# Comparison between Techniques of Imbalanced Multi-class Dataset

Hanaa S. Abdalaziz and Fakhreldeen A. Saeed

**Abstract**— Learning from Imbalanced Multi-class datasets is a challengeable problem that exists in a wide variety of real-world applications. Meanwhile, the imbalance problem for binary class datasets has been well surveyed and studied, Imbalanced Multi-class datasets have received less attention. The Imbalanced Multi-class problem belongs to supervised machine learning tasks where each instance should be assigned to one of ***N*** different classes with unequal sample sizes. It owns inherent complex characteristics that introduce more obstacles and issues to be considered during the learning process and require new principles, algorithms, and more tools. In this paper, we provide a review of the development of research in learning from Imbalanced Multi-class datasets. Our aim is at providing a critical review that involves an analysis of the problem notion, the state-of-the-art approaches, structured solutions and the current performance evaluation metrics of the Imbalanced Multi-class learning algorithms as well. Furthermore, we highlight the major challenges in this field.

**Index Terms**—Imbalanced learning, Multi-class dataset, Hierarchical classification techniques, Assessment metrics.

## 1 INTRODUCTION

THERE are many real-world fields that produce Multi-class data which is imbalanced as well, such as medical diagnosis, bioinformatics, protein fold classification, weld flaw classification, text classification and intrusion detection. This justifies the importance of concerning of such kind of data. Dealing with class imbalance problems has been studied and well surveyed last years. A range of techniques, strategies and performance metrics have been established, but most effort were exerted to classify instances into one of two classes, which is called **Binary classification** [1]. More investigations is necessary to treat problems occurs during learning from Imbalanced Multi-class data in particular, where the problem is to classify instances into one of the more than two classes that suffer from imbalanced distribution of instances [2]. There are many problem hinder learning from such data, so it needs to be treated using special methods to obtain good classification results. Our goal in this paper is

• *H.S. Abdalaziz is Assistant Professor at Department of Computer Science, Al-Neelain University, Sudan. E-mail: hanaasameeh@gmail.com.*
• *F.A.Saeed is Associate Professor at Department of Computer Science, Al-Neelain University, Sudan. E-mail: fakhry00@gmail.com .*

to introduce a survey of the problem notion, clarify the problems of learning from such data and review the state-of-the-art of its suggested solutions. In addition, perform a valid analysis of these solutions and present a figured summary so as to structure and classify them as well as the recommended evaluation metrics in order to provide researchers with a good view to choose the most suitable. We first describe the notion of the Multi-class imbalanced learning problems in Section 2 meanwhile Section 3 presents the problems of learning from it. Section 4 details the integrated techniques of handling it. An abstract comparison of these solutions is introduced in section 5. Assessment metrics for imbalanced Multi-class learning are reviewed in Section 6. Finally, a conclusion is provided in Section 7.

## 2 IMBALANCED MULTI-CLASS DATASETS NOTION:

Classification of Imbalanced Multi-class datasets refers to the process of assigning instances into one of the more than two classes which suffer from imbalanced distribution of instances.

## 3 PROBLEMS OF LEARNING FROM IMBALANCED MULTI-CLASS DATASETS

The imbalance nature of the data affects the learning process in many aspects [1]. The situation becomes more severe when learning from Imbalanced Multi-class datasets; several boundaries have to be determined and constructed and they can be overlapped causing increasing in the probability of error while dealing with Imbalanced Multi-class because of the multi-class nature of data. Moreover, Zhou and Liu [3] stated that most of the techniques developed for balancing binary classification become powerless when dealing with Multi-class learning problems and some methods are not applicable directly such as random oversampling and undersampling techniques. In addition, the performance evaluation metrics that dedicated for two class scenario are not suitable as well for assessing the results of classification algorithms considering Imbalanced Multi-class data accurately, which reveals the need for more sophisticated evaluation metrics.

### 4.1 METHODS OF HANDLING IMBALANCED DATASET

Even this paper concentrates on learning from Imbalanced Multi-class data, it is so important to touch the different techniques that handle the binary imbalanced data as well as the balanced multi-class data, since the methods and techniques that are suggested to treat the imbalanced Multi-class data depends totally on them. Figure 1 introduces small summary about balancing techniques for Binary classification:

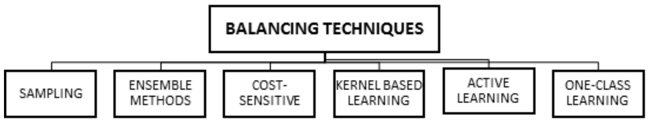


Figure1 Balancing Techniques for Binary Imbalanced Classification

### 4.2 METHODS OF HANDLING MULTI-CLASS DATASET

The following figures summaries the methods of handling Multi-class data.

