $\mathcal{F}_B$, respectively, and reporting the output at the corresponding time resolution. The overall classification result is produced at the finer time resolution by the *fuser-switch* classifier $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$, which is at least as frequent as that of classifiers in $\mathcal{B}$.

A. Classifier Enhancements: Overall Approach

Our overall approach is based on developing both fusers $\mathcal{F}_A$ and $\mathcal{F}_B$ that are superior to the corresponding individual classifiers, and then utilizing a fuser-switch classifier $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$ with suitable ρ value, as illustrated in Fig. 1. The superiority is characterized with probability 1 by utilizing the statistical

$\mathcal{C} = \{C_1, C_2, \ldots, C_{n_C}\}$	: generic set of classifiers/fusers
C^*, $\hat{C}$ and $\tilde{C}$	: expected best, empirical best and ML estimate of classifier C, respectively
F_C	: false alarm probability of classifier C
$\mathcal{D}_C(F)$	: detection probability of classifier C at false alarm probability F
$\mathcal{E}_C$	: enhancement region of C at F
$P_C = (F_C, \mathcal{D}_C(F_C)) = (F_C, \mathcal{D}_C)$	: operating point of classifier C
I_C and $\hat{I}_C$	: expected and empirical error of classifier C, respectively
$\mathbf{F}_C$	: function class of ML method for classifier C
$\mathcal{A} = \{A_1, A_2, \ldots, A_{n_A}\}$	: set of coarse resolution classifiers
$\mathcal{B} = \{B_1, B_2, \ldots, B_{n_B}\}$	: set of fine resolution classifiers
t_A and t_B	: interval of coarse and fine time resolution, respectively
$\rho = \frac{1/t_A}{(1/t_A+1/t_B)}$	: rate fraction of $\mathcal{A}$
$\mathcal{F}_A$ and $\mathcal{F}_B$	: fuser of classifier sets $\mathcal{A}$ and $\mathcal{B}$, respectively
$\Gamma_{A,B}$	: generic switched version of classifiers A and B
$\Gamma_{\tilde{\mathcal{F}}_A,\tilde{\mathcal{F}}_B}$	: fuser-switch of computed fusers of classifier sets $\mathcal{A}$ and $\mathcal{B}$
$\Gamma_{\hat{\mathcal{F}}_A,\hat{\mathcal{F}}_B}$	: fuser-switch of empirical best fusers of classifier sets $\mathcal{A}$ and $\mathcal{B}$
$\Gamma_{\mathcal{F}_A^*,\mathcal{F}_B^*}$	: fuser-switch of expected best fusers of classifier sets $\mathcal{A}$ and $\mathcal{B}$

TABLE I: Notation

independence conditions and known detection and false alarm probabilities in the ideal case using the classical results (Section IV), and with confidence probability $1-\delta$ based on the isolation property under more general conditions using more recent finite sample results (Section V).

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 overall approach is to utilize combinations of fusers and switching to improve the performance within each set such that the resultant operating point is within the intersection of enhancement regions of $\mathcal{A}$ and $\mathcal{B}$, as illustrated in Fig. 2(b). For instance, outputs of the two fusers are switched by $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$ so that their detection and false alarm probabilities are linear combinations with coefficients that reflect the rate fraction ρ. For a suitably high ρ, the operating point of $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$ can be made to lie within the intersection of enhancement regions of both fusers with certain probability at the expense of more higher quality data.

B. Fused Classifiers

The fusers $\mathcal{F}_A$ and $\mathcal{F}_B$ are considered to be I, J-superior to the corresponding individual classifier subsets if for $i \in I$ and $j \in J$,

$$D_{\mathcal{F}_A} \geq D_{A_i} \quad \text{and} \quad F_{\mathcal{F}_A} \leq F_{A_i}$$

$$D_{\mathcal{F}_B} \geq D_{B_j} \quad \text{and} \quad F_{\mathcal{F}_B} \leq F_{B_j}$$

with certain probability; they are simply called superior if the subsets are the entire classifier sets. When all classifiers have an identical operating point, the above are true with probability 1 for all $i = 1, 2, \ldots, n_A$ and $j = 1, 2, \ldots, n_B$, under the statistical independence condition (Section IV-A). They are similarly satisfied but with a confidence probability if a fuser $\mathcal{F} : [0,1]^n \mapsto [0,1]$ is chosen from a class $\mathbf{F}_{\mathcal{F}}$, as in the case of ML methods, under the isolation property [15] such that it contains $\mathcal{F}_i, i = 1, 2, \ldots, n$, where $\mathcal{F}_i(X) = x_i$ for $X = (x_1, x_2, \cdots, x_n)$. Analytical bounds are derived for the confidence probability of $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$ using the generalization equations of the classifiers in Section V.

