Evidential Reasoning combine with Mass-based Similarity for Imbalanced Classification

Anh Hoang
School of Business Information Technology, College of Technology and Design, University of Economics Ho Chi Minh City, Vietnam
Correspondence: Anh Hoang, Anhh@ueh.edu.vn
Communication: received 22/02/2025, revised 26/03/2025, accepted 08/04/2025
DOI: 10.32913/mic-ict-research.v2025.n2.1358

***Abstract*: This paper introduces an integrated approach to address imbalanced classification problems using the Dempster-Shafer theory of evidence and mass-based dissimilarity measurement. Traditional distance-based and density-based classifiers often struggle with skewed datasets, leading to two key challenges: (1) misclassification bias, where conventional classifiers treat all instances equally despite the dominance of majority-class samples, and (2) variability in instance densities, which affects similarity assessments. To overcome these limitations, we introduce EMass, a novel classifier that replaces distance- and density-based similarity calculations with mass-based measures. Each neighbor of a considering instance is treated as an independent source of prior knowledge, represented through a basic probability assignment (BPA). We then apply Dempster's rule of combination to aggregate multiple sources of information, producing a comprehensive probability estimate for classification. Experimental evaluations are conducted on 60 imbalanced data sets, comparing 12 algorithms using key performance metrics, including the F1 score, the Brier score, and the area under the curve (AUC). The results show the effectiveness of the proposed EMass classifier in improving the classification performance for imbalanced data sets.**

Keywords: *Imbalance classification, Mass-based similarity, Multi-source evidence, Predictive modeling, Knowledge discovery.*

I. Introduction

Predicting classification model involves determining to which predefined categorical class a specific instance belongs. To date, numerous classification approaches have been proposed and applied successfully to a wide range of domains, such as Decision tree (DT) classifiers or DT classifier that integrates building and pruning [1], DT with AdaBoost [2]; Classification based on association rules (CBA) or multiple association rules (CMAR) [3], predictive association rules (CPAR) [4]; Bayesian classifiers applied Bayes' Theorem [5]; Lazy learning algorithms for classification, k-nearest neighbor algorithm (k-NN) [6]; Discriminative pattern mining for effective classification [7]; the Ensemble methods, such as random forest classifiers [8], rainforest framework [9], or XGBoost (XGB) [10]; Support vector machine (SVM) classifiers, Gaussian RBF kernel SVM (RBF_SVM) [11], or using SVM with hierarchical clustering for classifying large dataset [12]; and applying Neural network with feed-forward or back-propagation classifier [13]; Naïve associative classifier (NAC) [14], Assisted classification of imbalanced datasets (ACID) [15], and Rule-based evolutionary online learning systems (LCSs) [16] were specially designed for mixed and incomplete classification problems.

However, most of these methods are designed to focus merely on balanced datasets; therefore, they may not work effectively on imbalanced datasets where the class distributions are skewed. The challenge of classification with imbalanced datasets remains a critical and complex issue in machine learning, data mining, and knowledge discovery, as it significantly impacts model performance and decision-making accuracy. When datasets have a disproportionate ratio of objects in their classes, some classes, called minority classes, have only very small numbers of data instances, while the others, called majority classes, contain large numbers of data instances. Classification with such imbalanced data sets has been a challenging task for most of the conventional learning methods mentioned above, which assume a relatively balanced class distribution and uniform misclassification costs [17]. Furthermore, a disproportionate distribution of data between classes in imbalanced datasets also poses another important issue when distance- or density-based classification methods are applied due to context-dependent variation, as briefly discussed in Section II.2.

The above challenges need either to restructure the data set or to specialize the algorithm to address the imbalanced classification problem. Previous studies often focus on four groups of methods to fulfill these requirements: resampling the data space, algorithmic modifications, cost-

sensitive learning, and ensemble methods. First, resampling techniques consist of oversampling, such as the synthetic minority oversampling technique (SMOTE) and its variations based on how synthetic samples are generated; and the undersampling technique, such as the TOMEK links algorithm Both techniques transform and restructure the data space to decrease the impacts of their class distribution. Second, algorithms-oriented approaches customize existing algorithms [18] or develop new algorithms for the class imbalance task. Third, cost-sensitive learning solutions, such as the BEE-Miner algorithm [19] and the integration of TANBN with the cost-sensitive classification algorithm [20], aimed at incorporating the algorithm-oriented into the data-level methods to minimize the total misclassification costs, instead of trying to optimize the accuracy. Finally, yet importantly, ensemble learning models, for example multi-imbalance software [21], are operated by adapting the previous ensemble learning algorithms or embedding a cost-sensitive algorithm in the ensemble learning procedure. Notice that all of these learning approaches consume more resources to learn the parameters for the majority class than the minority class. The reason is because the treatments for all instances are the same.

The mass-based similarity measurement and the Dempster-Shafer theory of evidence strongly motivated us to study the problem of class imbalance in advanced. For example, in [22], the authors developed a multiclassifier approach that integrated the support vector machine and the Dempster-Shafer theory to support risk analysis. In [23], the authors introduced an algorithm for making decisions based on kernel density estimation and using a pairwise learning method for binary classification problems. In [24], the Dempster-Shafer theory of evidence (DST) based classifier was proposed with a new rule of evidential reasoning for combining multiple sources of information to construct a multiattribute classifier. To our knowledge, our study has been tested for the first time. The Dempster-Shafer framework is used for the imbalanced classification problem. Through this work, we propose a novel classifier approach for classifying imbalanced datasets, herein referred to as EMass. Each neighbor of the considering instance is treated as an independent evidence source, contributing evidence toward determining the target variable. The Dempster rule of combination is then applied to aggregate these multiple sources of evidence, enabling a more robust decision-making process. By incorporating mass-based measures, the proposed EMass approach handle the disadvantages of traditional distance- and density-based classifiers. Unlike conventional methods, this new classifier relaxes the constraints imposed by distance axioms and assumptions of data independence, improving its adaptability to complex and imbalanced datasets.

The main contributions of this article are summarized as follows:

- We introduce a novel classification algorithm based on the Dempster-Shafer theory of evidence to address the challenges of imbalanced classification.
- Instead of based our work on distance functions that have weaknesses in the situation of a variety of data densities, we introduce a new similarity measure that makes use of the mass-based dissimilarity measurement[1].

The remaining parts of this paper are structured as follows. Section 2 summarizes preliminaries on the imbalanced classification problem, the mass-based dissimilarity measure, and a brief introduction to the Dempster-Shafer theory of evidence. In Section 3, we propose a new EMass method for classifying the imbalanced data sets. First, the confidence of each data point that belongs to its class is estimated. Second, the similarity between the considering instance and its neighbor is evaluated using an estimated dissimilarity measure. Third, each neighboring data point is treated as an independent information source, contributing evidence to predict the target variable. The Dempster rule of combination is then applied to integrate all multiple sources of evidence for final decision making. Section 4 presents and analyzes the results obtained from experiments conducted on 60 imbalanced data sets. Lastly, Section 5 summarizes the conclusions and provides recommendations for future research directions. The notation used in this study is clarified in the following.

- AUC: Area Under The Curve
- BPA: Basic Probability Assignment
- DST: Dempster-Shafer Theory of evidence
- EMass: Evidential Mass-based approach
- k-LMN: k Lowest Mass Neighbors
- SVM: Support Vector Machine

II. Problem Statement and Literature Review

We systematically specify the imbalanced classification problem under the definition of a classification task in Section II.1. The literature review is briefly presented in Section II.2. This covers the dissimilarity measures based on mass approach, and the Dempster-Shafer theory of evidence.

[1]The term *mass* used in this paper in the sense of mass-based dissimilarity introduced in [25]. It is not related to the concept of mass functions used in the Dempster-Shafer theory, which will be referred to as basic probability assignments in the present paper.

1. Imbalanced classification problem statement

In general, classification is a supervised learning task in machine learning, data mining, and knowledge discovery, where the purpose is to predict the target variable of unseen instances based on patterns learned from independent variables. Given the labeled dataset $X = (x, y)$ including n instances $(x_i, y_i), x_i \in \mathbb{R}^d, y_i \in \Omega, 1 \leq i \leq n$, d is the number of dimensions, where $\Omega = \{l_1, l_2, ..., l_M\}$ is the set of M labels, or the frame of discernment.

The goal of classification in predictive modeling is to determine the most probable label $l_i \in \Omega$ for a new unlabeled instance x_t. Almost all classification approaches are based on measuring the similarity between x_t and any or a subset of instances within the dataset, making the definition of similarity a crucial aspect of this research field. However, classification becomes significantly more challenging in the presence of imbalanced datasets, where certain classes dominate others in terms of the number of instances. This imbalance distorts similarity assessments between heterogeneous classes, rendering traditional similarity measures ineffective in such scenarios.

2. Related work

a) Dissimilarity measures based on Mass

Distance functions such as the Manhattan distance and the Euclidean distance are generally accepted to measure the dissimilarity between data points. As a result, numerous distance-based classifiers have been proposed and utilized in a variety of applications. However, these data-independent measurements have shortcomings due to data-dependent properties and assumptions on the distance measures.

In this part, we thoroughly review the measurement of dissimilarity based on mass [25] that rectifies the distance- or density-based algorithms. A mass-based measure may directly replace distance functions, and the remaining parts of the previous algorithms are maintained. It is of interest to note that calculations based on mass inherit their properties from psychological studies in which two instances in a sparse region of space are more likely to be of interest to each other than two instances with the same distance between points in a dense region [26, 27]. For example, most neighbor-based classifiers, such as the k-NN model, face difficulty classifying instances based on the distance measurement, especially when data sets have different data point densities. This problem can be overcome by using mass-based dissimilarity functions. Figure 1 shows the results of the dissimilarity based on the mass values, referring to equation (2). In this example, we selected a hierarchical structure for the dataset X that includes six instances. Then, the mass-based dissimilarity values can be obtained as shown in Figure 1.

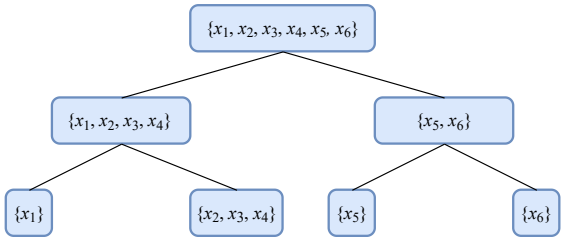

Figure 1. Mass-based dissimilarity measures:

$$\begin{aligned}
&\text{mass}_e(x_1, x_1) = \text{mass}_e(x_5, x_5) = \text{mass}_e(x_6, x_6) = 1/6\\
&\text{mass}_e(x_2, x_3) = \text{mass}_e(x_3, x_4) = \text{mass}_e(x_2, x_4) = 3/6\\
&\text{mass}_e(x_1, x_2) = \text{mass}_e(x_1, x_3) = \text{mass}_e(x_1, x_4) = 4/6\\
&\text{mass}_e(x_2, x_3) = \text{mass}_e(x_2, x_4) = \text{mass}_e(x_3, x_4) = 4/6\\
&\text{mass}_e(x_1, x_5) = \text{mass}_e(x_1, x_6) = \text{mass}_e(x_2, x_5) = 6/6\\
&\text{mass}_e(x_2, x_6) = \text{mass}_e(x_3, x_5) = \text{mass}_e(x_3, x_6) = 6/6
\end{aligned}$$

The following definitions outline the two key steps involved in studying and implementing the mass-based dissimilarity calculation introduced in our previous study [28]. The proposed approach applies the sampling technique to avoid missing any distribution in the dataset. Let H_j be a partition structure of the dataset. H_j contains nested nodes, in which the data points in the same node are more similar than the data points in the different nodes.

