Journal of Engineering Research

Volume 7 | Issue 1 Article 26

2023

Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial

Mahmoud Fahmy Amin

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial

Mahmoud M. Fahmy

Professor, Computer and Control Engineering Department, Faculty of Engineering, Tanta University, Tanta, Egypt
e-mail: m.fahmy@f-eng.tanta.edu.eg

***Abstract*: The confusion matrix is a specific table used in machine learning to describe and assess the performance of a classification model (e.g., an artificial neural network) for a set of test data whose actual distinguishing features are known. The confusion matrix for an n-class classification problem is square, with n rows and n columns. The rows represent the class actual samples (instances), which are the classifier inputs, and the columns represent the class predicted samples, which are the classifier outputs. Binary class classifiers have been presented in a previous paper, where in this paper, we are concerned with three-class classification performance measures. We also clarify the concept with numerical examples to make it close to the reader mind.**

***Keywords*— Machine Learning; Confusion matrix; Accuracy; Recall; Three-class classifier, Specificity; Precision; True Negative; False Positive.**

1. INTRODUCTION

The performance of a classifier is measured using different measures such as accuracy, recall and precision. The classifier is said to be efficient if its performance scores are high. These measures are widely used in the literature in different problems [1-6, 8, 9, 10]. In our previous work [7], we defined the binary classification performance measures. In this current paper, we will discuss the three-class classifiers. They have the same concepts as in the binary classification, however, the equations will be based on three classes instead of just two..

Formally, comparing the actual classifications to the predicted classifications reveals four distinct outcomes:

- The actual classification is positive, as is the predicted classification. This is known as a 'true positive,' abbreviated TP, because the classifier correctly identified the positive sample. • The actual classification is negative, and the predicted classification is negative. This is a "true negative" (TN) result because the classifier correctly identified the negative sample.
- The predicted classification is positive, while the actual classification is negative. This is a 'false positive' (FP) result because the classifier incorrectly identified the negative sample as positive.
- The predicted classification is negative, while the actual classification is positive. This is a 'false negative' (FN) result because the classifier incorrectly identified the positive sample as negative.

These four outcomes, with the above interpretation, pertain in fact to the positive class, provided this class is particularly important and deserves emphasis; it accommodates what can be called 'relevant' samples, while the negative class is regarded as 'irrelevant'. The outcomes TP, TN, FP, and FN are of primary importance and are referred to as the 'building blocks' because they are used to formulate all performance measures.

2. THREE-CLASS CLASSIFICATION

The concepts of the confusion matrix and relevant performance measures discussed in [7] for binary classification can readily be extended to multiple-class classification. As a starting step in this generalization process, we consider a 3-class classification problem (n= 3). The confusion matrix has three rows and three columns, as shown in Fig. 1. The classes are labeled A, B, and C. A matrix cell at the intersection of the i[th] row and j[th] column is designated C_{ij}. The cell values indicated in Fig. 1 are just examples for classification results; there is a total of N = 150 tested samples in the nine cells.

The first question to answer is: How can we determine the four building blocks: TP (true positive), TN (true negative), FP (false positive), and FN (false negative for each of the three classes?

For class A, we define:

- TP_A: Class-A samples classified correctly as class A. This is the value of cell c_{11} alone, at the intersection of row A and column A. See Fig. 1.
- TN_A: Not class-A samples (i.e. samples of class B or class C) classified correctly or incorrectly as not class A. This is the sum of the four cells c_{22}, c_{23}, c_{32}, and c_{33}, the part of the matrix remaining after removing row A and column A.
- FP_A: Not class-A samples classified incorrectly as class A. This is the sum of the two cells c_{21}, and c_{31}, the portion of column A remaining after removing cell c_{11} (TP_A).
- FN_A: Class-A samples classified incorrectly A as not class A. This is the sum of the two cells c_{12}, and c_{13}, the portion of row A remaining after removing cell c_{11} (TP_A).

		Predicted			
		A	B	C	
Actual	A	c_{11} 32	c_{12} 10	c_{13} 8	←Row A
	B	c_{21} 9	c_{22} 38	c_{23} 4	←Row B
	C	c_{31} 12	c_{32} 9	c_{33} 28	←Row C
		Column A	Column B	Column C	

Fig. 1: Confusion matrix for 3-class classification

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

Predicted

		A	B	C
Actual	A	TP_A	FN_A	FN_A
	B	FP_A	TN_A	TN_A
	C	FP_A	TN_A	TN_A

Fig. 2: Building blocks for class A in 3-class classification

		A	B	C
Actual	A	TN_B	FP_B	TN_B
	B	FN_B	TP_B	FN_B
	C	TN_B	FP_B	TN_B

Fig. 3: Building blocks for class B in 3-class classification

		A	B	C
Actual	A	TN_C	TN_C	FP_C
	B	TN_C	TN_C	FP_C
	C	FN_C	FN_C	TP_C

Fig. 4: Building blocks for class C in 3-class classification

Figure 2 shows, on the confusion matrix, the building blocks TP_A, TN_A, FP_A, and FN_A, for class A. Through similar arguments, we define the building blocks for classes B and C. See Figs. 3 and 4. For convenience, when considering one class, we regard this class as positive and the other two classes as negative.

Numerically, for class A,

$TP_A = c_{11} = 32$
$TN_A = c_{22} + c_{23} + c_{32} + c_{33} = 38 + 4 + 9 + 28 = 79$
$FP_A = c_{21} + c_{31} = 9 + 12 = 21$
$FN_A = c_{12} + c_{13} = 10 + 8 = 18$

For class B,

$TP_B = c_{22} = 38$
$TN_B = c_{11} + c_{13} + c_{31} + c_{33} = 32 + 8 + 12 + 28 = 80$
$FP_B = c_{12} + c_{32} = 10 + 9 = 19$
$FN_B = c_{21} + c_{23} = 9 + 4 = 13$

For class C,

$TP_C = c_{33} = 28$
$TN_C = c_{11} + c_{12} + c_{21} + c_{22} = 32 + 10 + 9 + 38 = 89$
$FP_C = c_{13} + c_{23} = 8 + 4 = 12$
$FN_C = c_{31} + c_{32} = 12 + 9 = 21$

Note particularly that $TP_A = 32$, $TP_B = 38$, and $TP_C = 28$ are the diagonal elements of the confusion matrix of Fig. 1. The building blocks for the three classes A, B, and C are summarized in Fig. 5. The sum TP +TN + FP + FN for each class is, as expected, equal to the total number a samples, N= 150. As is clear in Fig. 1, $N_A = 50$, $N_B = 51$, and $N_C = 49$.

Of interest is the fact that a 3x3 confusion matrix can be decomposed into three 2x2 confusion matrices (analogous to those defined for binary classification). Using the values of the building blocks for classes A, B, and C indicated in Fig. 5, the 3x3 confusion matrix of Fig. 1 is decomposed into three component 2x2 confusion matrices for classes A, B, and C, respectively, as shown in Fig. 6.