C. Switching Classifiers

The conditions for $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$ to be in the intersection of enhancement regions of the best individual classifiers are

$$D_{\Gamma_{\mathcal{F}_A,\mathcal{F}_B}} \geq \max\{D_{A^*}, D_{B^*}\}$$

and

$$F_{\Gamma_{\mathcal{F}_A,\mathcal{F}_B}} \leq \min\{F_{A^*}, F_{B^*}\},$$

where $D_{A^*} = \max_i D_{A_i}$, $D_{B^*} = \max_j D_{B_j}$, $F_{A^*} = \min_i F_{A_i}$, and $F_{B^*} = \min_j D_{B_j}$.

This detection probability condition depends on the rate fraction ρ, and is given by

$$\rho_D D_{\mathcal{F}_A} + (1-\rho_D) D_{\mathcal{F}_B} \geq \max\{D_{A^*}, D_{B^*}\}$$

or equivalently $\rho_D \geq \frac{\max\{D_{A^*}, D_{B^*}\} - D_{\mathcal{F}_B}}{D_{\mathcal{F}_A} - D_{\mathcal{F}_B}}$. The corresponding false alarm probability condition is given by

$$\rho_F \leq \frac{\min\{F_{A^*}, F_{B^*}\} - F_{\mathcal{F}_B}}{F_{\mathcal{F}_A} - F_{\mathcal{F}_B}}.$$

These individual detection and false alarm probabilities correspond to different time resolutions, which should be taken into account when comparing them. In general, these two conditions may not be simultaneously satisfied, but they are under some statistical independence conditions, as shown in the next section.

IV. Statistical Independence Conditions

We now consider simple cases to illustrate the basic concepts behind the performance of $\Gamma_{\mathcal{F}_A,\mathcal{F}_B}$ when the classifiers are statistically independent, namely, for two classifiers C_i and

C_j, the joint detection and false alarm probabilities are given by $D_{C_i}D_{C_j}$ and $F_{C_i}F_{C_j}$, respectively.

A. Identical Classifiers

We consider classifiers with identical detection and false alarm probabilities, denoted by $D_A > 1/2$ and $F_A < 1/2$, respectively for $\mathcal{A}$, and $D_B > 1/2$ and $F_B < 1/2$, respectively for $\mathcal{B}$; their operating points are denoted by $P_A = (F_A, D_A)$ and $P_B = (F_B, D_B)$. We use the majority fusers M_A and M_B for $\mathcal{A}$ and $\mathcal{B}$, respectively. We apply the classical Condorcet Jury theorem [19] separately for the detection and false alarm probabilities. For set $\mathcal{C} = \mathcal{A}, \mathcal{B}$, and classifier $C \in \mathcal{C}$, we have

$$D_{M_C} > D_C \text{ and } F_{M_C} < F_C, \text{ and furthermore,}$$
$$D_{M_C} \to 1 \text{ and } F_{M_C} \to 0 \text{ as } n_C \to \infty.$$

The first property is referred to as the *increasing probability* which ensures that the fuser's detection probability is larger and false alarm probability is smaller than individual classifier, respectively, when n_C is larger. The second weaker property is referred to as the *asymptotic improvement* which ensures that the detection probability goes to 1 and false alarm probability goes to 0 when n_C goes to infinity. The first is often called the non-asymptotic part and the second is often called the asymptotic part of the jury theorem.

The application of the *increasing probability* property to the detection and false alarm probabilities results in the fusers' operating point in the enhancement regions of the corresponding classifier set. Then, the detection and false alarm probabilities of the fuser-switch classifier are given by