Definition 1: Let $S(x_t, x_i|H_j)$ denote the smallest node that contains two instances x_t, x_i with respect to the hierarchical structure H_j, defined by:

$$S(x_t, x_i|H_j) := \underset{S \in H_j, \{x_t, x_i\} \subseteq S}{\text{argmax}} \text{dep}(S|H_j) \tag{1}$$

where $\text{dep}(S|H_j)$ denotes the depth required to separate node S from the initial node within H_j.

Definition 2: Let $\text{mass}(x_t, x_i|H_j)$ denote the dissimilarity function between x_t and x_i, depending on H_j, defined by:

$$\text{mass}(x_t, x_i|H_j) := E_{\mathcal{H}}[P(p|p \in S(x_t, x_i|H_j))]$$

i.e., the expected value of the chance that a random data point $p \in X$ applies to the smallest node $S(x_t, x_i|H_j)$, which is calculated by equation (1) over all possible H_j, where all H_j is denoted by $\mathcal{H}$.

Practically, we select h hierarchical partitioning structures H_j for a given data set X. Therefore, the $\text{mass}(x_t, x_i|H_j)$ can be estimated as the mean of the cardinality of node $S(x_t, x_i|H_j)$ over the set $\mathcal{H}$.

$$\text{mass}_e(x_t, x_i) = \frac{1}{h} \sum_{j=1}^{h} \frac{|S(x_t, x_i|H_j)|}{|H_j|} \tag{2}$$

Definition 3: Let k-LMN(x_t) denote the set of the k-lowest mass-based dissimilarity neighbors around the query instance x_t, defined by:

$$k\text{-LMN}(x_t) = \{x'_1, x'_2, ..., x'_k\}, \quad k \leq |X| \quad \text{(cardinality of } X\text{)} \tag{3}$$

where $x'_i = \underset{x \in X \setminus \{x'_1, x'_2, ..., x'_{i-1}\}}{\text{argmin}} \text{mass}_e(x_t, x), \quad x'_i \in X, \quad i = 1, \ldots, k.$

b) Dempster-Shafer theory (DST)

The theory of the belief function, also known as the evidence theory or DST [29, 30], provides a suitable framework for presenting and reasoning with incomplete information.

Let Ω be a *framework for discernment*, and let 2^Ω be its set of powers. In DST, a basic probability assignment (BPA) concept is defined as a function $m : 2^\Omega \rightarrow [0, 1]$ that satisfies the conditions below:

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad \text{and} \quad m(\emptyset) = 0 \tag{4}$$

where quantity $m(A)$ is a measure of the belief committed exactly to A, supported by multiple sources of evidence. Individual subset $A \subseteq \Omega$ such that $m(A) > 0$ is named a focal element of the m function. The total ignorance degree is represented by $m(\Omega)$.

A useful and important operation in DST is the Dempster rule of combination [30], which is used as a powerful framework to combine multiple sources of evidence in many applications [31]. Consider 2-pieces of evidence, on the same Ω, represented by two BPAs $m_1(.)$ and $m_2(.)$. Dempster's rule of combination is applied to combine them to achieve a new BPA, denoted by $(m_1 \oplus m_2)$ (also called the orthogonal sum of m_1 and m_2), which is defined by:

$$(m_1 \oplus m_2)(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)} \quad \text{for } A \neq \emptyset \tag{5}$$

Note that Dempster's rule of combination is only applicable to BPAs m_1 and m_2 that verify the condition $\sum_{B \cap C = \emptyset} m_1(B) m_2(C) < 1$.

III. Proposed Evidential Mass-based Approach: EMass

Distance- and density-based classifiers, such as k-Nearest Neighbors (k-NN), Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), Bagging and AdaBoost, struggle with challenges posed by skewed class distributions. These methods treat all instances equally, despite the fact that the majority class overwhelmingly dominates the dataset, leading to biased classification performance. Then, misclassification could happen due to the neighbors pick-up that does not contribute to the imbalanced classification problem. Misclassification occurs when all classes have the same error rate or probability of being misclassified. In this section, we will present our proposed so-called evidential mass-based approach (EMass for short) to handle this misclassification error. The EMass approach is graphically summarized in Figure 2 and detailed in the following subsections. Notice that the query x_t presents an unknown instance for evaluating the model.

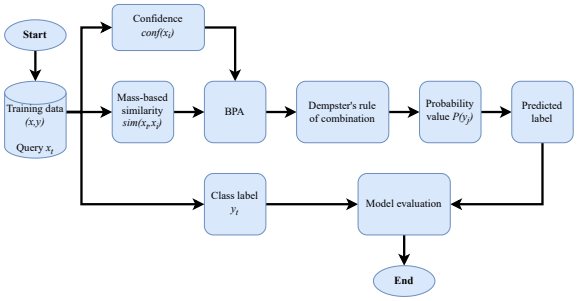

Figure 2. EMass flowchart

1. Symbols and notations

$X = \{x_i, y_i\}$	$\triangleq$	Labeled dataset of n instances
$\Omega = \{l_1, l_2, ..., l_M\}$	$\triangleq$	Set of M labels
mass()	$\triangleq$	Mass-based dissimilarity function
conf()	$\triangleq$	Confidence function
sim()	$\triangleq$	Similarity function
$w()$	$\triangleq$	Weighting function

2. Confidence estimation

We first define the confidence of an instance x_i, denoted by conf(x_i), that provides the degree of confidence in the belonging of the instance x_i to class label y_i. This conf(x_i) is calculated by equation (6), according to Bayes' theorem.

$$\text{conf}(x_i) = P(y_i|x_i) = \frac{P(y_i) \times P(x_i|y_i)}{\sum_{j=1}^{n} P(y_j) \times P(x_i|y_j)} \tag{6}$$

where n is the number of instances, $P(y_i)$ represents the prior probability of class label y_i, and $P(x_i|y_i)$ is the likelihood probability.

The process of estimating the confidence of a data instance is summarized in Algorithm 1 below.

Algorithm 1 *train* Model(x, y)

Input: learning data (x, y)
Output: conf, $\text{mass}_{\max}$

```
conf initialization
mass_max ← 0
for i = 1 to n do
    conf[i] ← compute the confidence, // see the equation (6)
    for j = i + 1 to n do
        mass ← mass_e(x_i, x_j), // refer to [28]
        mass_max ← max(mass, mass_max)
    end for
end for
return conf, mass_max
```

3. Similarity measures based on mass

To define the similarity among data instances, we apply dissimilarity measures based on the mass, as introduced in [28]. The similarity between two data instances should be maximum when they belong to the same leaf node in the hierarchical structure. In contrast, it should be minimal when these instances are in the root node of the partition structure. Formally, we define the similarity between any $x_i, x_t \in X$ as follows:

$$\text{sim}(x_i, x_t) = 1 - \frac{\text{mass}_e(x_i, x_t)}{\text{mass}_{\max}} \tag{7}$$

where $\text{mass}_e(x_i, x_t)$ is the dissimilarity measure based on the mass between two instances as introduced in [28] and $\text{mass}_{\max} = \max_{1 \le i,j \le n}\{\text{mass}_e(x_i, x_j)\}$ denotes the maximum value of all the estimated $\text{mass}_e(x_i, x_j)$.

Those classification methods, such as the NB classifier and the Bayesian belief networks, compute the conditional probability to classify a new instance. However, these methods might not perform accurately due to the skewed distribution of the classes. In addition, the EMass approach calculates the likelihood of an event in which the neighbors of the query instance belong to its context set. In addition, as an uncertain dataset contains noise that makes it deviate from the original values, confidence estimation plays a significant role in addressing imbalanced classification problems where the minority class lacks information. The Dempster rule is applied to merge the multiple sources of evidence contributed by each neighbor of the considered data point to address this limitation.

First, it should be noted that a piece of evidence will allocate a larger BPA value for the query instance to a specific class when its neighbor has a higher confidence or a higher degree of belief that it belongs to the considering class. In other words, evidence with a higher posterior chance will have more belief than evidence with a lower posterior probability value. Second, a piece of evidence will assign more BPA values or a higher degree of belief, for the query instance, to a specific class when the query instance and its neighbor have a higher similarity. We now define the production that satisfies the two mentions above in the following equation.

$$w(x_i, x_t) = \text{conf}(x_i) \times \text{sim}(x_i, x_t) \tag{8}$$

where $\text{conf}(x_i)$ is the confidence of instance x_i estimated by the posterior chance of class label y_i given x_i, and $\text{sim}(x_i, x_t)$ means the similarity between x_i and x_t.

For an unseen query data point x_t, a context set k-LMN(x_t) consists of k-neighbors around x_t having the lowest mass-based dissimilarities measurement. Each instance in k-LMN(x_t) is considered an evidence source that provides a piece of information, which is represented by a BPA defined on the set Ω as in the next subsection. Such evidence supports the prediction of the class label of x_t.

4. Basic probability assignment and evidence combination

We query each i-th neighbor x_i of the considering instance x_t as an evidence source that provides information supporting the hypothesis that the class label of x_t is the same as the class label of x_i, namely, $l_{j(i)} \in \Omega$. Then, based on the observations mentioned above, we formulate the BPA representing the piece of evidence associated with x_i, denoted by m_i, as follows:

$$m_i(\{l_{j(i)}\}) = \beta \times w(x_i, x_t), \tag{9}$$

$$m_i(\Omega) = 1 - \beta \times w(x_i, x_t) \tag{10}$$

$$m_i(A) = 0, \quad \forall A \in 2^{\Omega} \setminus \{\Omega, \{l_{j(i)}\}\} \tag{11}$$

where β is a hyperparameter and $0 < \beta < 1$.

Once the BPA function for the neighbors k $x_i \in k$-LMN(x_t) is defined, Dempster's rule is then applied to combine these k BPA m_i to obtain the overall BPA, denoted by m_t, to determine the label of x_t as follows

$$m_t(.) = \bigoplus_{1 \le i \le k} m_i(.)$$

In particular, we have

$$m_t(\{l'_j\}) = \bigoplus_{1 \le i \le k} m_i(l'_j), \quad l'_j \in \Omega \tag{12}$$

5. Decision making

To make decisions on the label of x_t, the so-called pignistic transformation [32] is applied to transform the combined BPA m_t into the probability distribution as done in [33, 34] as follows:

$$P_t(l'_j) = m_t(\{l'_j\}) + \frac{m_t(\{\Omega\})}{|\Omega|}, \quad l'_j \in \Omega \tag{13}$$

Based on this probability, we make the final decision to predict the class label using equation (14).

$$\hat{y_t} = \underset{l'_j \in \Omega}{\operatorname{argmax}} P_t(l'_j) \tag{14}$$

In summary, the pseudocode of the proposed method is presented in Algorithm 2.

