Analysis of multi-class classification performance metrics for remote sensing imagery imbalanced datasets

Análisis de métricas de rendimiento de clasificación multi-clase para conjuntos de datos desbalanceados de imágenes de percepción remota

GONZALEZ-RAMIREZ, Andrea^{†*}1, LOPEZ, Josue¹, TORRES-ROMAN, Deni¹ and YAÑEZ-VARGAS, Israel²

¹CINVESTAV del IPN, Unidad Guadalajara, Laboratorio de Telecomunicaciones
²Universidad Politécnica de Juventino Rosas, Departamento de Ingeniería en Redes y Telecomunicaciones

ID 1st Author: Andrea, González-Ramírez / ORC ID: 0000-0001-9961-4763, CVU CONACYT ID: 1001207
ID 1st Coauthor: Josué, Lopez / ORC ID: 0000-0001-9703-3973, CVU CONACYT ID: 486202
ID 2nd Coauthor: Deni, Torres / ORC ID: 0000-0002-9813-7712, CVU CONACYT ID: 20075
ID 3rd Coauthor: Israel, Yañez-Vargas / ORC ID: 0000-0001-5749-8442, CVU CONACYT ID: 295711

DOI: 10.35429/JQSA.2021.22.8.11.17Received March 25, 2021; Accepted June 30, 2021

Abstract	Resumen
Remote sensing imaging datasets for classification generally present high levels of imbalance between classes of interest. This work presented a study of a set of performance evaluation metrics for an imbalance dataset. In this work, a support vector machine (SVM) was used to perform the classification of seven classes of interest in a popular dataset called Salinas-A. The performance evaluation of the classifier was performed using two types of metrics: 1) Metrics for multi-class classification, and 2) Metrics based on the binary confusion matrix. In the results, a comparison of the scores of each metric is developed, some being more optimistic than others due to the bias that they present given the imbalance. In addition, our case study helps to conclude that the Matthews correlation coefficient (MCC) presents the lowest bias in imbalanced cases and is regarded to be robust metric. These results can be extended to any imbalanced dataset taking into account the equations developed by Luque.	Los conjuntos de datos de imágenes de percepción remota para la clasificación generalmente presentan altos niveles de desbalance entre las clases de interés. Este trabajo presentó un estudio de un conjunto de métricas de evaluación del desempeño para un conjunto de datos desbalanceados. En este trabajo, se utilizó una máquina de vectores de soporte (SVM) para realizar la clasificación de siete clases de interés en un conjunto de datos popular llamado Salinas-A. La evaluación del desempeño del clasificador se realizó utilizando dos tipos de métricas: 1) Métricas para clasificación multiclase y 2) Métricas basadas en la matriz de confusión binaria. En los resultados se desarrolla una comparativa de las puntuaciones de cada métrica, algunas siendo más optimistas que otras por el sesgo que presentan ante el desbalance. Además, nuestro caso de estudio ayuda a concluir que el coeficiente de correlación de Matthews (MCC) presenta el sesgo más bajo en los casos de desbalance y se considera una métrica robusta. Estos resultados se pueden extender para cualquier conjunto de datos desbalanceado teniendo en cuenta las ecuaciones desarrolladas por Luque.
Remote sensing images, Metrics, Imbalanced datasets	Imágenes de percepción remota, Métricas, Conjuntos de datos desbalanceados

Citation: GONZALEZ-RAMIREZ, Andrea, LOPEZ, Josue1, TORRES-ROMAN, Deni and YAÑEZ-VARGAS, Israel. Analysis of multi-class classification performance metrics for remote sensing imagery imbalanced datasets. Journal of Quantitative and Statistical Analysis. 2021. 8-22: 11-17

* Correspondence to Author (e-mail: andrea.gonzalez@cinvestav.mx)
[†] Researcher contributing as first author.

1. Introduction

Monitoring and analysis of the land surface through remote sensing image processing has become an important task in recent years. (Cheng, 2016, Wang, 2016). Therefore, the use and design of classification algorithms, such as Support Vector Machine (SVM) (Malgani, 2004, Liu, 2017), Random Forest (RF) (Belgiu, 2016) and Artificial Neural Networks (ANN) (Long, 2017), among others, has become an essential task to study these images. One of the main aspects that can influence the performance of a classifier is the imbalance in the input dataset. This happens when one of the classes in the dataset has a greater number of samples than the others. (Chawla 2005, Krawczyk 2016).

