Enhancing Confusion Entropy (CEN) for binary and multiclass classification

Rosario Delgado[1], J. David Núñez-González [2]*

1 Department of Mathematics, Universitat Autònoma de Barcelona, Campus de la UAB, Cerdanyola del Vallès, Spain, **2** Department of Mathematics, University of the Basque Country (UPV/EHU), Leioa, Spain

* josedavid.nunez@ehu.eus

Abstract

Different performance measures are used to assess the behaviour, and to carry out the comparison, of classifiers in Machine Learning. Many measures have been defined on the literature, and among them, a measure inspired by Shannon's entropy named the Confusion Entropy (CEN). In this work we introduce a new measure, MCEN, by modifying CEN to avoid its unwanted behaviour in the binary case, that disables it as a suitable performance measure in classification. We compare MCEN with CEN and other performance measures, presenting analytical results in some particularly interesting cases, as well as some heuristic computational experimentation.

Introduction

Machine Learning is the subfield of Computer Science, as well as the branch of Artificial Intelligence, whose objective is to develop techniques that allow computers to learn. It has a wide range of applications, such as search engines or pattern recognition. Examples are: medical diagnosis, fraud detection, stock market analysis, classification of DNA sequences, recognition of speech and written language, images, games and robotics.

Machine learning tasks are typically grouped into two broad categories: Supervised and Unsupervised Learning. Classification falls in the former, since it deals with some input variables (features or characteristics) and an output variable (the class), and uses an algorithm to infer the class of (that is, to classify) a new case from its known features. Different models are used to build classifiers. Decision Trees (J48, Random Forest), Rules (Decision Table, JRip, ZeroR), Neural Networks (Multilayer Perceptron, Extreme Learning Machines, RBFN), Support Vector Machines, and Bayesian Networks (Naive Bayes, TAN) are some, although not the only ones, approximations to supervised classification.

Once a classifier is built, a performance measure is needed in order to assess its behaviour and to compare it with other classifiers. In the binary case, in which the class variable has only two labels or classes, there are several classical measures that have been widely used: Accuracy, Sensitivity, Specificity and F-score, only to mention some of the most commonly used. Not of all them allow a natural extension to the multi-class case (more than two labels), and only few measures have been specially designed for multi-class classification, which is a more complex scenario. Accuracy, by far the simplest and widespread performance measure

in classification, extends seamlessly its definition in the binary case to multi-class classification. Another well known performance measure, formerly introduced in the binary case but that extends without problems, is Matthew's Correlation Coefficient (MCC), introduced by Matthews in [1].

In this work, whose seed is [2], we focus on a different performance measure, named Confusion Entropy (CEN), which measures the uncertainty generated by classification, and has been recently introduced by Wang et al. in [3] as a novel measure for evaluating classifiers based on the concept of Shannon's entropy. CEN measures generated entropy from misclassified cases considering not only how the cases of each fixed class have been misclassified into other classes, but also how the cases of the other classes have been misclassified as belonging to this class, as well as entropy inside well-classified cases. Given a set of non-negative numbers, say $\{n_1, \ldots, n_r\}$, the Shannon's entropy generated by the set can be defined as the sum $\sum_{i=1}^{r} -p_i \log(p_i)$, with $p_i = \frac{n_i}{n}$ if $n = \sum_{i=1}^{r} n_i$, where log can be, as usual, the logarithm in base 2.

CEN is compared in [3] with Accuracy and other measures, showing a relative consistency with them: higher Accuracy tends to result in lower Confusion Entropy. This performance measure, which is more discriminating for evaluating classifiers than Accuracy, specially when the number or cases grows, has also been studied in [4], where the authors show the strong monotone relation between CEN and MCC, and that both, MCC and CEN, improve over Accuracy.

There are some works in the recent literature using Confusion Entropy. For example, in [5] the authors propose a novel splitting criterion based on CEN for learning decision trees with higher performance; experimental results on some data sets show that this criterion leads to trees with better CEN value without reducing accuracy. The authors of [6] and [7] use CEN, among other performance measures, to compare several common data mining methods used with highly imbalanced data sets where the class of interest is rare. Other works propose modifications of this measure, as [8], in which a Confusion Entropy measure based on a probabilistic confusion matrix is introduced, measuring if cases are classified into true classes and separated from others with high probabilities. A similar approach to that of [8] is followed in [9] to analyze the probability sensitivity of the Gaussian processes in a bankruptcy prediction context, by means of a probabilistic confusion entropy matrix based on the model estimated probabilities. In the context of horizontal collaboration, the system global entropy is introduced in [10] analogously to CEN (see also [11] and [12]), and it is used in the collaborative part of a clustering algorithm, which is iterative with the optimization process continuing as long as the system global entropy is not stable.

It is remarkable that CEN shows to have a weakness in the binary case that invalidates it as a suitable performance measure: in some situations CEN gets values larger than one, unlike what happens in the multi-class case, in which CEN ranges between zero and one. CEN is a measure of the "overall" entropy associated to the confusion matrix, that can be thought as generated by two sources: entropy within the main diagonal, and the one generated by the values outside it, corresponding to misclassification. We will show that CEN is more sensible to the later. A second but not least important point in the weakness of the behaviour of CEN is its lack of monotonicity when the overall entropy does increase (or decrease) monotonously. Along the paper we will show different situations to stand out these items.

Our aim is to introduce an enhanced CEN measure, that we denote by MCEN, and compare it with CEN, MCC and Accuracy. This new measure will show to be highly correlated with CEN. Two aspects deserve to be highlighted:

1. definitions of probabilities involved in the construction of CEN have been modified in MCEN to improve interpretability as real probabilities,

2. weakness of CEN in the binary case (out-of-range and lack of monotonicity) are overcome with MCEN.

The paper is structured as follows: first we introduce the Modified Confusion Entropy MCEN and deal with the multi-dimensional perfectly symmetric and balanced case, which is deeply studied, performing a cross comparison between CEN, MCEN, Accuracy and MCC. The general binary case is treated next, focusing on different families of matrices and carrying out the corresponding cross comparisons. Next part is devoted to study the Z_A family of confusion matrices. Then, we compare CEN, MCEN, Accuracy and MCC with two recently introduced measures: the Probabilistic Acuracy PACC ([13]) and the Entropy-Modulated Accuracy EMA ([14]). Finally, some experiments performed in the binary setting to compare CEN with MCEN through four real database sets are included in the Supporting Information file. These experiments show that their behaviour is mostly analog, but when it is not the case, MCEN is the one that behaves more according to entropy generated by misclassification. The paper finishes with a conclusion section.

Methods

Given a multi-class classifier learned from a training dataset, with $N \geq 2$ classes labelled {1, 2, . . ., N}, we apply it in order to classify cases from a testing dataset, that is, to infer the class of the cases from their known features or characteristics. Since for the cases in the testing dataset we actually know the class to which they belong, we can construct the $N \times N$ confusion matrix $C = (C_{i,j})_{i,j=1, \ldots, N}$, which collects the results issued by the classifier over the testing dataset. $C_{i,j}$ is the number of cases of class i that have been classified as belonging to class j. We denote by S the sum of values of the matrix, that is, the total number of cases in the testing dataset, $S = \sum_{i=1}^{N} \sum_{j=1}^{N} C_{i,j}$.

We introduce notations OUT(C) and IN(C), respectively, to denote the Shannon's entropy generated by the elements of outside (respectively, inside) the main diagonal of matrix C. That is, while IN is the entropy generated by the well classified cases, OUT is generated by misclassification.

In [3] the misclassification probability of classifying class-i cases as being of class j "subject to class j", denoted by $P_{i,j}^{j}$, is introduced as:

$$P_{i,j}^{j} = \frac{C_{i,j}}{\sum_{k=1}^{N}(C_{j,k} + C_{k,j})}, \quad i,j = 1,...,N, \ i \neq j, \tag{1}$$

that is, $P_{i,j}^{j}$ is "almost" the relative frequency class-i cases that are classified as being of class j among all cases that are of class j or that have been classified as being of class j. But not exactly. The reason is that class-j cases that have been correctly classified, whose number is $C_{j,j}$, are counted twice in the denominator.

Analogously, the misclassification probability of classifying class-i cases as being of class-j "subject to class i", with analogous interpretation, denoted by $P_{i,j}^{i}$, is defined in the same paper by:

$$P_{i,j}^{i} = \frac{C_{i,j}}{\sum_{k=1}^{N}(C_{i,k} + C_{k,i})}, \quad i,j = 1,...,N, \ i \neq j. \tag{2}$$

Then, the Confusion Entropy associated to class j is defined in [3] by:

$$\mathrm{CEN}_j = -\sum_{k=1, k\neq j}^{N} \left(P_{j,k}^j \log_{2(N-1)}(P_{j,k}^j) + P_{k,j}^j \log_{2(N-1)}(P_{k,j}^j) \right) \tag{3}$$

with the convention $a \log_b(a) = 0$ if $a = 0$. Finally, the overall Confusion Entropy associated to the confusion matrix C is defined as a convex combination of the Confusion Entropy of the classes as follows:

$$\mathrm{CEN} = \sum_{j=1}^{N} P_j \, \mathrm{CEN}_j, \tag{4}$$

where the non-negative weights P_j, summing 1, are

$$P_j = \frac{\sum_{k=1}^{N}(C_{j,k} + C_{k,j})}{2\sum_{k,\ell=1}^{N} C_{k,\ell}}. \tag{5}$$

Note that CEN is an invariant measure; if we multiply all elements of the confusion matrix by a constant we obtain the same result. The same convenient and useful property holds with Accuracy, MCC and the modified Confusion Entropy measure MCEN, that we will introduce below. As MCC lives in [−1, 1] while Accuracy, CEN and MCEN range in [0, 1], we scale MCC and introduce $\mathrm{MCC}^* = \frac{1-\mathrm{MCC}}{2} \in [0, 1]$. Besides, since Accuracy usually has an inverse relationship with both CEN and MCEN, we choose to consider $\mathrm{ACC}^* = 1 - \mathrm{Accuracy}$ instead of Accuracy itself.

For $N > 2$, CEN ranges between 0 and 1, 0 is attained with perfect classification (the off-diagonal elements of matrix C being zero), while 1 under complete misclassification, symmetry and balance in C, that is, if all diagonal elements in C are zero, and the off-diagonal elements take all the same value. In the binary case ($N = 2$), although CEN remains to be 0 with perfect classification, and is 1 under complete misclassification with symmetry, in intermediate scenarios we can also obtain CEN = 1 and even higher values. That is, in some cases CEN is out-of-range. See, for example, the confusion matrices in Table 1, which have already been considered in [4]. The lack of monotonicity when the situation monotonously goes from perfect classification to completely symmetric and balanced misclassification, as showed by the sequence of matrices in Table 1, represents a great inconvenience of CEN in the binary case, and is our main motivation for introducing a modified version of it.

Definition

Instead of (1), we propose to introduce the probability of classifying class-i cases in class j "subject to class j", as

$$\tilde{P}_{i,j}^j = \frac{C_{i,j}}{\sum_{k=1}^{N}(C_{j,k} + C_{k,j}) - C_{j,j}}, \quad i,j = 1, ..., N, \ i \neq j.$$

Table 1. Examples in the perfectly symmetric and balanced binary case with $S = 12$. Only CEN values.

	$\begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}$	$\begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}$	$\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$	$\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$	$\begin{pmatrix} 2 & 4 \\ 4 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 6 \\ 6 & 0 \end{pmatrix}$
CEN =	0.0000	0.5975	0.8617	1.0000	1.0566	1.0525	1.0000

that is, we overcome the fact that in (1) correctly classified class-j cases are counted twice in the denominator. With this definition, $\tilde{P}^j_{i,j}$ is really the relative frequency of class-i cases classified as belonging to class j among all cases that are of class j or that have been classified as being of class j. Analogously, we modify definition (2) in the same sense:

$$\tilde{P}^i_{i,j} = \frac{C_{i,j}}{\sum_{k=1}^{N}(C_{i,k} + C_{k,i}) - C_{i,i}} \, , \, , i,j = 1,...,N, \, i \neq j,$$

and $\tilde{P}^i_{i,j}$ is really the relative frequency of class-i cases classified in class j among all cases that are of class i or that have been classified as being of class i.

Next, we modify definition of the weights in (5) in the following way:

$$\tilde{P}_j = \frac{\sum_{k=1}^{N}(C_{j,k} + C_{k,j}) - C_{j,j}}{2\sum_{k,\ell=1}^{N} C_{k,\ell} - \alpha \sum_{k=1}^{N} C_{k,k}} \, ,$$

where

$$\alpha = \begin{cases} 1/2 & \text{if } N = 2 \\ 1 & \text{if } N > 2 \, . \end{cases}$$

Then, we define the Confusion Entropy associated to class j as in (3) by

$$\text{MCEN}_j = -\sum_{k=1,k\neq j}^{N} (\tilde{P}^j_{j,k} \log_{2(N-1)}(\tilde{P}^j_{j,k}) + \tilde{P}^j_{k,j} \log_{2(N-1)}(\tilde{P}^j_{k,j})) \, ,$$

and the modified Confusion Entropy as in formula (4), that is,

$$\text{MCEN} = \sum_{j=1}^{N} \tilde{P}_j \, \text{MCEN}_j \, . \tag{6}$$

Note that when $N > 2$, $\sum_{j=1}^{N} \tilde{P}_j = 1$, so the modified overall Confusion Entropy is also defined as a convex combination of the modified Confusion Entropy corresponding to the classes, while in the binary case ($N = 2$), it is just defined as a conical combination since although the weights $\tilde{P}_j$ are non-negative, they do not necessarily sum up to 1 (indeed, their sum is 1 if and only if all the diagonal elements of the confusion matrix C are zero, that is, if all cases have been misclassified).

We see from (4) and (6) that both measures CEN and MCEN, are decomposable along classes, which makes it easy to assess the effect on the behaviour of the classifier of a simple modification affecting just one class.

We can start performing a preliminary comparison of the behaviour of ACC*, MCC*, CEN and MCEN in the toy example in dimension 2 of Table 2. In this example, the baseline confusion matrix is constant with all its entries equal to 3. First, maintaining the total sum equal to $S = 12$ and the out-diagonal invariant, we reduce the entropy IN in Table 2(a). In the baseline case, the diagonal elements are the set {3, 3}, whose entropy is 1 (maximum value). The corresponding values of IN in case (a) are consigned in Table 2, in a decreasing order. Analogously for Table 2(b) but in this case changes have been introduced outside the main diagonal. We observe that while ACC* remains insensitive to changes in the arrangement of the elements of the matrix, since the sum of the main diagonal remains constant, MCC* only decreases with decreasing entropy OUT, while when IN decreases, its value increases. As far as their interpretation is concerned, both CEN and MCEN measure the overall entropy of the confusion

Table 2. Toy example: Binary case with $S = 12$. (a): Entropy reduction within the main diagonal, IN. (b) Entropy reduction outside the main diagonal, OUT. In brackets the relative reduction in each measure with respect to the baseline case. Entropy refers to IN in (a) and to OUT in (b).

	Baseline	(a)			(b)		
	$\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$	$\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$	$\begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}$	$\begin{pmatrix} 0 & 3 \\ 3 & 6 \end{pmatrix}$	$\begin{pmatrix} 3 & 2 \\ 4 & 3 \end{pmatrix}$	$\begin{pmatrix} 3 & 1 \\ 5 & 3 \end{pmatrix}$	$\begin{pmatrix} 3 & 0 \\ 6 & 3 \end{pmatrix}$
Entropy =	1.0000	0.9183	0.6500	0.0000	0.9183	0.6500	0.0000
		(8.17%)	(35.00%)	(100.00%)	(8.17%)	(35.00%)	(100.00%)
ACC* =	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
MCC* =	0.5000	0.5130	0.5625	0.6667	0.4881	0.4375	0.3333
CEN =	1.0000	0.9898	0.9575	0.8962	0.9591	0.8250	0.5000
		(1.02%)	(4.25%)	(10.38%)	(4.09%)	(17.50%)	(50.00%)
MCEN =	0.9057	0.9006	0.8848	0.8571	0.8590	0.7057	0.3343
		(0.56%)	(2.31%)	(5.37%)	(5.16%)	(22.08%)	(63.09%)

matrix, giving less weight to the IN entropy, that is, that generated by the well classified cases, than to OUT entropy, corresponding to misclassification. In this example we observe how their values are reduced when IN decreases, maintaining its constant sum, or when the one that is reduced is OUT, but in this second case the reduction is much more drastic, both for CEN and MCEN, and more sharply for the second. The main difference between CEN and MCEN in this sense is that the former is more sensitive to changes of IN entropy than MCEN, while less than CEN to that of OUT (observe the percentages in brackets in Table 2, which are the relative reduction in the measure with respect to that of the baseline case).