	TP	TN	FP	FN	
Class A	32	79	21	18	N = 150
Class B	38	80	19	13	N = 150
Class C	28	89	12	21	N = 150

Fig. 5: Building blocks for classes in 3-class classification of Fig. 1

Predicted

		A	Not A	
Actual	A	$TP_A = 32$	$FN_A = 18$	$N_A= 50$
	Not A	$FP_A = 21$	$TN_A =79$	$N_B+ N_C =100$

(a) Class A

Predicted

		B	Not B	
Actual	B	$TP_B = 38$	$FN_B = 13$	$N_B = 51$
	Not B	$FP_B = 19$	$TN_B = 80$	$N_A+ N_C = 99$

(b) Class B

Predicted

		C	Not C	
Actual	C	$TP_C = 28$	$FN_C = 21$	$N_C= 49$
	Not C	$FP_C = 12$	$TN_C = 89$	$N_A+ N_B =101$

(c) Class C

Fig. 6: Three component 2X2 confusion matrices for 3X3 confusion matrix of Fig. 1.

The cell values of the three 2X2 confusion matrices of Fig. 6 yield the same information as the three rows of the table of Fig. 5 — the building blocks of the individual classes. Here, in a sense, we can visualize that the 3-class classification problem is converted into three binary classification problems.

Example 1

A 3-class classification problem has the confusion matrix shown in Fig. 7, where the three classes are labeled K, L, and M.

(a) How man samples does the test set contain? How are these samples distributed among the three classes?
b) How many samples are correctly classified?
(c) How man samples are incorrectly classified as class K? Class L? Class M?
(d) Determine the values of the building blocks or each class. Construct the corresponding three 2X2 confusion matrices.

Predicted

		K	L	M
Actual	K	50	20	0
	L	0	60	4
	M	1	5	54

Fig. 7. Confusion matrix for Example 1

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

Solution

The test set contains

N= 50 + 20 + 0 + 0 + 60 + 4 + 1 +5 + 54 = 194 samples

These samples are distributed among the three classes as

N_K = 50 + 20 + 0 = 70

N_L = 0 + 60 + 4 = 64

N_M = 1+ 5 + 54 = 60

The number of samples correctly classified is

50 + 60 + 54 = 164

The number of samples incorrectly classified as class K is

0 + 1 =1

The number of samples incorrectly classified as class L is

20 + 5 = 25

The number of samples incorrectly classified as class M is

0 + 4 = 4

For class K, we have

TP_K = 50

TN_K = 60 + 4 + 5 + 54 = 123

FP_K = 0 + 1 = 1

FN_K = 20 + 0 = 20

For class L, we have

TP_L = 60

TN_L = 50 + 0 + 1 + 54 = 105

FP_L = 20 + 5 = 25

FN_L = 0 + 4 = 4

For class M, we have

TP_M = 54

TN_M = 50 + 20 + 0 + 60 = 130

FP_M = 0 + 4 = 4

FN_M = 1 + 5 = 6

The building blocks for the three classes K, L, and M are summarized in Fig. 8. The corresponding three 2X2 confusion matrices are shown in Fig. 9.

	TP	TN	FP	FN
Class K	50	123	1	20
Class L	60	105	25	4
Class M	54	130	4	6

Fig. 8. Values of building blocks for Example 1

Actual \ Predicted	K	Not K
K	50	20
Not K	1	123

Actual \ Predicted	L	Not L
L	60	4
Not L	25	105

Actual \ Predicted	M	Not M
M	54	6
Not M	4	130

Fig. 9. Three 2X2 confusion matrices for Example 1

Example 2

For the 3-class classification problem of example 1, assume the confusion of Fig. 7 is rearranged as shown in

(a) Fig. 10.a (b) Fig. 10.b

Verify that the form of Fig. 10.a is valid (equivalent to that of Fig. 7) while the form in Fig. 10.b is not, and justify your answer. Check that the building blocks for the three classes as obtained from Fig. 10.a are the same as those obtained in example 1.

Actual \ Predicted	K	M	L
K	50	0	20
M	1	54	5
L	0	4	60

(a)

Actual \ Predicted	K	M	L
K	50	20	0
M	1	5	54
L	0	60	4

(b)

Fig. 10. Rearrangement of Fig. 7 in Example 1

Solution

The form of Fig. 10.a is valid because it conveys the same information as Fig. 7; rows L and M interchanged with interchanging columns L and M. The building blocks for the three classes are calculated from Fig. 10.a as:

For class K, we have

TP_K = 50

TN_K = 54 + 5 + 4 + 60 = 123

FP_K = 1 + 0 = 1

FN_K = 0 + 20 = 20

For class L, we have

TP_L = 60

TN_L = 50 + 0 + 1 + 54 = 105

FP_L = 20 + 5 = 25

FN_L = 0 + 4 = 4

For class M, we have

TP_M = 54

TN_M = 50 + 20 + 0 + 60 = 130

FP_M = 0 + 4 = 4

FN_M = 1 + 5 = 6

These values of the building blocks are seen to be identical to the corresponding values obtained in Example 1. The matrix form of Fig. 10b, however, does not convey the same information as Fig. 7; rows L and are interchanged but columns L and M are not. For example, the number of class_M samples correctly classified is 5 in Fig. 10b while it is 54 in Fig. 7, and the number of class_L samples incorrectly classified as class M is 60 in Fig. 10b while it is 4 in Fig. 7. Therefore, Fig. 10b is not equivalent to Fig. 7.

3. PERFORMANCE MEASURES FOR THREE-CLASS CLASSIFICATION

In our previous work [7], we discussed in some detail a

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

group of performance measures for binary classification models based on the relevant confusion matrix. Such measures are here extended as 3-class classifiers. The definitions and their implications, as will be evident, are the same in principle, taking into account that the confusion matrix is of order 3X3 and the building blocks for the three classifiers are determined as explained in Section 2.

For classes A, B, and C, the confusion matrix takes the general form of Fig. 11, with the three component 2X2 confusion matrices of Fig. 12. The symbol E_{AB} represents class A samples incorrectly classified as class B, and so on.

		Predicted A	Predicted B	Predicted C	
Actual	A	TP_A	E_{AB}	E_{AC}	N_A
	B	E_{BA}	TP_B	E_{BC}	N_B
	C	E_{CA}	E_{CB}	TP_C	N_C

Fig. 11: General form of 3x3 confusion matrix

		Predicted A	Predicted Not A
Actual	A	TP_A	$FN_A = E_{AB} + E_{AC}$
	Not A	$FP_A = E_{BA} + E_{CA}$	$TN_A = TP_B + E_{BC} + E_{CB} + TP_C$

(a) Class A

		Predicted B	Predicted Not B
Actual	B	TP_B	$FN_B = E_{BA} + E_{BC}$
	Not B	$FP_B = E_{AB} + E_{CB}$	$TN_B = TP_A + E_{AC} + E_{CA} + TP_C$

(b) Class B

		Predicted C	Predicted Not C
Actual	C	TP_C	$FN_C = E_{CA} + E_{CB}$
	Not C	$FP_C = E_{AC} + E_{BC}$	$TN_C = TP_A + E_{AB} + E_{BA} + TP_B$

(a) Class C

Fig. 12: Three 2X2 confusion matrices for 3X3 confusion matrix of Fig. 11.