Algorithm 2 EMass approach

Input: learning data (x, y), neighbor number k, considering instance x_t

Output: target variable $\hat{y_t}$

conf, $\text{mass}_{\max} \leftarrow$ learned Model(x, y), // refer to Algorithm 1
$s \leftarrow$ indices of k-LMN(x_t)
Weighted values initialization
for *i = 1 to k* **do**
  $index \leftarrow s[i]$
  conf $\leftarrow$ conf$[index]$
  mass $\leftarrow$ $\text{mass}_e(x_t, x_{index})$, // refer equation (2)
  sim $\leftarrow$ $\text{sim}(x_t, x_{index})$, // refer equation (7)
  BPA$[x_i] \leftarrow$ using equations (9–11)
end for
BPA values combination using equation (12)
Pignistic transformation $\leftarrow$ compute probability values, // refer equation (13)
$\hat{y_t} \leftarrow$ predict class label, // refer equation (14)
Return target label $\hat{y_t}$

IV. Experimental results and discussion

To evaluate our approach in comparison with previous studies, we conducted experiments on 60 imbalanced datasets from the UCI Machine Learning Repository (available at https://archive.ics.uci.edu/). The F1 score and AUC values were used as key evaluation metrics. The experimental results were then subjected to statistical analysis to validate the effectiveness of the EMass model.

1. Experimental datasets

Table I summarizes the characteristics of the 60 imbalanced data sets that were pre-processed for the imbalanced classification task. These datasets were gathered from the UCI machine learning repository [35] and the knowledge extraction process based on the evolutionary learning (KEEL) source [36]. The 60 datasets were chosen from a variety of domains to conduct class imbalance experiments. The numbers of instances, features, and imbalance ratios are all in a wide range of values. The imbalance ratio (IR) of each dataset is defined as the proportion of positive instances to the numbers of negative instances. In this study, the IR spreads from 1.82 up to 100.14.

The original classification datasets were used to generate the unbalanced datasets, as shown in Table I. These data sets were initially collected for conventional binary or multiclass classification tasks. To address the class imbalance problem, the data sets were pre-processed by designating one of the minority classes as the positive class (class 1), while the remaining classes were grouped to form the negative class (class 0). There are data sets in the group that are different versions of each other. However, we consider the positive class to be different from the same original dataset. Therefore, the dependence between data sets did not influence the conclusions or the statistical analysis.

The "Glass" recognition dataset was first experimented with 12 competitively imbalanced classifiers. The investigators used the glass sample, which had nine attributes, as evidence to recognize the type of glass. In total, the data set contains 214 instances distributed across all types of glass. To create the data set "Glass1", glass type 1 was designated as a positive class, while the remaining types were grouped as a negative class. Similarly, we prepared data sets "Glass0", "Glass4", "Glass6", and "Glass-0-4_vs_5" using the same approach.

The Wisconsin Breast Cancer Data Set (abbreviated Wisconsin) consists of 683 medical diagnoses, each described by nine characteristics. In this dataset, the malignant class was designated as a positive class, while the benign class was treated as a negative class. The data set has an imbalance ratio (IR) of 1.86.

The Pima data set was originally collected by the US National Institute of Diabetes, Digestive, and Kidney Diseases to predict diabetes in patients based on eight characteristics. All participants were 21-year-old females. The dataset contains 768 samples, with 268 positive instances, resulting in an imbalance ratio (IR) of 1.87.

The Vehicle dataset, also known as Vehicle Silhouettes [37], was originally designed to classify silhouettes into one of four vehicle types based on a set of characteristics extracted using the Hierarchical Image Processing System (HIPS). This system computes a combination of scale-independent attributes, including scaled variance and skewness/kurtosis along the major and minor axes as classical

moment-based measures. In addition, it employs heuristic measures such as hollows, rectangularity, circularity, and compactness to improve classification accuracy. In this study, we evaluated all 12 competitive approaches in the Vehicle 1 and Vehicle 2 datasets.

The Newthyroid data set categorizes patients into three classes: normal, subnormal, and hyperfunction, using 15 categorical and six numerical features to diagnose potential hypothyroidism. For this experiment, the hyper-function and subnormal classes were grouped as the positive class, while the normal class was designated as the negative class, using five numerical features for classification.

The Segment dataset, also known as the Image Segmentation Data Set, was developed by the University of Massachusetts. It consists of data samples randomly selected from seven outdoor images, where each image was manually segmented at the pixel level to create a classification dataset. Each sample represents a 3×3 pixel region. For this class imbalance experiment, we use Segment0, which contains 2,308 instances, each described by 19 features.

The Yeast data set [38] contains information about yeast cells, originally designed to predict the site of protein localization among 10 possible categories. For this class imbalance experiment, positive and negative classes were defined by selecting appropriate classes from the 10 original categories based on the imbalance ratio (IR) values. For example, in the "Yeas-2_vs_4" dataset, class 2 is designated as the positive class, while class 4 serves as the negative class.

The Page-Blocks dataset consists of document layout blocks identified through a segmentation procedure. It contains 5,472 samples extracted from 54 different documents, each sample representing a single block. The original classification task aims to determine the block type, which falls into one of the following categories: text (0), horizontal line (1), graphic (2), vertical line (3), or picture (4). For this experiment, we prepared the data sets "Page-blocks0" and "Page-blocks-1-3_vs_4" to compare the performance of 12 classifiers.

The Led7digit data set represents a 7-segment display using 7 Boolean features, corresponding to 10 classes, each representing a decimal digit (0-9). The original task was to identify which digit is displayed. Since noise is introduced in these datasets, a pre-processing step is required to remove the noise** before classification. For this experiment, class 1 (representing the digit "1") was designated as the negative class, while the remaining nine classes (0, 2-9) were grouped as the positive class.

The Shuttle datasets share the same nine features but vary in the number of instances. Several versions of the data set were prepared for this experiment. "Shuttle-0_vs_4" contains 1,829 instances, where class 0 is designated as the positive class and class 4 as the negative class; "Shuttle-2_vs_4" includes 129 instances, with class 2 as the positive class and class 4 as the negative class; "Shuttle-6_vs_2-3" consists of 230 instances, where class 6 forms the positive class, while classes 2 and 3 are grouped as the negative class. All remaining classes were removed; "Shuttle-2_vs_5" contains 3,316 instances, with class 2 as the positive class and class 5 as the negative class.

The Dermatology data set consists of 34 characteristics, including one nominal attribute, while the remaining 33 characteristics are numerical. This data set presents a challenging classification task as different erythematosquamous diseases share common clinical characteristics, such as erythema and scaling, with only subtle variations. Initially, 12 clinical characteristics were used for the preliminary diagnosis. Subsequently, skin samples were examined to assess 22 histopathological features, with their values determined by microscopic analysis.

The Winequality data set characterizes the red and white varieties of Portuguese Vinho Verde wine. Due to privacy restrictions, only physicochemical and sensory attributes are available for analysis. The data set consists of ordered but imbalanced classes, as there are significantly more wines of normal quality compared to excellent or poor quality wines. As a result, class imbalance algorithms can be applied to effectively identify excellent and poor wines.

The Poker data set represents a five-card poker hand, where each card is drawn from a standard 52-card deck. Each card is characterized by two attributes, suit and rank, resulting in a total of 10 predictive features. The data set includes a class attribute that defines the type of poker hand, with the order of the cards significant.

The KDDCup99 dataset consists of 34 numerical attributes and 7 categorical attributes. For simplicity, these 41 features were reduced to four: service, duration, src-bytes, and dst-bytes. The data set was then categorized into subsets based on the **service attribute, including HTTP, SMTP, FTP, FTP data, and other protocols.

The Abalone data set contains physical measurements that are used to estimate the age of abalone. In this study, we applied machine learning algorithms to predict the age of new abalone samples. For the class imbalance experiment, each individual class was considered as a positive class with slightly varied proportions.

2. Implementation and evaluation metrics

We compared the EMass approach with the other 11 classifiers, including traditional learning algorithms (C4.5 DT, NB, *k*-NN), logistic regression (LR), tree-based algorithms

Table I
DESCRIPTIONS OF THE IMBALANCED DATASETS. **Idx.**, **#Inst.**, **#Ftr.**, AND **IR** REPRESENT INDEX OF DATASET, NUMBER OF INSTANCES, FEATURES, AND IMBALANCE RATE RESPECTIVELY.

Idx.	Dataset	#Inst.	#Ftr.	IR	Idx.	Dataset	#Inst.	#Ftr.	IR
1	Glass1	214	9	1.82	31	Glass-0-4_vs_5	92	9	10.50
2	Wisconsin	683	9	1.86	32	Ecoli-0-6-7_vs_5	220	6	10.58
3	Pima	768	8	1.87	33	Ecoli-0-1-4-7_vs_2-3-5-6	336	7	11.00
4	Ecoli-0_vs_1	220	7	1.89	34	Led7digit-0-2-4-5-6-7-8-9_vs_1	443	7	11.31
5	Iris0	150	4	2.06	35	Ecoli-0-1_vs_5	240	6	11.63
6	Glass0	214	9	2.10	36	Glass-0-6_vs_5	108	9	12.50
7	Vehicle2	846	18	2.88	37	Ecoli-0-1-4-7_vs_5-6	332	6	12.83
8	Vehicle1	846	18	2.90	38	Ecoli-0-1-4-6_vs_5	280	6	13.74
9	Glass-0-1-2-3_vs_4-5-6	214	9	3.20	39	Shuttle-c0-vs-c4	1829	9	13.87
10	Vehicle0	846	18	3.25	40	Page-blocks-1-3_vs_4	472	10	16.52
11	Ecoli1	336	7	3.36	41	Glass4	214	9	16.83
12	Newthyroid1	215	5	5.14	42	Dermatology-6	358	34	16.90
13	Newthyroid2	215	5	5.32	43	Winequality-white-9_vs_4	168	11	17.67
14	Ecoli2	336	7	5.46	44	Ecoli4	336	7	17.68
15	Segment0	2308	19	6.02	45	Zoo-3	101	16	19.20
16	Glass6	214	9	6.38	46	Poker-9_vs_7	244	10	19.50
17	Yeast3	1484	8	8.10	47	Shuttle-2_vs_4	129	9	20.50
18	Ecoli3	336	7	8.60	48	Shuttle-6_vs_2-3	230	9	22.00
19	Page-blocks0	5472	10	8.79	49	Yeast-2_vs_8	482	8	23.10
20	Yeast-0-2-5-6_vs_3-7-8-9	1004	8	9.14	50	Kddcup-guess_passwd_vs_satan	1642	38	29.98
21	Yeast-0-2-5-7-9_vs_3-6-8	1004	8	9.14	51	Abalone-3_vs_11	502	7	32.47
22	Yeast-2_vs_4	514	8	9.28	52	Yeast5	1484	8	32.73
23	Ecoli-0-6-7_vs_3-5	222	7	9.57	53	Ecoli-0-1-3-7_vs_2-6	281	7	39.42
24	Ecoli-0-1_vs_2-3-5	244	7	9.61	54	Abalone-21_vs_8	581	7	40.50
25	Ecoli-0-2-3-4_vs_5	202	7	9.63	55	Kddcup-land_vs_portsweep	1061	38	49.52
26	Ecoli-0-2-6-7_vs_3-5	224	7	9.67	56	Poker-8-9_vs_6	1485	10	58.40
27	Ecoli-0-4-6_vs_5	203	6	9.68	57	Shuttle-2_vs_5	3316	9	66.67
28	Ecoli-0-3-4-7_vs_5-6	257	7	9.71	58	Kddcup-buffer_overflow_vs_back	2233	38	73.43
29	Ecoli-0-3-4-6_vs_5	205	7	9.79	59	Kddcup-land_vs_satan	1610	38	75.67
30	Vowel0	988	10	9.98	60	Kddcup-rootkit-imap_vs_back	2225	38	100.14

for class imbalance (RF), linear support vector machine (Linear_SVM), SVM with RBF kernel (RBF_SVM), ensemble learning (Bagging, AdaBoost, and XGBoost), and the evidential algorithm (mPE*k*NN).