We can extend this comparison to matrices of type $M_A = \begin{pmatrix} 1 & 50 \\ A & 1 \end{pmatrix}$, with $A = 1, \ldots, 100$,

for example. Their main diagonal stays constant. Fig 1 shows the behaviour of CEN, MCEN, ACC* and MCC* as OUT increases. We can observe that indeed, CEN is less correlated with this entropy than MCEN. The same can be observed from the correlations matrix given in Table 3.

Instead, if we consider matrices $W_A = \begin{pmatrix} 50 & 1 \\ 1 & A \end{pmatrix}$, with $A = 1, \ldots, 100$, the values outside

the main diagonal stay constant. Fig 2 shows the behaviour of CEN, MCEN, ACC* and MCC* as IN increases. CEN shows more correlation with this entropy than MCEN (see Table 4), although IN is less correlated (and in an inverse sense that could not be appreciated in the toy example of Table 2) than OUT, both with CEN and MCEN.

The perfectly symmetric and balanced case

In this section we consider the case in which $C_{i,j} = F$ for all $i, j = 1, \ldots, N$, $i \neq j$ and $C_{i,i} = T$,

with $T \geq 0$, $F > 0$, that is, $C = \begin{pmatrix} T & F & \ldots & F & F \\ F & T & \ldots & F & F \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ F & F & \ldots & T & F \\ F & F & \ldots & F & T \end{pmatrix}$.
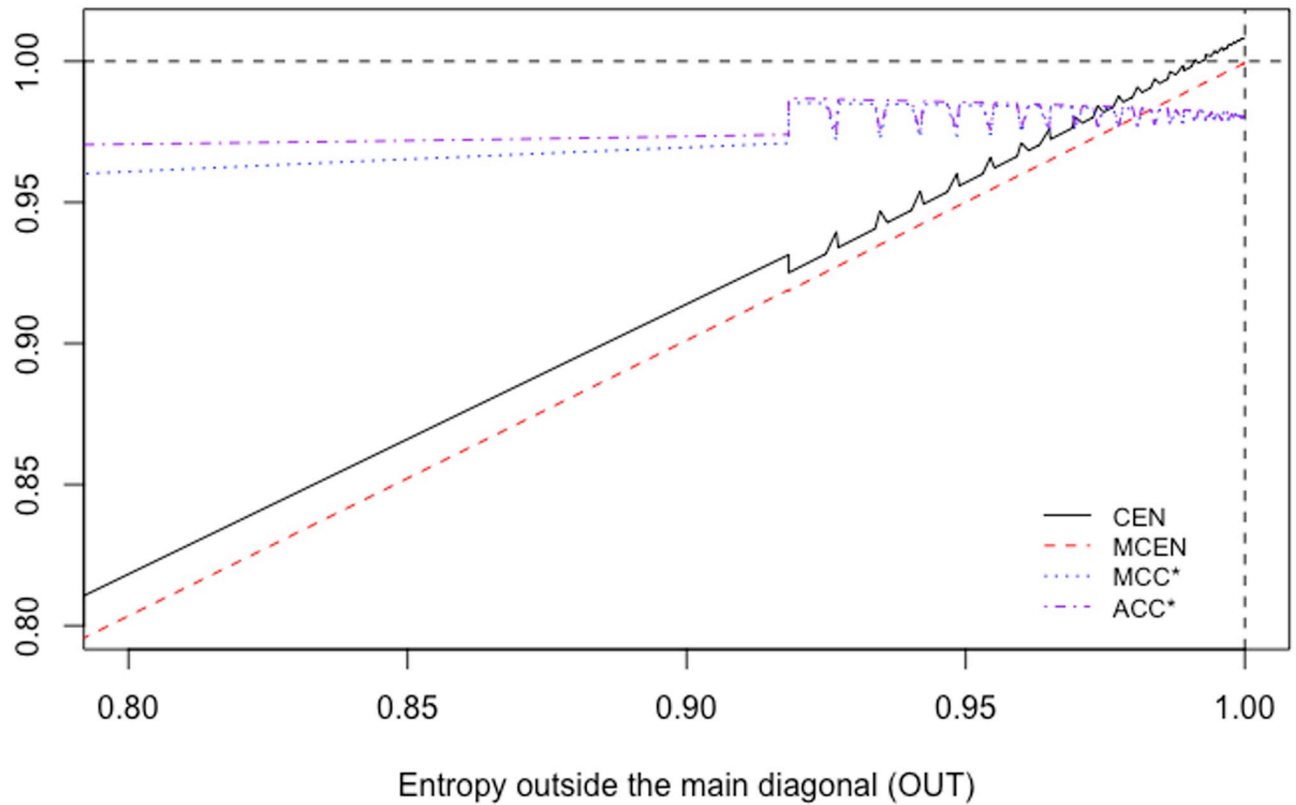
Fig 1. CEN, MCEN, ACC* and MCC* for matrix M_A, as function of entropy outside the diagonal.

Proposition 1 *In the perfectly symmetric and balanced case,*

$$\text{If } N > 2, \quad \text{CEN} = \frac{2\,(N-1)}{\delta}\,\log_{2(N-1)}(\delta), \quad \text{MCEN} = 2\,(N-1)\tilde{\delta}\,\log_{2(N-1)}(\tilde{\delta}),$$

$$\text{If } N = 2, \quad \text{CEN} = \frac{1}{1+\gamma}\,\log_2(\delta), \quad \text{MCEN} = \frac{1}{1+\frac{3}{4}\gamma}\,\log_2(\tilde{\delta}),$$

(7)

where

$$\gamma = \frac{T}{F} \geq 0, \quad \delta = 2\,(N-1) + 2\,\gamma > 0 \quad and \quad \tilde{\delta} = 2\,(N-1) + \gamma > 0,$$

$$\text{ACC}^* = \frac{N-1}{\gamma + (N-1)} \quad and \quad \text{MCC}^* = \frac{N}{2\,(\gamma + (N-1))} = \frac{N}{2\,(N-1)}\,\text{ACC}^*.$$

Table 3. Correlation matrix (Pearson) for the measures of the family of matrices M_A, $A = 1, \ldots, 100$.

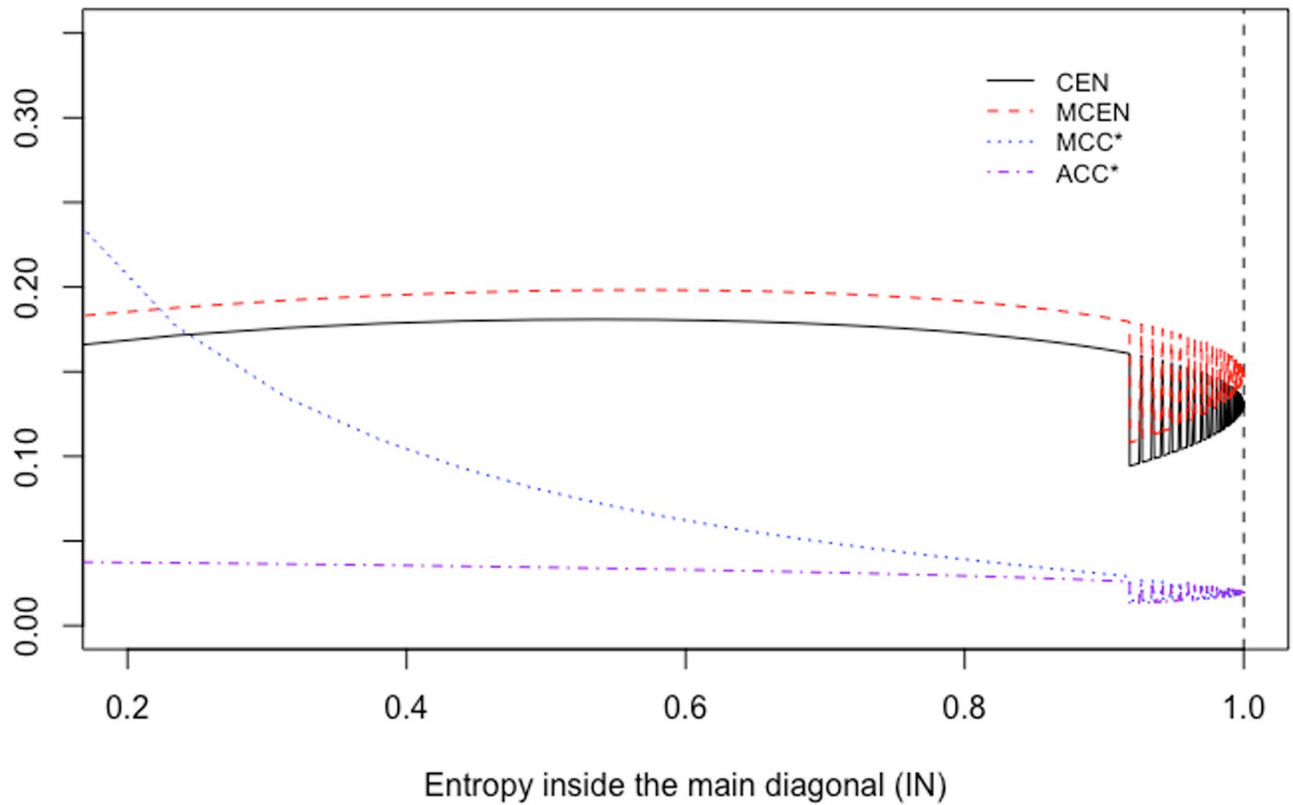|	CEN	MCEN	MCC*	ACC*	OUT
CEN	1.0000000	0.9999334	0.9229026	0.7783573	0.9999320
MCEN		1.0000000	0.9233945	0.7855300	0.9999963
MCC*			1.0000000	0.7340543	0.9241870
ACC*				1.0000000	0.7852756
OUT					1.0000000

Fig 2. CEN, MCEN, ACC* and MCC* for matrix W_A, as function of entropy inside the diagonal.

Note that ACC*, MCC*, CEN and MCEN depend on the matrix values T and F only through its ratio γ. In (7) (case $N > 2$), CEN and MCEN have the same expression except that CEN depends on δ, which is function of 2γ, while MCEN does on $\tilde{\delta} = \delta - \gamma$, which is the same function but of γ. Therefore,

$$\text{if } N > 2, \quad \text{MCEN}(2\,\gamma) = \text{CEN}(\gamma)\,,$$

where in the notation we highlight the dependency of CEN and MCEN on γ.

Corollary 1 *In the perfectly symmetric and balanced case, we have that:*

Table 4. Correlation matrix (Pearson) for the measures of the family of matrices W_A, $A = 1, \ldots, 100$.

	CEN	MCEN	MCC*	ACC*	IN
CEN	1.0000000	0.9995962	0.5499231	0.9672182	-0.6062876
MCEN		1.0000000	0.5355098	0.9609698	-0.5857654
MCC*			1.0000000	0.7340543	-0.9241870
ACC*				1.0000000	-0.7852756
IN					1.0000000

- *For any $N > 2$, CEN, MCEN, ACC* and MCC* are monotonically decreasing functions of $\gamma \geq 0$, with*

$$\lim_{\gamma \to +\infty} \text{CEN}(\gamma) = \lim_{\gamma \to +\infty} \text{MCEN}(\gamma) = \lim_{\gamma \to +\infty} \text{ACC}^*(\gamma) = \lim_{\gamma \to +\infty} \text{MCC}^*(\gamma) = 0 \,,$$

$$\text{CEN}(0) = \text{MCEN}(0) = \text{ACC}^*(0) = 1, \ \text{MCC}^*(0) = \frac{N}{2(N-1)} \to \frac{1}{2} \ as \ N \to +\infty,$$

and if $\gamma > 0$, MCC* < ACC* < CEN < MCEN.

- *Nevertheless, when $N = 2$, we have that although MCEN and ACC* = MCC* remain to be monotonically decreasing as functions of $\gamma \geq 0$, CEN does not. Indeed, CEN achieves its global maximum when $\gamma = \frac{e}{2} - 1$, which is $\text{CEN}\left(\frac{e}{2} - 1\right) \approx 1.06148 > 1$. More specifically,*

$$\text{CEN}(0) \quad = \text{CEN}(1) = 1, \ \text{CEN}(\gamma) > 1, \ for \ all \ 0 < \gamma < 1, \ \lim_{\gamma \to +\infty} \text{CEN}(\gamma) = 0,$$

$$\text{MCEN}(0) \quad = 1, \quad \lim_{\gamma \to +\infty} \text{MCEN}(\gamma) = 0,$$

$$\text{ACC}^*(0) \quad = \text{MCC}^*(0) = 1, \quad \lim_{\gamma \to +\infty} \text{ACC}^*(\gamma) = \lim_{\gamma \to +\infty} \text{MCC}^*(\gamma) = 0 \,.$$

Moreover, there exists $\gamma_0 \approx 5.78$ such that

$$\text{MCC}^* \quad = \text{ACC}^* < \text{MCEN} < \text{CEN} \ \ if \ \ 0 < \gamma < \gamma_0,$$

$$\text{MCC}^* \quad = \text{ACC}^* < \text{MCEN} = \text{CEN} \ \ if \ \ \gamma = \gamma_0, \ and$$

$$\text{MCC}^* \quad = \text{ACC}^* < \text{CEN} < \text{MCEN} \ \ if \ \ \gamma > \gamma_0 \,.$$

Proof 1 *The proofs of both Proposition 1 and Corollary 1 are straightforward, and then omitted. However, it is worth mentioning that in order to prove CEN < MCEN in case $N > 2$ we use that function $f(x) = \frac{1}{x} \log_b(x)$ is strictly decreasing for any base $b > 1$ (in our case, $b = 2(N - 1) \geq 4$), and $x > e$. We apply that fact to see that $f(x_0) > f(x_1)$ with $x_0 = 2(N - 1) + \gamma < x_1 = 2(N - 1) + 2\gamma$, since $x_0 \geq 4 > e$.*

The same property of function f allows to prove that both CEN and MCEN are monotonically decreasing as functions of γ, with $x = \delta = 2(N - 1) + 2\gamma$ and $x = \tilde{\delta} = 2(N - 1) + \gamma$, respectively, being both $> e$ for any $\gamma \geq 0$. Note that since for $N = 2$ the expression of CEN as function of δ is as in case $N > 2$, the monotonous decrease fails since $x = \delta = 2 + 2\gamma < e$ for $\gamma < \frac{e}{2} - 1$.

The rest of proofs are also omitted.

Remark 1 *Note that if $N = 2$, CEN exhibits the unwanted behaviour, not showed by MCEN, of being out-of-range $[0, 1]$, which despairs for $N > 2$ (see Figs 3 and 4).*

Remark 2 *Consider the particular case in which $T = F$, that is, $\gamma = 1$. In other words, the confusion matrix is constant, say $\begin{pmatrix} 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}$. Then, $\text{ACC}* = \frac{N-1}{N}$ and $\text{MCC}* = \frac{1}{2}$. Moreover, $\delta = 2N$ and $\tilde{\delta} = 2N - 1$.*

If $N > 2$, $\text{CEN} = \left(1 - \frac{1}{N}\right) \log_{2(N-1)}(2N)$ and $\text{MCEN} = \left(1 - \frac{1}{2N-1}\right) \log_{2(N-1)}(2N - 1)$.

If $N = 2$, $\text{CEN} = 1$ and $\text{MCEN} = \frac{4}{7} \log_2(3) < 1$.