Due to the class imbalance in the image data sets, some metrics present high levels of bias, which generates optimistic results in the evaluation of the classifiers (Luque, 2019). Hence, the study of imbalanced data sets has become an important problem in the classification, since it implies the determination of the most appropriate performance metrics that will be used for the performance evaluation of the classifier, therefore, a selection adequate evaluation metrics is an important key (Luque 2019, Boughorbel 2017, Chicco 2020).

There are many metrics that are useful for testing the capability of any single or multi-class classifier, which are useful for comparing performance and analyzing the behavior of the same model (Luque, 2019, Grandini, 2020, Chicco 2020). In this work, an analysis of the bias for a set of classification metrics is presented, both in multi-classes and in those based on the binary confusion matrix. In section 5 a comparison of a set of metrics is made in a case of study of hyperspectral remote sensing images, using SVM.

2. Contribution

This work presents a study for a set of performance evaluation metrics, in which the bias that occurs in multiclass metrics and the information obtained through the binary confusion matrix for an imbalanced data set will be analyzed. For the following study, the Matthews correlation coefficient (MCC) is considered a robust metric, as MCC exhibits the lowest bias in cases of data imbalance (Luque, 2019).

3. Problem statement

Let $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ be an imbalanced multidimensional array the input to a multi-class classifier, where I, J denote the spatial resolution and K spectral resolution, that produce a prediction $\hat{\mathbf{Y}} \in \mathbb{C}^{I \times J}$, where $\mathbb{C}$ the set of the classes of interest, and let $\mathbf{Y} \in \mathbb{C}^{I \times J}$ be the corresponding labels indicating the actual class of each element of $\mathbf{x}$, analyse the behavior of a set of performance evaluation metrics at different imbalance levels.

3.1. Mathematical definition

Given a multi-class confusion matrix $\mathbf{M} \in \mathbb{R}^{C \times C}$, obtained from the true labels $\mathbf{Y}$ and the prediction $\hat{\mathbf{Y}}$, where C denotes the number of classes of interest, and the element $m_{ij} \in \mathbf{M}$ is the number of samples that belong to class i -th, but that are classified as members of class j -th, i.e., $y_n = i$ and $\hat{y}_n = j$, for $n = 1, \dots, N$.

In this work, we use the mathematical definition of imbalance given by (Luque, 2019). The imbalance coefficient δ_c , for a given class c , is featured with a value in the $[-1, 1]$ range, where 0 means that classes are perfectly balanced. The c -class imbalance coefficient can be computed by

$$\delta_c = 2 \frac{\sum_{j=1}^C m_{cj}}{\sum_{i=1}^C \sum_{j=1}^C m_{ij}} - 1. \quad (1)$$

A performance evaluation metric μ is a function that assigns to each confusion matrix, a real value on the set $\mathbb{R}$ i.e.

$$\mu: \mathbf{Y} \times \hat{\mathbf{Y}} \rightarrow \mathbb{R}. \quad (2)$$

In general, the result of this function is in the range $[0, 1]$, where 0 means the perfect misclassification, and 1 the perfect classification.

4. Metrics for multi-class classification

4.1. Overall Accuracy

The Overall Accuracy (OA) is the ratio between the number of correctly classified elements and the overall number of samples, and it is computed from multi-class confusion matrix by

$$OA = \frac{\sum_{i=1}^C m_{ii}}{\sum_{i=1}^C \sum_{j=1}^C m_{ij}}. \quad (3)$$

This metric is one of the most used for classification performance evaluation. Nevertheless, if the dataset is imbalanced, the OA is not a reliable measure, as it produces optimistic results (Chicco, 2020).

4.2. Balanced Accuracy

An alternative metric to reduce the impact of imbalanced classes in performance evaluation is Balanced Accuracy (BA), computed by the average of another well-known metric, the Sensitivity (SNS) per class, i.e.

$$BA = \frac{\sum_{c=1}^C SNS_c}{C}, \quad (4)$$

where $SNS_c = \frac{m_{cc}}{\sum_{j=1}^C m_{cj}}$. SNS measures the proportion of the number of elements correctly classified from an individual class.