The total number of tested samples is

$$N_A + N_B + N_C = P_A + P_B + P_C = N \tag{1}$$

The class actual samples are

$$N_A = TP_A + E_{AB} + E_{AC} = TP_A + FN_A \tag{2a}$$
$$N_B = TP_B + E_{BA} + E_{BC} = TP_B + FN_B \tag{2b}$$
$$N_C = TP_C + E_{CA} + E_{CB} = TP_C + FN_C \tag{2c}$$

The class predicted samples are

$$P_A = TP_A + E_{BA} + E_{CA} = TP_A + FP_A \tag{3a}$$
$$P_B = TP_B + E_{AB} + E_{CB} = TP_B + FP_B \tag{3b}$$
$$P_C = TP_C + E_{AC} + E_{BC} = TP_C + FP_C \tag{3c}$$

The first measure to consider is the accuracy of the classification model as a whole. It is defined as the ratio of the number of correctly classified samples to the total number of tested samples. Referring to Fig. 11, it is found that

$$\text{Accuracy} = \frac{TP_A + TP_B + TP_C}{N} \tag{4}$$

where the numerator is the sum of the three diagonal elements of the confusion matrix.

Note that the model accuracy can also be determined from the three 2x2 confusion matrices of Fig. 12, where TP_A, TP_B, and TP_C already appear at the intersection of the first row and first column in Figs. 12a, b, and c, respectively.

For example, in Fig. 1 (or Fig. 6), where $TP_A = 32$, $TP_B = 38$, $TP_C = 28$, and N = 150, we have

$$\text{Accuracy} = \frac{32 + 38 + 28}{150} = 0.653$$

Next, we consider the other performance measures.

5.1 Precision, recall, and specificity for three-class classification

The forms of the three expressions in Table 1 defining the precision, recall (sensitivity), and specificity for either class in binary classification are also applicable for the classes of 3-class classification. That is,

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

To clarify, consider the confusion matrix of Fig. 1 and the building blocks of Fig. 5. The class precisions are

$$\text{Precision}_A = \frac{TP_A}{TP_A + FP_A} = \frac{TP_A}{P_A} = \frac{32}{32+21} = 0.604$$

$$\text{Precision}_B = \frac{TP_B}{TP_B + FP_B} = \frac{TP_B}{P_B} = \frac{38}{38+19} = 0.667$$

$$\text{Precision}_C = \frac{TP_C}{TP_C + FP_C} = \frac{TP_C}{P_C} = \frac{28}{28+12} = 0.7$$

The class recalls are

$$\text{Recall}_A = \frac{TP_A}{TP_A + FN_A} = \frac{TP_A}{N_A} = \frac{32}{32+18} = 0.64$$

$$\text{Recall}_B = \frac{TP_B}{TP_B + FN_B} = \frac{TP_B}{N_B} = \frac{38}{38+13} = 0.745$$

$$\text{Recall}_C = \frac{TP_C}{TP_C + FN_C} = \frac{TP_C}{N_C} = \frac{28}{28+21} = 0.571$$

The class specificities are

$$\text{Specificity}_A = \frac{TN_A}{TN_A + FP_A} = \frac{TN_A}{N_B + N_C} = \frac{79}{79+21} = 0.79$$

$$\text{Specificity}_B = \frac{TN_B}{TN_B + FP_B} = \frac{TN_B}{N_A + N_C} = \frac{80}{80+19} = 0.808$$

$$\text{Specificity}_C = \frac{TN_C}{TN_C + FP_C} = \frac{TN_C}{N_A + N_B} = \frac{89}{89+12} = 0.88$$

Figure 13 summarizes these results.

	Precision	Recall	Specificity
Class A	0.604	0.64	0.79
Class B	0.667	0.745	0.808
Class C	0.7	0.571	0.881

Fig. 13: Precision, recall (sensitivity), and specificity for classes in 3-class classification of Fig. 1

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

A glance at Fig. 6 reveals that the above results can be obtained from the three 2X2 confusion matrices of classes A, B, and C as is done in binary classification.

In terms of the notation of Fig. 11, Fig. 12 indicates that class A has

$$\text{Precision}_A = \frac{\text{TP}_A}{\text{TP}_A + \text{E}_{BA} + \text{E}_{CA}} \quad (5)$$

$$\text{Recall}_A = \frac{\text{TP}_A}{\text{TP}_A + \text{E}_{AB} + \text{E}_{AC}} \quad (6)$$

$$\text{Specificity}_A = \frac{\text{TN}_A}{\text{TN}_A + \text{E}_{BA} + \text{E}_{CA}} \quad (7)$$

Expressions similar to (8), (9), and (10) are written for classes B and C.

Example 3

In Example 1, determine the model accuracy as well as the precision recall, and specificity for classes K, L, and M.

Solution

From the confusion matrix of Fig. 7 and Eq. (4),

$$\text{Accuracy} = \frac{50+60+54}{194} = 0.845$$

Using the values of the building blocks of Fig. 8 and Eqs. (5), (6), and (7), we obtain:

For class K,

$$\text{Precision}_K = \frac{50}{50+1} = 0.98$$

$$\text{Recall}_K = \frac{50}{50+20} = 0.714$$

$$\text{Specificity}_K = \frac{123}{123+1} = 0.992$$

For class L,

$$\text{Precision}_L = \frac{60}{60+25} = 0.706$$

$$\text{Recall}_L = \frac{60}{60+25} = 0.938$$

$$\text{Specificity}_L = \frac{105}{105+25} = 0.808$$

For class M,

$$\text{Precision}_M = \frac{54}{54+4} = 0.931$$

$$\text{Recall}_M = \frac{54}{54+6} = 0.9$$

$$\text{Specificity}_M = \frac{130}{130+4} = 0.97$$

Figure 14 summarizes these results.

	Precision	Recall	Specificity
Class K	0.98	0.714	0.992
Class L	0.706	0.938	0.808
Class M	0.931	0.9	0.97

Fig. 14: Precision, recall, and specificity for classes in example 3

The same values for precision, recall, and specificity are obtained from the three 2x2 confusion matrices of Fig. 9. Having formulated the precision, recall, and specificity of each individual class, it is required to define such measures for the 3-class classification model as a whole. To this end, we perform averaging operations (as we did in binary classification). We have three sorts of averaging; namely, macro-, micro-, and weighted-averages.