As a consequence, we can easily check that if $N > 2$, MCC* < ACC* < CEN < MCEN, with $\lim_{N \to +\infty} \text{ACC}^* = \lim_{N \to +\infty} \text{CEN} = \lim_{N \to +\infty} \text{MCEN} = 1$, while if $N = 2$, MCC* = ACC* < MCEN < CEN.
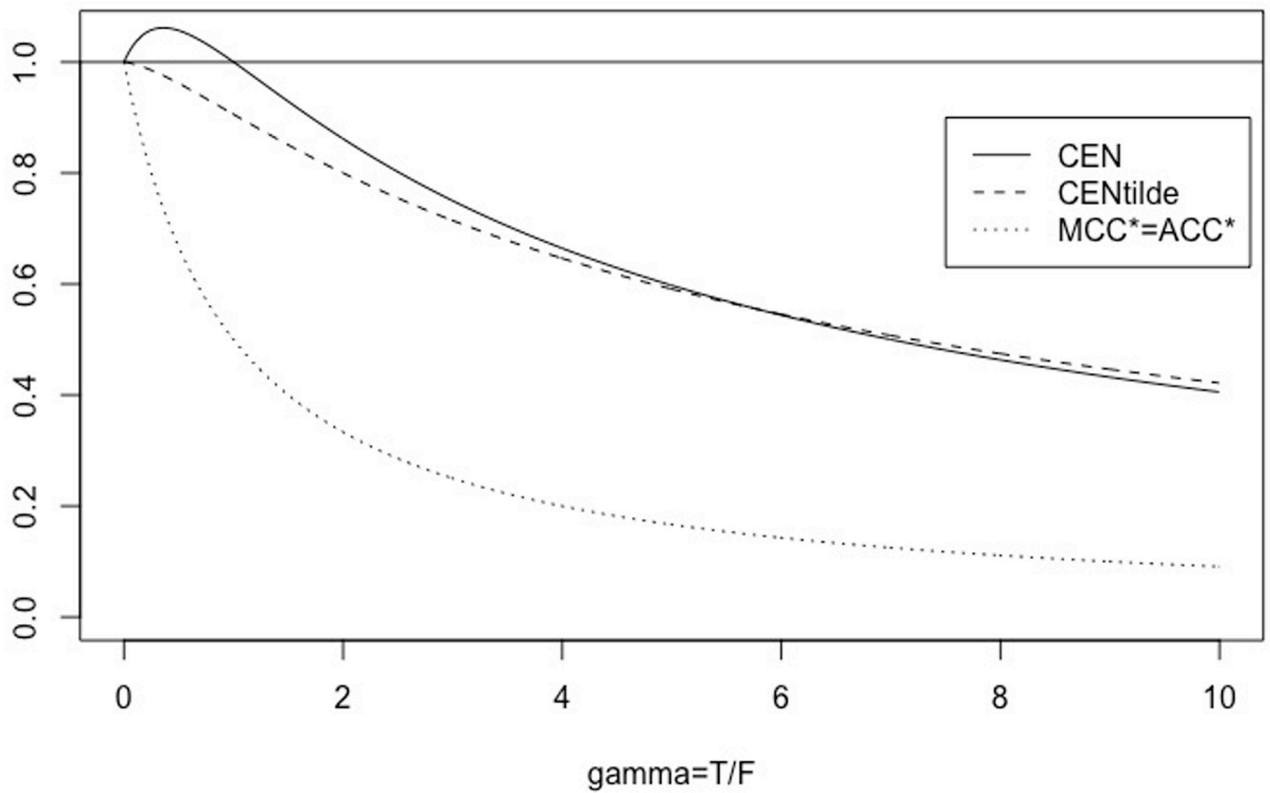
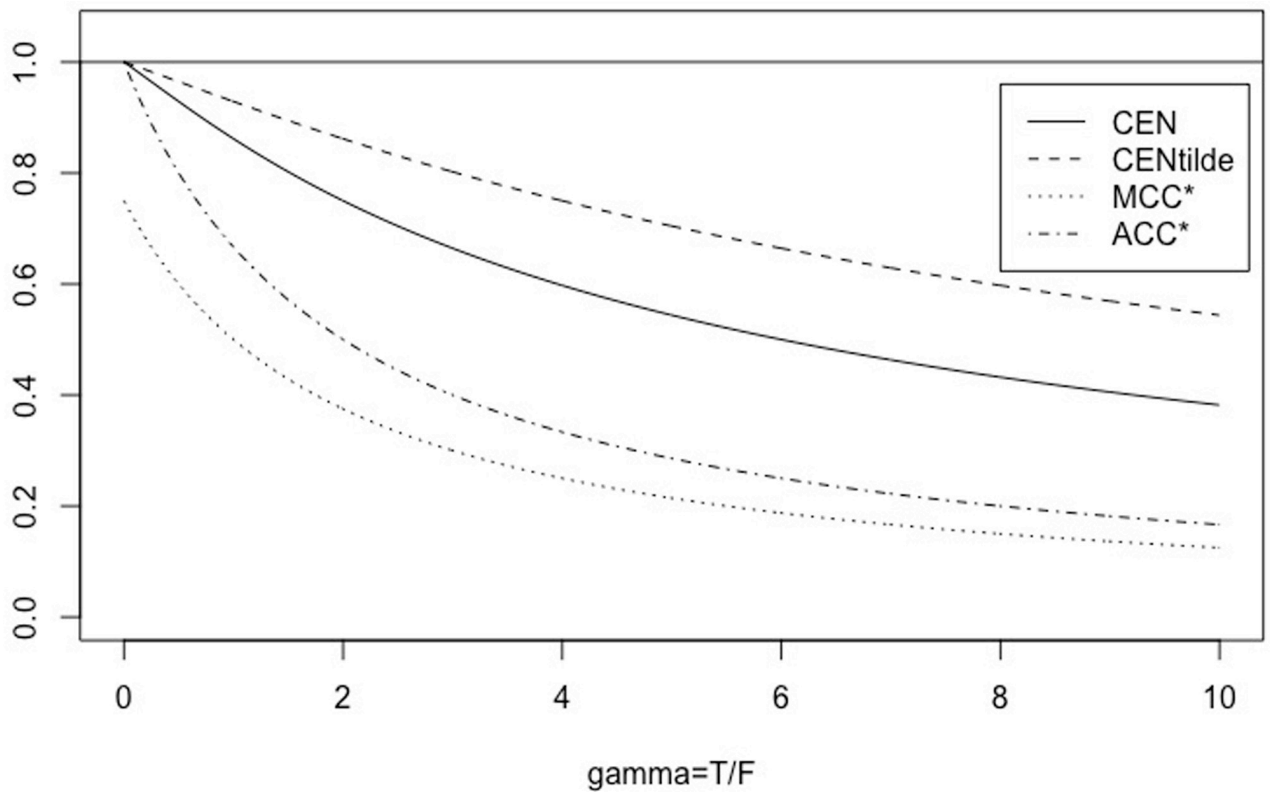**Fig 3. The symmetric case.** CEN, MCEN, ACC* and MCC* for $\gamma \in [0, 10]$, with $N = 2$.

https://doi.org/10.1371/journal.pone.0210264.g003

Fig 4. The symmetric case. CEN, MCEN, ACC* and MCC* for $\gamma \in [0, 10]$, with $N = 3$.

https://doi.org/10.1371/journal.pone.0210264.g004

The particular pathological case of matrices Z_A will be studied in the multi-class setting, but before we consider in some detail the binary case.

The general binary case

The binary case ($N = 2$) can be studied in more detail. We will use the following notation for the confusion matrix in the most general setting, taking class 1 as reference:

$$C = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix}, \qquad (8)$$

where TP is the true positive or number of class-1 cases that have been correctly classified, and the same for the true negative number of cases TN with class 2. On the other hand, FP denotes false positives or number of class-2 cases that have been miscllassified, and FN false negatives.

Proposition 2 *If the confusion matrix C is given by* (8), *we have that with $S = TP + TN + FP + FN$,*

$$
\begin{aligned}
\mathrm{CEN} &= \frac{(FN + FP)\log_2(S^2 - (TP - TN)^2)}{2\,S} - \frac{FN\log_2(FN) + FP\log_2(FP)}{S}, \\
\mathrm{MCEN} &= \frac{2(FN + FP)\log_2((S - TN)(S - TP))}{3S + (FN + FP)} - \frac{4(FN\log_2(FN) + FP\log_2(FP))}{3S + (FN + FP)}, \\
\mathrm{ACC}^* &= \frac{FP + FN}{S} \quad and \quad \mathrm{MCC}^* = \frac{1 - \mathrm{MCC}}{2}, \\
with \quad \mathrm{MCC} &= \frac{TP\,TN - FP\,FN}{\sqrt{(TP + FN)\,(FP + TN)\,(TP + FP)\,(TN + FN)}}.
\end{aligned}
\qquad (9)
$$

To carry out a deeper study, we have to consider particular situations; is what we do in the subsections below, where different particular scenarios have been introduced and developed.

The perfectly symmetric and balanced case. Table 5 below shows some examples of 2×2 confusion matrices of type $\begin{pmatrix} T & F \\ F & T \end{pmatrix}$, that is, in which $TP = TN = T$ and $FP = FN = F$. All of them correspond to $S = 12$ and have already been considered in [4]. This is a particular case of the previously considered setting, and Proposition 1 and Corollary 1 apply here. We can observe again the anomalous behaviour of CEN, in contrast with the other measures.

The symmetric but unbalanced family U_A. Consider the particular case of a confusion matrix of type $U_A = \begin{pmatrix} 1 & A \\ A & 0 \end{pmatrix}$, with $A > 0$. Both class-1 and class-2 cases are mainly misclassified if $A > 1$. Entropy out of the main diagonal is 1 and within the diagonal is 0, regardless of the value of A. When $0 < A < 1$, say for example that $A = 1/B$ with $B > 1$, then matrix U_A is equivalent to $\begin{pmatrix} B & 1 \\ 1 & 0 \end{pmatrix}$, that is, corresponds to an unbalanced scenario in which class 2 is

Table 5. Examples in the perfectly symmetric and balanced binary case with $S = 12$.

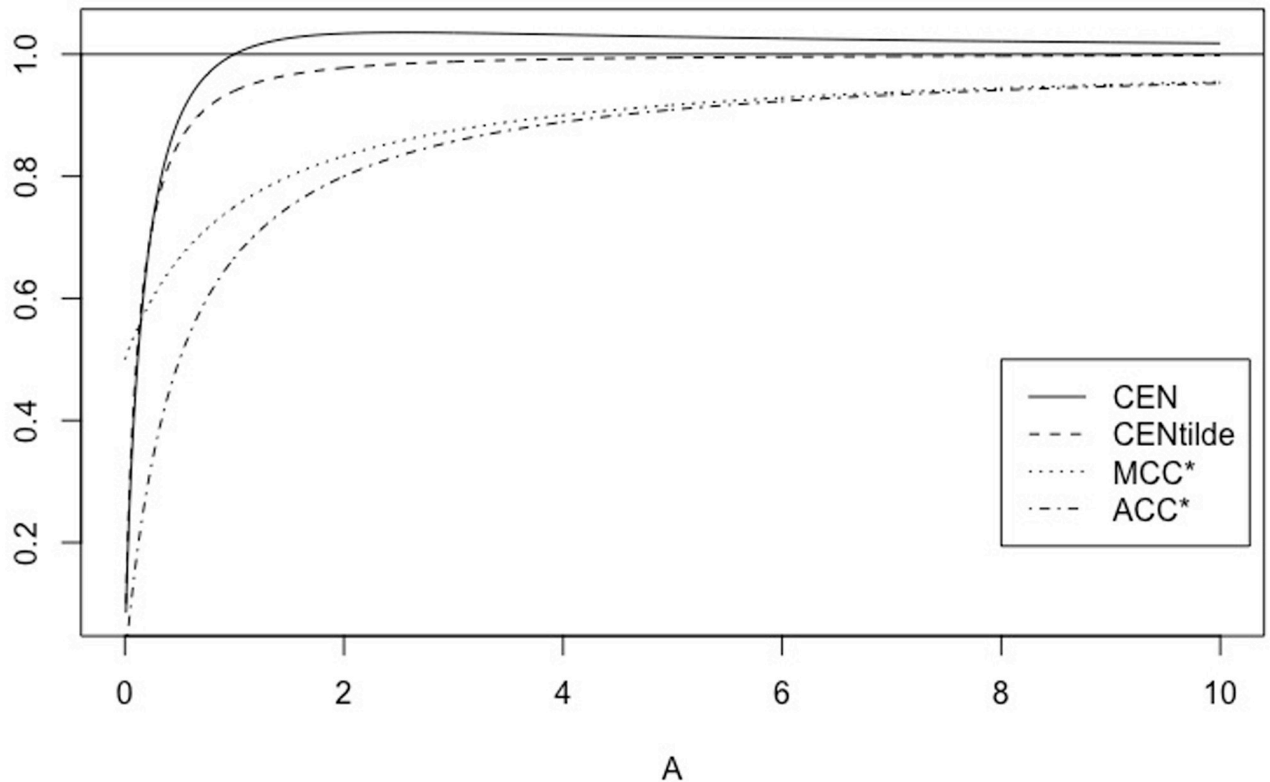|	$\begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}$	$\begin{pmatrix} 5 & 1 \\ 1 & 5 \end{pmatrix}$	$\begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}$	$\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$	$\begin{pmatrix} 2 & 4 \\ 4 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 6 \\ 6 & 0 \end{pmatrix}$
ACC* = MCC* =	0.0000	0.1667	0.3333	0.5000	0.6667	0.8333	1.0000
CEN =	0.0000	0.5975	0.8617	1.0000	1.0566	1.0525	1.0000
MCEN =	0.0000	0.5910	0.8000	0.9057	0.9614	0.9891	1.0000

Fig 5. Famlily U_A. CEN, MCEN, ACC* and MCC* for $A \in (0, 10]$.

underrepresented and class-1 cases are mainly well classified. We can observe some properties of CEN, MCEN, ACC* and MCC* (see Fig 5) in Proposition 3, which is derived from Proposition 2.

Proposition 3 *For confusion matrix U_A with $A > 0$, we have*:

$$\text{CEN}(A) = \frac{A \log_2((2A+1)^2 - 1) - 2A \log_2(A)}{2A+1},$$

$$\text{MCEN}(A) = \frac{4A \log_2(2A(2A+1)) - 8A \log_2(A)}{3(2A+1) + 2A},$$

$$\text{ACC}^*(A) = \frac{2A}{2A+1}, \quad \text{MCC}^*(A) = \frac{2A+1}{2(A+1)}.$$

As a consequence:

CEN(A) < 1 *if* A < 1, CEN(1) = 1, CEN(A) > 1 *if* A > 1, MCEN(A) < 1 *and* ACC*(A) < MCC*(A) < 1, *for all* A > 0, MCEN, ACC* *and* MCC* *are monotonically increasing functions of* A > 0, CEN *is not, and achieves its global maximum when* A ≈ 2.54, *which is* > 1,

$$\lim_{A \to 0} \text{CEN}(A) = \lim_{A \to 0} \text{MCEN}(A) = \lim_{A \to 0} \text{ACC}^*(A) = 0 < \lim_{A \to 0} \text{MCC}^*(A) = 0.5,$$

$$\lim_{A \to +\infty} \text{CEN}(A) = \lim_{A \to +\infty} \text{MCEN}(A) = \lim_{A \to +\infty} \text{ACC}^*(A) = \lim_{A \to +\infty} \text{MCC}^*(A) = 1.$$

Moreover, there exists $A_0 \in (0, 1)$ (indeed, $A_0 \approx 0.24$) such that

$$\text{MCEN}(A) < \text{CEN}(A) \quad \text{if} \quad A > A_0,$$
$$\text{MCEN}(A_0) = \text{CEN}(A_0),$$
$$\text{MCEN}(A) > \text{CEN}(A) \quad \text{if} \quad 0 < A < A_0.$$

The overall entropy associated to the four elements of the confusion matrix, which results to be $-\frac{2A}{2A+1} \log\left(\frac{A}{2A+1}\right)$, increases to 1 when $A \to +\infty$ and decreases to 0 when $A \to 0$, and both CEN and MCEN, are sensible to this fact. Note that the lack of monotonicity of CEN(A) as A (and then, as the overall entropy) monotonically increases, is an anomalous behaviour that MCEN has managed to overcome. Moreover, MCEN ranges between 0 and 1. We can also observe this phenomenon in the examples in Table 6.