4.3. Balanced Accuracy Weighted

The Balanced Accuracy Weighted (BAW) is a metric used for imbalanced classes, in which the SNS_c of each class is weighted by its relative frequency w_c . The formula for BAW is

$$BAW = \frac{\sum_{c=1}^C SNS_w c}{\sum_{c=1}^C w_c}, \quad (5)$$

where $SNS_w c = \frac{m_{cc}}{\sum_{j=1}^C m_{cj}} w_c$ and $w_c = \frac{\sum_{i=1}^C \sum_{j=1}^C m_{ij}}{C \sum_{j=1}^C m_{cj}}$ (P, 2021). This metric is an efficient performance indicator, since the recalls are weighted by a relative frequency to the classes size (P, 2021, Grandini, 2020).

4.4. Cohen's Kappa Coefficient

The Kappa Coefficient is currently one of the most popular metrics in machine learning for classification performance evaluation (Chicco, 2020). This metric measures the inter-rater concordance, as the degree of agreement among raters. It is computed by

$$K = \frac{qs - \sum_{c=1}^C p_c t_c}{s^2 - \sum_{c=1}^C p_c t_c}, \quad (6)$$

where $q = \sum_{c=1}^C m_{cc}$ denotes the overall number of elements correctly predicted, $s = \sum_{i=1}^C \sum_{j=1}^C m_{ij}$ is the total number of samples, $p_c = \sum_{j=1}^C m_{cj}$ the number of times class c was predicted, and $t_c = \sum_{i=1}^C m_{ic}$ the number of times class c truly occurs (Grandini, 2020). Kappa coefficient is high sensitive to the marginal totals (Chicco, 2020).

4.5. Matthews Correlation Coefficient

The Matthews Correlation Coefficient (MCC) is generally considered a balanced performance evaluation metric. All the elements of the confusion matrix are included in the numerator and denominator of its formula (Eq. 7), so this metric is less biased by imbalanced datasets than other metrics (Chicco, 2020).

$$MCC = \frac{qs - \sum_{c=1}^C p_c t_c}{\sqrt{(s^2 - \sum_{c=1}^C p_c^2)(s^2 - \sum_{c=1}^C t_c^2)}} \quad (7)$$

The main disadvantage is that MCC is undefined for extreme cases, for instance, when a whole row or column of the confusion matrix is zero. Generally MCC and Kappa are used in their normalized version to be in the range $[0, 1]$, computed by $MCC_n = \frac{MCC+1}{2}$ and $Kn = \frac{K+1}{2}$ respectively.

5. Metrics based on the binary confusion matrix

For the particular case of binary classification, the elements of the confusion matrix are defined as the True Positives (TP), is the case where the prediction is true and the actual output is also true, $TP = m_{11}$, the False Negatives (FN), is the case where the prediction is false and the actual output is true, $FN = m_{12}$, the False Positives (FP), is when the prediction is true and the actual output is false, $FP = m_{21}$, and the True Negatives (TN), is the cases where the prediction is false and the actual output is false., $TN = m_{22}$.

In the work of Luque (Luque, 2019), a classical metric based on the binary confusion matrix $\mu(TP, TN, FP, FN)$ is expressed as a function $\mu(\lambda_{PP}, \lambda_{NN}, \delta)$, where λ_{PP} represents the ratio of the correctly classified positive elements $m_{PP} = TP$ and the total of truly positive elements $m_P = TP + FN$, i.e., $\lambda_{PP} = \frac{m_{PP}}{m_P}$.

And λ_{NN} denotes the ratio of the correctly classified negative $m_{NN} = FP$ and the total of truly negative elements $m_N = FP + TN$, i.e., $\lambda_{NN} = \frac{m_{NN}}{m_N}$.

Considering the balanced case, $\mu_b = \mu(\lambda_{PP}, \lambda_{NN}, 0)$, it is possible to define the impact of imbalance by the bias of a metric B_μ as

$B_\mu = \mu - \mu_b.$

(8)

Metrics	$\mu(TP, TN, FP, FN)$
PRC(PPV)	$\frac{TP}{TP+FP}$
NPV	$\frac{TN}{TN+FN}$
ACC	$\frac{TP+TN}{TP+FN+TN+FP}$
F ₁	$2 \frac{PRC \cdot \frac{TP}{TP+FN}}{PRC + \frac{TP}{TP+FN}}$
GM	$\sqrt{\frac{TP}{TP+FN} \cdot \frac{TN}{TN+FN}}$
MCCn	$\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
MKn	$PPV + NPV - 1$

Table 1 Classification performance metrics as classical metric. (Luque, 2019)

