Journal of Theoretical and Applied Information Technology
31st August 2017. Vol.95. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

NEW HIERARCHICAL MODEL FOR MULTICLASS IMBALANCED CLASSIFICATION

[1] HANAA S. ABDALAZIZ, [2]FAKHRELDEEN A. SAEED
[1]Asstt Prof., Alneelain University, Department of Computer Science, Sudan
[2]Assoc Prof., University of Jeddah, Department of IT, KSA
E-mail: [1]hanaasameeh@gmail.com, [2]fakhry00@gmail.com

ABSTRACT

Multiclass imbalanced datasets exist in a wide variety of real-world applications where each instance should be assigned to one of N different classes that suffer from imbalanced distribution of instances. The misclassification of such instances is much expensive because they are the most intended. Another fact is that there is a significant concentration on the binary class imbalance problem, while multiclass datasets have been received less consideration. The main aim of this paper is at getting a more precise assignment of the few or the rare examples to their minority classes via presenting a novel hierarchical model based on Support Vector Machine (SVM) and MultiSVM. The model works using a new Algorithm (we call it Grouping Algorithm, it is not clustering) to create new balanced artificial groups from the original imbalanced classes, then heals the multiclass situation and carries out classification process through hierarchical steps. The model is tested with and without adding weights during classification process as well as the support vector machine, so results of the four machines are compared. The experiments are performed on nine Multiclass imbalanced datasets from U.C.I Repository from different fields and characteristics. When applying the proposed hierarchical model without weight, it achieves the best results in 4 out of 9 datasets in terms of Accuracy and kappa. When empowered with the weight it presents the best of 6 of 9 datasets in terms of G-mean, 4 of the 9 datasets considering Mean F-Measure(MFM) but they vary regarding the OVERALL ACCURACY. The experiments also demonstrate that the proposed model performs well even when increasing the number of classes.

Keywords: *Imbalanced Multiclass dataset, Imbalanced learning, Hierarchical classification.*

1. INTRODUCTION

Multiclass imbalanced datasets are kind of data that presents more challenges to learn from. While the binary imbalance learning solutions are well surveyed and established, the ones of learning from Multiclass imbalanced datasets are not yet. The Multiclass imbalance problem belongs to the supervised machine learning tasks, where each instance should be assigned to one of N different classes that have unequal sample sizes. It is kind of data that owns more complex characteristics that introduces more obstacles and issues to be considered during learning process and requires more sophisticated tools and more practical techniques that do not suffer from implementation complexity as majority of the previously introduced ones, so one objective of this paper is to get a solution that meets these requirements. Another fact is that the utilized evaluation metrics vary significantly across the Multiclass data and class imbalance literatures, so far, no single metric is totally agreed to assess the performance of each learning machine and could be applied over such Multiclass imbalanced data. Their suitability differs from a dataset to another. The importance of this paper rises from the considering this type of data which is produced from many real sensitive applications and fields in our life, such as the medical diagnosis, fraud detection, risk management in telecommunications, intrusion detection...etc[1][2][3][4].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Support Vector Machine (SVM) is a strong classic machine learning tool that has been widely used and it maintained magnificent results that stand on a solid mathematical ground. It is very effective tool even when trained by small sizes of samples. So, the main aim of this paper is at getting more precise assignment of the few or the rare examples to their minority classes, through developing a simple model for the classification process basing on Support Vector Machine (SVM) and Multiclass SVM. Then, investigate the overall performance through suitable assessment metrics empirically. So, the contribution of this paper can be briefed in building this model which is based on a new Grouping algorithm for the dataset classes while not depending on the similarities between instances such as the way the clustering technique works, instead, the algorithm originates new balanced artificial groups from the original imbalanced classes. So, this model does not use any fixed hierarchy based on features and/or classes, but, in order to group the heterogeneous different classes, the model gets the benefit of the black box of the nature of the Support Vector Machine. This algorithm provides no computational complexity or algorithmic modification or even data distribution adjustment as a preprocessing step for the classification process, so, it is different from common hierarchical methods which use supervised learning.

The rest of this paper is structured as follows: the subsection 2.1 addresses the Problems of learning from Multiclass Imbalanced Datasets, followed by subsection 2.2 Methods of handling Multiclass Imbalanced data. Section 3 illustrates the proposed hierarchical SVM model. Section 4 presents the chosen benchmarks and the experiments setting up. Evaluation Performance metrics and end results are shown in section 5 and 6 respectively. Finally, the conclusion in section 7.