There are many aspects and methods that can be considered when evaluating the performance of a system for class imbalance problems, such as memory, consumption time, accuracy, F-series score, Brier score, and AUC values. However, for the application of our approach, the AUC results are considered the most important factors to assess the performance when comparing those models. Moreover, the F1 score and Brier score of all testing models are included. For all experiments, we employ a 10-fold cross-validation approach across various methods and datasets. The results presented represent the average results obtained from these validations.

In addition, most imbalance classifiers often demonstrate binary-class problems as multiclass problems, which can be simplified to binary problems. In general, the positive class is considered the minority class, and the negative class is labeled the priority class. Thus, the results of a binary class problem are shown by a confusion matrix. Then, this confusion matrix was used to calculate the F1 score, Brier score, ROC-AUC, and PR-AUC values. Tenfold cross-validation is mainly applied to estimate the skill of all 12 competitive models on the new data. Then, all classifiers are ranked in each dataset in the chosen evaluation metrics.

3. Results and Discussions

Figure 3 a illustrates the comparison of the F1 score in 12 competitive models, evaluated using a tenfold cross-validation in 60 unbalanced datasets. In general, the EMass approach demonstrated superior performance, achieving the highest average rank (7.100) and the best average F1 score (0.857). A detailed comparison of F1 scores is provided in Table A1.

Figure 3b presents a comparison of the average Brier score across 12 models, evaluated using a tenfold cross-validation on 60 unbalanced datasets. The proposed approach outperformed all other models, achieving the highest average rank (9.067) and the lowest average Brier score (0.034), indicating superior performance. It is important to note that a lower Brier score signifies better predictive accuracy. Additionally, the *k*-NN model also achieved competitive results in the Brier score evaluation. Detailed comparisons of Brier scores are provided in Table A2.

Figure 3c illustrates the ROC-AUC comparison among 12 models, evaluated on 60 imbalanced datasets. The results indicate that the EMass model achieved the highest average ROC-AUC value (0.959), while the RBF-SVM model

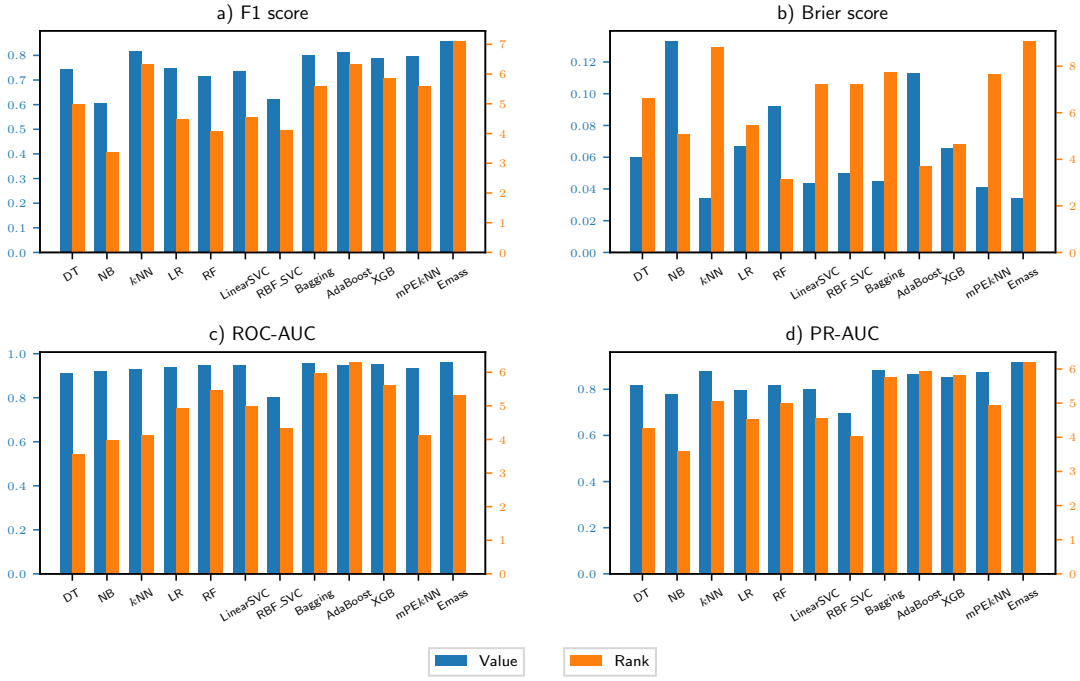

Figure 3. Plots of results comparisons for 12 tested models on 60 imbalanced datasets. Note that a higher average rank indicates better performance.

recorded the lowest average ROC-AUC value (0.800). In terms of average rankings, AdaBoost obtained the best performance (6.283), outperforming all other models. Detailed ROC-AUC measurements are provided in Table A3.

Figure 3d presents the average PR-AUC comparison between 12 models, evaluated using ten-fold cross-validation on the same datasets. In general, the EMass approach outperformed all competing models, achieving the highest mean PR-AUC value (0.915) and the best average ranking (6.183) in Table A4 reports a detailed comparison of PR-AUC results.

4. Wilcoxon signed ranks test results

We used the Wilcoxon signed ranks test [39] as a nonparametric statistical analysis to experiment with the pairwise comparison in terms of the F1 score, Brier score and AUC results. This analysis carries out multiple pairwise comparisons between the EMass approach and each of the other methods. The SPSS software package was used on all the experimental results and the output is presented in Table II. Note that $\mathbf{R}^+$ is defined as the sum of positive signed ranks and $\mathbf{R}^-$ as the sum of negative signed ranks for the results of each comparison method.

It can be seen in Table II that the RBF_SVM model achieved the best results $\mathbf{R}^-$ in terms of the F1 score (**1362.5**) and the PR-AUC metric (**883.0**) compared to the other methods. However, the proposed approach outperforms this model according to the multiple method comparison results presented in Table A1 and Table A4 There are significant differences between EMass and RBF_SVM with a confidence level higher than 99.9% (p-value = 0.001).

In terms of the Brier score, k-NN is the best algorithm from pairwise comparison with EMass due to the highest value $\mathbf{R}^-$ (**758.0**). However, the EMass approach outperforms it according to the results of the comparison of multiple methods in Table A2. However, RF is the worst model among the other methods. In the ROC-AUC metric, the DT model obtained the best $\mathbf{R}^-$ value (**772.0**) compared to the other algorithms, and the test reports a p-value = 0.001. This means that the confidence level is higher than 99.9% for the pairwise method comparison result.

V. Conclusion

In this study, we address the imbalanced classification problem through the lens of neighbor-based algorithms, mass-based similarity measurements, and evidential reasoning methods. Our research highlights that misclassification errors can be mitigated by considering each neighbor as a source of evidence, rather than relying solely on the query instance. To overcome the limitations of distance-based and density-based classifiers, we introduce a mass-based model to measure the similarity between instances. Additionally, we define instance confidence to mitigate the impact of skewed class distributions, weighting this confidence by

Table II
WILCOXON SIGNED RANKS TEST RESULTS.

EMass	F1 results			Brier results			ROC-AUC results			PR-AUC results		
	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value
DT	150.0	840.0	0.001	1075.0	200.0	0.001	174.0	**772.0**	0.001	156.0	790.0	0.001
NB	74.0	1054.0	0.001	1592.0	119.0	0.001	259.5	643.5	0.016	77.0	869.0	0.001
k-NN	256.5	373.5	0.338	727.0	**758.0**	0.894	106.0	455.0	0.002	191.5	438.5	0.043
LR	111.5	923.5	0.001	1572.0	258.0	0.001	349.5	511.5	0.294	152.0	751.0	0.001
RF	119.0	1057.0	0.001	1712.0	58.0	0.001	447.0	499.0	0.754	235.5	754.5	0.002
LinearSVM	120.0	915.0	0.001	1291.0	539.0	0.006	340.0	521.0	0.241	209.0	737.0	0.001
RBF_SVM	122.5	**1362.5**	0.001	1202.0	628.0	0.035	238.0	752.0	0.003	152.0	**883.0**	0.001
Bagging	228.0	592.0	0.014	1058.0	373.0	0.002	399.5	266.5	0.296	247.0	419.0	0.177
AdaBoost	206.5	496.5	0.029	1741.0	89.0	0.001	438.0	265.0	0.192	238.0	465.0	0.087
XGB	284.5	705.5	0.014	1664.0	166.0	0.001	400.5	460.5	0.697	290.0	571.0	0.069
mPEkNN	186.5	633.5	0.003	1027.0	458.0	0.014	114.5	480.5	0.002	184.5	481.5	0.020

mass-based similarity measures between the query instance and its neighbors. Each neighbor is treated as a piece of evidence supporting the label prediction, and Dempster's rule of combination is applied to aggregate these pieces of evidence for final decision-making. As a result, we propose a new classification algorithm for imbalanced data, called EMass.

Traditional classification methods predominantly rely on distance functions, which impose restrictive assumptions such as data independence and adherence to distance axioms that are often unrealistic in real-world datasets. In contrast, our approach leverages mass-based similarity measurements rather than distance-based functions, allowing greater flexibility in handling uncertain and dependent data. EMass follows Dempster-Shafer Theory (DST) to model and integrate multiple pieces of evidence, making it more robust for reasoning under uncertainty in imbalanced datasets.

Our empirical results demonstrate that EMass outperforms 11 competitive models in 60 imbalanced data sets, evaluated using the F1 score, the Brier score, and the AUC values. However, we recognize that the proposed approach is currently limited to datasets with complete numerical variables. Furthermore, the data sets used in our experiments have a restricted number of instances (maximum of 5,472) and features (maximum of 38). In modern classification problems, data sets often contain significantly more instances and features, as well as incomplete and mixed data types. For future research, our objective is to improve mass-based similarity measures, extend the approach to handle incomplete data, and improve adaptability for diverse data types to further increase its applicability in real-world classification tasks.

ACKNOWLEDGMENT

This research is funded by the University of Economics Ho Chi Minh City (UEH), Vietnam. [Grant number: CTD-2024-03]Its results were partially presented at Seminar, School of Business Information Technology, College of Technology and Design, UEH (see: https://bit.ueh.edu.vn/tham-luan-hoi-thao-khoa-hoc-cap-khoa-seminar-nam-2023).

REFERENCES

[1] R. Rastogi and K. Shim, "Public: A decision tree classifier that integrates building and pruning," *Data Mining and Knowledge Discovery*, vol. 4, pp. 315–344, 2000.

[2] J. Hatwell, M. M. Gaber, and R. M. A. Azad, "Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–25, 2020.

