Beyond Balanced Accuracy: Balanced Matthews' Correlation Coefficient

Sébastien Guesné

Senior Chemoinformatician

sebastien.guesne@lhasalimited.org

Ninth Joint Sheffield Conference on Chemoinformatics – 20th June 2023

Overview of the presentation

- Introduction
  - Definitions and an example
  - The issue
- Take home message
- From accuracy to balanced accuracy
- From Matthew's Correlation Coefficient (MCC) to balanced MCC
- Use cases
- Tale of two MCCs
- Final words

Lhasa Limited

Definitions

Confusion matrix

Actual \ Predicted	+	-
+	TP	FN
-	FP	TN

Positive prevalence

$$Pre^{+} = \frac{TP + FN}{N}$$

Negative prevalence

$$Pre^{-} = \frac{FP + TN}{N}$$

Total number of instances

$$N = TP + FP + TN + FN$$

Sensitivity

$$Sen = \frac{TP}{TP + FN}$$

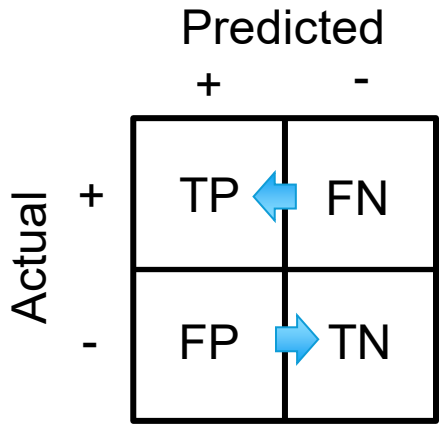
Specificity

$$Spe = \frac{TN}{TN + FP}$$

Lhasa Limited

Definitions

Model performance against test sets improves
if and only if
the variation of sensitivity and the variation of specificity is greater than zero



$$\Delta Sen \; and \; \Delta Spe > 0$$

Lhasa Limited

An example

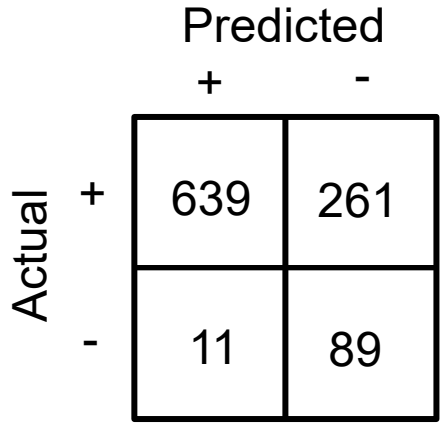
Internal test set			
	Predicted +	Predicted -
Actual +	639	261
Actual -	11	89

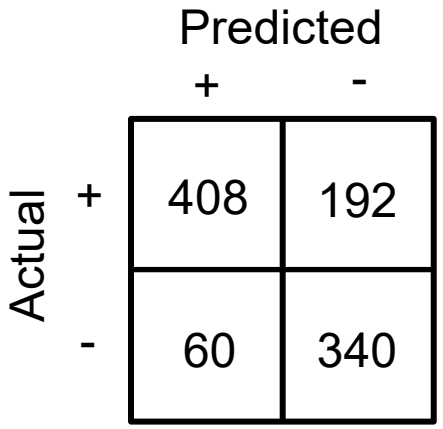
External test set			
	Predicted +	Predicted -
Actual +	612	288
Actual -	15	85

	Internal test set	External test set	Variation
Pos. Prevalence	0.900	0.900	
Sensitivity	0.710	0.680	Decrease
Specificity	0.890	0.850	Decrease
MCC	0.377	0.329	**Decrease**
Accuracy	0.728	0.697	**Decrease**
Bal. accuracy	0.800	0.765	Decrease
Pos. Predictivity	0.983	0.976	Decrease
Neg. Predictivity	0.254	0.228	Decrease
Kappa	0.284	0.239	Decrease