The asymmetric family V_A. Consider the particular case of confusion matrices of type $V_A = \begin{pmatrix} 1 & A \\ 1 & 0 \end{pmatrix}$, with $A > 0$. This is an asymmetric and unbalanced case in which class 2 is systematically misclassified and is underrepresented if $A > 1$. Class 1 is also mainly misclassified if $A > 1$. As $A \to +\infty$, entropy out the diagonal, which is $-\frac{A}{A+1} \log\left(\frac{A}{A+1}\right)$, decreases to zero. Entropy within diagonal is zero, while the overall entropy of the elements of matrix V_A is $\log(A+2) - \frac{A}{A+2} \log(A)$, which tends to 0 as $A \to +\infty$. When $0 < A < 1$ with $A = 1/B$, $B > 1$, matrix V_A is equivalent to $\begin{pmatrix} B & 1 \\ B & 0 \end{pmatrix}$, which corresponds to an almost balanced but asymmetric scenario in which class 1 is mainly well classified but class 2 is not. As B increases ($A \to 0$), entropy out the diagonal also drops to zero. Some properties of CEN, MCEN, ACC* and MCC* are given in Proposition 4 (see also Fig 6).

Proposition 4 *For confusion matrix V_A with $A > 0$, we have:*

$$\text{CEN}(A) = \frac{(A+1) \log_2((A+2)^2 - 1) - 2A \log_2(A)}{2(A+2)},$$

$$\text{MCEN}(A) = \frac{2(A+1) \log_2((A+1)(A+2)) - 4A \log_2(A)}{3(A+2) + (A+1)},$$

$$\text{ACC}^*(A) = \frac{A+1}{A+2}, \quad \text{MCC}^*(A) = \frac{1 + \sqrt{\frac{A}{2(A+1)}}}{2}.$$

As a consequence, there exists $A_1 \in (1, 2)$ ($A_1 \approx 1.414$) such that:
$\text{CEN}(A) > 1$ if $1 < A < A_1$, $\text{CEN}(1) = \text{CEN}(A_1) = 1$, $\text{CEN}(A) < 1$ if $A \notin [1, A_1]$, $\text{MCEN}(A) < 1$, $\text{ACC}^*(A) < 1$, $\text{MCC}^*(A) < 1$ and $\text{MCEN}(A) < \text{CEN}(A)$ for all $A > 0$,

Table 6. Examples in the binary case for famlily U_A.

|	$\begin{pmatrix} 10^3 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 10^2 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 10 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 10 \\ 10 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 10^2 \\ 10^2 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 10^3 \\ 10^3 & 0 \end{pmatrix}$
A =	$1/10^3$	$1/10^2$	$1/10$	1	10	10^2	10^3
ACC* =	0.00200	0.01961	0.16667	0.66667	0.952381	0.995025	0.9995002
MCC* =	0.50050	0.50495	0.54545	0.75000	0.954545	0.995050	0.9995005
CEN =	0.01194	0.08488	0.45495	1.00000	1.017859	1.002167	1.0002210
MCEN =	0.01459	0.09964	0.48263	0.93999	0.997778	0.9998483	0.9999856
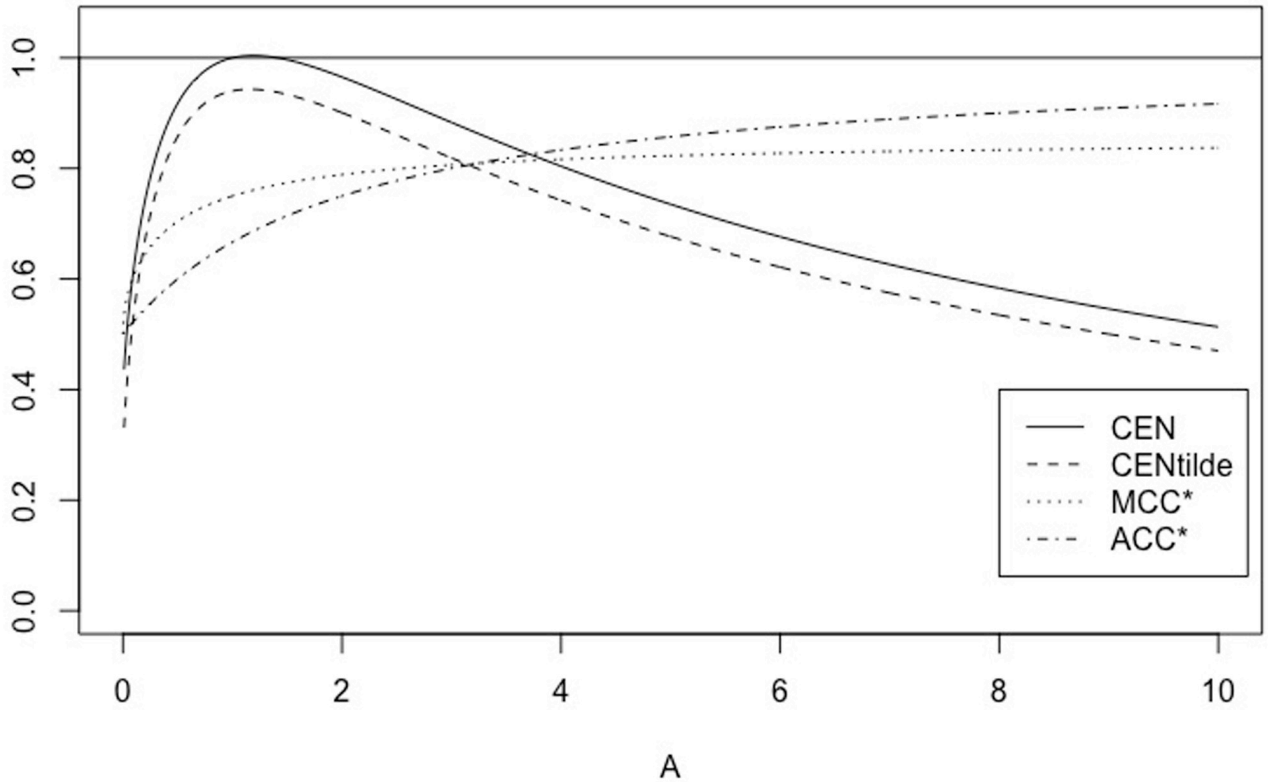
Fig 6. Family V_A. CEN, MCEN, ACC* and MCC* for $A \in (0, 10]$.

$$\lim_{A \to 0} \mathrm{MCC}^*(A) = \lim_{A \to 0} \mathrm{ACC}^*(A) = \frac{1}{2} > \lim_{A \to 0} \mathrm{CEN}(A) = \frac{\log_2(3)}{4} > \lim_{A \to 0} \mathrm{MCEN}(A) = \frac{2}{7},$$

$$\lim_{A \to +\infty} \mathrm{ACC}^*(A) = 1 > \lim_{A \to +\infty} \mathrm{MCC}^*(A) = \frac{2 + \sqrt{2}}{4} > \lim_{A \to +\infty} \mathrm{CEN}(A) = \lim_{A \to +\infty} \mathrm{MCEN}(A) = 0.$$

Note that as in previous cases, CEN(A) does not stay always (that is, for any $A > 0$) restricted to [0, 1], while MCEN does. See Fig 6 and some examples in Table 7.

Apart from the fact that CEN is out-of-range for some values of A, its behaviour is similar to that of MCEN, both decreasing with entropy, while nor ACC* nor MCC* are sensitive to the decrease of entropy when $A \to +\infty$.

Table 7. Examples in the binary case for famlily V_A.

	$\begin{pmatrix} 10^3 & 1 \\ 10^3 & 0 \end{pmatrix}$	$\begin{pmatrix} 10^2 & 1 \\ 10^2 & 0 \end{pmatrix}$	$\begin{pmatrix} 10 & 1 \\ 10 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 5 & 6 \\ 5 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 10 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 10^2 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 10^3 \\ 1 & 0 \end{pmatrix}$
A =	$1/10^3$	$1/10^2$	1/10	1	1.2	10	10^2	10^3
ACC* =	0.5002	0.5025	0.5238	0.6667	0.6875	0.9167	0.9902	0.9990
MCC* =	0.5112	0.5352	0.6066	0.7500	0.7611	0.8371	0.8518	0.8535
CEN =	0.4019	0.4361	0.6217	1.0000	1.0041	0.5133	0.0934	0.0128
MCEN =	0.2921	0.3309	0.5387	0.9400	0.9429	0.4702	0.0866	0.0121

The symmetric but unbalanced family $X_{A,r}$. Now we introduce the family of confusion matrices $X_{A,r} = \begin{pmatrix} A & rA \\ rA & 1 \end{pmatrix}$, with $A, r > 0$. Both class-1 and class-2 cases are mainly misclassified if $A, r > 1$. Overall entropy of $X_{A,r}$ is $-\frac{A}{(2r+1)A+1} \log\left(\frac{A}{(2r+1)A+1}\right) - \frac{2rA}{(2r+1)A+1} \log\left(\frac{rA}{(2r+1)A+1}\right)$, which drops to 0 when $A \to 0$, and when $A \to +\infty$ converges to $\log(2r+1) - \frac{2r}{2r+1} \log(r)$, which in turn converges to 1 as $r \to +\infty$. Fixed $A > 0$, overall entropy converges to 1 as $r \to +\infty$, and as $r \to 0$, it converges to $-\frac{A}{A+1} \log\left(\frac{A}{A+1}\right)$, which in turn converges to 0 both when $A \to 0$ and when $A \to +\infty$.

When $0 < A < 1$, $A = 1/B$ with $B > 1$, matrix $X_{A,r}$ is equivalent to $\begin{pmatrix} 1 & r \\ r & B \end{pmatrix}$. We have some properties of CEN, MCEN, ACC* and MCC* in Proposition 5 below. Moreover, for $r = 0.5$, 5 Figs 7 and 8 show how the measures evolve as function of A, while Figs 9 and 10 show their plots as function of r, fixed $A = 0.5, 10$.

Proposition 5 *For confusion matrix $X_{A,r}$ with $A, r > 0$ we have:*

$$\text{CEN}(A) = -\frac{rA}{(2r+1)A+1} \log_2\left(\frac{r^2 A}{4(r+1)(rA+1)}\right),$$

$$\text{MCEN}(A) = -\frac{4rA}{(8r+3)A+3} \log_2\left(\frac{r^2 A}{(2r+1)(2rA+1)}\right),$$

$$\text{ACC}^*(A) = \frac{2rA}{(2r+1)A+1}, \quad \text{MCC}^*(A) = \frac{2r^2 A + rA + r}{2(r+1)(rA+1)}.$$
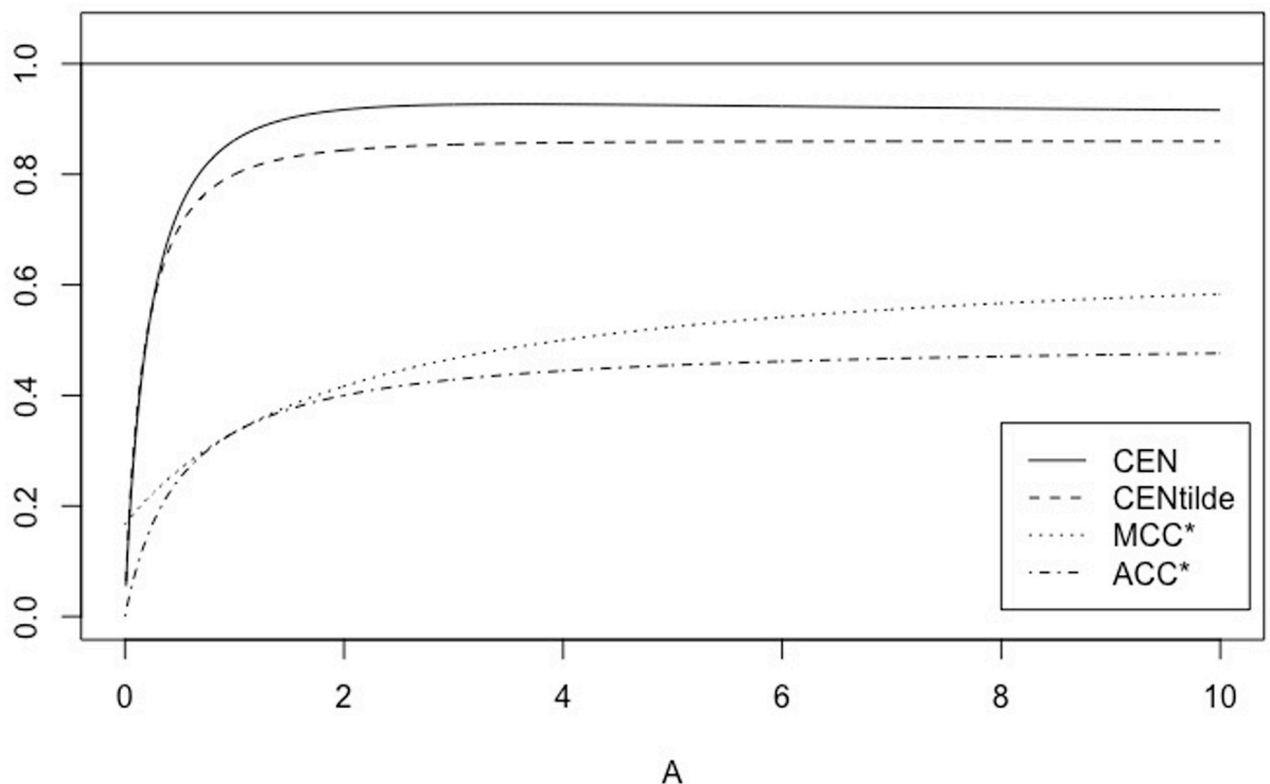

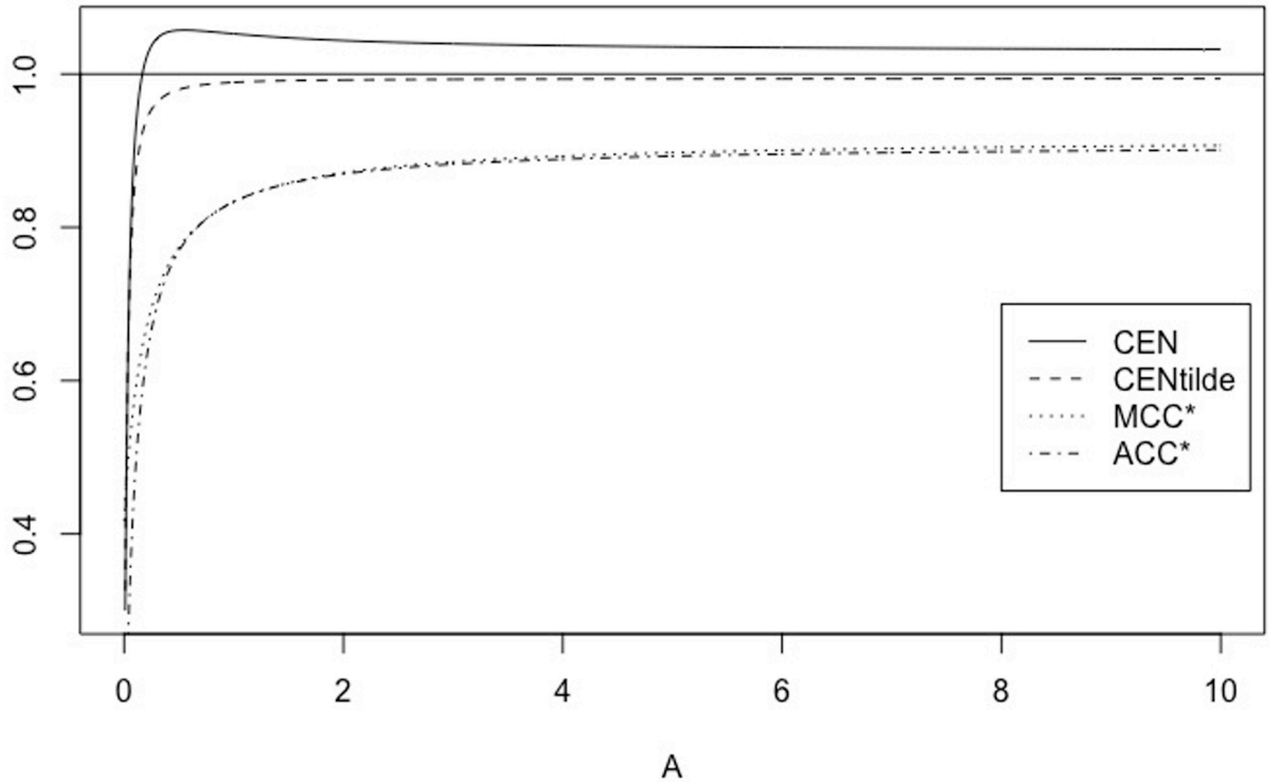**Fig 7. Family $X_{A,r}$.** CEN, MCEN, ACC* and MCC* as function of $A > 0$ for $r = 0.5$.