[3] F. Thabtah, P. Cowling, and Y. Peng, "Mcar: multi-class classification based on association rule," in *The 3rd ACS/IEEE International Conference on Computer Systems and Applications*. IEEE, 2005, p. 33.

[4] Z. Hao, X. Wang, L. Yao, and Y. Zhang, "Improved classification based on predictive association rules," in *2009 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2009, pp. 1165–1170.

[5] D. Heckerman, D. Geiger, and D. M. Chickering, "Learning bayesian networks: The combination of knowledge and statistical data," *arXiv preprint arXiv:1302.6815*, 2013.

[6] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning k for k-NN classification," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 3, pp. 1–19, 2017.

[7] H. Cheng, X. Yan, J. Han, and S. Y. Philip, "Direct discriminative pattern mining for effective classification," in *2008 IEEE 24th International Conference on Data Engineering*. IEEE, 2008, pp. 169–178.

[8] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, "Improved random forest for classification," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4012–4024, 2018.

[9] J. Gehrke, R. Ramakrishnan, and V. Ganti, "Rainforest—a framework for fast decision tree construction of large datasets," *Data Mining and Knowledge Discovery*, vol. 4, pp. 127–162, 2000.

[10] C. Wang, C. Deng, and S. Wang, "Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost," *Pattern Recognition Letters*, vol. 136, pp. 190–197, 2020.

[11] M. Ring and B. M. Eskofier, "An approximation of the Gaussian RBF kernel for efficient classification with SVMs," *Pattern Recognition Letters*, vol. 84, pp. 107–113, 2016.

[12] H. Yu, J. Yang, and J. Han, "Classifying large data sets using svms with hierarchical clusters," in *Proceedings of the Ninth*

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 306–315.
[13] W. M. Wang, J. Wang, Z. Li, Z. Tian, and E. Tsui, "Multiple affective attribute classification of online customer product reviews: A heuristic deep learning method for supporting kansei engineering," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 33–45, 2019.
[14] Y. Villuendas-Rey, C. F. Rey-Benguría, Á. Ferreira-Santiago, O. Camacho-Nieto, and C. Yáñez-Márquez, "The naïve associative classifier (nac): a novel, simple, transparent, and accurate classification model evaluated on financial data," *Neurocomputing*, vol. 265, pp. 105–115, 2017.
[15] Y. Villuendas-Rey, M. D. Alanis-Tamez, C. F. R. Benguría, C. Yáñez-Márquez, and O. Camacho-Nieto, "Medical diagnosis of chronic diseases based on a novel computational intelligence algorithm," *Journal of Universal Computer Science*, pp. 775–796, 2018.
[16] A. Orriols-Puig and E. Bernadó-Mansilla, "Evolutionary rule-based systems for imbalanced data sets," *Soft Computing*, vol. 13, pp. 213–225, 2009.
[17] Y. SUN, A. K. C. WONG, and M. S. KAMEL, "Classification of imbalanced data: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 04, pp. 687–719, 2009.
[18] A. G. de Sá, A. C. Pereira, and G. L. Pappa, "A customized classification algorithm for credit card fraud detection," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 21–29, 2018.
[19] P. Tapkan, L. Özbakır, S. Kulluk, and A. Baykasoğlu, "A cost-sensitive classification algorithm: BEE-Miner," *Knowledge-Based Systems*, vol. 95, pp. 99–113, 2016.
[20] D. Gan, J. Shen, B. An, M. Xu, and N. Liu, "Integrating TANBN with cost sensitive classification algorithm for imbalanced data in medical diagnosis," *Computers & Industrial Engineering*, vol. 140, p. 106266, 2020.
[21] C. Zhang, J. Bi, S. Xu, E. Ramentol, G. Fan, B. Qiao, and H. Fujita, "Multi-imbalance: An open-source software for multi-class imbalance learning," *Knowledge-Based Systems*, vol. 174, pp. 137–143, 2019.
[22] Y. Pan, L. Zhang, X. Wu, and M. J. Skibniewski, "Multi-classifier information fusion in risk analysis," *Information Fusion*, vol. 60, pp. 121–136, 2020.
[23] C. Zhu, B. Qin, F. Xiao, Z. Cao, and H. M. Pandey, "A fuzzy preference-based Dempster-Shafer evidence theory for decision fusion," *Information Sciences*, vol. 570, pp. 306–322, 2021.
[24] X. Xu, D. Zhang, Y. Bai, L. Chang, and J. Li, "Evidence reasoning rule-based classifier with uncertainty quantification," *Information Sciences*, vol. 516, pp. 192–204, 2020.
[25] K. M. Ting, Y. Zhu, M. Carman, Y. Zhu, and Z.-H. Zhou, "Overcoming key weaknesses of distance-based neighbourhood methods using a data dependent dissimilarity measure," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Singapore: Springer, 2016, pp. 1205–1214.
[26] A. Tversky, "Features of similarity." *Psychological Review*, vol. 84, no. 4, p. 327, 1977.
[27] C. L. Krumhansl, "Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density," *American Psychologist*, vol. 5, pp. 445–463, 1978.
[28] A. Hoang, T. N. Mau, D.-V. Vo, and V.-N. Huynh, "A mass-based approach for local outlier detection," *IEEE Access*, vol. 9, pp. 16 448–16 466, 2021.
[29] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," in *"Classic Works of the Dempster-Shafer Theory of Belief Functions"*. Berlin Heidelberg: Springer, 2008, pp. 57–72.
[30] G. Shafer, *A Mathematical Theory of Evidence*. New Jersey: Princeton University Press, 1976.
[31] T. N. Mau, Q. Le, D.-V. Vo, D. Doan, and V.-N. Huynh, "Integrating machine learning and evidential reasoning for user profiling and recommendation," *Journal of Systems Science and Systems Engineering*, vol. 32, pp. 393–412, 2023.
[32] P. Smets, "Decision making in the TBM: the necessity of the pignistic transformation," *International Journal of Approximate Reasoning*, vol. 38, no. 2, pp. 133–147, 2005.
[33] E. Ramasso and R. Gouriveau, "Prognostics in switching systems: Evidential markovian classification of real-time neuro-fuzzy predictions," in *2010 Prognostics and System Health Management Conference*. IEEE, 2010, pp. 1–10.
[34] Z. Su, T. Denoeux, Y. Hao, and M. Zhao, "Evidential k-NN classification with enhanced performance via optimizing a class of parametric conjunctive t-rules," *Knowledge-Based Systems*, vol. 142, pp. 7–16, 2018.
[35] M. Lichman *et al.*, "UCI Machine Learning Repository," 2013.
[36] I. Triguero, S. González, J. M. Moyano, S. García, J. Alcalá-Fdez, J. Luengo, A. Fernández, M. J. del Jesús, L. Sánchez, and F. Herrera, "Keel 3.0: An open source software for multi-stage analysis in data mining," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 1238–1249, 2017.
[37] J. P. Siebert, "Vehicle recognition using rule based methods," *Turing Institute Research Memorandum TIRM-87-0.18*, 1987.
[38] K. Nakai and M. Kanehisa, "Expert system for predicting protein localization sites in gram-negative bacteria," *Proteins: Structure, Function, and Bioinformatics*, vol. 11, no. 2, pp. 95–110, 1991.
[39] F. Wilcoxon, "Individual comparisons by ranking methods," in *Breakthroughs in Statistics*. New York: Springer, 1992, pp. 196–202.

Anh Hoang received the B.S. degree in Telecommunication engineer from the Department of Electrical, Electronic, and Information Engineering, Hanoi University of Transport and Communication, in 2007, and the M.S. degree in Computer Science from the National Taiwan University of Science and Technology (NTUST), Taiwan, in 2010. He completed Ph.D. program at the Graduate School of Advanced Science and Technology (major in Knowledge Science), Japan Advanced Institute of Science and Technology (JAIST), Japan, in 2021. He is currently teaching at School of Business Information Technology, College of Technology and Design, University of Economics Ho Chi Minh City (UEH), Vietnam. His research interests are related to AI/Machine Learning, Data Science/Data Mining/Data Analytics, CyberSecurity, and Business Intelligence/ Business Analytics
Email: Anhh@ueh.edu.vn

APPENDIX: DETAILED COMPARISON RESULTS FOR 12 TESTED METHODS ON 60 IMBALANCED DATASETS.

Table A1
F1 SCORES COMPARISON.