Lhasa Limited

The issue





	Internal test set	External test set	Variation
Pos. Prevalence	0.900	0.600	
Sensitivity	0.710	0.680	Decrease
Specificity	0.890	0.850	Decrease

Lhasa Limited

The issue

Internal test set		Predicted +	Predicted -
Actual	+	639	261
	-	11	89

External test set		Predicted +	Predicted -
Actual	+	408	192
	-	60	340

	Internal test set	External test set	Variation
Pos. Prevalence	0.900	0.600	
Sensitivity	0.710	0.680	Decrease
Specificity	0.890	0.850	Decrease
MCC	0.377	0.52	**Increase**
Accuracy	0.728	0.748	**Increase**
Bal. accuracy	0.800	0.765	Decrease
Pos. Predictivity	0.983	0.872	Decrease
Neg. Predictivity	0.254	0.639	**Increase**
Kappa	0.284	0.502	**Increase**

Lhasa Limited

Take home message

When a change/shift of prevalence occurs:

- Comparing metrics such as MCC or accuracy might not be a good idea

- If possible, use the same test set or maintain the same prevalence.
- If not possible, use the balanced accuracy (or sensitivity and specificity)
- Or if balanced accuracy (or sensitivity and specificity) is (are) not appropriate for the study, well that's a "cul de sac" or is something wrong with the data?

- Unless there is a way beyond balanced accuracy…

This is the topic of my talk…

In the literature

International Conference on Computational Science

↳ ICCS 2020: **Computational Science – ICCS 2020** pp 74–87 | Cite as

Home › Computational Science – ICCS 2020 › Conference paper

On Model Evaluation Under Non-constant Class Imbalance

Jan Brabec, Tomáš Komárek, Vojtěch Franc & Lukáš Machlica

Conference paper | First Online: 15 June 2020

2162 Accesses | **7** Citations | **5** Altmetric

International Symposium on Intelligent Data Analysis

↳ IDA 2020: **Advances in Intelligent Data Analysis XVIII** pp 457–469 | Cite as

Home › Advances in Intelligent Data Analysis XVIII › Conference paper

Master Your Metrics with Calibration

Wissam Siblini, Jordan Fréry, Liyun He-Guelton, Frédéric Oblé & Yi-Qing Wang

Conference paper | Open Access | First Online: 22 April 2020

18k Accesses | **8** Citations | **6** Altmetric

ELSEVIER

Pattern Recognition

Volume 91, July 2019, Pages 216-231

The impact of class imbalance in classification performance metrics based on the binary confusion matrix

Amalia Luque [a], Alejandro Carrasco [b], Alejandro Martín [a], Ana de las Heras [a]

Conferences > 2013 Humaine Association Conf...

Facing Imbalanced Data--Recommendations for the Use of Performance Metrics

Publisher: IEEE Cite This PDF

László A. Jeni ; Jeffrey F. Cohn ; Fernando De La Torre **All Authors**

Short Communication

DESCRIBING THE VALIDITY OF CARCINOGEN SCREENING TESTS

J. A. COOPER II*, R. SARACCI† AND P. COLE†

*From the *Division of Cancer Cause and Prevention, National Cancer Institute, Bethesda, Maryland, 20014, USA and the †Unit of Epidemiology and Biostatistics, International Agency for Research on Cancer, 69372 Lyon Cédex 2, France*

Received 28 August 1978 Accepted 16 October 1978

JOURNAL OF MAGNETIC RESONANCE IMAGING 33:505 (2011)