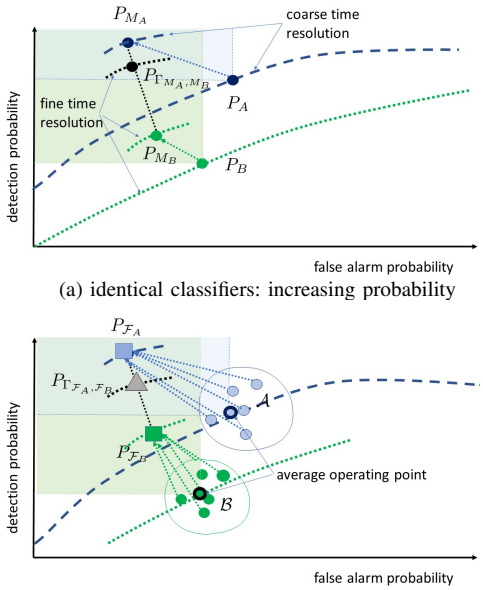
$$D_{\Gamma_{M_A,M_B}} = \rho D_{M_A} + (1-\rho) D_{M_B} \text{ and}$$
$$F_{\Gamma_{M_A,M_B}} = \rho F_{M_A} + (1-\rho) F_{M_B},$$

respectively. Also, for its operating point, we have $\left(F_{\Gamma_{M_A,M_B}}, D_{\Gamma_{M_A,M_B}}\right) \to (0,1)$ as $n_A \to \infty$ and $n_B \to \infty$. This case is illustrated in Fig. 3(a). When classifiers are non-identical, the majority and linear fusers are shown to be superior to some of the individual classifiers [20] but not necessarily the best individual classifier, as shown in Fig. 3(b).

B. Non-Identical Classifiers

For classifiers with different detection and false alarm probabilities, we use Bayes fuser [10], [21] that minimizes the expected error, which is the sum of missed detection and false alarm probabilities. Bayes fuser takes a linear form wherein the decisions are combined with linear weights and compared with a threshold. The weights are computed based on the ratios of various probabilities. For a set of classifiers $\mathcal{C} = \{C_1, C_2, \ldots, C_n\}$, the optimal fuser has the linear form: output of 1 if $\sum_{i=1}^{n} w_i C_i(X) \geq \tau$ and 0 otherwise, where $w_i = \log\left(\frac{D_{C_i}(1-F_{C_i})}{(1-D_{C_i})F_{C_i}}\right)$, and $\tau = \sum_{i=1}^{n} \log\frac{1-F_{C_i}}{1-D_{C_i}}$. The Bayes fuser does not guarantee the superiority over the best classifier, and can be analyzed by examining two general cases of the Condorcet Jury Theorem.

- Strong Competence: Consider that the false alarm probability $F_{C_i} < 1/2$ and detection probability $D_{C_i} > 1/2$ for all C_i, which indicate that the performance of classifiers are all better than random. A generalization of the jury theorem under strong competence and conditional independence together imply the *asymptotic improvement* but not the *increasing probability* [22].
- Average Competence: Consider that the average of detection (false alarm) probabilities is slightly greater (smaller) than half, or converges to a value above (below) 1/2. Then, the jury theorem generalizations in [23] show that average competence and conditional independence together implies the *asymptotic improvement* but not the *increasing probability*.



(a) identical classifiers: increasing probability

(b) dissimilar classifiers: increasing average probability

Fig. 3: Statistical independence leads to asymptotic improvement.

When the competence levels of the classifiers are known, the simple majority rule may not be the best decision rule. There are various works on identifying the optimal decision rule that maximizes the group correctness probability.

The classical results and their generalization provide insights into the performance of fusers under the statistical independence conditions, which are not guaranteed to be satisfied in several applications, including the scenario in Section VI. Furthermore, the detection and false alarm probabilities are not precisely known when ML solutions are obtained using finite training samples; rather, they are probabilistically specified by the generalization equations of ML classifiers [13].

V. Machine Learned Classifiers

For classifiers based on ML methods, the detection and false alarm probabilities are characterized using generalization equations that provide a confidence probability bound $\delta(\epsilon, l)$ for ensuring that their generalization errors are within a precision parameter ϵ of the optimal, based on l training examples [16]. We now derive the generalization equations for the computed fuser-switch classifier $\Gamma_{\tilde{\mathcal{F}}_A,\tilde{\mathcal{F}}_B}$ based on l_A

and l_B training examples for the classifier sets $\mathcal{A}$ and $\mathcal{B}$, respectively.