Idx.	Dataset	DT	NB	*k*NN	LR	RF	LinearSVM	RBF_SVM	Bagging	AdaBoost	XGB	mPE*k*NN	**EMass**
1	Glass1	0.722	0.591	0.581	0.526	0.692	0.591	0.491	0.688	0.667	0.647	0.650	0.778
2	Wisconsin	0.952	0.962	0.961	0.990	0.944	0.981	0.971	0.953	0.981	0.952	0.981	0.981
3	Pima	0.639	0.684	0.565	0.624	0.656	0.624	0.677	0.703	0.651	0.646	0.639	0.614
4	Ecoli-0_vs_1	0.963	0.783	1.000	1.000	0.963	1.000	1.000	0.963	0.963	0.963	0.963	1.000
5	Iris0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	Glass0	0.667	0.556	0.545	0.483	0.692	0.516	0.407	0.692	0.741	0.769	0.581	0.688
7	Vehicle2	0.819	0.583	0.833	0.927	0.819	0.911	0.509	0.819	0.953	0.952	0.800	0.962
8	Vehicle1	0.615	0.523	0.483	0.693	0.504	0.710	0.513	0.667	0.585	0.673	0.558	0.647
9	Glass-0-1-2-3_vs_4-5-6	0.846	0.786	0.889	0.897	0.769	0.889	0.517	0.929	1.000	0.889	0.929	1.000
10	Vehicle0	0.822	0.609	0.882	0.946	0.643	0.946	0.622	0.643	0.939	0.909	0.909	0.878
11	Ecoli1	0.733	0.343	0.667	0.690	0.759	0.733	0.714	0.710	0.640	0.769	0.690	0.714
12	New-thyroid1	1.000	0.923	0.909	1.000	1.000	1.000	0.909	1.000	1.000	1.000	1.000	1.000
13	Newthyroid2	0.941	1.000	1.000	0.933	0.941	0.933	0.875	0.941	1.000	0.941	1.000	1.000
14	Ecoli2	0.783	0.339	0.818	0.690	0.741	0.667	0.783	0.692	0.783	0.818	0.750	0.720
15	Segment0	0.970	0.615	0.969	0.992	0.590	0.992	0.558	0.955	0.985	0.962	0.961	0.984
16	Glass6	0.750	0.875	0.857	0.762	0.778	0.889	0.000	0.857	0.857	0.875	0.800	0.857
17	Yeast3	0.850	0.272	0.800	0.694	0.698	0.687	0.764	0.791	0.778	0.840	0.708	0.735
18	Ecoli3	0.636	0.516	0.615	0.556	0.533	0.615	0.667	0.667	0.706	0.769	0.706	0.778
19	Page-blocks0	0.589	0.500	0.805	0.736	0.631	0.688	0.275	0.585	0.813	0.850	0.768	0.822
20	Yeast-0-2-5-7-9_vs_3-6-8	0.741	0.230	0.851	0.656	0.704	0.702	0.741	0.741	0.755	0.667	0.741	0.755
21	Yeast-0-2-5-6_vs_3-7-8-9	0.615	0.313	0.700	0.471	0.582	0.516	0.571	0.615	0.576	0.566	0.595	0.615
22	Yeast-2_vs_4	0.710	0.818	0.842	0.667	0.846	0.692	0.750	0.667	0.880	0.645	0.696	0.900
23	Ecoli-0-6-7_vs_3-5	0.857	0.000	1.000	0.600	0.600	0.667	0.667	1.000	0.800	0.857	0.857	0.800
24	Ecoli-0-1_vs_2-3-5	0.267	0.000	1.000	0.500	0.500	0.444	0.800	0.800	0.571	0.500	1.000	1.000
25	Ecoli-0-2-3-4_vs_5	0.615	0.444	0.889	0.727	0.667	0.727	0.727	0.889	0.750	0.889	0.889	0.889
26	Ecoli-0-2-6-7_vs_3-5	0.889	0.000	1.000	0.533	0.667	0.667	0.615	0.889	0.800	0.727	0.857	0.857
27	Ecoli-0-4-6_vs_5	0.800	0.750	0.857	0.750	0.857	0.750	0.857	0.667	0.857	0.800	0.857	0.857
28	Ecoli-0-3-4-7_vs_5-6	0.588	0.400	0.909	0.625	0.714	0.769	0.833	0.800	1.000	0.625	0.667	0.769
29	Ecoli-0-3-4-6_vs_5	0.462	0.727	0.889	0.667	0.889	0.727	0.800	0.600	0.800	0.500	0.889	0.889
30	Vowel0	0.875	0.792	1.000	0.724	0.731	0.724	0.828	0.875	0.958	0.875	1.000	1.000
31	Glass-0-4_vs_5	1.000	1.000	0.667	0.800	1.000	0.800	0.000	1.000	1.000	1.000	1.000	0.667
32	Ecoli-0-6-7_vs_5	0.615	1.000	0.857	0.500	0.800	0.800	0.889	0.889	1.000	1.000	0.800	0.667
33	Ecoli-0-1-4-7_vs_2-3-5-6	0.600	0.444	0.667	0.533	0.533	0.533	0.625	0.667	0.571	0.667	0.500	0.600
34	Led7digit-0-2-4-5-6-7-8-9_vs_1	0.769	0.667	0.800	0.625	0.647	0.769	0.769	0.769	0.625	0.710	0.783	0.783
35	Ecoli-0-1_vs_5	0.833	0.833	0.923	0.875	1.000	0.857	0.923	0.833	0.833	0.769	0.923	0.923
36	Glass-0-6_vs_5	1.000	1.000	1.000	1.000	0.286	0.400	0.000	1.000	1.000	1.000	0.667	1.000
37	Ecoli-0-1-4-7_vs_5-6	0.444	0.750	0.667	0.571	0.400	0.429	0.600	0.444	0.727	0.444	0.667	0.750
38	Ecoli-0-1-4-6_vs_5	0.600	0.889	1.000	0.571	0.889	0.571	0.889	0.857	0.800	0.667	0.889	1.000
39	Shuttle-c0-vs-c4	1.000	0.967	0.983	0.983	1.000	1.000	0.968	1.000	1.000	1.000	0.983	1.000
40	Page-blocks-1-3_vs_4	0.947	0.526	0.800	0.857	0.581	0.783	0.455	0.947	0.947	0.947	0.778	0.941
41	Glass4	0.400	0.000	1.000	0.667	0.667	0.800	0.170	0.750	0.667	0.750	0.800	1.000
42	Dermatology-6	1.000	0.333	1.000	1.000	1.000	1.000	0.444	1.000	1.000	1.000	1.000	1.000
43	Winequality-white-9_vs_4	0.500	0.000	0.000	0.667	0.000	0.667	0.250	0.667	0.000	0.400	0.000	0.667
44	Ecoli4	1.000	0.522	1.000	0.857	0.667	0.800	0.800	1.000	0.833	1.000	0.923	1.000
45	Zoo-3	0.333	0.667	0.000	0.667	0.400	0.667	0.400	0.000	0.667	0.667	1.000	0.667
46	Poker-9_vs_7	0.222	0.667	0.667	0.000	0.333	0.000	0.800	0.667	0.667	0.500	0.500	0.667
47	Shuttle-c2-vs-c4	1.000	1.000	1.000	1.000	1.000	1.000	0.074	1.000	1.000	1.000	1.000	1.000
48	Shuttle-6_vs_2-3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.800
49	Yeast-2_vs_8	0.169	0.800	0.800	0.800	0.667	0.533	0.615	0.727	0.667	0.727	0.667	0.800
50	Kddcup-guess_passwd_vs_satan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
51	Abalone-3_vs_11	1.000	1.000	1.000	1.000	1.000	0.889	0.889	1.000	1.000	1.000	1.000	1.000
52	Yeast5	0.593	0.168	0.875	0.615	0.485	0.471	0.500	0.593	0.632	0.750	0.700	0.609
53	Ecoli-0-1-3-7_vs_2-6	0.286	0.667	1.000	0.286	0.500	0.286	0.333	0.667	0.667	0.000	0.667	0.667
54	Abalone-21_vs_8	0.571	0.381	0.571	0.889	0.348	0.800	0.727	0.800	0.667	0.444	0.500	0.750
55	Kddcup-land_vs_portsweep	1.000	0.010	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000
56	Poker-8-9_vs_6	0.025	0.000	0.000	0.035	0.000	0.020	0.000	0.000	0.030	0.067	0.000	1.000
57	Shuttle-2_vs_5	1.000	0.116	1.000	1.000	1.000	1.000	0.667	1.000	1.000	1.000	1.000	1.000
58	Kddcup-buffer_overflow_vs_back	1.000	0.941	1.000	0.941	0.842	1.000	0.842	1.000	1.000	1.000	0.933	1.000
59	Kddcup-land_vs_satan	1.000	1.000	0.667	1.000	1.000	0.667	0.400	1.000	1.000	1.000	0.667	1.000
60	Kddcup-rootkit-imap_vs_back	1.000	1.000	1.000	1.000	0.800	1.000	0.857	1.000	1.000	1.000	1.000	1.000
	Avg. Values	0.744	0.605	0.818	0.747	0.716	0.738	0.622	0.801	0.813	0.790	0.796	**0.857**
	Avg. Ranks	5.000	3.383	6.350	4.483	4.100	4.567	4.133	5.600	6.333	5.850	5.583	**7.100**

Table A2
BRIER SCORES COMPARISON.

Idx.	Dataset	DT	NB	*k*-NN	LR	RF	LinearSVM	RBF_SVM	Bagging	AdaBoost	XGB	mPE*k*NN	**EMass**
1	Glass1	0.174	0.354	0.171	0.239	0.194	0.201	0.220	0.166	0.229	0.164	0.264	0.172
2	Wisconsin	0.034	0.029	0.025	0.012	0.044	0.016	0.013	0.032	0.163	0.059	0.015	0.015
3	Pima	0.192	0.172	0.240	0.190	0.212	0.173	0.168	0.176	0.239	0.183	0.221	0.241
4	Ecoli-0_vs_1	0.011	0.080	0.013	0.011	0.073	0.007	0.004	0.009	0.100	0.049	0.017	0.001
5	Iris0	0.000	0.000	0.000	0.033	0.001	0.000	0.000	0.000	0.000	0.038	0.000	0.000
6	Glass0	0.172	0.354	0.147	0.186	0.155	0.151	0.193	0.124	0.198	0.122	0.263	0.227
7	Vehicle2	0.087	0.169	0.072	0.042	0.170	0.041	0.194	0.091	0.189	0.055	0.079	0.024
8	Vehicle1	0.155	0.279	0.192	0.126	0.221	0.131	0.185	0.153	0.236	0.140	0.208	0.177
9	Glass-0-1-2-3_vs_4-5-6	0.094	0.120	0.062	0.055	0.097	0.067	0.251	0.056	0.135	0.079	0.046	0.003
10	Vehicle0	0.083	0.299	0.042	0.015	0.165	0.026	0.132	0.123	0.188	0.070	0.050	0.059
11	Ecoli1	0.097	0.657	0.096	0.100	0.126	0.081	0.082	0.090	0.168	0.092	0.131	0.111
12	New-thyroid1	0.002	0.008	0.013	0.003	0.037	0.008	0.047	0.004	0.051	0.040	0.006	0.002
13	Newthyroid2	0.023	0.000	0.005	0.023	0.025	0.012	0.043	0.016	0.030	0.050	0.006	0.000
14	Ecoli2	0.073	0.560	0.057	0.108	0.142	0.087	0.054	0.073	0.177	0.073	0.092	0.103
15	Segment0	0.008	0.161	0.009	0.002	0.115	0.004	0.042	0.013	0.084	0.040	0.010	0.004
16	Glass6	0.100	0.044	0.054	0.073	0.061	0.102	0.153	0.035	0.136	0.055	0.065	0.033
17	Yeast3	0.056	0.610	0.045	0.075	0.144	0.052	0.043	0.055	0.177	0.070	0.084	0.079
18	Ecoli3	0.079	0.208	0.056	0.081	0.127	0.067	0.055	0.070	0.162	0.062	0.072	0.062
19	Page-blocks0	0.075	0.090	0.032	0.067	0.123	0.072	0.074	0.091	0.210	0.056	0.041	0.034
20	Yeast-0-2-5-6_vs_3-7-8-9	0.137	0.091	0.059	0.156	0.183	0.067	0.059	0.118	0.234	0.129	0.069	0.071
21	Yeast-0-2-5-7-9_vs_3-6-8	0.076	0.532	0.033	0.099	0.144	0.038	0.030	0.067	0.216	0.094	0.056	0.055
22	Yeast-2_vs_4	0.062	0.034	0.032	0.092	0.110	0.029	0.027	0.052	0.189	0.076	0.046	0.024
23	Ecoli-0-6-7_vs_3-5	0.052	0.088	0.010	0.121	0.111	0.044	0.013	0.020	0.189	0.059	0.006	0.021
24	Ecoli-0-1_vs_2-3-5	0.104	0.040	0.002	0.070	0.110	0.025	0.005	0.037	0.181	0.071	0.000	0.000
25	Ecoli-0-2-3-4_vs_5	0.104	0.204	0.011	0.082	0.081	0.066	0.029	0.029	0.110	0.064	0.024	0.024
26	Ecoli-0-2-6-7_vs_3-5	0.064	0.108	0.005	0.093	0.119	0.047	0.012	0.027	0.186	0.068	0.020	0.024
27	Ecoli-0-4-6_vs_5	0.024	0.037	0.014	0.070	0.059	0.059	0.020	0.041	0.144	0.062	0.021	0.024
28	Ecoli-0-3-4-7_vs_5-6	0.079	0.240	0.015	0.088	0.117	0.031	0.013	0.036	0.164	0.084	0.022	0.042
29	Ecoli-0-3-4-6_vs_5	0.136	0.054	0.027	0.103	0.104	0.075	0.019	0.052	0.144	0.113	0.027	0.025
30	Vowel0	0.032	0.047	0.000	0.058	0.104	0.038	0.021	0.031	0.129	0.051	0.000	0.000
31	Glass-0-4_vs_5	0.000	0.000	0.023	0.053	0.017	0.008	0.098	0.000	0.000	0.039	0.012	0.026
32	Ecoli-0-6-7_vs_5	0.043	0.017	0.023	0.121	0.083	0.073	0.013	0.025	0.172	0.046	0.030	0.032
33	Ecoli-0-1-4-7_vs_2-3-5-6	0.075	0.074	0.049	0.098	0.119	0.049	0.041	0.055	0.187	0.084	0.057	0.061
34	Led7digit-0-2-4-5-6-7-8-9_vs_1	0.084	0.081	0.057	0.076	0.122	0.061	0.048	0.067	0.214	0.095	0.048	0.054
35	Ecoli-0-1_vs_5	0.042	0.029	0.021	0.035	0.050	0.093	0.022	0.035	0.087	0.074	0.021	0.021
36	Glass-0-6_vs_5	0.000	0.000	0.005	0.004	0.123	0.027	0.065	0.000	0.000	0.044	0.012	0.000
37	Ecoli-0-1-4-7_vs_5-6	0.073	0.034	0.038	0.069	0.100	0.047	0.039	0.053	0.111	0.078	0.039	0.023
38	Ecoli-0-1-4-6_vs_5	0.056	0.020	0.002	0.097	0.065	0.047	0.005	0.032	0.142	0.066	0.005	0.002
39	Shuttle-c0-vs-c4	0.000	0.005	0.003	0.006	0.009	0.000	0.003	0.000	0.000	0.034	0.003	0.000
40	Page-blocks-1-3_vs_4	0.011	0.086	0.026	0.030	0.095	0.044	0.060	0.009	0.011	0.042	0.026	0.007
41	Glass4	0.070	0.195	0.013	0.086	0.119	0.081	0.088	0.039	0.076	0.065	0.028	0.009
42	Dermatology-6	0.000	0.111	0.003	0.001	0.047	0.004	0.015	0.001	0.000	0.035	0.007	0.000
43	Winequality-white-9_vs_4	0.058	0.058	0.059	0.029	0.064	0.052	0.055	0.042	0.043	0.082	0.059	0.018
44	Ecoli4	0.000	0.155	0.005	0.015	0.086	0.005	0.005	0.006	0.096	0.037	0.015	0.000
45	Zoo-3	0.129	0.048	0.042	0.048	0.115	0.073	0.097	0.071	0.097	0.069	0.022	0.048
46	Poker-9_vs_7	0.116	0.026	0.016	0.140	0.118	0.039	0.024	0.050	0.021	0.071	0.020	0.012
47	Shuttle-c2-vs-c4	0.000	0.000	0.000	0.000	0.004	0.000	0.037	0.000	0.000	0.037	0.000	0.000
48	Shuttle-6_vs_2-3	0.000	0.000	0.000	0.031	0.038	0.003	0.001	0.000	0.000	0.036	0.000	0.009
49	Yeast-2_vs_8	0.178	0.021	0.027	0.246	0.169	0.028	0.026	0.092	0.195	0.076	0.029	0.025
50	Kddcup-guess_passwd_vs_satan	0.000	0.000	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.034	0.000	0.000
51	Abalone-3_vs_11	0.000	0.000	0.000	0.000	0.000	0.003	0.001	0.000	0.000	0.035	0.000	0.000
52	Yeast5	0.020	0.256	0.010	0.028	0.069	0.012	0.011	0.023	0.092	0.042	0.020	0.029
53	Ecoli-0-1-3-7_vs_2-6	0.053	0.018	0.004	0.075	0.064	0.016	0.014	0.026	0.069	0.049	0.010	0.018
54	Abalone-21_vs_8	0.037	0.093	0.028	0.009	0.096	0.018	0.021	0.040	0.156	0.057	0.031	0.019
55	Kddcup-land_vs_portsweep	0.000	0.906	0.000	0.000	0.003	0.001	0.005	0.000	0.000	0.034	0.000	0.000
56	Poker-8-9_vs_6	0.197	0.017	0.018	0.254	0.210	0.016	0.017	0.145	0.235	0.094	0.019	0.000
57	Shuttle-2_vs_5	0.000	0.070	0.001	0.000	0.005	0.000	0.005	0.000	0.000	0.034	0.001	0.000
58	Kddcup-buffer_overflow_vs_back	0.000	0.002	0.000	0.002	0.019	0.002	0.001	0.000	0.000	0.034	0.001	0.000
59	Kddcup-land_vs_satan	0.000	0.000	0.003	0.001	0.007	0.003	0.003	0.000	0.000	0.034	0.003	0.000
60	Kddcup-rootkit-imap_vs_back	0.000	0.000	0.000	0.001	0.026	0.001	0.000	0.000	0.000	0.034	0.000	0.000
	Avg. Values	0.060	0.133	**0.034**	0.067	0.092	0.044	0.050	0.045	0.113	0.066	0.041	**0.034**
	Avg. Ranks	6.633	5.083	8.817	5.483	3.150	7.233	7.233	7.750	3.717	4.650	7.667	**9.067**