Letter to the Editor

Standardizing Predictive Values in Diagnostic Imaging Research

Lhasa Limited

Before we start…

Actual \ Predicted	+	-
+	TP	FN
-	FP	TN

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

$$Sen = \frac{TP}{TP + FN}$$

$$Spe = \frac{TN}{TN + FP}$$

$$Pre^{+} = \frac{TP + FN}{N}$$

Lhasa Limited

Accuracy

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

$$Acc = \frac{TP + TN}{N}$$

Lhasa Limited

Accuracy

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

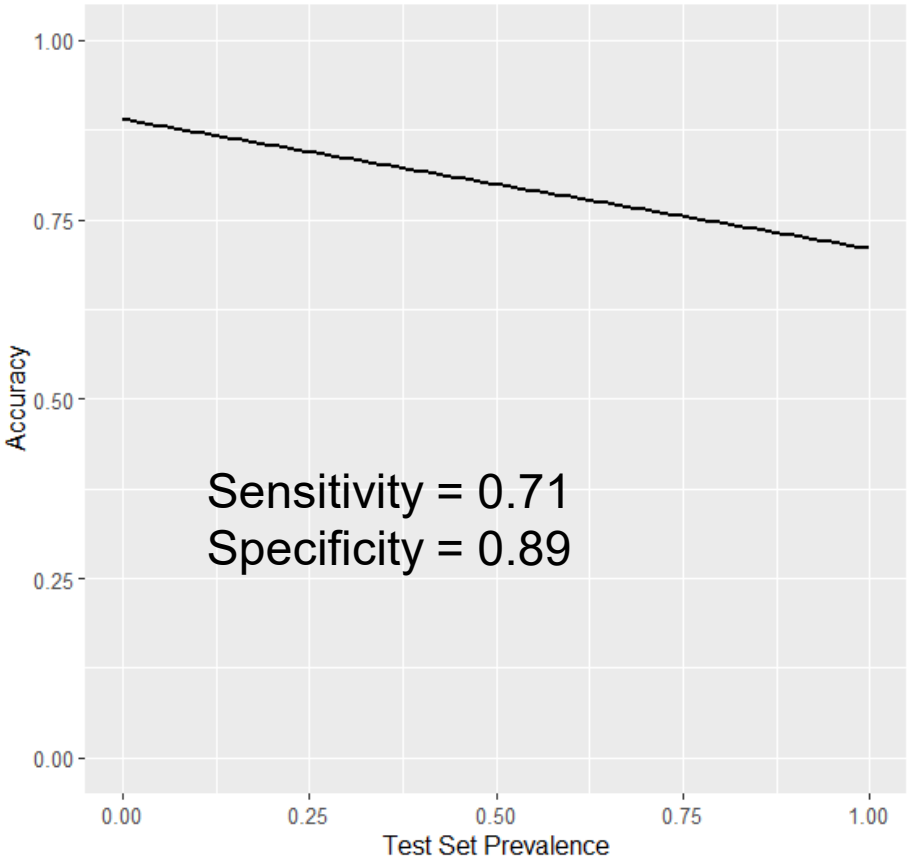
$$Acc = \frac{TP + TN}{N}$$

$$\Rightarrow Acc = Sen \times Pre^{+} + Spe \times (1 - Pre^{+})$$

Lhasa Limited

From accuracy to balanced accuracy

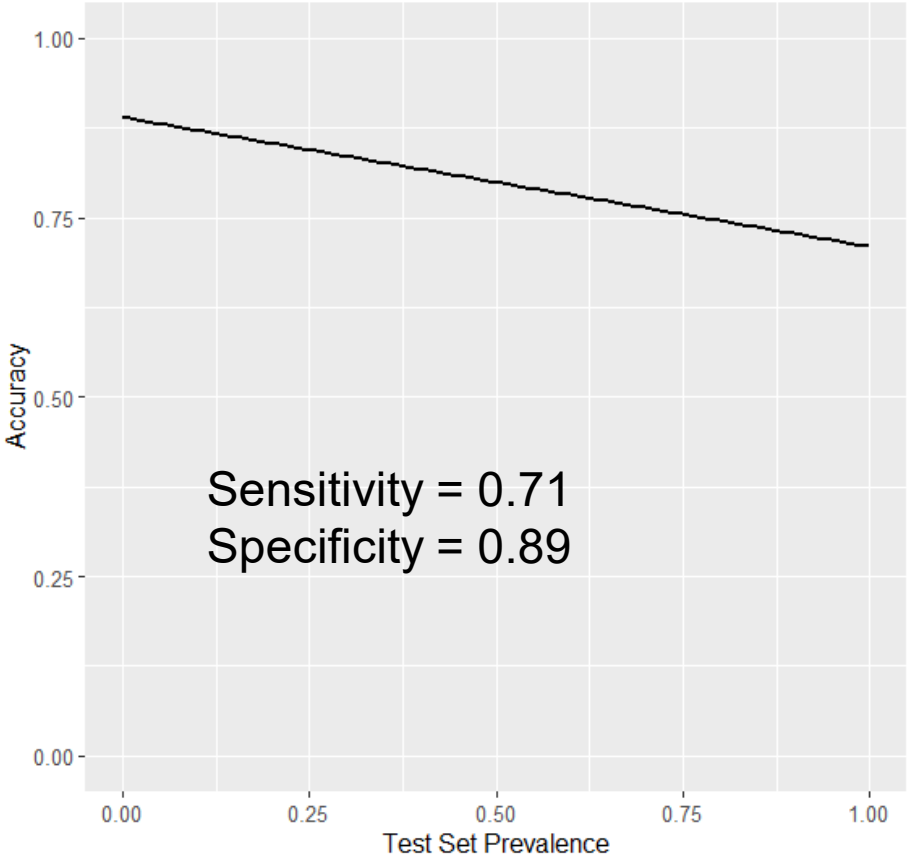