Fig 8. Family $X_{A,r}$. CEN, MCEN, ACC* and MCC* as function of $A > 0$ for $r = 5$.

As a consequence, $\ell_{\mathrm{CEN}}(r) = \lim_{A \to +\infty} \mathrm{CEN}(A) = \frac{r}{2\,r+1}\,\log_2\left(\frac{4\,(r+1)}{r}\right) > 0$, and there exists $r_0 < 1$ ($r_0 \approx 0.8$) such that for any $r > r_0$, there exists $A_r > 0$ such that $\mathrm{CEN}(A) < 1$ if $A < A_r$, $\mathrm{CEN}(A_r) = 1$, $\mathrm{CEN}(A) > 1$ if $A > A_r$ and

$$\ell_{\mathrm{CEN}}(r) \begin{cases} > 1 & \text{if } r > 1, \\ = 1 & \text{if } r = 1, \\ < 1 & \text{if } r_0 < r < 1. \end{cases}$$

If $r \leq r_0$, $\mathrm{CEN}(A) \leq 1$ for any $A > 0$ and $\ell_{\mathrm{CEN}}(r) < 1$.

On the other hand, for any $r > 0$,

$\mathrm{MCEN}(A) < 1$, $\mathrm{ACC}^*(A) < 1$ and $\mathrm{MCC}^*(A) < 1$, for all $A > 0$, MCEN, ACC * and MCC * are monotonically increasing functions of A, CEN is not, and has a global maximum, which is $>$ 1 if $r > r_0$, $\lim_{A \to 0} \mathrm{CEN}(A) = \lim_{A \to 0} \mathrm{MCEN}(A) = \lim_{A \to 0} \mathrm{ACC}^*(A) = 0$, $\lim_{A \to 0} \mathrm{MCC}^*(A) = \frac{r}{2(r+1)}$,

$$0 < \lim_{A \to +\infty} \mathrm{ACC}^*(A) = \frac{2r}{2r+1} < \lim_{A \to +\infty} \mathrm{MCC}^* = \frac{2r+1}{2(r+1)} = \ell_{\mathrm{MCC}^*}(r) < 1,$$

$$0 < \lim_{A \to +\infty} \mathrm{MCEN}(A) = \frac{4r}{8r+3}\log_2\left(\frac{2(2r+1)}{r}\right) = \ell_{\mathrm{MCEN}}(r) < 1, \lim_{r \to +\infty} \ell_{\mathrm{MCEN}}(r) = 1.$$
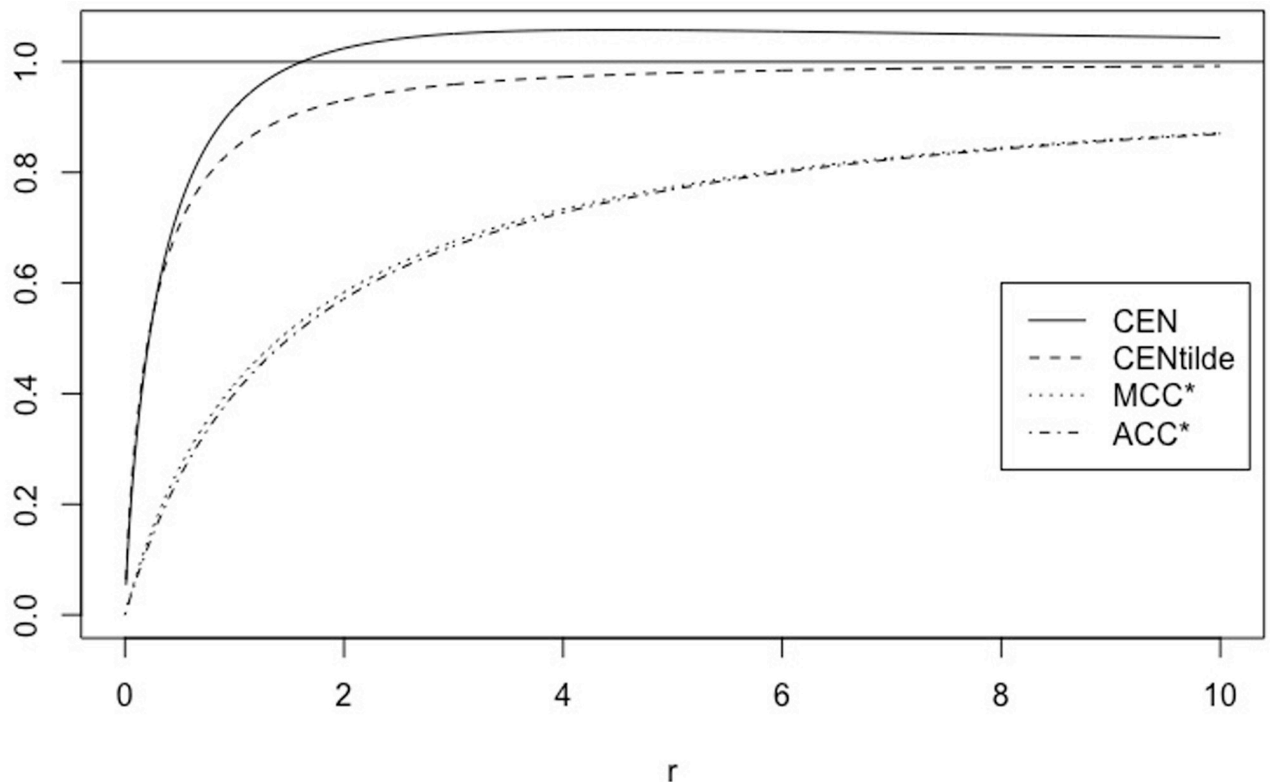
Fig 9. Family $X_{A,\,r}$. CEN, MCEN, ACC* and MCC* as function of $r > 0$ for $A = 0.5$.

Moreover, there exist $0 < r_3 < r_2 < r_1 < r_0 < 1$ ($r_3 \approx 0.13$, $r_2 \approx 0.15$, $r_1 \approx 0.23$) such that:

$$
\begin{cases}
\ell_{\mathrm{MCC}}(r) > \ell_{\mathrm{MCEN}}(r) > \ell_{\mathrm{CEN}}(r) & \text{if } 0 < r < r_3, \\
\ell_{\mathrm{MCC}}(r) = \ell_{\mathrm{MCEN}}(r) > \ell_{\mathrm{CEN}}(r) & \text{if } r = r_3, \\
\ell_{\mathrm{MCEN}}(r) > \ell_{\mathrm{MCC}}(r) > \ell_{\mathrm{CEN}}(r) & \text{if } r_3 < r < r_2, \\
\ell_{\mathrm{MCEN}}(r) > \ell_{\mathrm{MCC}}(r) = \ell_{\mathrm{CEN}}(r) & \text{if } r = r_2, \\
\ell_{\mathrm{MCEN}}(r) > \ell_{\mathrm{CEN}}(r) > \ell_{\mathrm{MCC}}(r) & \text{if } r_2 < r < r_1, \\
\ell_{\mathrm{MCEN}}(r) = \ell_{\mathrm{CEN}}(r) > \ell_{\mathrm{MCC}}(r) & \text{if } r = r_1, \\
\ell_{\mathrm{CEN}}(r) > \ell_{\mathrm{MCEN}}(r) > \ell_{\mathrm{MCC}}(r) & \text{if } r > r_1.
\end{cases}
$$

Finally, for any fixed, $A > 0$, while MCEN, ACC and MCC* are monotonically increasing functions of r, CEN is not, as can be seen in* Figs 9 *and* 10, *for two values of A. Given $A > 0$, there exists $r_A > r_0$ such that CEN$(A) > 1$ for all $r > r_A$.*

Note that although we do not specify it in the notations so as not to complicate them, the performance measures depend on both A and r in the case of this doubly indexed family $X_{A,\,r}$.

The asymmetric family $Y_{A,\,r}$. Finally, we consider another particular doubly indexed family of confusion matrices in the binary case, with the same overall entropy as $X_{A,\,r}$,

denoted by $Y_{A,\,r}$, with $A, r > 0$. We define this family by $Y_{A,\,r} = \begin{pmatrix} r\,A & r\,A \\ A & 1 \end{pmatrix}$. Class-2 is
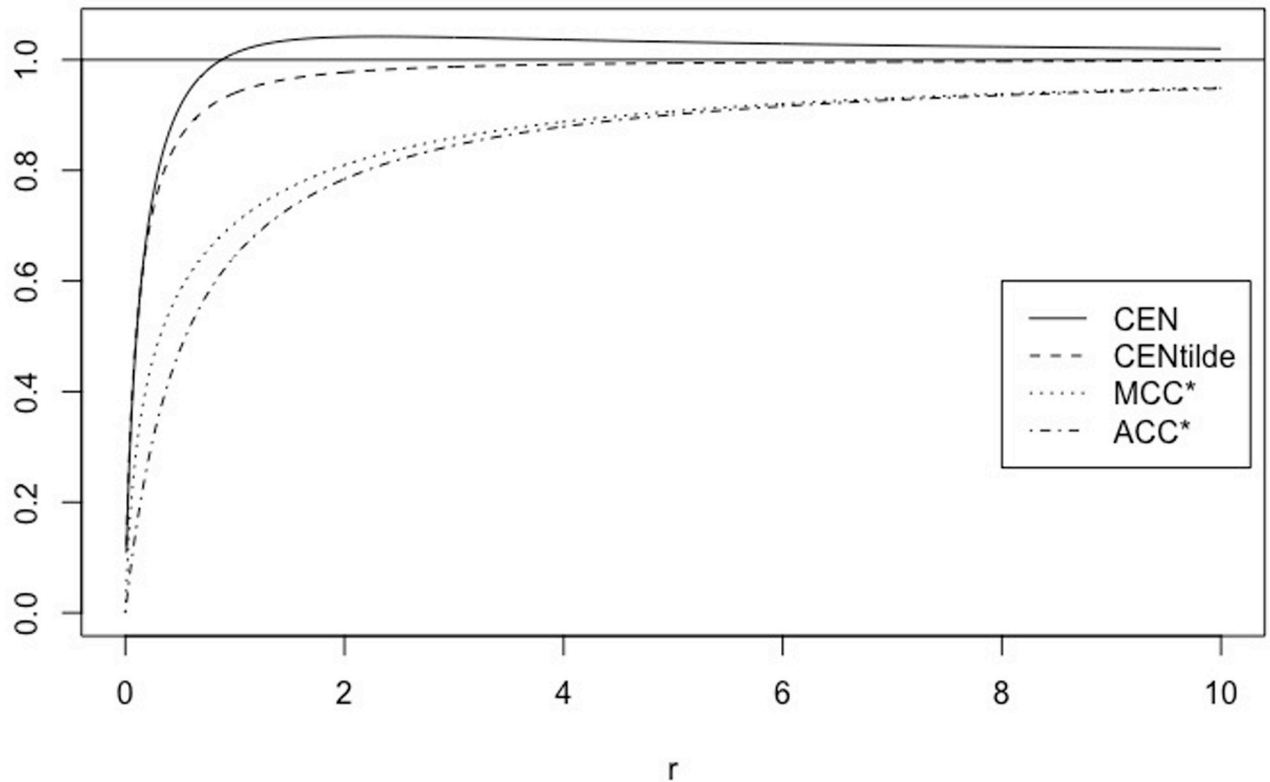
Fig 10. Family $X_{A, r}$. CEN, MCEN, ACC* and MCC* as function of $r > 0$ for $A = 10$.

underrepresented and mainly misclassified if $A, r > 1$, while class-1 cases are classified "at random", that is, a class-1 case has the same probability to be classified into any of the two classes. Although entropy is as for $X_{A, r}$, we will see that performance measures behave in a different way for this family of confusion matrices. When $0 < A < 1$, $A = 1/B$ with $B > 1$, then matrix $Y_{A, r}$ is equivalent to $\begin{pmatrix} r & r \\ 1 & B \end{pmatrix}$. In Proposition 6 we give some properties of CEN, MCEN, ACC* and MCC*. See in Fig 11 for $r = 0.1$, in Fig 12 for $r = 0.8$, and see Fig 13 for a plot of them as function of r, fixed $A = 10$.

Proposition 6 *For confusion matrix $Y_{A, r}$ with $A, r > 0$ we have*:

$$\text{CEN}(A) = \frac{(r+1)A \log_2(((r+1)A + 2)(3r+1)) + (r-1)A \log_2(A) - 2rA \log_2(rA)}{2((2r+1)A + 1)},$$

$$\text{MCEN}(A) = \frac{2((r+1)A \log_2(((r+1)A + 1)(2r+1)) + (r-1)A \log_2(A) - 2rA \log_2(rA))}{3((2r+1)A + 1) + (r+1)A},$$

$$\text{ACC}^*(A) = \frac{(r+1)A}{(2r+1)A + 1}, \quad \text{MCC}^*(A) = \frac{1 - \frac{r(1-A)}{\sqrt{2r(A+1)(r+1)(rA+1)}}}{2}.$$

As a consequence, $L_{\text{CEN}}(r) = \lim_{A \to +\infty} \text{CEN}(A) = \frac{1}{2(2r+1)} \log_2\left(\frac{((3r+1)(r+1))^{r+1}}{r^{2r}}\right) > 0$, and there

exists $R_0 < 1 (R_0 \approx 0.71)$ such that $L_{\text{CEN}}(r) \begin{cases} > 1 & \text{if } R_0 < r < 1, \\ = 1 & \text{if } r = R_0, 1, \\ < 1 & \text{if } r < R_0 \text{ or } r > 1. \end{cases}$ *Moreover, there exist 0*
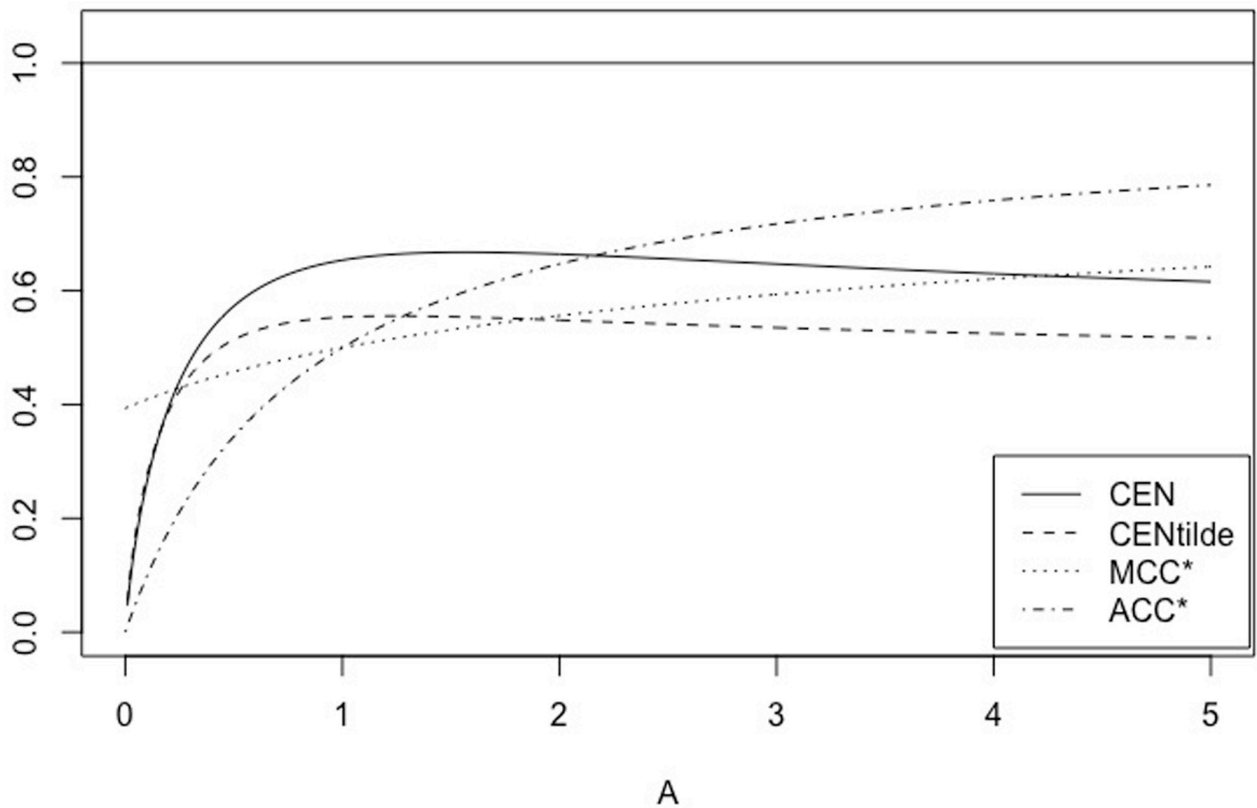
Fig 11. Family $Y_{A, r}$. CEN, MCEN, ACC* and MCC* as function of $A > 0$ for $r = 0.1$.