A. Classifier Generalization Equations

Let $C(X)$ be the classifier output based on input X that corresponds to output Y distributed according to an unknown distribution $\mathbb{P}_{X,Y}$. The classifier C is chosen from a function class $\mathbf{F}_C$ according to a cost criterion. Its *expected error* is defined as

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

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 $I_C = F_C + 1 - D_C$. Given the training set $(X_1, Y_1), (X_2, Y_2), \ldots, (X_l, Y_l)$, the *empirical error* is

$$\hat{I}_C = \sum_{i=1}^{l} (C(X_i) \oplus Y_i) \ .$$

The best classifiers that minimize the expected and empirical error are denoted by C^* and $\hat{C}$, respectively. Thus, for any $C \in \mathbf{F}_C$, $I_C \geq I_{C^*}$ 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

$$\hat{I}_{\tilde{C}} = \hat{I}_{\hat{C}} + \hat{\epsilon}_{\tilde{C}}; I_{\tilde{C}} = I_{C^*} + \epsilon_{\tilde{C}}$$

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

$$\mathbb{P}^l_{X,Y} \left\{ I_{\tilde{C}} - I_{C^*} > \epsilon + \epsilon_{\tilde{C}} \right\} < \delta_{\mathbf{F}_C}(\epsilon, l),$$

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

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

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

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

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

B. Fuser Improvement

For set of classifiers $\mathcal{C}$, with each classifier C_i chosen from a function class $\mathbf{F}_{C_i}$, the corresponding expected best, empirical best, and computed versions are denoted respectively by

$$\mathcal{C}^* = \left\{ C_1^*, C_2^*, \cdots, C_{n_C}^* \right\}, \quad \hat{\mathcal{C}} = \left\{ \hat{C}_1, \hat{C}_2, \cdots, \hat{C}_{n_C} \right\},$$

$$\tilde{\mathcal{C}} = \left\{ \tilde{C}_1, \tilde{C}_2, \cdots, \tilde{C}_{n_C} \right\}.$$

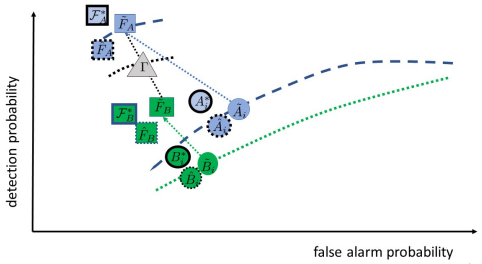


Fig. 4: Optimal expected C^*, optimal empirical $\hat{C}$, and computed $\tilde{C}$ versions for classifier/fuser $C \in \{A_i, B_i, \mathcal{F}_A, \mathcal{F}_B\}$.

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

$$\mathbf{F}_{C_i} = \mathbf{F}_{C_i^*} = \mathbf{F}_{\hat{C}_i} = \mathbf{F}_{\tilde{C}_i}.$$

These classifiers are fused by a fuser $\mathcal{F}_\mathcal{C} : \{0,1\}^{n_C} \mapsto \{0,1\}$ chosen from class $\mathbf{F}_{\mathcal{F}_\mathcal{C}}$. Let $\mathcal{F}^*_{\mathcal{C}^*}$, $\hat{\mathcal{F}}_{\hat{\mathcal{C}}}$, and $\tilde{\mathcal{F}}_{\tilde{\mathcal{C}}}$ represent the expected best, empirical best and computed fuser of the expected best, empirical best and computed classifiers, respectively. The *expected fuser improvement* is defined as

$$\Delta_{\mathcal{F}_\mathcal{C}} = \min_{i=1}^{n_C} I_{C_i} - I_{\mathcal{F}_\mathcal{C}},$$

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

$$\hat{\Delta}_{\mathcal{F}_\mathcal{C}} = \min_{i=1}^{n_C} \hat{I}_{C_i} - \hat{I}_{\mathcal{F}_\mathcal{C}}.$$

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

We consider the *best expected fuser improvement* as

$$\Delta^*_{\mathcal{F}^*_{\mathcal{C}^*}} = \min_{i=1}^{n_C} I_{C_i^*} - I_{\mathcal{F}^*_{\mathcal{C}^*}},$$

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

1) Fuser Generalization Equations: The closeness of the empirical best classifiers and fusers in terms of the best expected fuser improvement $\Delta^*_{\mathcal{F}^*_{\mathcal{C}^*}}$ is probabilistically guaranteed by the following result.

Theorem 5.1: Let the fuser $\tilde{\mathcal{F}}_{\tilde{\mathcal{C}}}$ be the computed fuser trained with $l_{\mathcal{F}_C}$ samples based on the computed classifiers set $\tilde{\mathcal{C}}$ each trained with l_C samples. Under the isolation properties of fuser and all classifier classes, we have

$$\mathbb{P}\left\{ I_{\tilde{\mathcal{F}}_{\tilde{\mathcal{C}}}} - \min_{i=1}^{n_C} I_{C_i^*} + \Delta^*_{\mathcal{F}^*_{\mathcal{C}^*}} > \epsilon + \tilde{\epsilon} \right\} < \delta_{\Delta_{\tilde{\mathcal{F}}_{\tilde{\mathcal{C}}}}}(\epsilon, l_C, l_{\mathcal{F}_C}),$$