$$Acc = Sen \times Pre^{+} + Spe \times (1 - Pre^{+})$$



Lhasa Limited

From accuracy to balanced accuracy

$$Acc = Sen \times Pre^{+} + Spe \times (1 - Pre^{+})$$



$$when\ Pre^{+} = 0.5\ then$$

$$Acc^{0.5} = \frac{1}{2} \times Sen + \frac{1}{2} \times Spe$$

$$= \frac{Sen + Spe}{2} = Bal\ Acc$$

Lhasa Limited

Balanced accuracy is an accuracy calibrated to a prevalence of 0.5.

$$Acc^{Pre^+} = Sen \times Pre^+ + Spe \times (1 - Pre^+)$$

Lhasa Limited

MCC

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Lhasa Limited

MCC

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

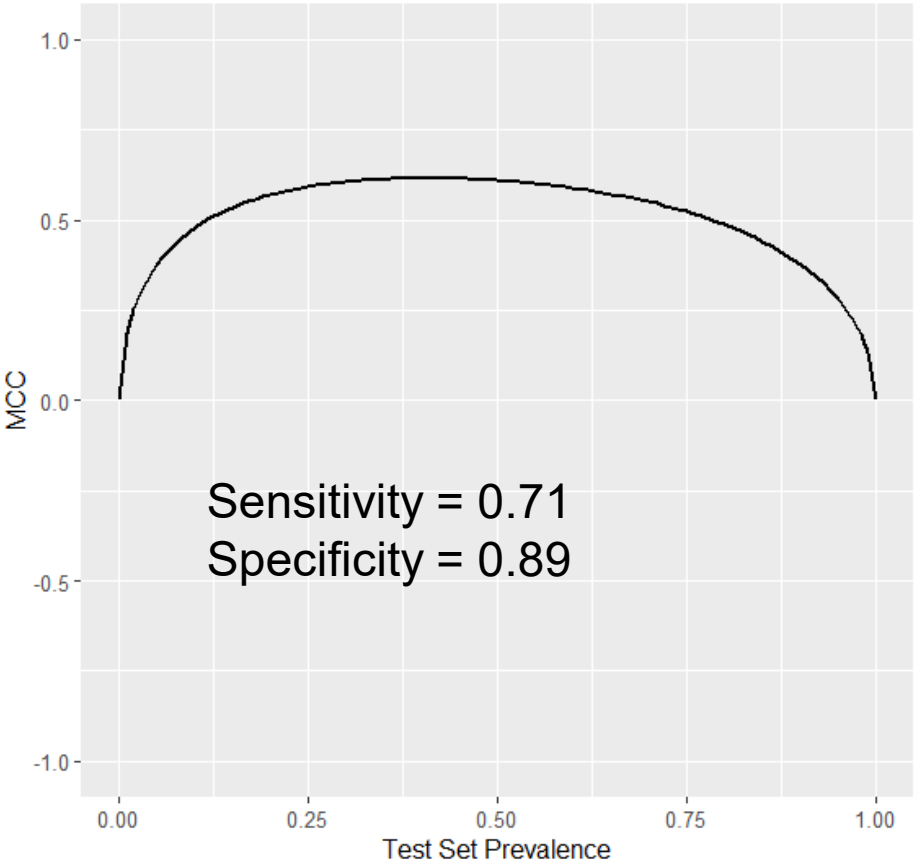
$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