2. MULTICLASS IMBALANCED

2.1 Multiclass Imbalanced Datasets problems

The imbalance nature of the data affects the learning process in many aspects [5] [6], so naturally, the situation becomes more severe when learning from multiclass imbalanced datasets; several boundaries have to be determined and constructed and they can be overlapped causing increasing in the probability of error while dealing with multiclass imbalanced because of the multiclass nature of data. Moreover, Zhou and Liu [7] stated that most of the techniques developed for balancing binary classification instances become powerless when dealing with multiclass learning problems [8] and some methods are not applicable directly such as random oversampling and undersampling techniques, so, the problem is worse if the multiclass data is imbalanced as well. In addition, the performance evaluation metrics that dedicated for two class scenarios are not suitable for assessing the results of classification algorithms considering Multiclass imbalanced data accurately, which reveals the need for more sophisticated evaluation metrics.

2.2 Methods of handling Multiclass imbalanced data

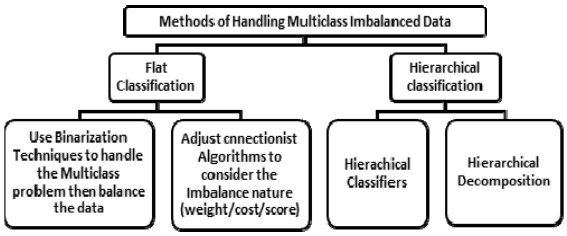

Figure 1:Methods of Handling Imbalanced Multiclass Data

The techniques which had been introduced to treat such data were naturally emanated from those dedicated for treating the multiclass balanced data and binary imbalanced data ones. So, they also could be subjoined to the traditional types of classification methods for Multiclass data: Flat and Hierarchical Classification methods as figure 1 shows , where Flat classification indicates to a single level of classes that examples should be assigned to, and 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 [9]. Regarding Flat Classification, it can be divided into two main methods – Figure 1- : The first one is using Binarization techniques that transform the multiclass data into binary imbalanced sub-datasets [10] then rebalance the data. The Binarizations strategies include: One Against One (OAO) and One Against All (OAA) [11], Error-Correcting Output-Coding (ECOC) [12]. Then as second step, many balancing strategies can be utilized such as various kinds of Sampling techniques [13][14][15], Boosting and ensemble techniques [16][17][18][19] and Kernel-based learning methods like Support vector machines (SVM) [20][21][22][23]. The second approach to handle the Multiclass imbalanced data is via adjusting the Extensible Algorithms – as Neural Networks, k-Nearest Neighbor, Naive Bayes

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classifiers and SVM – to consider both multiclass and imbalance together [24]. Here, the modification introduces costs/weights for minority instances during the classification process or changing the decision threshold considering the multiclass situation as well. This could be applied by utilizing cost sensitive methods to find an appropriate cost matrix with multiple classes and imbalance nature [19][7][25][26]. The Hierarchical classification techniques that are dedicated for treating imbalanced multiclass data often treat the imbalance problem initially, then lever the multiclass situation by turning the classification process into stages of levels. According to Beyan & Fisher's study et al.[9], the first type of these techniques is Hierarchical Classifiers, the classes were organized in a pre-defined hierarchy like a tree [12]. The classes at each parent node are divided into several 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 is performed via a simple classifier, usually a binary classifier. This type include Decision-tree algorithms[12] and The Decision Directed Acyclic Graph (DDAGs)[27]. Considering Hierarchical decomposition, the class hierarchy is formed regarding some factors such as the similarity of data or its classes[9]. Here, there is no pre-defined class hierarchy, it places the classes in a tree, usually a binary tree, utilizes a hierarchical division of the output space[28]. Binary Tree of Classifiers (BTC)[29][30][1] can be an example of this type.

3. THE PROPOSED HIERARCHICAL SVM MODEL

Figure 2 illustrates the structure of the proposed hierarchical model.