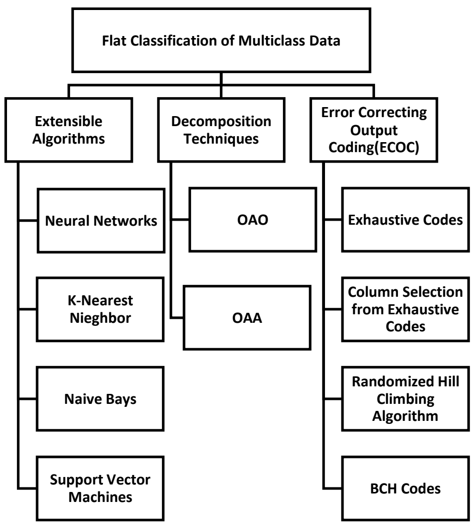

Figure 2.1: Flat Classifications Techniques of Multi-class Data

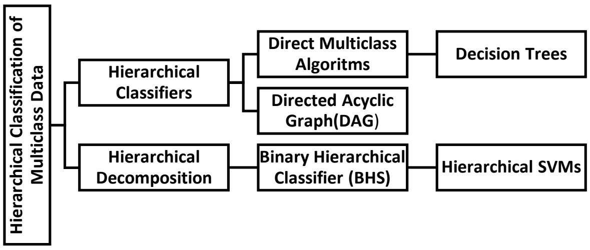


Figure 2.2: Hierarchical Classifications Techniques of Multi-class Data

### 4.3 METHODS OF HANDLING IMBALANCED MULTI-CLASS DATASET

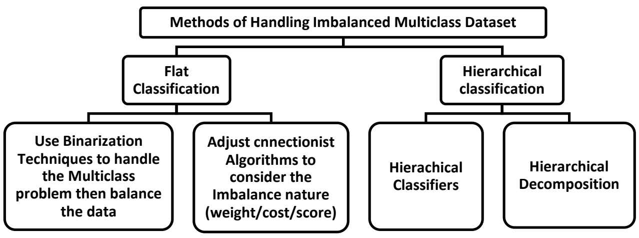


Figure 3. Methods of Handling Imbalanced Multi-class Dataset

The solutions introduced to treat such data were naturally emanated from those dedicated for treating the binary imbalanced data and those for the Multi-class ones. So, they also could be subjoined to the traditional types of classification methods for Multi-class data: Flat and Hierarchical Classification methods, where Flat classification – we intend in this paper- indicates to a single level of classes that examples should be assigned to, while the Hierarchical one refers to the presence of a number of levels of classes where each example could be assigned to some at any level [4]. Regarding Flat Classification, it can be divided into two main methods – Figure 3 - :

**The first** one is using Binarization techniques that transform the Multi-class data into binary imbalanced sub-datasets, then rebalance the data using one of balancing techniques which were indicated in Figure 2.1 before starting the classification process. **Fernandez** et al. [5] integrated OAO with SMOTE in their algorithm. Instead of using data-level methods in two steps: firstly, they deployed OAO. Then, whenever each one of these binary sub problems are imbalanced, an oversampling using the SMOTE algorithm was exploited before the pairwise learning process. Wang and Yao et al. [6] studied the effect of two kinds of multi-class imbalance problems; multi-minority and multi-majority on the performance of two basic resampling techniques. They both showed strong negative effects. Then they applied AdaBoost.NC to several real-world multi-class imbalance datasets and compared it to other three popular ensemble methods based on the correlation analysis and performance pattern analysis ensemble methods. AdaBoost.NC was better at recognizing minority class examples and balancing the performance among classes in terms of G-mean without using any class decomposition, meanwhile, using class decomposition (the one-against-all scheme in their experiments – OAA) did not provide any advantages in multi-class imbalance learning in their experiments. On other hand, **Chen** & **Lu** et al. [7] proposed an algorithm that used OAA, then applied some advanced sampling methods to further decompose each binary problem and rebalance the training set. **Zhao** & **Li** et al. [8] used OAA in addition to undersampling and SMOTE techniques to remedy the imbalanced distribution in their protein data. **Choon** & **Gilber**t et al. [9] proposed utilizing esnsemble methods for classification. They combined the eKISS Method rules of base classifiers to generate new classifiers. They had applied the PART rule-based machine learning technique to generate the base classifiers for their ensemble learning system to improve the coverage of examples from small protein classes .Then they deployed both OAA and OAO schemes to generate one new classifier per class, called the ensemble classifiers.