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

$$\delta_{\Delta_{\tilde{\mathcal{F}}_{\tilde{C}}}}(\epsilon, l_C, l_{\mathcal{F}_C}) = \delta_{\mathbf{F}_{\mathcal{F}_C}}(\epsilon/2, l_{\mathcal{F}_C}) + \sum_{i=1}^{n_C} \delta_{\mathbf{F}_{C_i}}(\epsilon/(2n_c), l_C)\,.$$

Proof: We first show that

$$\mathbb{P}\left\{ I_{\hat{\mathcal{F}}_{\hat{C}}} - \min_{i=1}^{n_C} I_{C_i^*} + \Delta^*_{\mathcal{F}^*_{C^*}} > \epsilon + \tilde{\epsilon} \right\} < \delta_{\Delta_{\hat{\mathcal{F}}_{\hat{C}}}}(\epsilon, l_C, l_{\mathcal{F}_C}).$$

The condition $\left\{ I_{\hat{\mathcal{F}}_{\hat{C}}} - \min_{i=1}^{n_C} I_{C_i^*} + \Delta^*_{\mathcal{F}^*_{C^*}} > \epsilon \right\}$ is the same as $\left\{ I_{\hat{\mathcal{F}}_{\hat{C}}} - I_{\mathcal{F}^*_{C^*}} > \epsilon \right\}$, which implies $I_{\hat{\mathcal{F}}_{\hat{C}}} - I_{\mathcal{F}^*_{C^*}} > \epsilon/2$ or, for some i, $I_{\hat{C}_i} - I_{C_i^*} > \epsilon/(2n_C)$. Thus, the probability in theorem statement is upper bounded by the sum of probabilities of the above terms, which are upper-bounded by $\delta_{\mathbf{F}_{\mathcal{F}_C}}(\epsilon/2, l_{\mathcal{F}_C})$ and $\delta_{\mathbf{F}_{C_i}}(\epsilon/(2n_c), l_C)$ for each i, respectively. □

With confidence $1 - \delta_{\Delta_{\hat{\mathcal{F}}_{\hat{C}}}}(\epsilon, l_C, l_{\mathcal{F}_C})$, this theorem ensures the expected error of the computed fuser $\tilde{\mathcal{F}}_{\tilde{C}}$ is within the precision parameter $\epsilon + \tilde{\epsilon}$ of the best expected error $\min_{i=1}^{n_C} I_{C_i^*}$ reduced by fuser improvement $\Delta^*_{\mathcal{F}^*_C}$. 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.

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

$$\hat{\Delta}_{\tilde{\mathcal{F}}_{\tilde{C}}} = \min_{i=1}^{n_C} \hat{I}_{\hat{C}_i} - \hat{I}_{\hat{\mathcal{F}}_{\hat{C}}},$$

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 $\mathbf{F}_{\mathcal{F}_C}$, both $\Delta^*_{\mathcal{F}^*_C}$ and $\hat{\Delta}_{\hat{\mathcal{F}}_{\hat{C}}}$ 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 $\tilde{C}_i$ 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

$$\tilde{\Delta}_{\tilde{\mathcal{F}}_{\tilde{C}}} = \min_{i=1}^{n_C} \hat{I}_{\tilde{C}_i} - \hat{I}_{\tilde{\mathcal{F}}_{\tilde{C}}},$$

which is not guaranteed to be non-negative even under isolation property of the class $\mathbf{F}_{\mathcal{F}_C}$, unlike the above two. It is shown to be closer to $\Delta^*_{\mathcal{F}^*_{C^*}}$ by using the empirically best estimates in [24], which provides the following result

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

where $\tilde{\epsilon} = \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{C}}}$, This result is extended to $\Gamma_{\tilde{\mathcal{F}}_A, \tilde{\mathcal{F}}_B}$ in the next section.

C. Fuser-Switch Generalization equations

We combine the results of the previous section to estimate the generalization equations of the computed fuser-switch $\Gamma_{\tilde{\mathcal{F}}_A, \tilde{\mathcal{F}}_B}$.