$< R_1 < R_0 < 1 < R_2 (R_1 \approx 0.5, R_2 \approx 1.4)$ *such that*

$$\begin{cases} \textit{if } r \in [R_0, 1], & \textit{there exists } A_r > 0 \textit{ such that } \mathrm{CEN}(A) < 1 \textit{ if } A < A_r, \\ & \mathrm{CEN}(A_r) = 1, \ \mathrm{CEN}(A) > 1 \textit{ if } A > A_r, \\ \textit{if } r \in (R_1, R_0) \cup (1, R_2), & \textit{there exist } 0 < A_r < B_r \textit{ such that } \mathrm{CEN}(A) < 1 \textit{ if } A < A_r \\ & \textit{or } A > B_r, \ \mathrm{CEN}(A_r) = \mathrm{CEN}(B_r) = 1, \\ & \mathrm{CEN}(A) > 1 \textit{ if } A \in (A_r, B_r), \\ \textit{if } r \notin (R_1, R_2), & \mathrm{CEN}(A) \leq 1 \textit{ for any } A > 0. \end{cases}$$

On the other hand, for any $r > 0$,

$\mathrm{MCEN}(A) < 1$, $\mathrm{ACC}^*(A) < 1$ *and* $\mathrm{MCC}^*(A) < 1$, *for all $A > 0$, ACC* and MCC* are monotonically increasing functions of A, CEN is not, and MCEN is or not, depending on the value of r,*

$$\lim_{A \to 0} \mathrm{CEN}(A) = \lim_{A \to 0} \mathrm{MCEN}(A) = \lim_{A \to 0} \mathrm{ACC}^*(A) = 0, \lim_{A \to 0} \mathrm{MCC}^*(A) = \frac{1 - \sqrt{\frac{r}{2(r+1)}}}{2},$$

$$\lim_{A \to +\infty} \mathrm{ACC}^*(A) = \frac{r+1}{2r+1} = L_{\mathrm{ACC}^*}(r), \ \lim_{A \to +\infty} \mathrm{MCC}^* = \frac{1 + \frac{1}{\sqrt{2(r+1)}}}{2} = L_{\mathrm{MCC}^*}(r), L_{\mathrm{MCEN}}(r) =$$

$$\lim_{A \to +\infty} \mathrm{MCEN}(A) = \frac{2}{3(2r+1) + (r+1)} \log_2\left(\frac{((2r+1)(r+1))^{r+1}}{r^{2r}}\right) < 1, L_{\mathrm{MCEN}}(r) < L_{\mathrm{CEN}}(r)$$

for all $r > 0$.
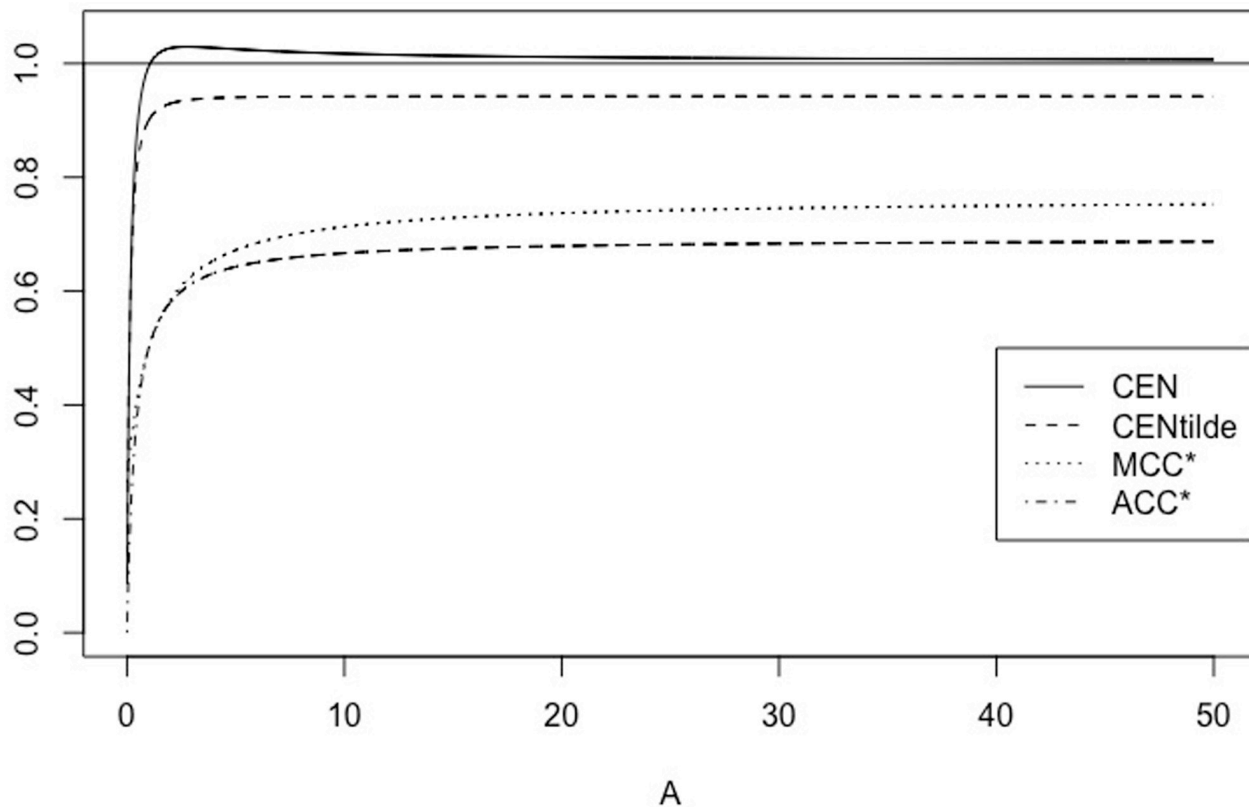
Fig 12. Family $Y_{A,\,r}$. CEN, MCEN, ACC* and MCC* as function of $A > 0$ for $r = 0.8$.

Note that $L_{\mathrm{ACC}^*}(r) < L_{\mathrm{MCC}^*}(r)$ *if and only if* $r > \frac{-1+\sqrt{5}}{4} > 0$.

Improving classification of the minority class while maintaining the imbalance between the classes. Up to now, we have evaluated binary confusion matrices with different balances of the two classes but not different classification results. Now let's do just the opposite. To help clarify the utility of MCEN in the evaluation of improvements in classification of the minority class while maintaining the same amount of imbalance, we consider two different examples.

Example 1: We introduce the family of confusion matrices $X^{\alpha}_{50,\,2} = \begin{pmatrix} 50 & 100 \\ 101-\alpha & \alpha \end{pmatrix}$,

with $\alpha = 1, 2, \ldots, 101$. Note that when $\alpha = 1$, the corresponding matrix belongs to the family $\{X_{A,\,r}\}$ with $A = 50$ and $r = 2$. Imbalance in classes stays fix. When $\alpha = 1$, the minority class is classified very badly, improving classification as α increases and reaching the perfect classification when $\alpha = 101$. Is MCEN able to detect this behaviour? Yes, it is. Unlike what happens with CEN, MCEN (as well as ACC* and MCC*) monotonically decreases when classification of the minority class improves (α increases). CEN incongruously first increases up to $\alpha = 18$ and then starts to decrease and behave like the other performance measures (see Fig 14).

Example 2: A similar phenomenon can be observed with family $Y^{\beta}_{100,\,1} = \begin{pmatrix} 100 & 100 \\ 101-\beta & \beta \end{pmatrix}$,

with $\beta = 1, 2, \ldots, 101$ (with $\beta = 1$ the corresponding matrix belongs to the family $\{Y_{A,\,r}\}$ with $A = 100$ and $r = 1$. As in Example 1, imbalance in classes is constant and when $\beta = 1$, the
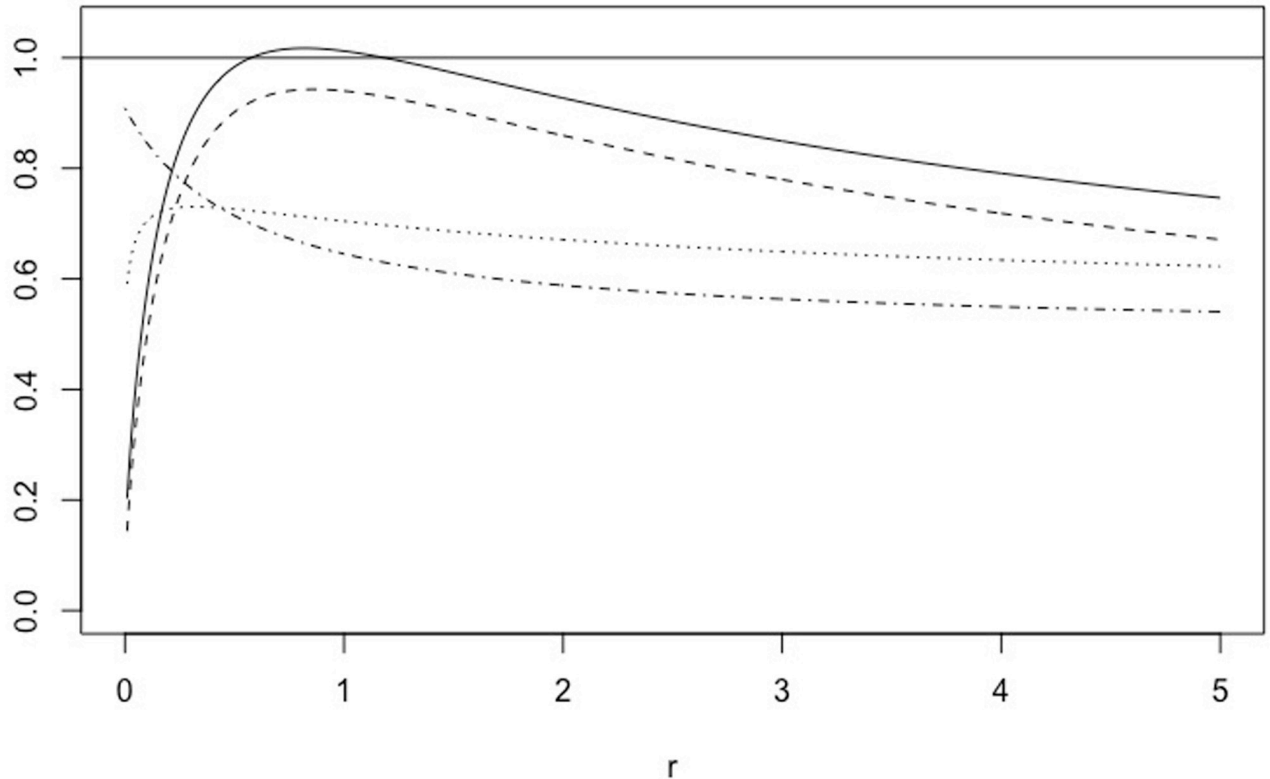
Fig 13. Family $Y_{A, r}$**.** CEN, MCEN, ACC* and MCC* as function of r for $A = 10$.

minority class is classified very badly, improving classification as β increases up to 101, when perfect classification is reached. MCEN as well as ACC* and MCC*, monotonically decrease when β increases, while CEN increases up to $\beta = 14$ and then starts to decrease and behave like the other performance measures (see Fig 15).

The Z_A family

As noted in [4], the behaviour of the Confusion Entropy CEN is rather diverse from that of MCC* and ACC* for the pathological case of the family of confusion matrices

$Z_A = (a_{i,j})_{i,j = 1, \ldots, N}$, defined by $a_{i,j} = \begin{cases} A & \text{if } i = N, j = 1 \\ 1 & \text{otherwise,} \end{cases}$, with $A > 0$. That is,

$Z_A = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ldots & \vdots \\ 1 & 1 & \ldots & 1 \\ A & 1 & \ldots & 1 \end{pmatrix}$. We want to study how MCEN behaves when applied to elements of
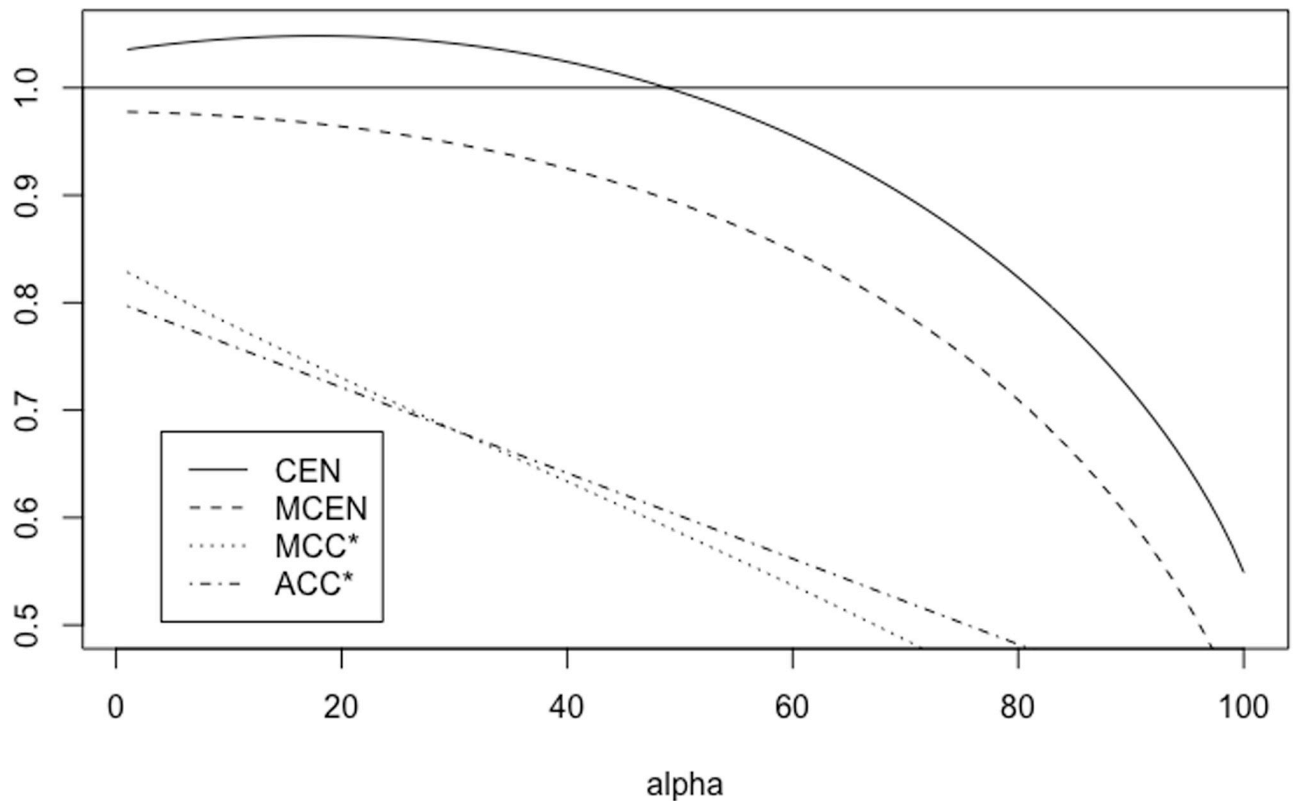
this family.