**Ghanem** & **Venkatesh** et al. [10] suggested a method called Multi-IM which derived its fundamentals from the probabilistic relational technique (PRMs- IM) that was designed for learning from imbalanced relational data for the two-class problem, in addition to All-and-One (A&O) approach to treat the imbalanced problem. Then an independent classifier was trained on each balanced subset. They used the weighted voting strategy as applied in PRMs-IM to combine classifiers to get the result for the parent classifier. **Liao** et al. [11] investigated a variety of oversampling and undersampling techniques used with OAA for a weld flaw classification problem in addition to three algorithms including minimum distance, nearest neighbors, and fuzzy nearest neighbors that were utilized as the classifiers. **Abdi** & **Hashemi** et al. [12] combined over-sampling (Mahalanobis distance-based over-sampling technique (MDO in short)) into boosting algorithm and called it MDOBoost. They over-sampled the minority classes via MDO considering the original minority class characteristics. MDO generates more similar minority class examples to original class samples more than SMOTE .They claimed that their classifier was able to construct larger decision regions and MDOBoost increased the generalization ability of a classifier. The study of **Platt** & **Cristianini** et al. [13] didn't consider the Binarization technique for handling the Multi-class situation, instead ,they deployed a balancing techniques (Dynamic sampling method (DyS)) for multilayer perceptrons (MLP) to deal with the Multi-class nature of the data, then combined the outputs of the ensemble as multi-class classifier. This study utilized the idea of using Codewords beside OAA; **Jeatrakul** et al. [60] suggested the One-Against-All technique with Data Balancing (OAA-DB) algorithm which was an extension of OAA and aimed at improving the weakness of OAA. It balanced the data utilizing combination of SMOTE and CMTNN and combined it with OAA. CMTNN worked as an under-sampling technique while SMOTE was applied as an over-sampling technique. The multi-binary classifier generated **K** outputs of **K** classes, each **K** output was converted to a binary bit to produce binary codewords of each testing example. A binary codeword was represented by the **K** bits class output of each testing instance to utilize it in the classification process. **Alejo** et al. [14] algorithm made the error function of neural networks cost-sensitive by incorporating the proportion of classes within the data set to confirm minority classes, after OAA was applied. [15], [16] also are studies depended on the sampling for balancing Multi-class data.

**The second** approach is adjusting the Extensible Algorithms [17] to consider both imbalance and Multi-class problems. Here, the modification introduces costs into classification process or moving decision threshold. This could be applied by utilizing cost sensitive methods to find an appropriate cost matrix with multiple classes and suit its imbalance nature such as these following studies: **Langford** et al. [18] combined two ideas; firstly, to enhance the performance of neural network on Multi-class imbalanced data, he deployed diverse random subspace ensemble learning with evolutionary search. In order to increase the performance of the learning and optimization of neural network, he exploited the minimum overlapping mechanism to provide diversity. Secondly, to optimize the misclassification, an evolutionary search technique was utilized cost under the guidance of imbalanced data measures. Some studies assign different misclassification costs through using SVMs classifier. The misclassification cost of the minority classes must be higher than the majority class's. So, SVMs could handle all imbalanced Multi-class data in one optimization formulation such as the study of **Landgrebe** and **Duin** et al. [19] who proposed a multi-class Weighted Support Vector Machines (WSVM) method to perform automatic recognition of activities in a smart home environment. This method supported analytic parameter selection of the + C and − C regularization parameters with a new criterion from the training data directly, on the basis of the proportion of class data. In empirical study **Wei & Lin** [20] compared the performance of MultiSVM that considered all classes at once with three methods based on binary classifications: "one-against-all," "one-against-one," and directed acyclic graph SVM (DAGSVM). They concluded that the "one-against-one" and DAG methods are more suitable for practical use. Additionally, Ensemble algorithms, Boosting techniques that modify the weight updating rule and/or loss function such that the minority examples were emphasized with higher weights, or high scores for most interested and confident instances could be deployed as well such as the study of **Wang** et al. [21] who developed a cost-sensitive boosting algorithm AdaC2.M1. They enhanced this algorithm by reducing its weight update parameter to minimize the overall training error of the combined regarding the misclassification costs. They utilized the Genetic Algorithm to get the efficient cost vectors for applying AdaC2.M1. Generally, Studies on misclassification cost can be categorized into two types: **Example-dependent cost** which assumes that the each example has its own misclassification cost, and **Class-dependent cost** which assumes that each class has its own misclassification cost [3]. According to **Zhou** [3], to utilize the rescaling approach, the consistency of the costs should be investigated firstly. The rescaling approach can be deployed directly, if the costs are consistent; otherwise it is better to apply rescaling after decomposing the multi-class problem into a series of two-class problems.

**The Hierarchical classification techniques** that are dedicated for treating imbalanced Multi-class data often handle the imbalance problem initially, then lever the Multi-class situation by turning the classification process into stages of levels. According to **Beyan & Fisher**'s study et al [4]. **The first type** of these techniques is Hierarchical Classifiers; the classes were organized in a pre-defined hierarchy like a tree. So, to get binary hierarchical classifier, it transforms the Multi-class problem into a binary hierarchically [22]. The classes at each parent node are divided into a number of clusters; one for each child nodes till only one class is obtained in the leaf nodes. The discrimination between the different child class clusters at each node of the tree is performed via a simple classifier, usually a binary classifier. So, to get the classification result of a new instance, follow a path from the root node to a leaf. As an instance for this approach, One-Against-Higher-Order (OAHO) [23] method was a hierarchy of classifiers based on the data distribution. OAHO constructed ***K-1*** classifiers for ***K*** classes in a list of $\{C_1, C_2\ldots, C_K\}$. The first classifier was trained using the samples of the first class $C_1$ against all the samples of all the other classes. Then, the second classifier was trained using the samples of the second class in the list $C_2$ against the samples of the higher ordered classes $\{C_3,\ldots,C_K\}$ and so on until the last classifier was trained for $C_{K-1}$ against $C_K$. To diminish the imbalanced situation, the classes were organized descendly according to the number of the samples in each class, in which the small classes were grouped together against the majority class. The problems were that misclassifications made by the top classifiers couldn't be improved by the lower classifiers and OAHO performance was sensitive to the classifier order. **Li** et al. [24] suggested automatic music genre classification approach where the taxonomy gave the relationship between the genres and the similarity matrix from linear discrimination was utilized to construct automatic taxonomies. **Wu** et al. [25] constructed a tree for handling the multi class nature of the data and a multi-class classifier