For a model with classes A, B, and C, the macro-averages for precision, recall, and specificity are

$$\text{Precision}_{macro} = \frac{\text{Precision}_A + \text{Precision}_B + \text{Precision}_C}{3} \quad (11)$$

$$\text{Recall}_{macro} = \frac{\text{Recall}_A + \text{Recall}_B + \text{Recall}_C}{3} \quad (12)$$

$$\text{Specificity}_{macro} = \frac{\text{Specificity}_A + \text{Specificity}_B + \text{Specificity}_C}{3} \quad (13)$$

The micro-averages are:

For precision,

$$\text{Precision}_{micro} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{\text{TP}_A + \text{TP}_B + \text{TP}_C + \text{FP}_A + \text{FP}_B + \text{FP}_C} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{\text{N}} \quad (14)$$

For recall,

$$\text{Recall}_{micro} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{\text{TP}_A + \text{TP}_B + \text{TP}_C + \text{FN}_A + \text{FN}_B + \text{FN}_C} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{\text{N}} \quad (15)$$

The denominator is N in Eqs. (14) and (15) because Figs. 2, 3, and 4 show that

$$\text{FP}_A + \text{FP}_B + \text{FP}_C = \text{FN}_A + \text{FN}_B + \text{FN}_C = \text{N} - (\text{TP}_A + \text{TP}_B + \text{TP}_C) \quad (16)$$

Note that Eq. (16) implies that the sum of false positives is equal to the sum of false negatives of. Eq. (8). Comparing Eqs. (4), (14), and (15), we find that

$$\text{Accuracy} = \text{Precision}_{micro} = \text{Recall}_{micro} \quad (17)$$

For specificity,

$$\text{Specificity}_{micro} = \frac{\text{TN}_A + \text{TN}_B + \text{TN}_C}{\text{TN}_A + \text{TN}_B + \text{TN}_C + \text{FP}_A + \text{FP}_B + \text{FP}_C} = \frac{\text{TN}_A + \text{TN}_B + \text{TN}_C}{2\text{N}} = \frac{\text{N} + \text{TP}_A + \text{TP}_B + \text{TP}_C}{2\text{N}} \quad (18)$$

The forms of the numerators and denominators in (18) are valid since, by (16),

$$\begin{aligned}\text{Numerator} &= \text{TN}_A + \text{TN}_B + \text{TN}_C = \text{N} - (\text{TP}_A + \text{FP}_A + \text{FN}_A) + N - (\text{TP}_B + \text{FP}_B + \text{FN}_B) + N - (\text{TP}_C + \text{FP}_C + \text{FN}_C) \\ &= 3\text{N} - (\text{TP}_A + \text{TP}_B + \text{TP}_C) - (\text{FP}_A + \text{FP}_B + \text{FP}_C) - (\text{FN}_A + \text{FN}_B + \text{FN}_C) \\ &= 3\text{N} - (\text{TP}_A + \text{TP}_B + \text{TP}_C) - N + (\text{TP}_A + \text{TP}_B + \text{TP}_C) - N + (\text{TP}_A + \text{TP}_B + \text{TP}_C) \\ &= \text{N} + \text{TP}_A + \text{TP}_B + \text{TP}_C\end{aligned}$$

and $\text{Denominator} = \text{TN}_A + \text{TN}_B + \text{TN}_C + (\text{FP}_A + \text{FP}_B + \text{FP}_C) = \text{N} + (\text{TP}_A + \text{TP}_B + \text{TP}_C) + N - (\text{TP}_A + \text{TP}_B + \text{TP}_C) = 2\text{N}$

Equation (18), by Eq. (14), is written as

$$\text{Specificity}_{\text{micro}} = \frac{1}{2} + \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{2\text{N}} = \frac{1}{2}(1 + \text{Precision}_{\text{micro}}) \quad (19a)$$

Or

$$\text{Precision}_{\text{micro}} = 2 * \text{Specificity}_{\text{micro}} - 1 \quad (19b)$$

Note that, unlike binary classification, $\text{Specificity}_{\text{micro}}$, is not equal to $\text{Precision}_{\text{micro}}$; that is, $\text{Specificity}_{\text{micro}}$ is not eligible to be included in (17).

The weighted -averages are

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

$$\text{Precision}_{\text{weighted}} = \frac{N_A * (\text{Precision}_A) + N_B * (\text{Precision}_B) + N_C * (\text{Precision}_C)}{N} = \frac{\frac{N_A}{P_A}*(TP_A)+\frac{N_B}{P_B}*(TP_B)+\frac{N_C}{P_C}*(TP_C)}{N} \quad (19)$$

$$\text{Recall}_{\text{weighted}} = \frac{N_A * (\text{Recall}_A) + N_B * (\text{Recall}_B) + N_C * (\text{Recall}_C)}{N} = \frac{TP_A+TP_B+TP_C}{N} \quad (20)$$

$$\text{Specificity}_{\text{weighted}} = \frac{N_A * (\text{Specificity}_A) + N_B * (\text{Specificity}_B) + N_C * (\text{Specificity}_C)}{N} = \frac{\frac{N_A}{N_B+N_C}*(TN_A)+\frac{N_B}{N_A+N_C}*(TN_B)+\frac{N_C}{N_A+N_B}*(TN_C)}{N} \quad (21)$$

The form of Eq. (20) indicates that $\text{Recall}_{\text{weighted}}$ readily be added to (17) so that we have:

$$\text{Accuracy} = \text{Precision}_{micro} = \text{Recall}_{micro} = \text{Recall}_{\text{weighted}} \quad (22)$$

For example, consider Fig. 1 along with Figs. 5 and 13. From Eqs. (11), (12), and (13),

$$\text{Precision}_{macro} = \frac{0.604 + 0.667 + 0.7}{3} = 0.657$$

$$\text{Recall}_{macro} = \frac{0.64 + 0.745 + 0.571}{3} = 0.652$$

$$\text{Specificity}_{macro} = \frac{0.79 + 0.808 + 0.881}{3} = 0.826$$

From Eqs. (14), (17) , and (19a),

$$\text{Precision}_{micro} = \frac{32 + 38 + 28}{150} = 0.653$$

$$\text{Recall}_{micro} = \text{Precision}_{micro} = 0.653$$

$$\text{Specificity}_{micro} = \frac{1}{2}(1 + \text{Precision}_{micro}) = \frac{1}{2}(1 + 0.653) = 0.827$$

From Eqs. (19), (21), and (22),

$$\text{Precision}_{weighted} = \frac{(50 * 0.604) + (51 * 0.667) + (49 * 0.7)}{150} = 0.657$$

$$\text{Recall}_{weighted} = \text{Recall}_{micro} = 0.653$$

$$\text{Specificity}_{weighted} = \frac{(50 * 0.79) + (51 * 0.808) + (49 * 0.811)}{150} = 0.826$$

Fig. 15 summarizes these results.

	Precision	Recall	Specificity
Macro-average	0.657	0.652	0.826
Micro-average	0.653	0.653	0.827
Weighted-average	0.657	0.653	0.826

Fig. 15: Average values of precision recall, and specificity for classifier of Fig. 1

Example 4

A three-class classifier with classes K, L, and M has the confusion matrix shown in Fig. 16. Use the macro-, micro-, and weighted-averaging to determine the classifier precisions, recalls, and specificities.