$$\Rightarrow MCC = \frac{Sen + Spe - 1}{\sqrt{\left[Sen + (1 - Spe)\frac{(1 - Pre^{+})}{Pre^{+}}\right]\left[Spe + (1 - Sen)\frac{Pre^{+}}{(1 - Pre^{+})}\right]}}$$

Lhasa Limited

From MCC to balanced MCC

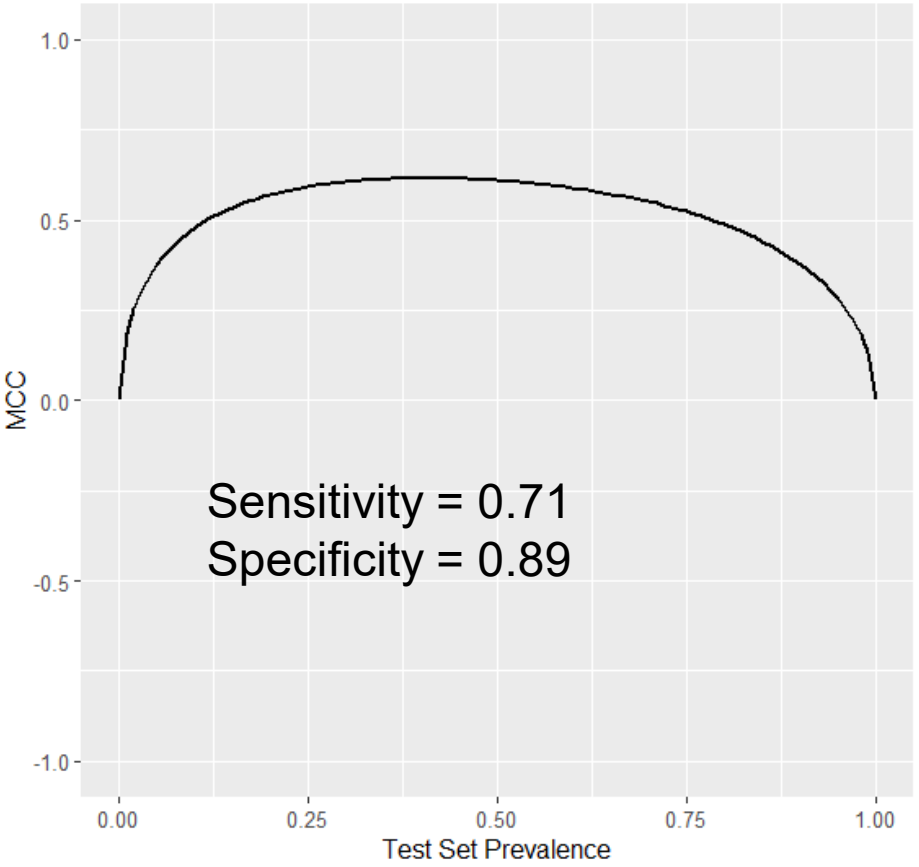
$$MCC = \frac{Sen + Spe - 1}{\sqrt{\left[Sen + (1 - Spe)\frac{(1 - Pre^{+})}{Pre^{+}}\right]\left[Spe + (1 - Sen)\frac{Pre^{+}}{(1 - Pre^{+})}\right]}}$$