Theorem 5.2: The computed fuser-switch $\Gamma_{\tilde{\mathcal{F}}_{\tilde{A}}, \tilde{\mathcal{F}}_{\tilde{B}}}$ uses the computed fusers for classifier classes $\mathcal{A}$ and $\mathcal{B}$ with $l_{\mathcal{F}_A}$ and $l_{\mathcal{F}_B}$ samples, respectively. The computed classifiers from $\mathcal{A}$ and $\mathcal{B}$ are trained with l_A and l_B samples, respectively. Under the isolation properties of fusers and all classifier classes, we have $\Delta^*_{\mathbf{F}_\Gamma} \geq 0$, and

$$\mathbb{P}\left\{ I_{\Gamma_{\tilde{\mathcal{F}}_{\tilde{A}}, \tilde{\mathcal{F}}_{\tilde{B}}}} - I_{\Gamma_{\mathcal{F}^*_{A^*}, \mathcal{F}^*_{B^*}}} + \Delta^*_{\mathbf{F}_\Gamma} > \epsilon + \tilde{\epsilon}_\Gamma \right\} < \delta_{\Delta_{\tilde{\mathcal{F}}_{\mathcal{A}}}}(\epsilon/2, l_A, l_{\mathcal{F}_A}) + \delta_{\Delta_{\tilde{\mathcal{F}}_{\mathcal{B}}}}(\epsilon/2, l_B, l_{\mathcal{F}_B})$$

where $\Gamma_{\mathcal{F}^*_{A^*}, \mathcal{F}^*_{B^*}}$ is the switched classifier based on expected best fusers of expected best classifiers, $\tilde{\epsilon}_\Gamma = \rho\tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{A}}} + (1-\rho)\tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{B}}}$ and $\Delta^*_{\mathbf{F}_\Gamma} = \rho\Delta^*_{\mathcal{F}^*_{A^*}} + (1-\rho)\Delta^*_{\mathcal{F}^*_{B^*}}$.

Proof: The condition

$$\left\{ I_{\Gamma_{\tilde{\mathcal{F}}_{\tilde{A}}, \tilde{\mathcal{F}}_{\tilde{B}}}} - I_{\Gamma_{\mathcal{F}^*_{A^*}, \mathcal{F}^*_{B^*}}} + \Delta^*_{\mathbf{F}_\Gamma} > \epsilon + \tilde{\epsilon}_\Gamma \right\} \text{ implies}$$

$$I_{\hat{\mathcal{F}}_{\tilde{A}}} - I_{\mathcal{F}^*_{A^*}} + \Delta^*_{\mathcal{F}^*_{A^*}} > \epsilon + \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{A}}} \quad \text{or}$$

$$I_{\hat{\mathcal{F}}_{\hat{B}}} - I_{\mathcal{F}^*_{B^*}} + \Delta^*_{\mathcal{F}^*_{B^*}} > (1-\rho)\left(\epsilon + \tilde{\epsilon}_{\hat{\mathcal{F}}_{\hat{B}}}\right).$$

Thus, the probability in the theorem is upper bounded by sum of the probabilities of the above two terms, which are upper-bounded by $\delta_{\Delta_{\tilde{\mathcal{F}}_{\mathcal{A}}}}(\epsilon/2, l_A, l_{\mathcal{F}_A})$ and $\delta_{\Delta_{\tilde{\mathcal{F}}_{\mathcal{B}}}}(\epsilon/2, l_B, l_{\mathcal{F}_B})$ obtained by applying Theorem 5.1 to classifiers $\mathcal{A}$ and $\mathcal{B}$. □

The result from the previous section is extended following the approach used in Theorems 5.1 and 5.2 to obtain

$$\mathbb{P}\left\{ \left| \Delta^*_{\mathcal{F}^*_{C^*}} - \tilde{\Delta}_{\tilde{F}_{\tilde{C}}} \right| > \epsilon + \tilde{\epsilon}_\Gamma \right\} < \delta_{\Delta_{\tilde{\mathcal{F}}_{\mathcal{A}}}}(\epsilon/2, l_A, l_{\mathcal{F}_A}) + \delta_{\Delta_{\tilde{\mathcal{F}}_{\mathcal{B}}}}(\epsilon/2, l_B, l_{\mathcal{F}_B}),$$

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

VI. Classification of Radiochemical Dissolution

We consider a practical problem of classifying events associated with the radiochemical dissolution of irradiated Np-237 targets used to produce Pu-238 material (detailed descriptions provided in [5], [17]). The target dissolution activities take place at a radiochemical processing facility, and the resultant gas effluents contain isotopes characteristic of the target material. The gamma spectra of the effluent measurements collected by a High Purity Germanium (HPGe) detector located at the facility's off-gas stack contain the related signatures, and identification of the target based on these measurements is an important component of the facility analytics needed for certain scenarios.