		Predicted K	Predicted M	Predicted L
Actual	K	10	2	4
Actual	M	6	11	7
Actual	L	22	13	266

Fig. 16: Confusion matrix for Example 4

Solution

The building blocks for classes K,L, and M are
For class K, we have

$TP_K = 10$
$TN_K = 11 + 7 + 13 + 266 = 297$
$FP_K = 6 + 22 = 28$
$FN_K = 2 + 4 = 6$

For class L, we have

$TP_L = 11$
$TN_L = 10 + 4 + 22 + 266 = 302$
$FP_L = 2 + 13 = 15$
$FN_L = 6 + 7 = 13$

For class M, we have

$TP_M = 266$
$TN_M = 10 + 2 + 6 + 11 = 29$
$FP_M = 4 + 7 = 11$
$FN_M = 22 + 13 = 35$

See Fig. 17.

	TP	TN	FP	FN
Class K	10	297	28	6
Class L	11	302	15	13
Class M	266	29	11	35

Fig. 17: Building blocks for classes in Example 4

The precisions, recalls, and. specificities of classes K, L, and M are

$$\text{Precision}_K = \frac{10}{10 + 28} = 0.263$$

$$\text{Recall}_K = \frac{10}{10 + 6} = 0.625$$

$$\text{Specificity}_K = \frac{297}{297 + 28} = 0.914$$

$$\text{Precision}_L = \frac{11}{11 + 15} = 0.423$$

$$\text{Recall}_L = \frac{11}{11 + 13} = 0.458$$

$$\text{Specificity}_L = \frac{302}{302 + 15} = 0.953$$

$$\text{Precision}_M = \frac{266}{266 + 11} = 0.96$$

$$\text{Recall}_M = \frac{266}{266 + 35} = 0..884$$

$$\text{Specificity}_M = \frac{29}{29 + 11} = 0.725$$

See Fig. 18.

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

	Precision	Recall	Specificity
Class K	0.263	0.625	0.914
Class L	0.423	0.458	0.953
Class M	0.96	0.884	0.725

Fig. 18: Precision, recall, and specificity for classes in Example 4

Using Eqs. (11), (12), and (13), the macro-averages are

$$\text{Precision}_{macro} = \frac{0.263 + 0.423 + 0.96}{3} = 0.549$$

$$\text{Recall}_{macro} = \frac{0.625 + 0.458 + 0.884}{3} = 0.656$$

$$\text{Specificity}_{macro} = \frac{0.914 + 0.953 + 0.725}{3} = 0.864$$

From the confusion matrix of Fig. 16,

$N_K = 10 + 2 + 4 = 16$
$N_L = 6 + 11 + 7 = 24$
$N_M = 22 + 13 + 266 = 301$
$N = 16 + 24 + 301 = 341$

Using Eqs. (14), (19a), and (22), the micro—averages are

$$\text{Precision}_{micro} = \frac{10 + 11 + 266}{341} = 0.842$$

$$\text{Recall}_{micro} = \text{Precision}_{micro} = 0.842$$

$$\text{Specificity}_{micro} = \frac{1}{2} + (1 + \text{Precision}_{micro}) = \frac{1}{2} + (1 + 0.842) = 0.921$$

Using Eqs. (19), (21), and (22), the weighted- averages are

$$\text{Precision}_{weighted} = \frac{(16 * 0.263) + (24 * 0.423) + (301 * 0.96)}{341} = 0.89$$

$$\text{Recall}_{weighted} = \text{Recall}_{micro} = 0.842$$

$$\text{Specificity}_{weighted} = \frac{(16 * 0.914) + (24 * 0.953) + (301 * 0.725)}{341} = 0.75$$

See Fig. 19.

	Precision	Recall	Specificity
Macro-average	0.549	0.656	0.864
Micro-average	0.842	0.842	0.921
Weighted-average	0.89	0.842	0.75

Fig. 19: Average values of precision, recall, and specificity for classifier in Example 4

Example 5

A three-class classifier with classes K, L, and M receives a group of tested samples, where

$N_K = 53$, $N_L = 110$, $N_M = 37$

The true positives and true negatives for the classes are found to be

$TP_K = 30$, $TP_L = 52$, $TP_M = 22$, $TN_K = 38$, $TN_L = 73$, $TN_M = 133$

Determine the macro-, micro-, anal weighted-average precisions, recalls, and specificities of the classifier.

Solution

The total number of tested samples is

$N = 53 + 110 + 37 = 200$

The 3x3 confusion matrix is decomposed into three 2x2 confusion matrices as shown in Fig. 20.

Predicted

Actual		K	Not K	
	K	$TP_K = 30$	FN_K	$N_K = 53$
	Not K	FP_K	$TN_K = 98$	$N_L + N_M = 147$

(a) Class K (N = 200)

Predicted

Actual		L	Not L	
	L	$TP_L = 52$	FN_L	$N_L = 110$
	Not L	FP_L	$TN_L = 73$	$N_K + N_M = 90$

(b) Class L (N= 200)

Predicted

Actual		M	Not M	
	M	$TP_M = 22$	FN_M	$N_M = 37$
	Not M	FP_M	$TN_M = 133$	$N_K + N_L = 163$

(c) Class M (N= 200)

Fig. 20: Three 2x2 confusion matrices in Example 5

We have

$FN_K = 53 – 30 = 23$
$FN_L = 110 – 52 = 58$
$FN_M = 37 – 22 = 15$

and

$FP_K = 200 - 53 – 98 = 49$
$FP_L = 200 - 110 – 73 = 17$
$FP_M = 200 - 37 – 133 = 30$

The precisions, recalls, and specificities of classes K, L, and M are obtained from Fig. 20. From Fig. 20a,

$$\text{Precision}_K = \frac{30}{30 + 49} = 0.38$$

$$\text{Recall}_K = \frac{30}{30 + 23} = 0.566$$

$$\text{Specificity}_K = \frac{98}{98 + 49} = 0.667$$

From Fig. 20b,

$$\text{Precision}_L = \frac{52}{52 + 17} = 0.754$$

$$\text{Recall}_L = \frac{52}{52 + 58} = 0.473$$

$$\text{Specificity}_L = \frac{73}{73 + 17} = 0.811$$

From Fig. 20c,

$$\text{Precision}_M = \frac{22}{22 + 30} = 0.423$$

$$\text{Recall}_M = \frac{22}{22 + 15} = 0.595$$

$$\text{Specificity}_M = \frac{133}{133 + 30} = 0.816$$

See Fig. 21.