at each parent node.
Considering **Hierarchical decomposition** which is **the second type** of hierarchical classification techniques [4], the class hierarchy is formed regarding some factors such as the similarity of data or its classes. Here, there is no pre-defined class hierarchy. As an example for this approach, the study of Cesa-Bianchi et al. [26] that utilized the similarity of classes to construct a hierarchy. Also, the study of **Ramanan** et al. [27] in which they proposed the Learning architecture (Unbalanced Decision Tree (UDT)) standing on Directed Acyclic Graph (DAG) and One-versus-All (OVA) approaches. At each decision node, The OVA based concept was implemented. Each decision node of UDT was considered an optimal classification model. The based classifier of the OVA which resulted the maximum performance measure was considered the optimal model for each decision node. Beginning with the root node, the optimal model evaluated one selected class against the rest. Then, from this level of the decision tree, the UDT removed the selected class moving to the next level. When the algorithm yields an output pattern it terminated at a level of the decision node. A hierarchical SVM was proposed by **Chen**, **Crawford** and **Ghosh** et al. [28] basing on class similarities the classes were partitioned into two subsets until one class label was found at a leaf node**.** **Kumar** et al. [29] organized classes in a hierarchy collecting similar classes together to transform the multi-class classification problem into a binary classification problem. For text mining, SVM based hierarchical clustering was used utilizing the similarities between features [30]. Moreover, in the previous mentioned study of **Beyan & Fisher** et al. [4] presented a hierarchical decomposition method which based on clustering and deployed outlier detection for classification. The hierarchy grounded on the similarities of data (i.e. clusters). Different data and feature subsets where employed to construct the hierarchy levels. Supposing that the minority class samples in each class were outliers by cardinality, or by their distance to class, Classification of minority class samples was done via Outlier detection center. **Hoens**, **Chawla** and **Zhou** et al. [31] suggested using Hellinger distance decision trees (HDDTs) to solve the class imbalance problem for decision trees without sampling. They compared different methods of building C4.4 and Hellinger distance decision trees for Imbalanced Multi-class datasets. **LUO** et al. [32] proposed a hierarchical classification method which was a simple bi-classifier with less features input made out most normal samples with an allowable low error rate for minority samples, then a complicated multi-classifier with more features input was constructed by learning the rest less imbalanced samples. To get accurate output for every class, they deployed complicated classifier of ANN ensembles. For classification process, two classifiers operated in parallel. When normal-class result had been acquired the simple classifier of the first layer was able to end the second one.

## 5 An Abstract comparison between Multi-class Imbalanced Solutions

### 5.1 ADVANTAGES:

Naturally, the pros and cons of each method is generated from each techniques that forms a part of the whole method that treat the Multi-class imbalanced data. For instance, SVMs is a very strong algorithm that has big generalization capability and as well as strong mathematical background, so it works very well, even with very small training sample sizes comparing with Binarization techniques, but according to **Wei & Lin** [33] the later techniques are more suitable for practical use, specifically when dealing with large scale problems and they are more accurate for rule learning algorithms. Considering the hierarchical decomposition, dividing the problem into smaller problems by the hierarchy results in selecting a smaller set of features (a more specific domain term features) to a sub-problem which increased the accuracy and efficiency. Many Studies such as [30], [34], [35], and [36] agreed that comparing hierarchical methods to Flat Classification techniques, the former can have better classification results.

### 5.2 DISADVANTAGES:

On one hand, The Binarization approach suffers from excessive testing time because of the need of combing the results of k (k-1)/2 binary classifiers. On the other hand, adding weights or scores modifying the kernel functions of the Extensible algorithms face the difficulty of constructing direct connections between the parameters. Moreover, during training time, a matrix of kernel values for every pair of examples must be computed noticing that SVM is slow and suffers from computational complexity in training according to the hyperplane it deals with and its kernel function , so regarding large-scale problems , learning can take a very long time when dealing with MultiSVM with scores. The hierarchical approach needs to proceed until a leaf node is reached to make a decision on any input pattern, so it also consume time depending on the path. In general, the characteristics of the dataset affects directly on how to decide the most suitable solution to handle each part of the problem of the data nature- Multi-class or imbalance- for any considering learning problem: The number of instance whether its large scale or small one, the number of its classes and number of attributes, The degree of the imbalance in instances distribution and other data complexity if exists.

## 6 EVALUATION METRICS FOR IMBALANCED AND MULTI-CLASS DATA:

Imbalance effects not only classification process, but also the way of evaluating its performance. Considering this situation, there are many sorts of evaluation metrics [37], [38]. Here are some of them that have been extended to suit the Multi-class situation: The **Threshold** metrics (e.g. accuracy, G-mean and F-measure), The **Ranking** methods (e.g. Receiver Operating Characteristics (ROC) analysis and AUC), The **Probabilistic** metrics (e.g. Root-mean-squared error) and for the hierarchical classification, there are many performance metrics that can be suit the Multi-class situation .They are classified into : distance-based, depth-dependent, semantics-based and hierarchy-based [39]. In addition there are Multi-criteria Measures, such as interestingness and comprehensibility [40].