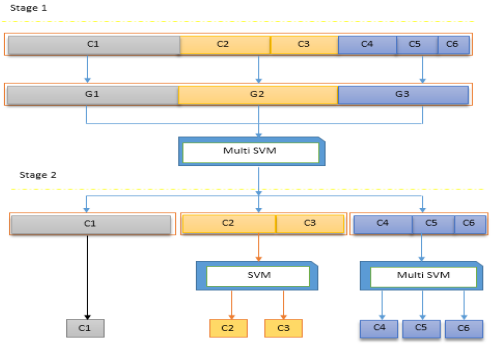

Figure 2: the structure of the proposed hierarchical model

3.1 How does the model work?

The model goes through two main stages:

STAGE ONE: Treat the Imbalance Situation: We decompose the classification stages into a series of sub-decisions stages. The dataset classes will be reorganized in new groups such that the differences between the number of the instances in the groups is almost or nearly balanced, regardless to the number of the classes in each group. So, a group may include just a class or more. We achieve the previous step through Grouping Algorithm that originates new artificial balanced groups. The algorithm goes through the following procedures:

1. Reorder the classes decently according to the number of the instances in each class, i.e. the classes' sizes $\{C_1, C_2, C_3 \ldots C_L\}$.
2. Starting from the last class C_L in the ordered list of the classes, add the number of its instances (C_L) to those belong to the former classes in the ordered classes list $\{C_L+C_{L-1}+\ldots+C_{L-N}= SUM\}$ till the accumulated summation (SUM) becomes bigger than the number of the instances of class $C_{L-(N+1)}$ (the class at the top of the ordered list that contains the biggest number of instances).
3. If the difference between the accumulated summation (SUM) and the number of the instances at class (C_{L-N}) is less than the difference between the number of the instance of that corresponding class (C_{L-N}) and $C_{L-(N+1)}$ then join all the classes starting from the last class C_L up to C_{L-N} in one group G_1 and the each one of the rest of the classes $\{ C_{L-(N+1)}, \ldots..C_1\}$ will be in an independent group, else join all the classes starting from the last class C_L up to $C_{L-(N-1)}$ in one group G1 and the each one of the rest of the classes $\{C_{L-N}, \ldots..C_1\}$ will be in an independent group.
4. Start new level in the hierarchy.
5. Repeat the previous procedure to the classes in G_1, noticing that the class C_{L-N} or $C_{L-(N+1)}$ will be the top of its ordered classes. Then repeat them in every new formed group till regroup all the dataset classes following the same way.

STAGE TWO: The Mutli-stages of Classification: After reorganizing the original dataset in new sub datasets, each one will be examined by an independent SVM machine. At each level in the hierarchy, if the SVM decides to assign some tested example to an internal group that contains two classes or more, a new SVM will be applied to that group to assign the example for one of the classes it contains.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Classes Grouping Algorithm

	Name	#Attributes	#Examples in each Class	# Total Example	IR
1	**Yeast**	8	244/429/463/44/35/51/163/30/20/5	1484	23.15
2	**New-Thyroid**	5	150/35/30	215	4.84
3	**Dermatology**	34	112/61/72/52/49/20	366	5.55
4	**Balance**	4	49/288/288	625	5.88
5	**Glass Identification**	9	70/76/17/13/9/29	214	8.44
6	**Thyroid**	21	666/17/37	720	36.94
7	**Ecoli**	7	143/77/2/2/35/20/5/52	336	71.5
8	**Page Blocks**	10	492/33/12/8/3	548	164
9	**Shuttle**	9	1706/338/123/6/2	2175	853

Figure 3 illustrates the Grouping algorithm, wrote in pseudo-code.

Classes Grouping Algorithm

Input: n: Number of class; x [n]: Array of Number of samples for each class
Output: New balanced Groups
1: repeat
2: Let j=0
3: Let y[0]=x[0]
4: **repeat**
5: Let j=j+1
6: Let y[j]=y[j-1]+x[j];
7: Let t=j
8: **until** y[j]<x[n-1]
9: **if** ((y[t]-x[n-1])>(x[n-1]-x[t])) **then**
10: **return** a new group including the considering class only and another group contains the rest of the classes
11: Let n=t
12: **else**
13 **return** a new group including the considering class as well as the rest of the classes
14: n=t+1;
15: **end if**
16: **until** t>1