The macro-averages are

$$\text{Precision}_{macro} = \frac{0.38 + 0.754 + 0.423}{3} = 0.519$$

$$\text{Recall}_{macro} = \frac{0.566 + 0.473 + 0.595}{3} = 0.545$$

$$\text{Specificity}_{macro} = \frac{0.667 + 0.811 + 0.816}{3} = 0.765$$

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

	Precision	Recall	Specificity
Class K	0.38	0.566	0.667
Class L	0.754	0.473	0.811
Class M	0.423	0.595	0.816

Fig. 21: Precision, recall, and specificity for classes in Example 5

The micro-averages are

$$\text{Precision}_{micro} = \frac{\text{TP}_\text{K} + \text{TP}_\text{L} + \text{TP}_\text{M}}{\text{N}} = \frac{30 + 52 + 22}{200} = 0.52$$

$$\text{Recall}_{micro} = \text{Precision}_{micro} = 0.52$$

$$\text{Specificity}_{micro} = \frac{1}{2} + (1 + \text{Precision}_{micro}) = \frac{1}{2} + (1 + 0.52)$$
$$= 0.76$$

The weighted- averages are

$$\text{Precision}_{weighted} = \frac{(53 * 0.38) + (110 * 0.754) + (37 * 0.423)}{200}$$
$$= 0.594$$

$$\text{Recall}_{weighted} = \text{Recall}_{micro} = 0.52$$

$$\text{Specificity}_{weighted} = \frac{(53 * 0.667) + (110 * 0.811) + (37 * 0.816)}{200}$$
$$= 0.774$$

See Fig. 22.

	Precision	Recall	Specificity
Macro-average	0.519	0.545	0.765
Micro-average	0.52	0.52	0.76
Weighted-average	0.594	0.52	0.774

Fig. 22: Average values of precision, recall, and Specificity for classifier in Example 5

As in binary classification, recall and specificity in three-class classification are referred to as TPR (true positive rate) and TNR (true negative rate), respectively, and FNR (false negative rate) is 1-TPR and FPR (false positive rate) is 1-TNR.

5.2 Average accuracy and balanced accuracy for three-class classification

When the datasets of the three classes are imbalanced, the accuracy of the classification model defined in Eq. (4) can be misleading, as we mentioned in [7] for binary classification. In this respect, two measures are defined: the average accuracy and the balanced accuracy.

The average accuracy is the arithmetic average of the accuracies of the three classes. That is, for classes A, B, and C,

$$\text{Average accuracy} = \frac{\text{Accuracy}_\text{A} + \text{Accuracy}_\text{B} + \text{Accuracy}_\text{C}}{3} = \frac{1}{3}\left(\frac{\text{TP}_\text{A}+\text{TN}_\text{A}}{\text{N}} + \frac{\text{TP}_\text{B}+\text{TN}_\text{B}}{\text{N}} + \frac{\text{TP}_\text{C}+\text{TN}_\text{C}}{\text{N}}\right)$$

$$= \frac{\text{TP}_\text{A}+\text{TN}_\text{A}+\text{TP}_\text{B}+\text{TN}_\text{B}+\text{TP}_\text{C}+\text{TN}_\text{C}}{3\text{N}} \quad (23)$$

where the numerator is the sum of true positives plus the sum of true negatives, and the denominator is 3 times the total number of tested samples.

		Predicted A	Predicted B	Predicted C	
Actual	A	4	0	1	$N_A = 5$
	B	10	65	16	$N_B = 91$
	C	1	1	9	$N_C = 11$

Fig. 23: Confusion matrix with imbalanced datasets

To illustrate, consider the 3x3 confusion matrix of Fig. 23. The datasets of the classes are seen to be imbalanced since the number of samples of class B, N_B = 91, is much greater than that of class A or class C.
The total number of tested samples is

$$N = N_A + N_B + N_C = 5 + 91 + 11 = 107$$

The classifier accuracy is

$$\text{Accuracy} = \frac{4 + 65 + 9}{3} = 0.729$$

The true positives and true negatives are

$$\text{TP}_\text{A} = 4, \text{TN}_\text{A} = 91, \text{TP}_\text{B} = 65, \text{TN}_\text{B} = 15, \text{TP}_\text{C} = 9, \text{TN}_\text{C} = 79$$

By Eq. (23),

$$\text{Average accuracy} = \frac{4 + 65 + 9 + 91 + 15 + 79}{3 * 107} = 0.819$$

The average accuracy can also be thought of in terms of the three component 2X2 confusion matrices shown in Fig. 24, from which we have,

$$\text{Accuracy}_\text{A} = \frac{4 + 91}{107} = 0.888$$
$$\text{Accuracy}_\text{B} = \frac{65 + 15}{107} = 0.748$$
$$\text{Accuracy}_\text{C} = \frac{9 + 79}{107} = 0.822$$

which is of course the same value calculated from Fig. 23.

		Predicted A	Predicted Not A	
Actual	A	4	1	$N_A = 5$
	Not A	11	91	$N_B + N_C = 102$

(a) Class A (N = 107)

		Predicted B	Predicted Not B	
Actual	B	65	26	$N_B = 91$
	Not B	1	15	$N_A + N_C = 16$

(a) Class B (N = 107)

		Predicted C	Predicted Not C	
Actual	C	9	2	$N_C = 11$
	Not C	17	79	$N_A + N_B = 96$

(a) Class C (N = 107)

Fig. 24: Three 2x2 confusion matrices for confusion matrix of Fig. 23

It's worth noting that the average accuracy for binary classification is nothing but the accuracy of the classifier. The balanced accuracy, from a rather different perspective, is the arithmetic average of recalls of the three classes. This means

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

that this measure is the macro-average recall of the model, Eq. (2). We thus write, for classes A, B, and C,

$$\text{Balanced accuracy} = \text{Recall}_{\text{macro}} = \frac{\text{Recall}_{\text{A}} + \text{Recall}_{\text{B}} + \text{Recall}_{\text{C}}}{3} \quad (24)$$

Note that definition (24) guarantees that all of the three rows of the confusion matrix take pare in calculating the balanced accuracy, and the three classes are regarded to be of the same weight (unity) and same importance. Remember that, in binary classification, Eq. (47), the balanced accuracy is likewise equal to $\text{Recall}_{\text{macro}}$, the arithmetic average of recalls of the two classes.
Referring to Fig. 23, recalls of the classes are

$$\text{Recall}_{\text{A}} = \frac{4}{5} = 0.8$$

$$\text{Recall}_{\text{B}} = \frac{65}{91} = 0.714$$

$$\text{Recall}_{\text{C}} = \frac{9}{11} = 0.818$$

By Eq. (24),

$$\text{Balanced accuracy} = \frac{0.8 + 0.714 + 0.818}{3} = 0.777$$

The balanced accuracy, like the average accuracy, can readily be derived from the component 2x2 confusion matrices of Fig. 24, where

$$\text{Recall}_{\text{A}} = \frac{4}{5}\ , \quad \text{Recall}_{\text{B}} = \frac{65}{91}, \quad \text{Recall}_{\text{C}} = \frac{9}{11}$$

which are, as is well expected, the same results obtained from Fig. 23.

We remark that the values of accuracy, average accuracy, and balanced accuracy will be very close fs each other when the datasets are balanced.