The threshold-metrics based on the concept of the Confusion Matrix which extended for multi-class data as Figure 4 illustrates. It also based on Sensitivity (also called True Positive Rate or Recall of minority class, and known as the ratio of correctly classified examples from the minority class) and Specificity (the ratio of correctly excluded examples from the majority classes). Mosley et al. [41] designed new performance measure specifically for model validation in the presence of multi-class imbalance that called **Class Balance Accuracy** or **Recall (j)** or **Acc (j).**It was defined as:

For any $C^k$ confusion matrix:

$$CBA = \frac{\sum_i^k \frac{c_{ii}}{\max(c_{i.}, c_{i.})}}{k}$$

Where $C^k$ denote a k×k confusion matrix or contingency table of actual class labels aligned by their model predictions, with cij representing the number of cases with true label i classified into group j and $c_{i.} = \sum_{j=1}^{k} c_{ij}$ .

**G-mean** adapted by Sun & Kamel et al. [21] to multi-class scenarios. It is defined as the geometric mean of the Recall values of all classes. Given a j-class problem:

$$\boldsymbol{G-mean} = \left(\prod_{i=1}^{j} Acc(i)\right)^{1/j}$$

$$or$$

$$\mathbf{G-mean} = \frac{\sum_{i=1}^{J} Acc(i)}{J}$$

It can offer the balanced performance among minority and majority classes effectively, as the recognition rate of every class or the accuracies are balanced. Considering cost-sensitive learning, it is natural to utilize misclassification costs for performance evaluation for Multi-class imbalanced problems [42], [3], [43]. For the evaluation of learning algorithms based on class decomposition, some works chose to take the average of any two-class performance measure for produced binary classifiers [9], [13], [2].

**Mean F−measure (MFM):** this measure aggregates both the Precision and the Recall of the minority class. So, it can be illustrated as the weighted average of the Precision and Recall [44].

$$\mathbf{F-measure(j)} = \frac{2.Recall(j).Precision(j)}{Recall(j) + Precision(j)}$$

$$\mathbf{MFM} = \frac{\sum_{j=1}^{K} F-measure(j)}{K}$$

**Kappa Statistic**: It is a measure that compares the accuracy of the system to the accuracy of a random system [45].

$$\mathbf{Kappa} = \frac{\text{Total Accuracy} - \text{RandomAccuracy}}{1 - \text{RandomAccuracy}}$$

Total accuracy is simply the sum of true positive and true negatives, divided by the total number of items.

$$\textbf{Total Accuracy} = \frac{\sum TP + \sum TN}{Total}$$

| Class | 0 | 1 | 2 | … | j |
|---|---|---|---|---|---|
| 0 | TP | FN | FN | FN | FN |
| 1 | FP | TN | FN | FN | FN |
| 2 | FP | FN | TN | FN | FN |
| : | FP | FN | FN | TN | FN |
| j | FP | FN | FN | FN | TN |

Figure 4. Confusion Matrix for Multi-class

Random Accuracy is defined as the sum of the products of reference likelihood and result likelihood for each class. That is,

$$\mathbf{RandomAccuracy} = \frac{(TN + FP) * (TN + FN) + (FN + TP) * (FP + TP)}{Total * Total}$$

Considering **Ranking Methods** for evaluation the scoring classifiers, Multi-class ROC graphs was proposed [19], it generates as many ROC curves as there are classes, where ROC curve originally is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is the sensitivity or recall. The false-positive rate is (1 - specificity) [45] but Multi-class ROC graphs are sensitive to the class skew [46], [47]. A pairwise approach is utilized by discounting some interactions, it approximates the multidimensional operating characteristic to obtain a tractable algorithm and can be extended to large numbers of classes to produce the Multi-class ROC by pairwise analysis [19]. A ROC surface is defined for the Q-class problem as well, in terms of a multi-objective optimization problem utilizing evolutionary algorithm [48]. Another ranking measure is Multi-class AUC which has been proposed to compute the weighted average of all the AUCs produced by the Multi-class ROC graph and a skew-sensitive version of this Multi-class AUC [43], where the Area Under the Curve is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one.

But under the Multi-class imbalanced learning scenario, the AUC values for two-class problems become multiple pairwise discriminability values [49]. To calculate such Multi-class AUCs, a probability estimation-based approach: First, the ROC curve for each reference class $w_i$ is generated and their respective AUCs are measured. Second, all of the AUCs are combined by a weight coefficient according to the reference class's prevalence in the data. It was also sensitive to the class [50]. Moreover, M-measure or (MAUC) is a generalization approach that aggregates all pairs of classes based on the inherent characteristics of the AUC [19]. It is the average of AUC of all pairs of classes, and defined as:

$$M = \frac{2}{c(c-1)} \sum_{i<j} A(i,j)$$

Where A (i, j) = [A (i|j) + A (j|i)] =2 for class pair (i , j). A (i, j) measures the separability between classes. A (i|j) is the probability that a randomly drawn example of class j will have a lower estimated probability of belonging to class i than a randomly drawn example of class i. It should be noted that AUC = A (i|j) = A (j|i) in the two-class scenario, but the equality does not hold when more than two classes exist. Unfortunately, MAUC is insensitive to class distributions and error costs [2], [9], [51]. Another extension of the AUC measure to the Multi-class case tended the volume under the ROC hypersurface that evaluates the VUS over the C-dimensional ROC surface [41].