Lhasa Limited

From MCC to balanced MCC

$$MCC = \frac{Sen + Spe - 1}{\sqrt{\left[Sen + (1 - Spe)\frac{(1 - Pre^{+})}{Pre^{+}}\right]\left[Spe + (1 - Sen)\frac{Pre^{+}}{(1 - Pre^{+})}\right]}}$$



$$when\ Pre^{+} = 0.5\ then$$

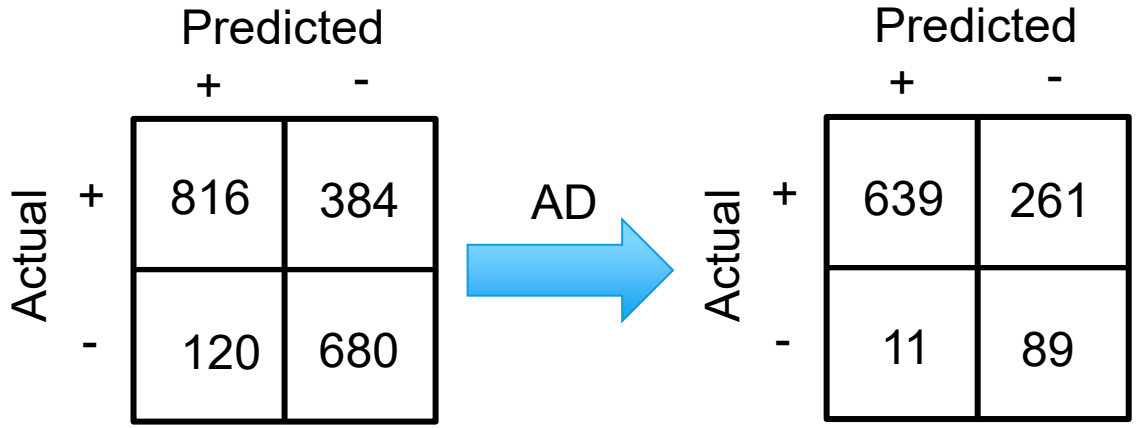
$$MCC^{0.5} = \frac{Sen + Spe - 1}{\sqrt{1 - (Sen - Spe)^2}} = Bal\ MCC$$

Lhasa Limited

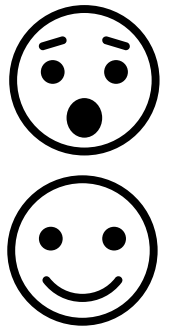
Use cases

- Applicability domain
- External and internal test sets (the issue at the beginning of the presentation)
- Non-stationary data streams
- Cluster splits

Tale of two MCCs



	Test set no AD	Test set with AD	Variation
Pos. Prevalence	0.600	0.900	
Sensitivity	0.680	0.710	Increase
Specificity	0.850	0.890	Increase
MCC	0.520	0.377	**Decrease**
Bal. MCC	0.538	0.610	**Increase**



"balanced or not balanced"

Lhasa Limited

Back to the issue

Actual \ Predicted	+	-
+	639	261
-	11	89

Actual \ Predicted	+	-
+	408	192
-	60	340

	Internal test set	External test set	Variation
Pos. Prevalence	0.900	0.600	
Sensitivity	0.710	0.680	Decrease
Specificity	0.890	0.850	Decrease
MCC	0.377	0.52	**Increase**
Accuracy	0.728	0.748	**Increase**
Bal. accuracy	0.800	0.765	Decrease
Bal. MCC	0.610	0.538	Decrease

Lhasa Limited

Final words

Mind the prevalence

Balanced metrics

Pros	Cons
Fair comparison	Does not represent the metric value at the real prevalence. If it is known the metric can be calibrated to that prevalence (the test dataset Under investigation has a different prevalence).