5.3 Fβ measure and F1 score for three-class classification

The forms of F_β measure and F1 score for three-class classification are not different from those formulated for binary classification. But the pertinent building blocks are calculated on the basis of a 3x3 confusion matrix; see Figs. 2, 3, and 4.

For each of the three classes A, B, and C, the F_β measure, where 0<β<1, is the weighted harmonic average of precision and recall. The Fil score, on the other hand, is the harmonic average of precision and recall; it is in essence the same as F_β when β =0.5. For class A, similar to Eqs. (54) and (55), we have

$$F_{\beta A} = \frac{\text{Precision}_A \times \text{Recall}_A}{\beta(\text{Recall}_A) + (1-\beta)\text{Precision}_A} = \frac{\text{TP}_A}{\text{TP}_A + \beta(\text{FP}_A) + (1-\beta)\text{FN}_A} \quad (25)$$

$$\text{F1}_A = \frac{2\times\text{Precision}_A \times\text{Recall}_A}{\text{Precision}_A + \text{Recall}_A} = \frac{\text{TP}_A}{\text{TP}_A + 0.5(\text{FP}_A + \text{FN}_A)} \quad (26)$$

Similar expressions are written for classes B and C. Again as in binary classification, Eqs. (56) and (57), the 3-class classification model has

$$F_{\beta\text{model}} = \frac{\text{Precision}_{\text{model}} \times \text{Recall}_{\text{model}}}{\beta(\text{Recall}_{\text{model}}) + (1-\beta)\text{Precision}_{\text{model}}} \quad (27)$$

$$\text{F1}_{\text{model}} = \frac{2\times\text{Precision}_{\text{model}}\times\text{Recall}_{\text{model}}}{\text{Precision}_{\text{model}} + \text{Recall}_{\text{model}}} \quad (28)$$

Equations (27) and (28) express the macro-, micro-, or weighted-average F_β and F1 of the model, respectively, where $\text{Precision}_{\text{model}}$ denotes, correspondingly, the model macro-, micro-, or weighted-average precision, and $\text{Recall}_{\text{model}}$ is similarly interpreted. Since, by Eq. (17), $\text{Recall}_{\text{model}}$ = $\text{Precision}_{\text{model}}$, Eqs. (27) and (28) imply that Eq. (58)- is also here valid. We can write

$$F_{\beta\text{micro}} = \text{F1}_{\text{micro}} = \text{Precision}_{\text{ micro}} \quad (29)$$

It turns out that

$$\text{Accuracy} = \text{Precision}_{\text{ micro}} = \text{Recall}_{\text{ micro}} = \text{Recall}_{\text{ weighted}} = F_{\beta\text{micro}} = \text{F1}_{\text{micro}} = \frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{\text{N}} \quad (30)$$

Six measures are thus defined by one and the same expression, $\frac{\text{TP}_A + \text{TP}_B + \text{TP}_C}{\text{N}}$. Note particularly that, in comparison with Eq. (60), the term $\text{Specificity}_{\text{ micro}}$ is absent in Eq. (30).

To illustrate, consider the 3X3 confusion matrix of Fig. 23. The necessary calculations are carried out for determining the class building blacks, class precisions and recalls, and model average precisions and recalls, and the results are listed in Figs. 25, 26, and 27, respectively.

	TP	TN	FP	FN
Class A	4	91	11	1
Class B	65	15	1	26
Class C	9	79	17	2

Fig. 25: Building blocks for classes with confusion matrix of Fig. 23

	Precision	Recall
Class A	0.267	0.8
Class B	0.985	0.714
Class C	0.346	0.818

Fig. 26: Precision and recall for classes with confusion matrix of Fig. 23

It follows that

$$F_{\beta macro}(\beta = 0.8 \text{ for example}) = \frac{\text{Precision}_{\text{macro}} \times \text{Recall}_{\text{macro}}}{\beta(\text{Recall}_{\text{macro}}) + (1-\beta)\text{Precision}_{\text{macro}}} = \frac{0.533 \times 0.777}{(0.8 \times 0.777) + (0.2 \times 0.533)} = 0.569$$

	Precision	Recall
Macro-average	0.533	0.777
Micro-average	0.729	0.729
Weighted-average	0.886	0.729

Fig. 27: Average values of precision and recall for model with confusion matrix of Fig. 23

$$\text{F1}_{\text{macro}} = \frac{2 \times \text{Precision}_{\text{macro}} \times \text{Recall}_{\text{macro}}}{\text{Precision}_{\text{macro}} + \text{Recall}_{\text{macro}}} = \frac{2 \times 0.533 \times 0.777}{0.533 + 0.777} = 0.632$$

$$F_{\beta\text{micro}} = \text{Precision}_{\text{ micro}} = 0.729 \quad (\text{independent of } \beta)$$

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

$$F1_{micro} = F_{\beta micro} = 0.729$$

$$F_{\beta weighted}(\beta = 0.8) = \frac{\mathrm{Precision}_{weighted} \times \mathrm{Recall}_{weighted}}{\beta(\mathrm{Recall}_{weighted}) + (1-\beta)\mathrm{Precision}_{weighted}}$$

$$= \frac{0.886 \times 0.729}{(0.8 \times 0.729) + (0.2 \times 0.886)} = 0.849$$

$$F1_{weighted} = \frac{2 \times \mathrm{Precision}_{weighted} \times \mathrm{Recall}_{weighted}}{\mathrm{Precision}_{weighted} + \mathrm{Recall}_{weighted}}$$

$$= \frac{2 \times 0.886 \times 0.729}{0.886 + 0.729} = 0.8$$

Figure 28 summarizes the average values of F_β and F1 for the model.

	F_β	F1
Macro-average	0.569	0.632
Micro-average	0.729*	0.729
Weighted-average	0.849	0.8

Fig. 28: Average values of Fβ and F1 for model with confusion matrix of Fig. 23

If we are interested in the values of F_β and F1 or each of the individual classes, then Eqs. (25) and (26) will give us

$$F_{\beta A}(\beta = 0.8) = \frac{0.267 \times 0.8}{(0.8 \times 0.8) + (0.2 \times 0.267)} = 0.308$$

$$F1_A = \frac{2 \times 0.267 \times 0.8}{0.267 + 0.8} = 0.4$$

$$F_{\beta B}(\beta = 0.8) = \frac{0.985 \times 0.714}{(0.8 \times 0.714) + (0.2 \times 0.985)} = 0.916$$

$$F1_A = \frac{2 \times 0.985 \times 0.714}{0.985 + 0.714} = 0.828$$

$$F_{\beta C}(\beta = 0.8) = \frac{0.346 \times 0.818}{(0.8 \times 0.818) + (0.2 \times 0.346)} = 0.391$$

$$F1_C = \frac{2 \times 0.346 \times 0.818}{0.346 + 0.818} = 0.486$$

5.4. Summary of results for three-class classification

The expressions of the performance measures for 3-class classification with classes A, B, and C are summarized in Table 2 for the individual classes and Table 3 for the whole classification model. In the rows of Table 2, the subscript 'class' refers to either class A, B, or C. In rows 13 and 14 of Table 3, the subscript 'model' refers to either macro-, micro-, or weighted-average. The points of similarity and dissimilarly between binary and 3-class classifications should be evident and well understood.