Figure 3:Classes Grouping Algorithm

4. BENCHMARK DATASETS

For the experimental setup, we ran 10 iterations of 10-fold cross-validation. Nine popular imbalanced data sets were selected from U.C.I. Repository. The datasets are from different fields such as biology, physics, medicine, etc. While choosing these datasets, we tried to cover the range of variety in the datasets properties. The selection was based on: A range of Imbalance Ratio (IR), Variation in number of Classes (#Class), A varying number of total examples (#Examples) and number of attributes (#Attributes). The following table shows the selected benchmark datasets with their characteristics:

Table 1:The Benchmark Datasets & their Statistics

Each dataset will be examined by four machines: SVM without weight, SVM with weight, the proposed model without weight and the proposed model with weight. In order to describe the Grouping algorithm details, a number of abbreviations and colored cells are used. Table 2 illustrates the meaning of each:

Table 2:Abbreviations & colored cells

HS	Highest number of sample
i	The class number in the descendly ordered list
S(i)	Summation of the classes {C1 to Ci}
S(t)	Summation of the classes {C1 to Ct}
The yellow cell	indicates to the biggest number of examples
The dark blue cell (t)	Indicates to the examples number of the corresponding class which will be tested either to be included alone in a group or to be joined to the rest of the classes in a group
The light blue cells	The Summation of the samples of the descendly ordered classes
The green cell	Indicates that the corresponding class cell will be separated in a new group
The red cell	Indicates that the corresponding class cell will be included with its following classes in a new group

Table 3, Table 4, Table 5, Table 6, and Table 7 clarify the steps of applying the Grouping Algorithm of the model over the YEAST dataset. It will be applied over the rest of the selected datasets in the same way.

Table 3:Classification of Applying the Grouping Algorithm over YEAST dataset classes: STEP1

	I	Class	Samples	Summation	HS-S(i)	HS-S(t)
	10	CYT	463	1484	-1021	
	9	NUC	429	1021	-558	
t	8	MIT	244	592	-129	219
	7	ME3	163	348	115	
	6	ME2	51	185	278	
	5	ME1	44	134	329	
	4	EXC	35	90	373	
	3	VAC	30	55	408	
	2	POX	20	25	438	
	1	ERL	5	5	458	

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4:Classification of Applying the Grouping Algorithm over YEAST dataset classes: STEP2

	i	Class	Samples	Summation	HS-S(i)	HS-S(t)
	8	MIT	244	592	-348	
t	7	ME3	163	348	-104	81
	6	ME2	51	185	59	
	5	ME1	44	134	110	
	4	EXC	35	90	154	
	3	VAC	30	55	189	
	2	POX	20	25	219	
	1	ERL	5	5	239	

Table 5:Classification of Applying the Grouping Algorithm over YEAST dataset classes: STEP3

	i	Class	Samples	Summation	HS-S(i)	HS-S(t)
	6	ME2	51	185	-134	
	5	ME1	44	134	-83	
	4	EXC	35	90	-39	
T	3	VAC	30	55	-4	21
	2	POX	20	25	26	
	1	ERL	5	5	46	

Table 6:Classification of Applying the Grouping Algorithm over YEAST dataset classes: STEP4

	i	Class	Samples	Summation	HS-S(i)	HS-S(t)
	3	VAC	30	55	-25	0
	2	POX	20	25	5	
	1	ERL	5	5	25	

Table 7:Classification of Applying the Grouping Algorithm over YEAST dataset classes: STEP5

	i	Class	Samples	Summation	HS-S(i)	HS-S(t)
	2	POX	20	25	-5	0
	1	ERL	5	5	15	

Figure 4 shows how the dataset 1 (YEAST) classes will be formed in multiple stages by the model:

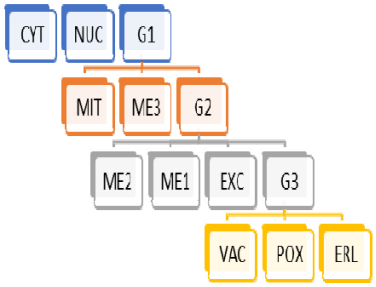

Figure 4:Applying the Grouping Algorithm over Dataset

5. PERFORMANCE EVALUATION

There are three families of evaluation metrics used in the context of classification. The threshold metrics (e.g. accuracy and F-measure), the ranking methods (e.g. Receiver Operating Characteristics (ROC) analysis and AUC), and the probabilistic metrics (e.g. Root-mean-squared error). The first class can have a multiple-or a single-class focus. The multiple-class focus metrics consider the overall performance of the learning algorithm on all the classes in the dataset. For results evaluation, we used:

Class Balance Accuracy or **Recall** (j) or **Acc** (j). It is defined as:

For any Ck confusion matrix:

$$CBA = \frac{\sum_k^k \frac{c_{kk}}{\max(c_{kj}, c_{jk})}}{k}$$

Where C_k denote a k×k confusion matrix or contingency table of actual class labels aligned by their model predictions, with c_{ij} 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. [19] to Multiclass scenarios. It is defined as the geometric mean of the Recall values of all classes. Given a j-class problem:

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

It can capture the balanced performance among classes effectively, as the recognition rate of every class is equally taken into account. Moreover, considering cost-sensitive learning, it is natural to use misclassification costs for performance evaluation for multiclass imbalanced problems [7].

Mean F–measure (MFM): this measure aggregates both the Precision and the Recall of the minority class. To be used, we calculate the hierarchical F-measure [31][32]:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$hF = \frac{2 * hP * hR}{hP + hR}$$

$$hP = \frac{\sum_i |\vec{P}_i \cap \vec{T}_i|}{\sum_i |\vec{P}_i|}$$

$$hR = \frac{\sum_i |\vec{P}_i \cap \vec{T}_i|}{\sum_i |\vec{T}_i|}$$

Where hP is the hierarchical precision and hR is the hierarchical recall. $\vec{P}_i$ is the hierarchical categories predicted for test example xi while $\vec{T}_i$ is the true categories of xi.

$$MFM = \frac{\sum_{i=1}^{|C|} F - \text{measure}(C_i)}{|C|}$$

Kappa Statistic: It is a measure that compares the accuracy of the system to the accuracy of a random system [33].

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

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

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

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

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

6. END RESULTS DISCUSSIONSS

Table 2, Table 3, Table 4 and Table 5 show the results of applying the four classification methods (SVM, SVM with weight, the new model without weight and the proposed model with weight) considering (Overall Accuracy, G-mean, MFM and Kappa) respectively. The highlighted cells in the tables refer to the best results obtained.

Table 8: Overall Accuracy of the Four Methods

	SVM	SVM with weight	New model	New model with weight	# Class	IR
new-thyroid	0.9444	0.9448	0.9075	0.9369	3	4.84
dermatology	0.2074	0.2074	0.3568	0.3568	6	5.55
balance	0.8735	0.4671	NA	NA	3	5.88
glass	0.4578	0.3918	0.5305	0.5202	6	8.44
yeast	0.312	0.0101	0.3696	0.1997	10	23.15
thyroid	0.925	0.3833	0.8490	0.2853	3	36.94
ecoli	0.4257	0	0.4938	0.3111	8	71.5
pageblocks	0.9161	0.9197	0.7609	0.7654	5	164
Shuttle	0.9936	0.9807	0.9856	0.9780	5	853
AVERAGE	0.64775	0.479725	0.656713	0.544175		

Table 9:G-mean of the Four Methods

	SVM	W-SVM	New Model	W-New Model
new-thyroid	0.8762	0.9304	0.8951	0.9432
dermatology	0.3426	0.3426	0.4524	0.4524
balance	0.6319	0.6145	NA	NA
glass	0.2966	0.4288	0.5408	0.5565
yeast	0.1	0.0428	0.6266	0.2846
thyroid	0.3333	0.2720	0.4688	0.4848
ecoli	0.125	0	0.6530	0.4296
pageblocks	0.2908	0.43158	0.4762	0.5327
Shuttle	0.6905	0.75558	0.9789	0.9858
AVERAGE	0.381875	0.40047	0.636475	0.5837

Table 10:MFM for the Four Methods

	SVM	W-SVM	New model	W-New model
new-thyroid	0.918893	0.9262	0.8964	0.9354
dermatology	NA	NA	NA	NA
balance	NA	0.4713	NA	NA
glass	NA	0.3969	NA	NA
yeast	NA	NA	NA	0.1147
thyroid	NA	0.2168	NA	0.2910
ecoli	NA	NA	NA	0.4360
pageblocks	NA	NA	0.5699	0.6115
Shuttle	NA	NA	0.9873	0.9813
AVERAGE	0.102099	0.223467	0.272622	0.374433

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 11:kappa for the Four Methods