Fig 14. Family $X^z_{50,2}$ with $\alpha = 1, 2, \ldots, 101$. CEN, MCEN, ACC* and MCC* as function of α.

Proposition 7

$$\textit{If } N > 2, \; \mathrm{CEN}(\mathrm{Z_A}) = \frac{1}{N^2 + A - 1}\Bigg((N-1)(N-2)\log_{2(N-1)}(2N)$$

$$+(2N + A - 3)\log_{2(N-1)}(2N + A - 1) - A\log_{2(N-1)}(A)\Bigg),$$

$$\mathrm{MCEN} = \frac{2}{2(N^2 + A - 1) - N}\Bigg((N-1)(N-2)\log_{2(N-1)}(2N-1)$$

$$+(2N + A - 3)\log_{2(N-1)}(2N + A - 2) - A\log_{2(N-1)}(A)\Bigg),$$

$$\textit{if } N = 2, \; \mathrm{CEN}(\mathrm{Z_A}) = \frac{1}{A + 3}\left((A+1)\log_2(A+3) - A\log_2(A)\right),$$

$$\mathrm{MCEN} = \frac{2}{2A + 5}\left((A+1)\log_2(A+2) - A\log_2(A)\right).$$

In general ($N \geq 2$),

$$\mathrm{MCC}^*(Z_A) = \frac{N(N^2 + 2(A-1)) - (N^2 + (A-1))}{2(N-1)(N^2 + 2(A-1))},$$

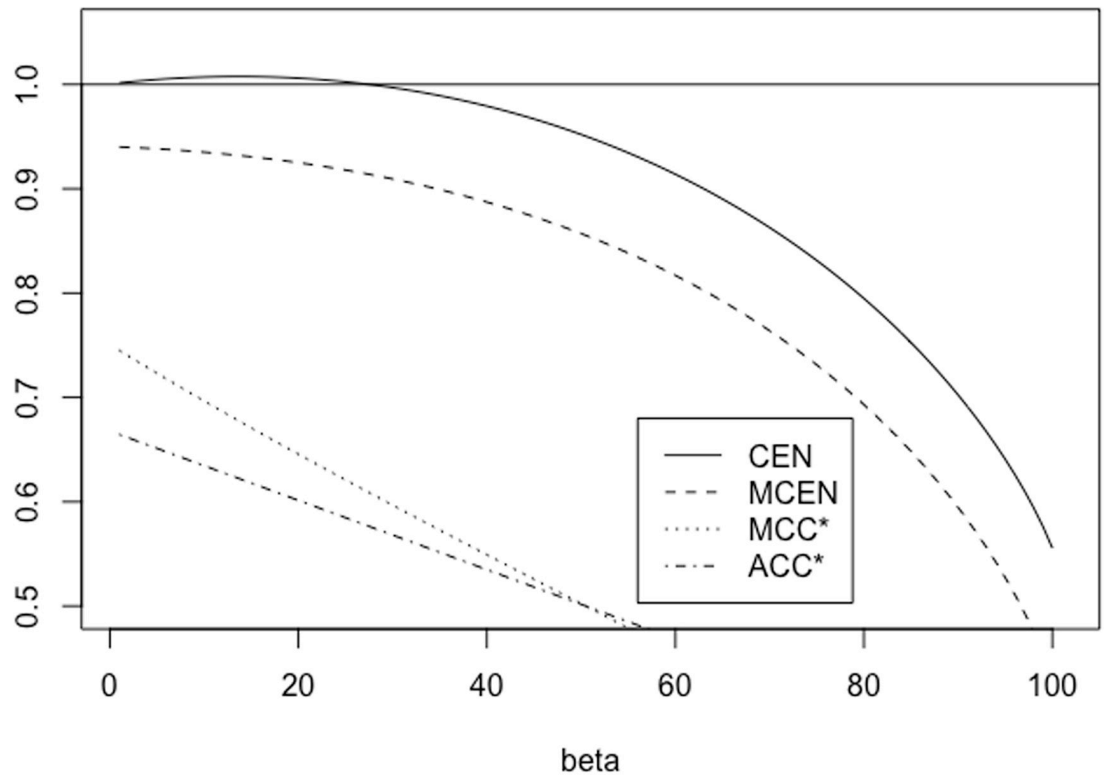$$\mathrm{ACC}^*(Z_A) = \frac{N^2 - N + (A-1)}{N^2 + (A-1)}$$

As a consequence,

Fig 15. Family $Y_{50,2}^{\beta}$ with $\beta = 1, 2, \ldots, 101$. CEN, MCEN, ACC* and MCC* as function of β.

- *If N = 2,*
  MCEN < CEN(Z_A) for all A>0,
  MCEN < 1 for all A > 0, and there exists $A_3 \in (1, 2)(A_3 \approx 1.85)$ such that

  $$\mathrm{CEN}(Z_1) = \mathrm{CEN}(Z_{A_3}) = 1,$$
  $$\mathrm{CEN}(Z_A) > 1 \ \textit{if} \ A \in (1, A_3) \ \ \textit{and} \ \ \mathrm{CEN}(Z_A) < 1 \ \textit{if} \ A \notin [1, A_3],$$

  $$\lim_{A \to 0} \mathrm{MCC}^*(A) = \frac{1}{4} < \lim_{A \to 0} \mathrm{ACC}^* = \frac{1}{3} < \lim_{A \to 0} \mathrm{MCEN}(A) = \frac{2}{5} < \lim_{A \to 0} \mathrm{CEN}(A) = \frac{\log_2(3)}{3},$$

  $$\lim_{A \to +\infty} \mathrm{CEN}(A) = \lim_{A \to +\infty} \mathrm{MCEN}(A) = 0 < \lim_{A \to +\infty} \mathrm{MCC}^* = \frac{3}{4} < \lim_{A \to +\infty} \mathrm{ACC}^* = 1 \,.$$

- *If N = 3 (we take this case as example of what happens with N > 2),*

  $$\lim_{A \to 0} \mathrm{MCC}^*(A) = \frac{13}{28} < \lim_{A \to 0} \mathrm{ACC}^* = \frac{5}{8} <$$

  $$< \lim_{A \to 0} \mathrm{CEN}(A) = \frac{2 \log_4(6) + 3 \log_4(5)}{8} < \lim_{A \to 0} \mathrm{MCEN}(A) = \frac{2}{13}(2 \log_4(5) + 3) < 1,$$

  $$\lim_{A \to +\infty} \mathrm{CEN}(A) = \lim_{A \to +\infty} \mathrm{MCEN}(A) = 0 < \lim_{A \to +\infty} \mathrm{MCC}^* = \frac{5}{8} < \lim_{A \to +\infty} \mathrm{ACC}^* = 1.$$
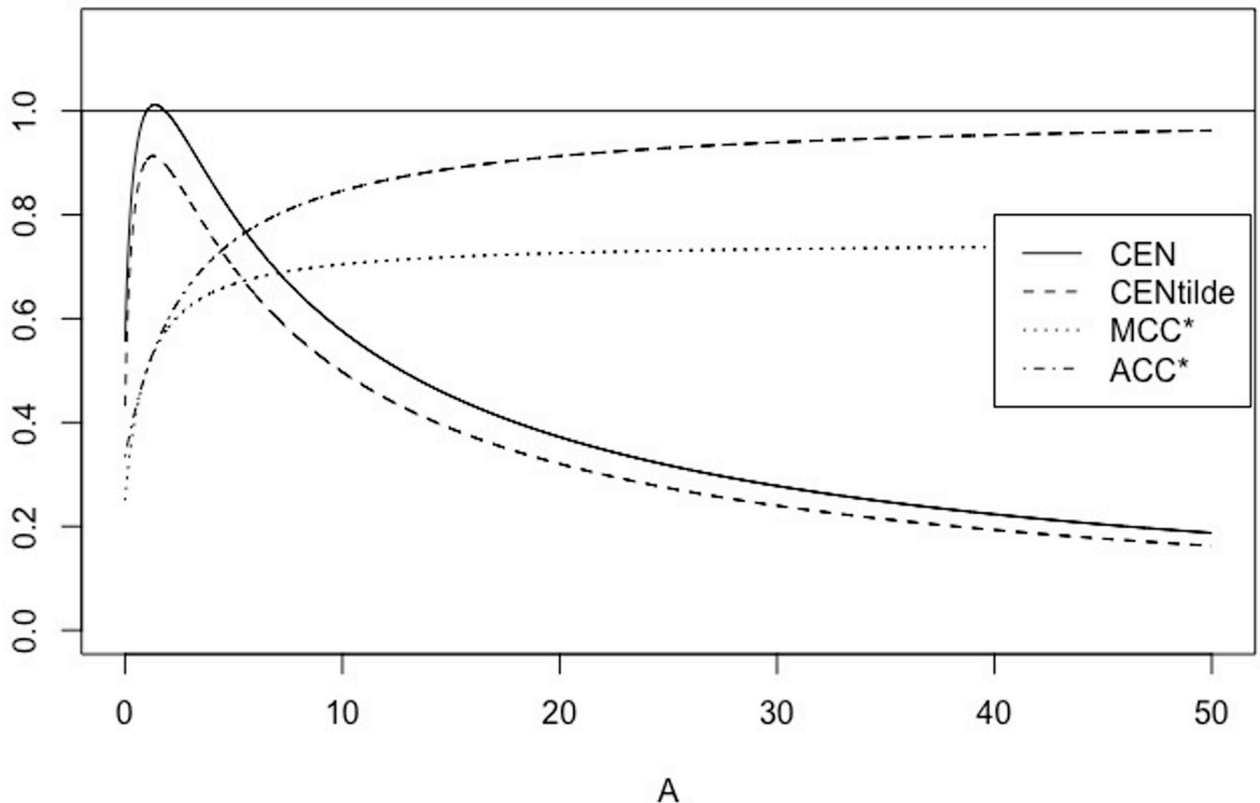
Fig 16. Family Z_A. CEN, MCEN, MCC* and ACC* as function of $A > 0$ for $N = 2$.

In Figs 16 and 17 we can observe this behaviour when $N = 2$ and $N = 3$, respectively.

Table 8 shows some examples of confusion matrices of the family Z_A, first with $N = 2$, and secondly with $N = 4$.

Note that CEN and MCEN exhibit a very different behaviour comparing with ACC* and MCC*, since the former are sensitive to the overall entropy associated to the elements of the matrix, which is $\log\left(N^2 + A - 1\right) - \frac{A}{N^2+A-1} \log\left(A\right)$. Entropy decreases to $\log(N^2 - 1)$ when $A \rightarrow 0$, and drops to 0 when $A \rightarrow +\infty$.

Comparing with other performance measures

Several works have considered the question of the introduction and comparison of different performance measures for classification, inspired, in one way or another, by Shannon's entropy. For example, in [13] the authors introduce a novel measure called PACC (Probabilistic Accuracy) in the multi-class setting, making a comparative study of it with other measures as Accuracy, MCC and CEN, among others.

Besides, Entropy-Modulated Accuracy (EMA), introduced in [14], is a performance measure of classification tasks based on the concept of *perplexity*, the latter being defined as the effective number of classes a classifier sees. The authors also introduce NIT (Normalized Information Transfer) factor, which is a correction of EMA. They compare both EMA and NIT factor with Accuracy and CEN, rejecting rankings of classifiers based in Accuracy and choosing more meaningful and interpretable classifiers. They show in some examples that MCC is
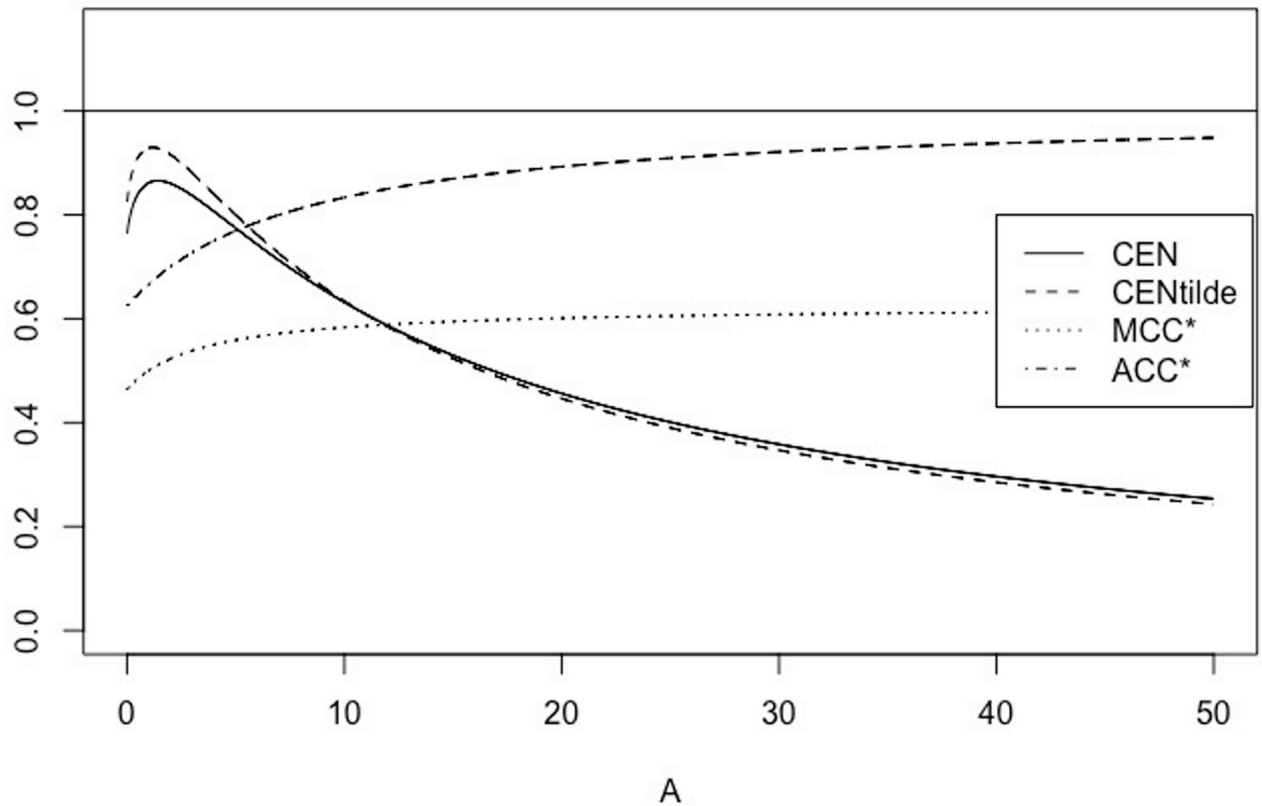
Fig 17. Family Z_A. CEN, MCEN, MCC* and ACC* as function of $A > 0$ for $N = 3$.

https://doi.org/10.1371/journal.pone.0210264.g017

highly correlated with Accuracy, while rankings obtained with CEN, EMA and NIT factor are comparable in some cases but disagree in others.

Although PACC, EMA and NIT factor are useful measures to assess classifiers, in our opinion none of them is completely satisfactory in grading the effectiveness of the classifier learning process, since all reflect some concrete feature of the classification process, being insufficient for covering all the aspects of this complex task, so they should be used cautiously and in a complementary way. That is, all the measures suffer from certain weaknesses that are evident in specific, more or less gimmicky situations. This comment extends also to both CEN and MCEN, although it should be noted that the latter solves the problems showed by CEN in the binary setting, as well as to MCC and Accuracy, the last one having been widely treated (see, for example, the Introduction section in [14]).

Let us exemplify this fact by going back to the toy example in Table 2. In Table 9 we add the calculated values of PACC* = 1-PACC and 1/NIT to that of Table 2. We use NIT factor (inverted to make it comparable with the other measures) instead of EMA since the probability distribution of classes in the validation set is not uniform. Note that our confusion matrices are transposed with respect to that in [14], and also that for the NIT factor we use formula (4). We have used the corrected definition provided by the authors, which had already acknowledged an erratum in Eq (4) in the comments of https://www.researchgate.net/publication/259743406_100_Classification_Accuracy_Considered_Harmful_The_Normalized_Information_Transfer_Factor_Explains_the_Accuracy_Paradox/.