Acknowledgement

This paper was written in memory of Prof Dr. Mahmoud Fahmy Amin, God bless his sole. Thanks for Prof. Dr. Amany Sarhan for help valuable help in writing and editing this tutorial and incorporating the references used in this tutorial, hoping that this material can help in a deeper understanding of this important topic.

Table 2: Performance measures for individual classes A, B, and C in 3-class classification

#	Measure for class
1	$\mathrm{Accuracy}_{class} = \frac{TP_{class} + TN_{class}}{N}$
2	$\mathrm{Precision}_{class} = \frac{TP_{class}}{TP_{class} + FP_{class}}$
3	$\mathrm{Recall}_{class} = \frac{TP_{class}}{TP_{class} + FN_{class}}$
4	$\mathrm{Specificity}_{class} = \frac{TN_{class}}{TN_{class} + FP_{class}}$
5	$F_{\beta class} = \frac{\mathrm{Precision}_{class} \times \mathrm{Recall}_{class}}{\beta(\mathrm{Recall}_{class}) + (1-\beta)\mathrm{Precision}_{class}} = \frac{TP_{class}}{TP_{class} + \beta(FP_{class}) + (1-\beta)FN_{class}}$
6	$F1_{class} = \frac{2 \times \mathrm{Precision}_{class} \times \mathrm{Recall}_{class}}{\mathrm{Precision}_{class} + \mathrm{Recall}_{class}} = \frac{TP_{class}}{TP_{class} + 0.5(FP_{class} + FN_{class})}$

Table 3: Performance measures for 3-class classification with classes A, B, and C

#	Measure for class
1	$\mathrm{Accuracy} = \frac{TP_A + TP_B + TP_C}{N}$
2	$\mathrm{Precision}_{macro} = \frac{\mathrm{Precision}_A + \mathrm{Precision}_B + \mathrm{Precision}_C}{3}$
3	$\mathrm{Recall}_{macro} = \frac{\mathrm{Recall}_A + \mathrm{Recall}_B + \mathrm{Recall}_C}{3}$
4	$\mathrm{Specificity}_{macro} = \frac{\mathrm{Specificity}_A + \mathrm{Specificity}_B + \mathrm{Specificity}_C}{3}$
5	$\mathrm{Precision}_{weighted} = \frac{N_A * (\mathrm{Precision}_A) + N_B * (\mathrm{Precision}_B) + N_C * (\mathrm{Precision}_C)}{N} = \frac{\frac{N_A}{P_A}*(TP_A) + \frac{N_B}{P_B}*(TP_B) + \frac{N_C}{P_C}*(TP_C)}{N}$
6	$\mathrm{Recall}_{weighted} = \frac{N_A * (\mathrm{Recall}_A) + N_B * (\mathrm{Recall}_B) + N_C * (\mathrm{Recall}_C)}{N} = \frac{TP_A + TP_B + TP_C}{N} = \mathrm{Accuracy}$
7	$\mathrm{Specificity}_{weighted} = \frac{N_A * (\mathrm{Specificity}_A) + N_B * (\mathrm{Specificity}_B) + N_C * (\mathrm{Specificit}}{N} = \frac{\frac{N_A}{N_B+N_C}*(TN_A) + \frac{N_B}{N_A+N_C}*(TN_B) + \frac{N_C}{N_A+N_B}*(TN_C)}{N}$
8	$\mathrm{Average\ accuracy} = \frac{\mathrm{Accuracy}_A + \mathrm{Accuracy}_B + \mathrm{Accuracy}_C}{3} = \frac{TP_A + TN_A + TP_B + TN_B + TP_C + TN_C}{3N}$
9	$\mathrm{Balanced\ accuracy} = \frac{\mathrm{Recall}_A + \mathrm{Recall}_B + \mathrm{Recall}_C}{3} = \mathrm{Recall}_{macro}$
10	$F_{\beta model} = \frac{\mathrm{Precision}_{model} \times \mathrm{Recall}_{model}}{\beta(\mathrm{Recall}_{model}) + (1-\beta)\mathrm{Precision}_{model}}$
11	$F1_{model} = \frac{2 \times \mathrm{Precision}_{model} \times \mathrm{Recall}_{model}}{\mathrm{Precision}_{model} + \mathrm{Recall}_{model}}$
12	$F_{\beta micro} = F1_{micro} = \mathrm{Precision}_{micro}$
13	$\mathrm{Accuracy} = \mathrm{Precision}_{micro} = \mathrm{Recall}_{micro} = \mathrm{Recall}_{weighted} = F_{\beta micro} = F1_{micro} = \frac{TP_A + TP_B + TP_C}{N}$

DOI: 10.21608/ERJENG.2023.296718

Journal of Engineering Research (ERJ)
Vol. 7 – No. 1, 2023
©Tanta University, Faculty of Engineering
ISSN: 2356-9441 https://erjeng.journals.ekb.eg/ e ISSN: 2735-4873

REFERENCES

[1] David M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation," Journal of machine learning technologies, volume 2, issue 1, pp-37-63, 2011.

[2] Hasnae Zerouaouib and Ali Idri, "Deep hybrid architectures for binary classification of medical breast cancer images," Biomedical Signal Processing and Control," volume 71, Part B, January 2022.

[3] H. Witten, E. Frank, M. A. Hall, and C. J. Pal. Data Mining. Morgan Kaufmann, fourth edition, 2017.

[4] Kai Ming Ting, "Precision and Recall," Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA, 2011.

[5] Kanvinde, Nandan, Abhishek Gupta, and Raunak Joshi. "Binary classification for high dimensional data using supervised non-parametric ensemble method." arXiv preprint arXiv:2202.07779, 2022.

[6] W. Siblini, J. Fréry, L. He-Guelton, F. Oblé, and Y. Q. Wang, "Master Your Metrics with Calibration. In Berthold, M., Feelders, A., and G., K., editors, Advances in Intelligent Data Analysis XVIII. IDA 2020. Springer, Cham. Lecture Notes in Computer Science, vol. 12080.

[7] Mahmoud M Fahmy Amin, "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial," Volume 6, Issue 5, December 2022.

[8] DM Ibrahim, NM Elshennawy, AM Sarhan, "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," Computers in biology and medicine 132, 104348.

[9] AM Gab Allah, AM Sarhan, NM Elshennawy, "Classification of brain mri tumor images based on deep learning PGGAN augmentation," Diagnostics 11 (12), 2343, 2022.

[10] E Shalaby, N ElShennawy, A Sarhan, "Utilizing deep learning models in CSI-based human activity recognition," Neural Computing and Applications, 1-18, 2022.

[11] https://www.turing.com/kb/precision-recall-method

DOI: 10.21608/ERJENG.2023.296718