	SVM	W-SVM	New Model	W-New Model
new-thyroid	0.868	0.88	0.855	0.902
dermatology	0.089	0.089	0.259	0.249
balance	0.765	0.311	NA	NA
glass	0.195	0.237	0.427	0.434
yeast	0	0	0.329	0.128
thyroid	0	0	0.373	0.022
ecoli	0	0	0.441	0.204
pageblocks	0.289	0.422	0.467	0.501
shuttle	0.982	0.947	0.98	0.97
AVERAGE	0.302875	0.321875	0.516375	0.42625

Figure 3, Figure 4, Figure 5 and Figure 6 demonstrate the results of applying the four methods (SVM, SVM with weight, the new model without weight and the proposed model with weight).

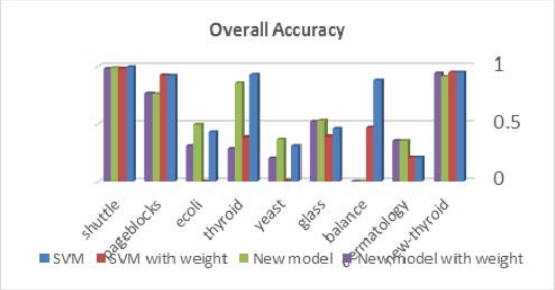

Figure 5: Overall Accuracy of the four methods

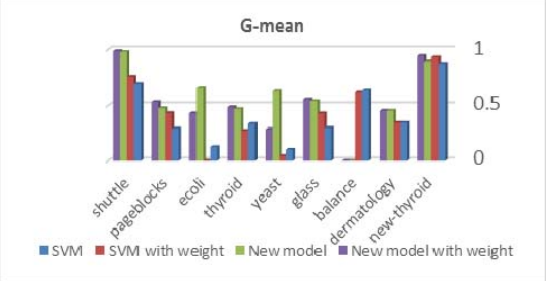

Figure 6:G-mean of the four methods

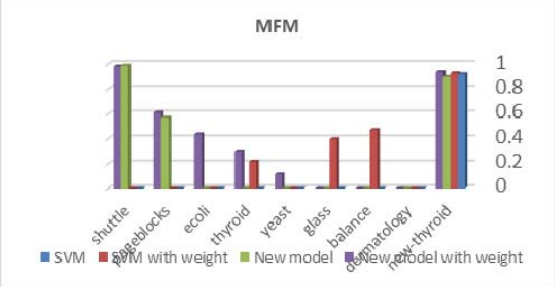

Figure 7:MFM for the four methods

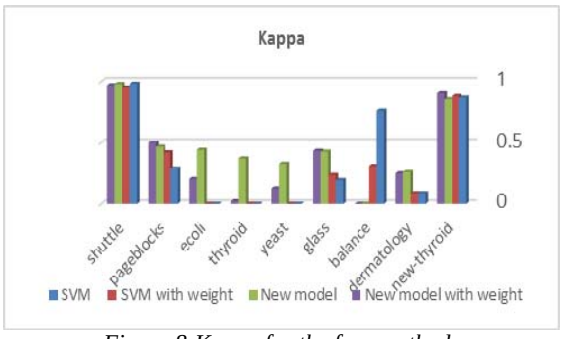

Figure 8:Kappa for the four methods

Generally, the results demonstrate that the performance of the proposed hierarchical method produces the best results. When applying the proposed hierarchical model without weight, it achieves the best results in 4 out of 9 datasets in terms of Accuracy and kappa. When empowered with weight it presents the best on 6 of 9 datasets in terms of G-mean, 4 of 9 datasets considering MFM but they vary regarding the OVERALL ACCURACY. The high performance in terms of G-mean also shows that it is good at the classification of minority class while as good as other methods for classification of majority class (can be infer from G-mean and kappa results). Average results over the 9 datasets also show that the proposed method is the best method for the four metrics. Regarding the Overall Accuracy, we notice that the model works better as the number of the classes increases; considering the datasets Yeast, Ecoli, Glass Dermatology which have 10,8,6,6 classes respectively, the results are better when comparing with the datasets New-Thryroid, Thyroid, Pageblocks and Shuttle which have 3,3,5 classes respectively - less number of classes-.