The measurements are aggregated every 6 and 24 hours, and processed and analyzed to estimate 15 isotope counts, which are used as classifier features. These features and results produced for this paper are different from various feature selections reported in [5], [17] and the isotope ratios features reported in [18]. The overall approach of using fusers and different aggregation windows is common to all these works, and resulted in significant performance improvements, which motivated the analytical formulation of this paper. The 15 isotopes consist of five iodine isotopes, I-131, I-132, I-133,

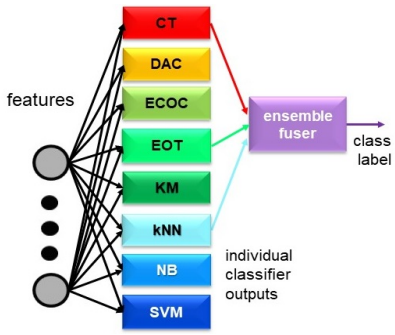


Fig. 5: Ensemble fuser combines outputs of classifiers CT, EOT and kNN with lowest error among eight classifiers [17].

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. The counts are inherently complex since they are Poisson distributed both when the target is being dissolved and at other times, albeit with different parameters. In particular, the measurements corresponding to small targets are challenging to discriminate from background since both are Poisson distributed with (typically unknown) parameters that may be only slightly different.

A. Fuser Performance

We fuse eight classifiers (as in [17]) for Pu/Np target signatures with the lower quality measurements every 6 hours and higher quality measurements every 24 hours, using the proposed fuser-switch method. Eight classifiers are trained using different machine learning methods, 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 $\mathcal{C} = \{$ 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 the complex, unknown features based on finite samples [14].

The non-smooth classifiers CT, EOT and kNN perform significantly better than others, and they are fused using the ensemble fuser (EOT-F), which satisfies the isolation property. The ROC plots and OPs of classifiers trained using measurements at 24 hours time resolution (Set $\mathcal{A}$) and those at 6 hours time resolution (Set $\mathcal{B}$), together with their corresponding fuser, are shown in Figures 6(a) and (b), respectively. For measurements at 24 hours time resolution, the data have much higher quality as indicated by improved performance of the corresponding classifiers as well as fuser as shown in Fig. 6(a). Here, the fuser's OP is inside the enhancement regions of classifiers under both coarser and finer time resolutions. The fuser provides an overall superior performance to most classifiers in these examples but that may not be asserted in general since the guarantee in Theorem 5.1 is probabilistic.

B. Fuser-Switch Performance

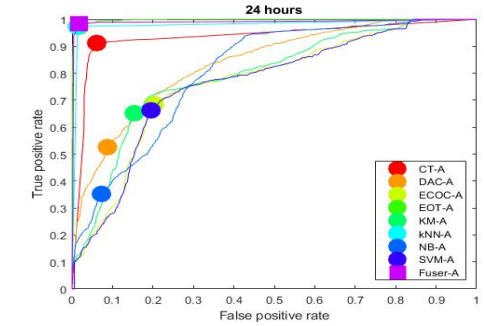
The OPs of classifiers, fuser and fuser-switch at finer time resolution are shown for all classifiers and top three classifiers in Figures 7(a) and (b), respectively. To highlight the effectiveness of fusers, OPs of the switched versions of classifier i, i.e., Γ_{A_i,B_i} where $A_i \in \mathcal{A}$ and $B_i \in \mathcal{B}$, are also plotted in these figures. For the top three classifiers and their fuser, OPs of switched-classifiers and fuser-switch are in their corresponding enhancement regions as shown in Fig. 7(b). However, the performance of the switched-classifier is mixed for other classifiers: it is true for DAC and KM, partially true for SVM and ECOC with higher detection and also higher false alarm rates, and NB with lower false alarm rate but also lower detection rate. By selecting the top three classifiers, fusing them, and combining them using fuser-switch method, performance superior to the best of classifiers, switched-classifiers, and fuser is achieved. Thus, fusers perform superior to all classifiers, and fuser-switch performs superior to all fusers and switched-classifiers, namely, with higher detection rate and lower false alarm rate at finer time resolution and provides output at time resolution finer than first class.

Theorems 5.1 and 5.2 provide analytical insights into performance of classifiers and fusers, as well as practical guidance for the selection of classifiers to be fused. Among all classifiers, those with lower training errors are associated with smaller $\tilde{\epsilon}$ values in Theorem 5.2, and hence stronger performance guarantees. As a result, they are used as inputs to fusers and subsequently in fuser-switch which provides superior performance to all classifiers, generic switched-classifiers, and fusers at the finer time resolution. The classifiers with higher training error are associated with larger $\tilde{\epsilon}$ and weaker performance guarantees, thereby are not included.