The third sort of evaluation metrics used with the Probabilistic Classifiers, such as **RMSE** or **RMSD** which is used to measure the differences between values (sample and population values) predicted by the classifier and the values actually observed or estimated [52]. The RMSD of predicted values $\hat{y}_t$ for times t of the variable $y_t$ is computed for n different predictions [53]:

$$RMSD = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - \hat{y}_t)^2}{n}}$$

Additionally, **Cosine Similarity** measures the similarity between two output categories as well as using **The Ranking Loss** which tends the order of the predicted score among C categories. They can also be deployed for probabilistic performance evaluation for multi-class [54]. A Bayesian framework is proposed for inferring on the posterior **Balanced Accuracy** [55]. The Balanced accuracy, i.e., by the arithmetic mean of class-specific accuracies is given by: $1/l.\sum_{i=1}^{l}\theta$i where θi is the (latent) accuracy of the classifier on class i.

Finally, Brier score is also utilized for our tended problem [56].

## 7 Conclusion:

In this paper, we discussed the problem of learning from Imbalanced Multi-class data, which is very critical fundamental issue in knowledge discovery and data engineering field through defining its fundamental nature, the state-of-the-art solutions used to address both. We structured these solutions and studies, then introduced several major assessment techniques used to evaluate this problem so as to serve as a comprehensive review for existing and future studies.

## References

[1] Haibo He and Yunqian Ma, "Imbalanced Learning: Foundations, Algorithms, and Applications", First Edition. The Institute of Electrical and Electronics Engineers, Inc. Published 2013 by John Wiley & Sons.

[2] Haibo He and Edwardo A. Garcia, "Learning from Imbalanced Data", IEEE Transactions on Knowledge and Data Engineering, VOL. 21, NO. 9: September 2009.

[3] Zhi-Hua Zhou, Xu-Ying Liu ,"On Multi-Class Cost-Sensitive Learning" National Key Laboratory for Novel Software Technology, Nanjing University, China, , AAAI'06

proceeding of the 21st national conference on Artificial intelligence,Vol1, pg567-572, ISBN: 978-1-57735-281-5.

[4] Robert Fisher & Cigdem Beyan , "Classifying Imbalanced Data Sets Using Similarity Based Hierarchical Decomposition", Published in: Journal Pattern Recognition © ACM, Volume 48 Issue 5, May 2015, Pages 1653-1672, Elsevier Science Inc. New York, NY, USA, doi:10.1016/j.patcog.2014.10.032.

[5] Alberto Fernandez, Mara Jose Del Jesus, and Francisco Herrera, "Imbalanced Multi-class data-sets with linguistic fuzzy rule based classification systems based on pairwise learning", Computational Intelligence for Knowledge-Based System Design, 6178:89-98, 2010.

[6] Shuo Wang, Xin Yao, "Multi-Class Imbalance Problems: Analysis and Potential Solutions", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, Vol. 42, pp. 1119 - 1130. 10.1109/TSMCB.2012.2187280.

[7] K. Chen, B. L. Lu, and J. Kwok, "Efficient classification of multi-label and imbalanced data using min-max modular classifiers," in Proceedings of World Congress on Computation Intelligence - International Joint Conference on Neural Networks, pp. 1770–1775, 2006.

[8] Xing-Ming Zhao, Xin Li, Luonan Chen, and Kazuyuki Aihara," Protein classification with imbalanced data", Proteins: Structure, Function, and Bioinformatics, 70:1125-1132, 2008.

[9] Aik Choon Tan, David Gilbert, and Yves Deville, "Multi-class protein fold classification using a new ensemble machine learning approach, In Genome Informatics, vol 14, pages 206-217, 2003.

[10] Amal S. Ghanem, Svetha Venkatesh, Geoff West," Multi-Class Pattern Classification in Imbalanced Data", 20th International Conference on Pattern Recognition, 2010 IEEE, DOI: 10.1109/ICPR.2010.706 · Source: DBLP.

[11] T. Warren Liao. "Classification of weld flaws with imbalanced class data". Expert Systems with Applications, 35(3):1041-1052, 2008.

[12] Abdi L. and Hashemi S, "To Combat Imbalanced Multi-class Problems by Means of Over-sampling and Boosting Techniques", Springer, Dec 2015, Vol 19 Issue 12, pp 3369-3385.

[13] J. C. Platt, N. Cristianini and J. Shawe-Taylor, "Large margin DAGs for Multi-class classification", In Advances in Neural Information Processing System s, vol 12, pgs: 547-553, MIT 2000.

[14] Roberto Alejo, Jose M. Sotoca, R. M. Valdovinos, and Gustavo A. Casan, "The multi-class imbalance problem: Cost functions with modular and non-modular neural networks", In The Sixth International Symposium on Neural Networks, volume 56, pages 421-431, 2009.