The results also demonstrate that using the suggested hierarchical model fails in imbalance multiclass learning in a certain situation. Considering dataset 3 (Balance), it is incapable of applying the Grouping algorithm to redistribute the instances in new artificial groups. Regarding the way of the algorithm works, the new groups are identical with the original classes. Therefore, in this case the model is not applicable (NA) for such dataset. The reason for this seems to be the little number of classes that could not be decomposed into different new groups of nearly balanced numbers of examples. To more consolidate the model performance, different weights are added to the minority classes during learning process, but we noticed that they do not provide very high advantages for its performance. For example, the Recall of Yeast dataset when applying the model without supporting it with weights is better from its

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

value when supplying the model with them. The same situation we get considering Ecoli dataset in terms of Recall, Precision and F-measure. We also observe that the Recall and F-measure are very similar in Glass dataset –as another instance -. Conversely, the MultiSVM is improved significantly when adding the weights during learning. The optimal weights are in various ranges for different problems. They are decided by the proportion of the corresponding class examples within the whole data set. It can be given as:

Weight of $class_i$ = total sample/ (number of class * sample of $class_i$)

Regarding another perspective, the model performance dose not affected by the increasing the number of the dataset features; the results of applying the model over the New-Thyroid and Thyroid datasets -which are similar in the classes number but differ in both imbalance ratio and the number of the features - show that using the suggested hierarchical model provides advantages in both datasets despite the difference in their features number. It is important to mention that this model - naturally- owns the flaws of hierarchical classification models that cannot produce their final classification result unless the path from the root to the final leaf is passed, which may consume more time.

7. CONCLUSION

This study is set out to investigate how does the performance of the classifier that deal with multiclass imbalanced dataset to classify rare data (samples) can be improved. In order to perform this research; two questions are constructed: How can we get more precise assignment of few or rare samples of minority classes? What are the most suitable evaluation metrices can be used? The study presents a novel hierarchical model based on a new Grouping Algorithm for rebalancing the dataset classes and SVM and MultiSVM for classification process which is carried out through the levels of the hierarchy. Interestingly the new model performs well even when increasing the number of classes. This means that the proposed method is more successful than utilizing a Support Vector Machine even when it is powered with weight during the classification process. There are number of suggestions that this study can introduce for future work. First, other data mining tools of classifications such as Neural Networks or ensemble techniques could be examined instead of using SVM in the model. Secondly, real life data can be examined so as to better demonstrate the model performance. Thirdly, large scale of data can be tested as well. Finally, another strategy of grouping the classes can be set and tried instead of basing on the concept of the least difference between the created groups.

REFERENCES

[1] K. Chen, B. Lu, and J. T Kwok, "Efficient Classification of Multi-label and Imbalanced Data using Min-Max Modular Classifiers," *2006 IEEE Int. Jt. Conf. Neural Netw. Proc.*, no. March 2016, pp. 1770–1775, 2006.

[2] X. Zhao, X. Li, L. Chen, and K. Aihara, "Protein classification with imbalanced data," *Proteins Struct. Funct. Genet.*, vol. 70, no. 4, pp. 1125–1132, Mar. 2008.

[3] A. Tan, D. Gilbert, and Y. Deville, "Multiclass protein fold classification using a new ensemble machine learning approach.," *Genome Inform.*, vol. 14, no. July, pp. 206–217, 2003.

[4] T. W. Liao, "Classification of weld flaws with imbalanced class data," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1041–1052, Oct. 2008.

[5] Y. M. Haibo He, *Imbalanced Learning.* Hoboken, NJ, USA: John Wiley & Sons, Inc., 2013.

[6] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009.

[7] Z. H. Zhou and X. Y. Liu, "On Multiclass cost-sensitive learning," *Comput. Intell.*, vol. 26, no. 3, pp. 232–257, Aug. 2010.

[8] Z. H. Zhou and X. Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan. 2006.

[9] C. Beyan and R. Fisher, "Classifying imbalanced data sets using similarity based hierarchical decomposition," *Pattern Recognit.*, vol. 48, no. 5, pp. 1653–1672, May 2015.

[10] G. Ou and Y. L. Murphey, "Multiclass pattern classification using neural networks," *Pattern Recognit.*, vol. 40, no. 1, pp. 4–18, Jan. 2007.

[11] R. Rifkin and A. Klautau, "In defense of one-vs-all classification," *J. Mach. Learn. Res.*, vol. 5, pp. 101–141, 2004.