Table A3
ROC-AUC RESULTS COMPARISON.

Idx.	Dataset	DT	NB	*k*-NN	LR	RF	LinearSVM	RBF_SVM	Bagging	AdaBoost	XGB	mPE*k*NN	**EMass**
1	Glass1	0.841	0.677	0.813	0.626	0.800	0.709	0.330	0.855	0.791	0.877	0.892	0.961
2	Wisconsin	0.982	0.994	0.987	0.993	0.989	0.993	0.988	0.991	0.997	0.991	0.987	0.987
3	Pima	0.804	0.825	0.701	0.815	0.812	0.799	0.806	0.830	0.813	0.808	0.744	0.722
4	Ecoli-0_vs_1	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	Iris0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	Glass0	0.825	0.781	0.832	0.807	0.896	0.815	0.287	0.925	0.940	0.912	0.841	0.895
7	Vehicle2	0.938	0.809	0.936	0.986	0.972	0.986	0.744	0.958	0.995	0.997	0.934	0.978
8	Vehicle1	0.815	0.679	0.740	0.892	0.663	0.891	0.690	0.856	0.836	0.875	0.754	0.823
9	Glass-0-1-2-3_vs_4-5-6	0.867	0.955	0.960	0.981	0.969	0.988	0.024	0.988	1.000	0.929	0.964	1.000
10	Vehicle0	0.947	0.807	0.984	0.998	0.839	0.995	0.857	0.943	0.998	0.993	0.980	0.979
11	Ecoli1	0.883	0.847	0.845	0.934	0.884	0.940	0.945	0.892	0.891	0.890	0.858	0.895
12	New-thyroid1	1.000	1.000	1.000	1.000	1.000	1.000	0.982	1.000	1.000	1.000	1.000	1.000
13	Newthyroid2	0.986	1.000	1.000	1.000	1.000	1.000	0.896	1.000	1.000	1.000	1.000	1.000
14	Ecoli2	0.860	0.861	0.892	0.890	0.915	0.874	0.895	0.902	0.957	0.909	0.889	0.893
15	Segment0	0.990	0.987	0.983	1.000	0.987	1.000	0.971	0.986	0.998	0.990	0.983	0.985
16	Glass6	0.850	0.971	0.864	0.982	0.989	0.986	0.050	0.982	0.975	0.993	0.864	0.986
17	Yeast3	0.947	0.966	0.938	0.968	0.980	0.962	0.977	0.973	0.967	0.973	0.935	0.972
18	Ecoli3	0.916	0.950	0.858	0.927	0.965	0.935	0.979	0.949	0.950	0.975	0.867	0.918
19	Page-blocks0	0.954	0.931	0.940	0.973	0.938	0.960	0.880	0.955	0.988	0.985	0.944	0.942
20	Yeast-0-2-5-6_vs_3-7-8-9	0.843	0.851	0.869	0.853	0.899	0.854	0.832	0.867	0.843	0.879	0.862	0.842
21	Yeast-0-2-5-7-9_vs_3-6-8	0.959	0.954	0.985	0.960	0.950	0.961	0.969	0.973	0.960	0.959	0.981	0.983
22	Yeast-2_vs_4	0.981	0.980	0.900	0.963	0.993	0.975	0.988	0.994	0.998	0.986	0.897	0.904
23	Ecoli-0-6-7_vs_3-5	0.988	0.976	1.000	0.976	0.984	0.976	1.000	1.000	1.000	0.992	1.000	1.000
24	Ecoli-0-1_vs_2-3-5	0.888	1.000	1.000	0.979	1.000	0.989	1.000	1.000	1.000	0.936	1.000	1.000
25	Ecoli-0-2-3-4_vs_5	0.973	1.000	1.000	0.953	0.986	0.959	1.000	1.000	0.993	1.000	1.000	1.000
26	Ecoli-0-2-6-7_vs_3-5	1.000	0.951	1.000	0.976	1.000	0.970	1.000	1.000	0.976	0.994	1.000	1.000
27	Ecoli-0-4-6_vs_5	0.833	0.991	1.000	0.982	0.991	0.982	1.000	0.982	1.000	0.939	1.000	1.000
28	Ecoli-0-3-4-7_vs_5-6	0.964	0.966	0.998	0.983	0.996	1.000	1.000	0.991	1.000	0.991	0.987	0.985
29	Ecoli-0-3-4-6_vs_5	0.845	0.986	0.986	0.959	1.000	0.986	1.000	0.980	0.993	0.885	0.986	0.983
30	Vowel0	0.932	0.976	1.000	0.979	0.958	0.977	0.993	0.978	0.996	0.972	1.000	1.000
31	Glass-0-4_vs_5	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000
32	Ecoli-0-6-7_vs_5	1.000	1.000	0.991	0.981	1.000	1.000	1.000	1.000	1.000	1.000	0.994	0.981
33	Ecoli-0-1-4-7_vs_2-3-5-6	0.761	0.941	0.742	0.828	0.938	0.812	0.884	0.935	0.952	0.886	0.742	0.817
34	Led7digit-0-2-4-5-6-7-8-9_vs_1	0.918	0.962	0.896	0.954	0.964	0.934	0.924	0.960	0.957	0.958	0.899	0.936
35	Ecoli-0-1_vs_5	0.857	0.882	0.929	0.993	1.000	0.895	0.993	1.000	0.997	0.836	0.929	0.929
36	Glass-0-6_vs_5	1.000	1.000	1.000	1.000	0.810	1.000	0.333	1.000	1.000	1.000	1.000	1.000
37	Ecoli-0-1-4-7_vs_5-6	0.684	0.958	0.794	0.832	0.932	0.803	0.806	0.910	0.968	0.869	0.794	0.897
38	Ecoli-0-1-4-6_vs_5	0.868	0.981	1.000	0.962	0.995	0.962	1.000	0.971	1.000	0.844	1.000	1.000
39	Shuttle-c0-vs-c4	1.000	0.965	0.983	0.967	1.000	1.000	1.000	1.000	1.000	1.000	0.983	1.000
40	Page-blocks-1-3_vs_4	0.994	0.934	0.939	0.996	0.924	0.995	0.935	1.000	1.000	0.994	0.939	1.000
41	Glass4	0.625	0.750	1.000	0.936	0.974	0.949	0.071	0.981	0.801	0.955	1.000	1.000
42	Dermatology-6	1.000	0.979	1.000	1.000	1.000	1.000	0.993	1.000	1.000	1.000	1.000	1.000
43	Winequality-white-9_vs_4	0.734	0.891	0.500	1.000	0.984	0.969	0.844	0.938	0.969	0.891	0.500	1.000
44	Ecoli4	1.000	0.984	1.000	1.000	0.989	1.000	1.000	1.000	0.995	1.000	0.995	1.000
45	Zoo-3	0.671	0.750	1.000	0.658	0.763	1.000	0.132	0.921	0.697	1.000	1.000	0.750
46	Poker-9_vs_7	0.686	0.543	0.989	0.564	0.894	0.479	1.000	1.000	0.500	0.830	0.989	1.000
47	Shuttle-c2-vs-c4	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000
48	Shuttle-6_vs_2-3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
49	Yeast-2_vs_8	0.941	0.902	0.888	0.897	0.911	0.865	0.872	0.841	0.967	0.972	0.888	0.888
50	Kddcup-guess_passwd_vs_satan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
51	Abalone-3_vs_11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
52	Yeast5	0.992	0.993	0.934	0.992	0.993	0.993	0.993	0.994	0.990	0.994	0.931	0.930
53	Ecoli-0-1-3-7_vs_2-6	1.000	0.991	1.000	0.982	0.982	0.946	0.964	1.000	1.000	0.938	1.000	0.991
54	Abalone-21_vs_8	0.896	0.895	0.695	0.975	0.888	0.941	0.939	0.849	0.793	0.857	0.695	0.798
55	Kddcup-land_vs_portsweep	1.000	1.000	1.000	1.000	1.000	1.000	0.000	1.000	1.000	1.000	1.000	1.000
56	Poker-8-9_vs_6	0.500	0.325	0.495	0.504	0.406	0.690	0.253	0.432	0.521	0.531	0.495	1.000
57	Shuttle-2_vs_5	1.000	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	1.000	1.000
58	Kddcup-buffer_overflow_vs_back	1.000	0.999	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
59	Kddcup-land_vs_satan	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	1.000
60	Kddcup-rootkit-imap_vs_back	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Avg. Values	0.912	0.922	0.930	0.939	0.945	0.945	0.800	0.958	0.949	0.950	0.932	**0.959**
	Avg. Ranks	3.567	3.967	4.133	4.933	5.467	4.983	4.317	5.967	**6.283**	5.600	4.117	5.300