Metrics	$\mu(\lambda_{PP}, \lambda_{NN}, \delta)$
PRC(PPV)	$\frac{\lambda_{PP}(1+\delta)}{\lambda_{PP}(1+\delta)+(1-\lambda_{NN})(1-\delta)}$
NPV	$\frac{\lambda_{NN}(1-\delta)}{\lambda_{NN}(1-\delta)+(1-\lambda_{PP})(1+\delta)}$
ACC	$\lambda_{PP} \frac{1+\delta}{2} + \lambda_{NN} \frac{1-\delta}{2}$
F ₁	$\frac{2\lambda_{PP}(1+\delta)}{(1+\lambda_{PP})(1+\delta)+(1-\lambda_{NN})(1-\delta)}$
GM	$\sqrt{\lambda_{PP} \cdot \lambda_{NN}}$
MCCn	$\frac{1}{2} \left(\frac{\lambda_{PP} + \lambda_{NN} - 1}{\sqrt{[\lambda_{PP} + (1-\lambda_{NN})\frac{1-\delta}{1+\delta}][\lambda_{NN} + (1-\lambda_{PP})\frac{1+\delta}{1-\delta}]}} + 1 \right)$
MKn	$\frac{1}{2} \left(\frac{1+\delta}{(1+\delta) + \frac{1-\lambda_{NN}}{\lambda_{PP}}(1-\delta)} + \frac{1-\delta}{(1-\delta) + \frac{1-\lambda_{PP}}{\lambda_{NN}}(1+\delta)} \right)$

Table 2 Classification performance metrics as a function of imbalance. (Luque, 2019)

Metrics	$\mu_b(\lambda_{PP}, \lambda_{NN})$
PRC(PPV)	$\frac{\lambda_{PP}}{\lambda_{PP} + (1-\lambda_{NN})}$
NPV	$\frac{\lambda_{NN}}{\lambda_{NN} + (1-\lambda_{PP})}$
ACC	$\frac{\lambda_{PP} + \lambda_{NN}}{2}$
F ₁	$\frac{2\lambda_{PP}}{2 + \lambda_{PP} - \lambda_{NN}}$
GM	$\sqrt{\lambda_{PP} \cdot \lambda_{NN}}$
MCCn	$\frac{1}{2} \left(\frac{\lambda_{PP} + \lambda_{NN} - 1}{\sqrt{[\lambda_{PP} + (1-\lambda_{NN})][\lambda_{NN} + (1-\lambda_{PP})]}} + 1 \right)$
MKn	$\frac{1}{2} \left(\frac{1}{1 + \frac{1-\lambda_{NN}}{\lambda_{PP}}} + \frac{1}{1 + \frac{1-\lambda_{PP}}{\lambda_{NN}}} \right)$

Table 3 Classification performance metrics as a function of balance. (Luque, 2019)

We lead this approach to the multi-class case, starting from the multi-class confusion matrix, generating multiple binary confusion matrices M_c using the equations 9a to 9d. It is important to highlight that the results given by the transformation of the multi-class confusion matrix to binary confusion matrices are not necessarily equal to the results when using multiple binary classifiers.

$TP_c \cong m_{cc}$

(9a)

$FP_c \cong \sum_{j=1}^C m_{cj} \text{ for } j \neq c$

(9b)

$FN_c \cong \sum_{i=1}^C m_{ic} \text{ for } i \neq c$

(9c)

$TN_c \cong \sum_{i=1}^C \sum_{j=1}^C m_{ij} \text{ for } i \neq c \text{ and } j \neq c$

(9d)

Hence, the equations for each metric presented in Tables 2 y 3 can be used to evaluate classification performance given a class of reference. In this analysis, only one of the four metrics that do not present bias, studied in (Luque, 2019), was considered for comparison purposes. In addition, an overall performance can be given by the average of the metric for each class.

The classification metrics studied under this approach are: Precision (PRC), Negative Predictive Value (NPV), Accuracy (ACC), F₁ score (F₁) , Geometric Mean (GM), the Normalized versions of Matthews Correlation Coefficient (MCCn), and Markedness (MKn).

6. Results

In this work, we use an imagery dataset, to study imbalance with the performance evaluation metrics aforementioned. We use popular hyperspectral dataset for experimentation in this work, are available at Hyperspectral Remote Sensing Scenes .