[12] T. G. Dietterich and G. Bakiri, "Solving Multiclass Learning Problems via Error-Correcting Output Codes," *Jouranal Artifical Intell. Res.*, vol. 2, pp. 263–286, 1995.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[13] S. J. Yen and Y. S. Lee, "Cluster-based under-sampling approaches for imbalanced data distributions," *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 5718–5727, Apr. 2009.

[14] L. Abdi and S. Hashemi, "To Combat Multiclass Imbalanced Problems by Means of Over-Sampling Techniques," in *IEEE Transactions on Knowledge and Data Engineering*, 2016, vol. 28, no. 1, pp. 238–251.

[15] B. Debowski, S. Areibi, G. Grewal, and J. Tempelman, "A dynamic sampling framework for Multiclass imbalanced data," in *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, 2012, vol. 2, pp. 113–118.

[16] C. Li, "Classifying imbalanced data using a bagging ensemble variation (BEV)," *Proc. ACM Southeast Reg. Conf.*, pp. 203–208, 2007.

[17] N. V Chawla, A. Lazarevic, L. O. Hall, and K. Bowyer, "SMOTEBoost: improving prediction of the minority class in boosting," *Princ. Knowl. Discov. Databases, PKDD-2003*, pp. 107–119, 2003.

[18] S. Wang and X. Yao, "Multiclass Imbalance Problems: Analysis and Potential Solutions.," *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 42, no. 4, pp. 1–13, 2012.

[19] Y. Sun, M. S. Kamel, and Y. Wang, "Boosting for learning multiple classes with imbalances class distribution," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2006, pp. 592–602.

[20] Y. Zhang, P. Fu, W. Liu, and G. Chen, "Imbalanced data classification based on scaling kernel-based support vector machine," *Neural Comput. Appl.*, vol. 25, no. 3–4, pp. 927–935, Sep. 2014.

[21] Jie Xu, B. Zou, and Hanlei Shen, "Learning performance of Multiclass support vector machines based on Markov sampling," in *2015 11th International Conference on Natural Computation (ICNC)*, 2015, pp. 74–80.

[22] S. Zou, Y. Huang, Y. Wang, J. Wang, and C. Zhou, "SVM learning from imbalanced data by GA sampling for protein domain prediction," in *Proceedings of the 9th International Conference for Young Computer Scientists, ICYCS 2008*, 2008, pp. 982–987.

[23] D. Tomar and S. Agarwal, "An effective Weighted Multiclass Least Squares Twin Support Vector Machine for Imbalanced data classification," *Int. J. Comput. Intell. Syst.*, vol. 8, no. 4, pp. 761–778, Jun. 2015.

[24] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni, "Incremental Algorithms for Hierarchical Classification," *J. Mach. Learn. Res.*, vol. 7, pp. 31–54, 2006.

[25] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognit.*, vol. 40, no. 12, pp. 3358–3378, Dec. 2007.

[26] N. Abe, B. Zadrozny, and J. Langford, "An Iterative Method for Multiclass Cost-sensitive Learning," in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2004, pp. 3–11.

[27] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification," in *Advances in Neural Information Processing Systems*, 2000, pp. 547–553.

[28] M. Aly and <malaa@caltech Edu>, "Survey on multiclass classification methods," *Neural Netw*. pp. 1–9, 2005.

[29] S. Kumar, J. Ghosh, and M. M. Crawford, "Hierarchical Fusion of Multiple Classifiers for Hyperspectral Data Analysis," *Pattern Anal. Appl.*, vol. 5, no. 2, pp. 210–220, 2002.

[30] P. HAO, J. CHIANG, and Y. TU, "Hierarchically SVM classification based on support vector clustering method and its application to document categorization," *Expert Syst. Appl.*, vol. 33, no. 3, pp. 627–635, Oct. 2007.

[31] C. J. Van Rijsbergen, *Information Retrieval*, 2nd ed. Butterworth-Heinemann Newton, MA, USA, 1979.

[32] A. Kosmopoulos, I. Partalas, E. Gaussier, G. Paliouras, and I. Androutsopoulos, *Evaluation measures for hierarchical classification: a unified view and novel approaches*, vol. 29, no. 3. 2014.

[33] T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, p. 40, 2004.