[15] Shuo Wang, "Ensemble Diversity For Class Imbalance Learning", A thesis submitted to The University of Birmingham for the degree of PHD, School of Computer Science ,College of Engineering and Physical Sciences ,The University of Birmingham ,July 2011..

[16] Xu-Ying Liu, "Learning from combination of data chunks for Imbalanced Multi-class data", Neural Networks (IJCNN), 2014 International Joint Conference,6-11 July 2014, pg:1680 – 1687,ISBN: 978-1-4799-6627-1

[17] T. G. Dietterich and G. Bakiri. "Solving Multi-class learning problems via error correcting output codes", Journal of Artificial Intelligence Research 2, pgs 263-286, 1995.

[18] N. Abe, B. Zadrozny, and J. Langford, "An iterative method for multi-class cost sensitive learning," in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Seattle, WA, USA), pp. 3–11, ACM, 2004.

[19] Thomas C.W. Landgrebe, Robert P.W. Duin," Approximating the Multi-class ROC

by pairwise analysis ", Pattern Recognition Letters Volume 28, Issue 13, 1 October 2007, Pages 1747–1758, doi :10.1016/j.patrec.2007.05.001

[20] Chih-Wei Hsu and Chih-Jen Lin, ”A Comparison of Methods for Multi-class Support Vectore Machines”, Neural Networks, IEEE Transactions on (Vol13 , Issue: 2 ), Pg(s): 415–425,Mar 2002 , DOI:10.1109/72.991427

[21] Y. Sun, M. S. Kamel, and Y. Wang, “Boosting for learning multiple classes with imbalanced class distribution” ,in Proceedings of International Conference on Data Mining, pp. 592–602, IEEE Computer Society, (Washington, DC, USA), 2006.

[22] N. Cesa-Bianchi, C. Gentile, L. Zaniboni, Incremental algorithms for hierarchical classification, The Journal of Machine Learning Research 7 (2006) 31-54.

[23] Y. Murphey, H. Wang, G. Ou, and L. Feldkamp, “OAHO: an effective algorithm for multi-class learning from imbalanced data,” in International Joint Conference on Neural Networks (IJCNN), Aug. 2007, pp. 406–411.

[24] T. Li, M. Ogihara, "Music genre classification with taxonomy", Proceedings of IEEE Int. Conf. on Acoustics Speech and Signal Processing 197-200 2005.

[25] F. Wu, J. Zhang, V. Honavar, "Learning classifiers using hierarchically structured class taxonomies", in: Proceedings of the Symposium on Abstraction, Reformulation, and Approximation, Springer 3607 313-320, 2005.

[26] N. Cesa-Bianchi, C. Gentile, L. Zaniboni. "Incremental algorithms for hierarchical classification", The Journal of Machine Learning Research 7. 2006, Vol. 7, pp. 31-54.

[27] A.Ramanan, S.Suppharangsan, and M.Niranjan, "Unbalanced Decision Trees for Multi-class Classification", 2007, IEEE, International Conference on Industrial and Information Systems, 9-11 Aug. 2007, ISSN : 2164-7011,doi :10.1109/ICIINFS.2007.4579190

[28] Y. Chen, M. M. Crawford, J. Ghosh, "Integrating Support Vector Machines in a hierarchical output space decomposition framework", in: Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing 2 (2004) 949-952.

[29] S. Kumar, J. Ghosh, M. M. Crawford, "Hierarchical fusion of multiple classifiers for hyperspectral data analysis", Pattern Analysis and Applications 5 (2002) 210-220

[30] H. Chiang, Y. K. Tu, "Hierarchically SVM classification based on support vector clustering method and its application to document categorization", Expert Systems with Applications 33 (2007) 627-635.

[31] T. Ryan Hoens, Qi Qian, Nitesh V. Chawla, and Zhi-Hua Zhou, "Building Decision Trees for the Multi-class Imbalance Problem", Springer Berlin Heidelberg, Advances in Knowledge Discovery and Data Mining,Volume 7301 of the series Lecture Notes in Computer Science,2012, pp 122-134, doi: 10.1007/978-3-642-30217-6_11

[32] Bing LUO & Yun ZHANG, " Hierarchical Classification for Imbalanced Multiple Classes in Machine Vision Inspection ", Fourth International Conference on Image and Graphics , 0-7695-2929-1/07 © 2007 IEEE , DOI 10.1109/ICIG.2007.81

[33] Chih-Wei Hsu and Chih-Jen Lin, ”A Comparison of Methods for Multi-class Support Vectore Machines”, Neural Networks, IEEE Transactions on (Vol13 , Issue: 2 ), Pg(s): 415–425,Mar 2002 , DOI:10.1109/72.991427

[34] B. Epshtein, S. Ullman, "Feature hierarchies for object classification", in: Proceedings of International Conference on Computer Vision (ICCV), 220-227, 2005.