In addition, a Support Vector Machine (SVM) multiclass classifier is used in this work for quantitative analysis of the classification performance evaluation metrics described in Section 4, (Cortes, 2009).

6.1. Study Case

6.1.2. SalinasA scene

Figure 1 SalinasA scene. (http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

Figure 1, shows a small sub-scene collected by the AVIRIS sensor over Salinas Valley, California. It has 86×83 pixels and 204 bands.

Figure 2 Groundtruth SalinasA scene. (http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes).

Figure 2 and Table 4 describe the Salinas ground truth, where each of its 6 classes is shown with their labels and corresponding samples (M Graña, MA Veganzons, B Ayerdi).

	Class	Samples	Training	Test
1	Brocoli-GW1	391	37	354
2	Corn-SGW	1343	143	1200
3	Lettuce-R4	616	70	546
4	Lettuce-R5	1525	130	1395
5	Lettuce-R6	674	69	605
6	Lettuce-R7	799	85	714

Table 4 Ground truth classes for the Salinas-A scene and their respective samples number (http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)

Table 4 describes the classes of the hyperspectral image, where we can observe the samples of each class, with the idea of visualizing that the data is imbalanced, where classes 2 and 4 have a greater number of samples than the others, while class 1 has the smallest number of samples.

Figure 3 shows the multi-class confusion matrix obtained from the SVM classifier implemented in this work, where 90% of the samples were used as tests and 10% of the samples as training, in which we can observe the imbalance of the classes as well as the performance of the classifier.

		Predictions				
Actual	353	0	0	1	0	0
	0	1083	26	20	60	11
	0	152	360	34	0	0
	0	5	0	1390	0	0
	0	0	0	0	603	2
	0	0	0	0	17	697
	0	0	0	0	17	697

Figure 3 Confusion matrix. *Python, own.*

With the results of the previously seen multiclass confusion matrix, we obtained the following results (See Table 5), applying the multiclass classification metrics, described in section 3, where we can observe the performance evaluation of the classifier.

Metric	Performance
OA	0.931865393
BA	0.921386044
BAW	0.918517413
AA	0.977288464
Kn	0.957131663
MCCn	0.957576207

Table 5 Metrics multi-class classification. *Python, own.*

One of the important points of this work is the analysis of metrics based on the binary confusion matrix, described in section 4. In which we perform a conversion from the multi-class confusion matrix to multiple binary confusion matrices per class using equations (9a), (9b), (9c), (9b), to apply the equations of Table 2 considering equation (1) and we obtained the results shown in Table 6, where the performance evaluation can be observed of each of the metrics.

Class	PRC	NPV	ACC	F1	GM	MCCn	MKn
1	0.9999	0.9998	0.9998	0.9986	0.9985	0.9992	0.9999
2	0.8734	0.9673	0.9431	0.8877	0.9228	0.9249	0.9203
3	0.9326	0.958	0.956	0.7725	0.7168	0.8814	0.9453
4	0.9619	0.9985	0.9875	0.9789	0.9891	0.9852	0.9802
5	0.8868	0.9995	0.9836	0.9385	0.9922	0.9656	0.9431
6	0.9817	0.9959	0.9938	0.9789	0.9868	0.9876	0.9888
Average	0.9394	0.9865	0.9773	0.9259	0.9343	0.9573	0.9629

Table 6 Classification performance evaluation metrics. Python, own.

We develop out the same process using the equations in Table 3 where the variable in equation (1) is $\delta_c = 0$, this to apply equation 8 described in section 4, where we can show the result of the bias of each of the metrics in Table 7.

Class	PRC	NPV	ACC	F1	GM	MCCn	MKn
1	2e-6	0.0026	0.0012	1e-5	-	0.0006	0.0013
2	-0.081	0.0598	0.0136	-0.04	-	-0.005	-0.011
3	-0.058	0.2133	0.1293	-0.019	-	0.0348	0.0775
4	-0.022	0.0021	-0.003	-0.011	-	-0.005	-0.01
5	-0.095	0.0029	-0.006	-0.051	-	-0.024	-0.046
6	-0.015	0.0192	0.0073	-0.007	-	0.001	0.0021
Average	-0.045	0.05	0.0239	-0.021	-	0.0004	0.0024

Table 7 Bias. Python, own

7. Conclusions

This paper presents the analysis of a set of classification performance evaluation metrics, under imbalanced data sets.