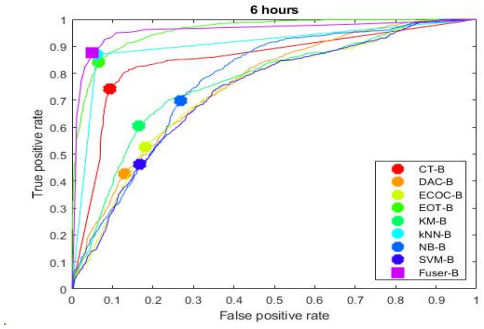
VII. Conclusion and Future Works

In practical scenarios involving multiple sensor systems, the quality of measurements may vary significantly [6], [25]. We studied their simplified abstraction using two measurement streams of different feature qualities and time resolutions. The proposed fuser-switch method utilizes classifiers for each type, and fuses and switches between them to provide classification decisions at a finer time resolution with overall superior detection and false alarm probabilities. For the practical case of ML classifiers, this superior performance is guaranteed with a confidence probability specified by their generalization equations. These results provide analytical foundations for performance results achieved in classifying target dissolution events at a radiochemical processing facility. Our results show that both fusion and switching parts are essential for the superior performance, as indicated by both analytical and experimental results.

This work addresses only a small part of the broader spectrum of sensor measurements quality. Future directions include more complicated fusers across time and feature dimensions, compared to the simple switching studied here. It would be of future interest to use this approach in other application areas.

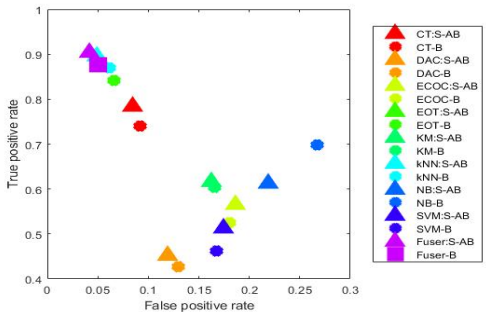


(a) 24 hours

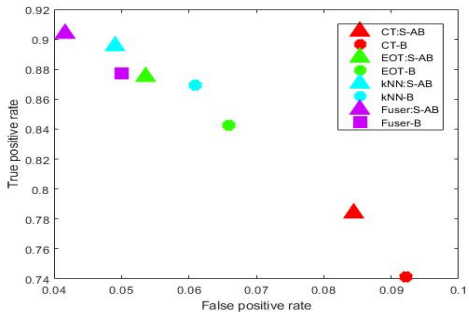


(b) 6 hours

Fig. 6: ROC plots and operating points of classifiers and fusers at 24 and 6 hours time resolutions. Notation: $\mathcal{A}$-classifier: big circle; $\mathcal{B}$-classifier: small circle; fuser: square.



(a) 6 hours: all classifiers



(b) 6 hours: top three classifiers

Fig. 7: Operating points of classifiers, fuser and fuser-switch at 6 hours time resolution. Notation: $\mathcal{B}$-classifier: small circle; fuser: square; switch: triangle.

Acknowledgements

Authors are grateful for the detailed and constructive comments of anonymous reviewers which significantly improved the presentation and clarity of the results in this paper.

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. 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.

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

[7] 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.

[8] 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.

[9] M. de Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, 1785, in French.

[10] C. K. Chow, "Statistical independence and threshold functions," *IEEE Trans. Electronic Computers*, vol. EC-16, pp. 66–68, 1965.

[11] 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

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

[15] 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.

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

[17] 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.

[18] 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.

[19] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, vol. 5, pp. 559–583, 1989.

[20] A. Tangian, *Analytical theory of democracy*. Cham, Switzerland: Springer, 2020, vol. 1 and 2.

[21] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer-Verlag, 1997.

[22] J. Paroush, "Stay away from fair coins: A condorcet jury theorem," *Social Choice and Welfare*, 1998.

[23] B. Grofman, G. Owen, and S. L. Feld, "Thirteen theorems in search of the truth," *Theory and decision*, vol. 15, no. 3, p. 61 278, 1983.

[24] 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.

[25] S. S. Iyengar and R. R. Brooks, Eds., *Distributed Sensor Networks*. Chapman and Hall, 2005.