[35] P. X. Huang, B. J. Boom, B. F. Fisher, "Underwater Live Fish Recognition using a Balanced-Guaranteed Optimized Tree", in: Proceedings of 11th Asian Conference on Computer Vision (ACCV), 2012

[36] C. Freeman, D. Kulic, O. Basir, "Joint feature selection and hierarchical classifier

design", IEEE International Conference on Systems Man and Cybernetic (2011) 1728-1734
[37] C. Ferri, J. Haernandez-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," Pattern Recognition Letters, vol. 30, pp. 27–38, 2009.
[38] R. Caruana and A. Niculescu-Mizil, "Data mining in metric space: An empirical analysis of supervised learning performance criteria," in Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data mining, (Seattle, WA), pp. 69–78, 2004.
[39] In C. Drummond, W. Elazmeh, N. Japkowicz, and S.A. Macskassy, editors," A review of performance evaluation measures for hierarchical classifiers " ,Evaluation Methods for Machine Learning II: papers from the AAAI-2007 Workshop, AAAI Technical Report WS-07-05, pages 182-196. AAAI Press, July 2007.
[40] Nathalie Japkowicz, Mohak Shah," Evaluating Learning Algorithms: A Classification Perspective", 32 Avenue of the Americas, NK,,10013-2473,Cambridge Press, the book first published 2011, ISBN 978-0-521-19600-0, USA.
[41] Lawrence Mosley, Iowa State University," A balanced approach to the multi-class imbalance problem" PhD thesis 2013.
[42] X.Y. Liu and Z.H. Zhou, "Training Cost-Sensitive Neural Networks with Methods Addressing the Class Imbalance Problem", IEEE Trans. Knowledge and Data Eng., vol. 18, no. 1, pp. 63-77, Jan. 2006.
[43] N. Abe, B. Zadrozny, and J. Langford, "An Iterative Method for Multi-Class Cost-Sensitive Learning," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 3-11, 2004.
[44] Van Rijsbergen, C, J, 1979, "Information Retrieval", (2nd ed), Butterworth
[45] T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts", ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 40–49, 2004.
[46] T. Fawcett, "ROC Graphs: Notes and Practical Considerations for Data Mining Researchers," Technical Report HPL-2003, 4, HP Labs, 2003.
[47] T. Fawcett, "An Introduction to ROC Analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861-874, 2006.
[48] Richard M. Everson, Jonathan E. Fieldsend," Multi-class ROC analysis from a multi-objective optimisation perspective", Pattern Recognition Letters - Special issue: ROC analysis in pattern recognition, Volume 27 Issue 8, June 2006 Pages 918-927 , doi:10.1016/j.patrec.2005.10.016
[49] D.J. Hand and R.J. Till, "A Simple Generalization of the Area under the ROC Curve to Multiple Class Classification Problems," Machine Learning, vol. 45, no. 2, pp. 171-186, 2001.
[50] F. Provost and P. Domingos, "Well-Trained Pets: Improving Probability Estimation Trees," CDER Working Paper: IS-00-04, Stern School of Business, New York Univ., 2000.
[51] R. Alejo, J. A. Antonio, R. M. Valdovinos, J. H. Pacheco-Sánchez" , Assessments Metrics for Multi-class Imbalance Learning: A Preliminary Study", Springer Berlin Heidelberg, Volume 7914 of the series Lecture Notes in Computer Science pp 335-343, 2013.
[52] Thomas Landgrebe, and Robert P.W. Duin, "A simplified extension of the Area under the ROC to the Multi-class domain", 17th annual Symposium of the Pattern Recognition Association of South Africa, November 2006.
[53] Hyndman, Rob J. Koehler, Anne B.,Koehler, 2006, "Another Look at Measures of Forecast Accuacy", International journal of forecasting, Vol 22, pp 679-688, doi: 10.1016/j.ijforecast.2006.03.001
[54] Sidney D΄Mello, Arthur Graesser, Bjoern Schuller, Jean-Claude Martin," Affective Computing and Intelligent Interaction" :

Fourth International conference,AC|| 2001 emphis,TN,USA, October 2011, Proceedings Part1..

[55] Henry Carrillo1 , Kay H. Brodersen Jose A. Castellanos," Probabilistic performance evaluation for Multi-class classification using the posterior balanced accuracy", ROBOT2013: First Iberian Robotics Conference Vol. 252 of the series Advances in Intelligent Systems and Computing pp 347-361

[56] Wei Cheng and Eyke H¨ullermeier," Probability Estimation for Multi-Class Classification based on Label Ranking ", Machine Learning and Knowledge Discovery in Databases Vol. 7524 of the series Lecture Notes in Computer Science, pp 83-98.

**Hanaa S. Abdalaziz** received the B.Sc. degree in Computer Science from Al-neelain University, Sudan, in 2003, was a member of Computer Science Department in Al-Neelain University in 2005, She was received the M.Sc. degree in Computer Sceince from Al-Neelain University, Sudan, in 2006. She was received the PhD. degree in Computer Science from Sudan University for Science and Technology. Her research interests include Information Security, Machine Learning, Data Mining and Data Science.

**Fakhreldeen A. Saeed** received the B.Sc., M.Sc. and PhD degree in Computer Science from Al-neelain University, Sudan, in 2002, 2005 and 2008 respectively; He is currently an associate professor at Al-neelain University. His research interests include Web Security, Data Mining, Big Data and Data Science.