In this paper, we used SVM as a classifier in different experiments, where 90% of the samples for testing and 10% of the samples for training were considered. As an output from the classifier, we obtained a multiclass confusion matrix observed in Figure 3. Analyzing the results obtained, we can see that for class 2 and 3 there was a higher proportion of errors than in other classes when performing the classification. Since class 2 is one of the classes with highest number of samples, some of the metrics would penalize the result more.

Classes 1,4,5,6 were very high classification (over 0.97) therefore, so a metric such as AA that averages the accuracies for each class presents very positive results. While OA is mostly reflected in the classification errors of classes 2 and 3. On the other hand, the BA and BAW metrics are less optimistic than the OA due to the fact that it mainly considers class 2 errors due to the imbalance. Besides, MCC and K are highly correlated and their scores are higher than OA, BA and BAW. These metrics usually generate high scores only if the classifier was able to correctly predict most of the data (Chicco, 2020).

At the same time, we propose to perform a performance evaluation by class, and from the multiclass confusion matrix, we obtained multiple binary confusion matrices with equations 9a-9d. The metrics based on binary confusion matrices from Table 2 are used. The results are shown in Table 6.

Classes 2 and 3, present a low proportion of TP, hence, the classification scores for classes are penalized by metrics such as PRC and F1, while other metrics such as NPV and ACC do not depend only on TP, but also on TN, which produces higher scores for our case study.

In Table 7, we have the bias of each metric by class, where we can see that the F1 metric presents less bias on average than PRC, NPV and ACC, the F1 metric is a good metric when we perform an evaluation by class. Since GM does not depend on the imbalance, i.e., an unbiased metric, therefore it is a good point of comparison for other metrics.

The MCCn metric turns out to be the least biased and the most reliable for evaluating imbalanced data, (Luque, 2019). In the results, MKn shows to be one of the least biased metrics after MCCn, as a mentioned in (Luque, 2019).

References

Belgiu, Mariana & Drăguț, Lucian. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 114. 24-31. 10.1016/j.isprsjprs.2016.01.011.

- Boughorbel, S. & Jarray, Fethi & El-Anbari, Mohammed. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLOS ONE. 12. e0177678. 10.1371/journal.pone.0177678.
- Chawla, Nitesh. (2005). Data Mining for Imbalanced Datasets: An Overview. 10.1007/0-387-25465-X_40.
- Cheng, Gong & Han, Junwei. (2016). A Survey on Object Detection in Optical Remote Sensing Images. ISPRS Journal of Photogrammetry and Remote Sensing. 117. 11-28. 10.1016/j.isprsjprs.2016.03.014.
- Chicco, Davide & Jurman, Giuseppe. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 21. 10.1186/s12864-019-6413-7.
- Cortes, Corinna & Vapnik, Vladimir. (2009). Support-vector networks. Chem. Biol. Drug Des.. 297. 273-297. 10.1007/%2FBBF00994018.
- P. (2021, 6 enero). How to Improve Class Imbalance using Class Weights in Machine Learning. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2020/10/improve-class-imbalance-class-weights/>
- Grandini, Margherita & Bagli, Enrico & Visani, Giorgio. (2020). Metrics for Multi-Class Classification: an Overview.
- Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5, 221-232.
- Liu, Peng & Choo, Kim-Kwang Raymond & Wang, Lizhe & Huang, Fang. (2017). SVM or deep learning? A comparative study on remote sensing image classification. Soft Computing. 21. 10.1007/s00500-016-2247-2.
- Long, Yang & Gong, Yiping & Xiao, Zhifeng & Liu, Qing. (2017). Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing. PP. 1-13. 10.1109/TGRS.2016.2645610.
- Luque, A. & Carrasco, Alejandro & Martín, Alejandro & de Las Heras, Ana. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognition. 91. 6829. 10.1016/j.patcog.2019.02.023.
- Melgani, Farid & Bruzzone, Lorenzo. (2004). Classification of Hyperspectral Remote Sensing Images with Support Vector Machines. Geoscience and Remote Sensing, IEEE Transactions on. 42. 1778 - 1790. 10.1109/TGRS.2004.831865.
- Wang, Weixing & Yang, Nan & Zhang, Yi & Wang, Fengping & Cao, Ting & Eklund, Patrik. (2016). A Review of Road Extraction from Remote Sensing Images. Journal of Traffic and Transportation Engineering (English Edition). 3. 10.1016/j.jtte.2016.05.005.