Table A4
PR-AUC RESULTS COMPARISON.

Idx.	Dataset	DT	NB	k-NN	LR	RF	LinearSVM	RBF_SVM	Bagging	AdaBoost	XGB	mPEkNN	**EMass**
1	Glass1	0.795	0.472	0.779	0.372	0.632	0.432	0.242	0.656	0.537	0.807	0.850	0.921
2	Wisconsin	0.960	0.988	0.976	0.982	0.967	0.982	0.921	0.979	0.994	0.981	0.976	0.977
3	Pima	0.719	0.711	0.601	0.711	0.713	0.655	0.664	0.695	0.675	0.697	0.632	0.614
4	Ecoli-0_vs_1	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	Iris0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
6	Glass0	0.501	0.537	0.698	0.585	0.784	0.599	0.178	0.759	0.820	0.775	0.715	0.795
7	Vehicle2	0.879	0.708	0.917	0.964	0.904	0.966	0.659	0.914	0.988	0.995	0.910	0.980
8	Vehicle1	0.361	0.547	0.536	0.728	0.545	0.730	0.557	0.667	0.729	0.698	0.563	0.666
9	Glass-0-1-2-3_vs_4-5-6	0.913	0.898	0.966	0.958	0.937	0.979	0.201	0.980	1.000	0.948	0.974	1.000
10	Vehicle0	0.902	0.555	0.955	0.995	0.719	0.985	0.696	0.879	0.994	0.981	0.941	0.940
11	Ecoli1	0.777	0.573	0.710	0.811	0.676	0.832	0.753	0.783	0.731	0.785	0.761	0.816
12	New-thyroid1	1.000	1.000	1.000	1.000	1.000	1.000	0.930	1.000	1.000	1.000	1.000	1.000
13	Newthyroid2	0.944	1.000	1.000	1.000	1.000	1.000	0.901	1.000	1.000	1.000	1.000	1.000
14	Ecoli2	0.785	0.747	0.828	0.746	0.828	0.601	0.822	0.818	0.856	0.877	0.816	0.841
15	Segment0	0.983	0.954	0.977	1.000	0.954	1.000	0.896	0.971	0.994	0.988	0.977	0.987
16	Glass6	0.795	0.808	0.857	0.915	0.964	0.926	0.101	0.940	0.940	0.969	0.857	0.944
17	Yeast3	0.852	0.863	0.811	0.752	0.856	0.725	0.773	0.854	0.754	0.850	0.788	0.882
18	Ecoli3	0.851	0.727	0.756	0.714	0.772	0.675	0.881	0.803	0.741	0.911	0.829	0.807
19	Page-blocks0	0.818	0.621	0.860	0.814	0.782	0.692	0.617	0.772	0.901	0.891	0.864	0.887
20	Yeast-0-2-5-6_vs_3-7-8-9	0.442	0.418	0.676	0.459	0.761	0.474	0.584	0.671	0.413	0.686	0.603	0.606
21	Yeast-0-2-5-7-9_vs_3-6-8	0.603	0.656	0.877	0.632	0.889	0.654	0.911	0.863	0.760	0.810	0.851	0.872
22	Yeast-2_vs_4	0.883	0.869	0.864	0.894	0.947	0.886	0.919	0.959	0.983	0.919	0.834	0.886
23	Ecoli-0-6-7_vs_3-5	0.875	0.514	1.000	0.514	0.850	0.514	1.000	1.000	1.000	0.903	1.000	1.000
24	Ecoli-0-1_vs_2-3-5	0.438	1.000	1.000	0.708	1.000	0.792	1.000	1.000	1.000	0.597	1.000	1.000
25	Ecoli-0-2-3-4_vs_5	0.833	1.000	1.000	0.626	0.908	0.650	1.000	1.000	0.944	1.000	1.000	1.000
26	Ecoli-0-2-6-7_vs_3-5	1.000	0.455	1.000	0.579	1.000	0.544	1.000	1.000	0.796	0.944	1.000	1.000
27	Ecoli-0-4-6_vs_5	0.846	0.903	1.000	0.850	0.903	0.850	1.000	0.850	1.000	0.754	1.000	1.000
28	Ecoli-0-3-4-7_vs_5-6	0.775	0.675	0.983	0.835	0.963	1.000	1.000	0.938	1.000	0.938	0.920	0.843
29	Ecoli-0-3-4-6_vs_5	0.710	0.908	0.900	0.519	1.000	0.908	1.000	0.884	0.944	0.793	0.900	0.850
30	Vowel0	0.910	0.872	1.000	0.852	0.849	0.845	0.919	0.873	0.984	0.948	1.000	1.000
31	Glass-0-4_vs_5	1.000	1.000	1.000	1.000	1.000	1.000	0.054	1.000	1.000	1.000	1.000	1.000
32	Ecoli-0-6-7_vs_5	1.000	1.000	0.946	0.835	1.000	1.000	1.000	1.000	1.000	1.000	0.944	0.835
33	Ecoli-0-1-4-7_vs_2-3-5-6	0.461	0.658	0.672	0.656	0.687	0.643	0.691	0.712	0.795	0.647	0.672	0.635
34	Led7digit-0-2-4-5-6-7-8-9_vs_1	0.825	0.789	0.819	0.791	0.791	0.764	0.762	0.830	0.781	0.834	0.807	0.795
35	Ecoli-0-1_vs_5	0.878	0.880	0.939	0.966	1.000	0.883	0.966	1.000	0.981	0.780	0.939	0.939
36	Glass-0-6_vs_5	1.000	1.000	1.000	1.000	0.100	1.000	0.033	1.000	1.000	1.000	1.000	1.000
37	Ecoli-0-1-4-7_vs_5-6	0.472	0.759	0.735	0.771	0.605	0.668	0.682	0.561	0.821	0.523	0.735	0.874
38	Ecoli-0-1-4-6_vs_5	0.821	0.579	1.000	0.462	0.944	0.462	1.000	0.842	1.000	0.782	1.000	1.000
39	Shuttle-c0-vs-c4	1.000	0.953	0.985	0.969	1.000	1.000	0.999	1.000	1.000	1.000	0.985	1.000
40	Page-blocks-1-3_vs_4	0.950	0.739	0.904	0.966	0.467	0.958	0.687	1.000	1.000	0.950	0.896	1.000
41	Glass4	0.660	0.155	1.000	0.399	0.579	0.442	0.051	0.767	0.775	0.804	1.000	1.000
42	Dermatology-6	1.000	0.700	1.000	1.000	1.000	1.000	0.792	1.000	1.000	1.000	1.000	1.000
43	Winequality-white-9_vs_4	0.515	0.587	0.529	1.000	0.792	0.708	0.149	0.633	0.417	0.466	0.529	1.000
44	Ecoli4	1.000	0.663	1.000	1.000	0.930	1.000	1.000	1.000	0.955	1.000	0.944	1.000
45	Zoo-3	0.399	0.774	1.000	0.551	0.570	1.000	0.053	0.663	0.557	1.000	1.000	0.774
46	Poker-9_vs_7	0.332	0.517	0.875	0.052	0.149	0.037	1.000	1.000	0.515	0.542	0.792	1.000
47	Shuttle-c2-vs-c4	1.000	1.000	1.000	1.000	1.000	1.000	0.019	1.000	1.000	1.000	1.000	1.000
48	Shuttle-6_vs_2-3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
49	Yeast-2_vs_8	0.809	0.739	0.762	0.845	0.717	0.769	0.711	0.625	0.521	0.778	0.762	0.680
50	Kddcup-guess_passwd_vs_satan	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
51	Abalone-3_vs_11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
52	Yeast5	0.803	0.733	0.648	0.779	0.887	0.777	0.759	0.841	0.635	0.852	0.626	0.733
53	Ecoli-0-1-3-7_vs_2-6	1.000	0.750	1.000	0.250	0.250	0.125	0.167	1.000	1.000	0.083	1.000	0.750
54	Abalone-21_vs_8	0.849	0.652	0.563	0.849	0.627	0.824	0.721	0.811	0.642	0.580	0.563	0.759
55	Kddcup-land_vs_portsweep	1.000	1.000	1.000	1.000	1.000	1.000	0.002	1.000	1.000	1.000	1.000	1.000
56	Poker-8-9_vs_6	0.215	0.012	0.008	0.018	0.013	0.030	0.010	0.016	0.017	0.024	0.008	1.000
57	Shuttle-2_vs_5	1.000	1.000	1.000	1.000	1.000	1.000	0.359	1.000	1.000	1.000	1.000	1.000
58	Kddcup-buffer_overflow_vs_back	1.000	0.944	1.000	1.000	0.915	1.000	1.000	1.000	1.000	1.000	1.000	1.000
59	Kddcup-land_vs_satan	1.000	1.000	0.750	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.750	1.000
60	Kddcup-rootkit-imap_vs_back	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Avg. Values	0.819	0.776	0.878	0.795	0.819	0.800	0.697	0.880	0.865	0.852	0.875	**0.915**
	Avg. Ranks	4.250	3.600	5.067	4.517	5.000	4.567	4.017	5.767	5.917	5.817	4.933	**6.183**