The behaviour of PACC* showed in Table 9 is consistent with that of MCC*, increasing when IN entropy decreases (a) and decreasing when OUT decreases (b). However, the

Table 8. Examples with different matrices Z_A in cases $N = 2$ and $N = 4$.

	$\begin{pmatrix} 10 & 10 \\ 1 & 10 \end{pmatrix}$	$\begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 10 & 1 \end{pmatrix}$
A =	1/10	1/2	1	2	10
ACC* =	0.3548	0.4286	0.5000	0.6000	0.8462
MCC* =	0.2955	0.4167	0.5000	0.5833	0.7045
CEN =	0.6864	0.9174	1.0000	0.9932	0.5758
MCEN =	0.5806	0.8276	0.9057	0.8889	0.4972
Z_A =	$\begin{pmatrix} 10^2 & 10^2 & 10^2 & 10^2 \\ 10^2 & 10^2 & 10^2 & 10^2 \\ 10^2 & 10^2 & 10^2 & 10^2 \\ 1 & 10^2 & 10^2 & 10^2 \end{pmatrix}$	$\begin{pmatrix} 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 1 & 10 & 10 & 10 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 10 & 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 10^2 & 1 & 1 & 1 \end{pmatrix}$
A =	10^{-2}	10^{-1}	1	10	10^2
ACC* =	0.7335	0.7351	0.7500	0.8400	0.9652
MCC* =	0.4882	0.4894	0.5000	0.5441	0.5771
CEN =	0.8284	0.8391	0.8704	0.7132	0.2068
MCEN =	0.8883	0.9001	0.9309	0.7338	0.2016

behaviour of 1/NIT is consistent with that of CEN and MCEN, decreasing in both cases. Nevertheless, unlike what happens with CEN and MCEN, NIT factor does not distinguish among scenarios (a) and (b). This is because both EMA and NIT factor are invariants to permutations of the columns.

Another example is that of the *MEG mind reading challenge* organized by the PASCAL (Pattern Analysis, Statistical modeling and ComputAtional Learning) network in [15], already considered in [14]. We restrict our comparison to the group of the four most outstanding systems, denoted C_1 (Huttunen et al.), C_2 (Santana et al.), C_3 (Jylänki et al.) and C_4 (Tu & Sun), since for them, unlike what happens with the rest, we could access to the confusion matrices in [15]. The results are in Table 10, and from them we see that the most comparable rankings are that given by the NIT factor, CEN and MCEN, showing clusters $\{C_4, C_2\}$ and $\{C_1, C_3\}$, with very small differences inside the clusters, specially the second. The authors of the report [15] were specially interested in comparison C_1 vs. C_4, and 1/NIT factor, as well as CEN and MCEN, give the same ordering: C_4 is better (lower value) than C_1, in concordance with interpretability given in [14].

Table 9. Toy example of Table 2 revisited, adding PACC and the NIT factor.

	Baseline	(a)			(b)		
	$\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$	$\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}$	$\begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}$	$\begin{pmatrix} 0 & 3 \\ 3 & 6 \end{pmatrix}$	$\begin{pmatrix} 3 & 2 \\ 4 & 3 \end{pmatrix}$	$\begin{pmatrix} 3 & 1 \\ 5 & 3 \end{pmatrix}$	$\begin{pmatrix} 3 & 0 \\ 6 & 3 \end{pmatrix}$
ACC* =	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
MCC* =	0.5000	0.5130	0.5625	0.6667	0.4881	0.4375	0.3333
CEN =	1.0000	0.9898	0.9575	0.8962	0.9591	0.8250	0.5000
MCEN =	0.9057	0.9006	0.8848	0.8571	0.8590	0.7057	0.3343
PACC* =	0.5000	0.5071	0.5312	0.5833	0.4929	0.4687	0.4167
1/NIT =	2.0000	1.9992	1.9840	1.8371	1.9992	1.9840	1.8371

Table 10. Results for the first four systems of the *MEG mind reading challenge*. Confusion matrices have been obtained from [15].

System	ACC*	MCC*	CEN	MCEN	PACC*	1/NIT
C_1	0.3201	0.2010	0.4360	0.5694	0.3230	2.5877
C_2	0.3675	0.2286	0.4043	0.4981	0.3668	2.4715
C_3	0.3721	0.2319	0.4483	0.5645	0.3667	2.6151
C_4	0.3783	0.2369	0.4213	0.5279	0.3737	2.4545

https://doi.org/10.1371/journal.pone.0210264.t010

Table 11. Two toy examples. With $S = 30$ for $N = 2$, and with $S = 40$ for $N = 3$.

	$A = \begin{pmatrix} 10 & 0 \\ 10 & 10 \end{pmatrix}$		$B = \begin{pmatrix} 0 & 10 \\ 10 & 10 \end{pmatrix}$	$C = \begin{pmatrix} 10 & 0 & 0 \\ 10 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}$		$D = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 10 \\ 10 & 0 & 0 \end{pmatrix}$
ACC* =	0.3333	<	0.6667	0.2500	<	0.5000
MCC* =	0.2500	<	0.7500	0.1500	<	0.3500
CEN =	0.5283	<	1.0000	0.1981	<	0.3231
MCEN =	0.4000	<	0.9400	0.2000	<	0.3333
PACC* =	0.2917	<	0.7083	0.1944	<	0.5000
1/NIT =	1.6799	=	1.6799	1.5000	=	1.5000

https://doi.org/10.1371/journal.pone.0210264.t011

One more example to show the variability when performance measures are compared: in Table 11 we see that the NIT factor (equivalently, EMA), unlike the other measures, is not able to distinguish between classifiers whose confusion matrices are *A* and *B* in the binary case, nor between *C* and *D* in multi-class classification.

Supporting information file: Experiments and results

The advantages of using Modified Confusion Entropy MCEN measure against CEN have been tested on different binary classifiers, constructed from four available datasets from the UCI ML Repository (https://archive.ics.uci.edu). From each dataset we construct and assess eight different classifiers, five of which are Bayesian networks, while the rest are other standard machine learning procedures used in supervised classification problems.

Because of the comparisons carried out previously with different examples, we have to recognize the impossibility of deciding what measure of behaviour, of the considered ones, can allow to decide in the case that the rankings of classifiers obtained with CEN and MCEN were different. We decided, then, to use OUT entropy as such a reference when there is disparity; in case of a tie, we will use IN entropy to break it. This is what we will call "the criterion of entropy".

To compare rankings obtained from CEN and MCEN and that obtained by the criterion of entropy, we use both the Hamming distance and the degree of consistency indicator *c* (see [16]).

The results obtained with all the considered datasets heuristically reinforce that MCEN is more correlated with entropy than CEN. (see S1 File and Tables A-F in S1 File).

Conclusion

We introduced MCEN as a modification of the original Confusion Entropy performance measure CEN introduced in [3], both for binary and multi-class classification, proving some

properties. We compared this measure with CEN, MCC and Accuracy, showing that in the binary case, MCEN overcomes the unreliability of CEN in a twofold sense: the departure of the range where it should be (the interval [0, 1]), and the lack of monotonicity when the entropy increases or decreases. These features made CEN an inappropriate measure in the binary case, proving MCEN to be a good alternative, and we study different scenarios to highlight this fact. Moreover, while nor Accuracy nor MCC can distinguish among different misclassification distributions of cases in the confusion matrix, MCEN and CEN have an high level of discrimination.

First, we show that in the binary case (see Table 2), both CEN and MCEN are sensitive to the decreasing in the entropy within the main diagonal IN, an also to that outside the diagonal OUT, but while CEN is more sensitive than MCEN to IN, the opposite occurs with OUT. By contrast, ACC is insensitive as long as the sum of the diagonal and the total sum remain constant. Secondly, we consider the multi-class perfectly symmetric and balanced case in which the main diagonal elements are equal to T and the elements outside the diagonal are equal to F, which is analytically studied in detail, showing the output-of-range of CEN in the binary case when $\gamma = T/F \in (0, 1)$.

After that, se consider different particular situations in the binary setting, through the study of some families of confusion matrices. Family U_A is symmetric and unbalanced, showing the out-of-range of CEN for any $A > 1$, and in addition a lack of monotonicity that contrast with the behaviour of the overall entropy associated to the elements of the matrix. Family V_A is asymmetric and unbalanced, and also shows the out-of-range of CEN but only for A in the interval $(1, A_1)$, where $A_1 \approx 1.4$.

Two doubly indexed families have been considered in the binary case. CEN has an anomalous behaviour for family $X_{A, r}$, which is symmetric but unbalanced, for $r > r_0$ (with $r_0 \approx 0.8$) since it is not only out-of-range from a certain value of A, but its limit when $A \to +\infty$ is >1 if $r > 1$, showing lack of monotonicity. The same happens from a certain value of r, fixed A. Family $Y_{A,r}$ is also unbalanced but asymmetric. When r is in the interval $(R_0, 1)$ with $R_0 \approx 0.71$, CEN is not only out-of-range from a certain value of A, but its limit when $A \to +\infty$ is >1 if $r > 1$, showing lack of monotonicity. But there are other two intervals of values for r in which CEN>1 for A living in a certain bounded interval.

Besides evaluating binary confusion matrices with the same classification results for the minority class but different balances of the two classes, we compare through two examples the behaviour of MCEN with that of CEN, ACC* and MCC*, in evaluating improvements in classification of the minority class while maintaining the same amount of imbalance. We show that CEN is the only one that does not show a monotonous decrease as the classification improves, for which MCEN proves, also in this sense, that it outperforms CEN.

Finally, we also consider the multi-class family Z_A, which is asymmetric and unbalanced, and observe that in the binary case, CEN is out-of-range for $A \in (1, A_3)$, with $A_3 \approx 1.85$.

In all of these examples, MCEN behave appropriately. Comparing with the overall Shannon's entropy associated to the set of elements of the confusion matrix, both CEN and MCEN are sensitive to it but CEN sometimes does not show the same behaviour in terms of monotonicity than entropy. With respect to Accuracy and MCC, conveniently scaled, they show sometimes a behaviour in contradiction with Shannon's entropy, as for families V_A and Z_A.

A further comparison has been carried out with the Probabilistic Accuracy (PACC) introduced in [13], and the Entropy-Modulated-Accuracy EMA and the Normalized Information Transfer (NIT) factor, both introduced in [15]. We consider different examples in which sometimes PACC* = 1−PACC behaves consistently with MCC*, increasing when IN entropy decreases and decreasing when OUT decreases, while 1/NIT behaves in accordance with CEN and MCEN, decreasing in both cases, but with the handicap that unlike what happens with

CEN and MCEN, NIT factor does not distinguish between IN and OUT. But not always. Actually, no measure seems to be completely satisfactory since each one reflects a specific characteristic of the classification process, so they should be used in a complementary way and none can be taken as a gold standard to compare the others.

Finally, to make clear the improvement of MCEN over CEN, we carry out experimentation consisting in the comparison of the rankings of some classifiers obtained from four different real datasets by using both measures. Mostly the classifiers orderings match, but when they do not, it is the MCEN that most agrees with the criterion of entropy. To see that, we use both the Hamming distance and the degree of consistency indicator c. These results heuristically support the use of MCEN as a better alternative to CEN in the binary case, when a performance measure based in entropy is required.

Supporting information

S1 File. Supporting information: Experiments and results. Table A. Datasets used in the experiments. Table B. Classifiers used in the experiments. Table C. Results for the Breast cancer dataset. Table D. Results for the SPECT heart dataset. Table E. Results for the Congressional voting dataset. Table F in S1 File. Results for the MONK's Problems.
(PDF)

S2 File. Breast cancer dataset.
(CSV)

S3 File. Breast cancer description.
(PDF)

S4 File. SPECT dataset.
(CSV)

S5 File. SPECT description.
(PDF)

S6 File. UCB admissions dataset.
(CSV)

S7 File. UCB admissions description.
(PDF)

Acknowledgments

Author Contributions

Formal analysis: Rosario Delgado, J. David Núñez-González.

Investigation: Rosario Delgado, J. David Núñez-González.

Writing – original draft: Rosario Delgado, J. David Núñez-González.

Writing – review & editing: Rosario Delgado, J. David Núñez-González.

References

1. Matthews B.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et biophysica acta. Vol 405, Num 2, 442–451 (1975). https://doi.org/10.1016/0005-2795 (75)90109-9

2. Delgado, R., Núñez-González, D.: Enhacing Confusion Entropy (CEN) as measure for evaluating classifiers. In: Graña M. et al. (eds) International Joint Conference SOCO'18-CISIS'18-ICEUTE'18. Advances in Intelligent Systems and Computing, vol 771. Springer, Cham (2019).

3. Wei J.-M., Yuan X.-Y., Hu Q.-H., Wang S.-Q.: A novel measure for evaluating classifiers. Expert Systems with Applications, Vol 37, 3799–3809 (2010). https://doi.org/10.1016/j.eswa.2009.11.040

4. Jurman G., Riccadonna S., Furlanello C.: A Comparison of MCC and CEN Error Measures in Multi-Class Prediction. Plos One. Vol 7, Num 8, 1–8 (2012).

5. Jin, H., Wang, X.-N., Gao, F., Li, J., Wei, J.-M.: Learning Decision Trees using Confusion Entropy. Proceedings of the 2013 International Conference on Machine Learning and Cybernetics, Tianjin, 14-17 July (2013).

6. Roumani Y.-F., May J.-H., Strum D.-P.: Classifying highly imbalanced ICU data. Health Care Manag. Sci. Vol 16, 119–128 (2013). https://doi.org/10.1007/s10729-012-9216-9

7. Roumani Y.-F., Rouman Y., Nwankpa J.-K., Tanniru M.: Classifying readmissions to a cardiac intensive care unit. Annals of Operations Research, vol. 263 (1-2), 429–451 (2018). https://doi.org/10.1007/s10479-016-2350-x

8. Wang X.-N., Wei J.-M., Jin H., Yu G., Zhang H.-W.: Probabilistic Confusion Entropy for Evaluating Classifiers. Entropy, Vol 15, 4969–4992 (2013). https://doi.org/10.3390/e15114969

9. Antunes F., Ribeiro B., Pereira F.: Probabilistic modeling and visualization for bankruptcy prediction. Applied Soft Computing vol. 60, 831–843 (2017). https://doi.org/10.1016/j.asoc.2017.06.043

10. Sublime J., Grozavu N., Cabanes G., Bennani Y., Cornuéjols A.: From Horizontal to Vertical Collaborative Clustering using Generative Topographic Maps. International Journal of Hybrid Intelligent Systems, vol. 12(4), 245–256 (2015). https://doi.org/10.3233/HIS-160219

11. Sublime, J., Matei, B., Murena, P.-A.: Analysis of the influence of diversity in collaborative and multi-view clustering. 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 4126–4133 (2017).

12. Sublime J., Matei B., Cabanes G., Grozavu N., Bennani Y., Cornuéjols A.: Entropy based probabilistic collaborative clustering. Pattern Recognition, vol. 72, 144–157 (2017). https://doi.org/10.1016/j.patcog.2017.07.014

13. Sigdel M., Aygun R.: Pacc—A Discriminative and Accuracy Correlated Measure for Assessment of Classification Results. Machine Learning and Data Minning in Pattern Recognition. Vol 7988. LNCS. pp 281–295 (2013). https://doi.org/10.1007/978-3-642-39712-7_22

14. Valverde-Albacete F.J., Peláez-Moreno C.: 100% Classification Accuracy Considered Harmful: The Normalized Information Transfer Factor Explains the Accuracy Paradox. Plos One. Vol 9, Num 1, 1–10 (2014).

15. Klami, A., Ramkumar, P., Virtanen, S., Parkkonen, L., Hari, R., Kaski, S.: ICANN / PASCAL 2 Challenge: MEG Mind Reading –Overview and Results. In: Klami, A., editor, Proceedings of ICANN/PASCAL2 Challenge: MEG Mind Reading. Espoo, Aalto University Publication series SCIENCE + TECHNOLOGY 29/2011, pp. 3–19. http://urn.fi/URN:ISBN:978-952-60-4456-9

16. Huang J., Ling C.: Using AUC and Accuracy in Evaluating Learning Algoritms. IEEE Transactions on Knowledge and Data Engineering, vol. 17, 299–310 (2005). https://doi.org/10.1109/TKDE.2005.50