Lhasa Limited

Thanks for listening

shared **knowledge** • shared **progress**

Lhasa Limited
Granary Wharf House, 2 Canal Wharf
Leeds, LS11 5PS
Registered Charity (290866)

Company Registration Number 01765239

+44(0)113 394 6020
info@lhasalimited.org
www.lhasalimited.org

Supplementary material

Positive predictivity (PPV)

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

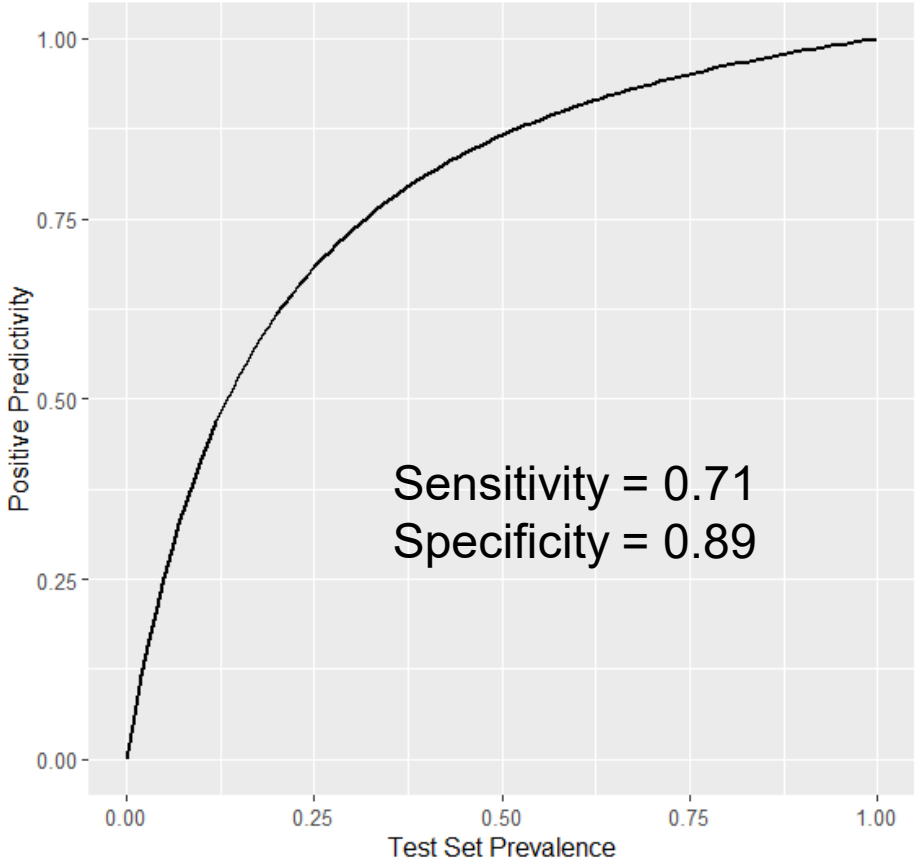
$$PPV = \frac{TP}{TP + FP}$$

$$\Rightarrow PPV = \frac{Sen \cdot Pre^{+}}{Sen \cdot Pre^{+} + (1 - Spe) \cdot (1 - Pre^{+})}$$

Lhasa Limited

From PPV to balanced PPV

$$PPV = \frac{Sen \cdot Pre^{+}}{Sen \cdot Pre^{+} + (1 - Spe) \cdot (1 - Pre^{+})}$$



$$when\ Pre^{+} = 0.5\ then$$

$$PPV^{0.5} = \frac{Sen}{Sen + 1 - Spe} = Bal\ PPV$$

Lhasa Limited

Negative predictivity (NPV)

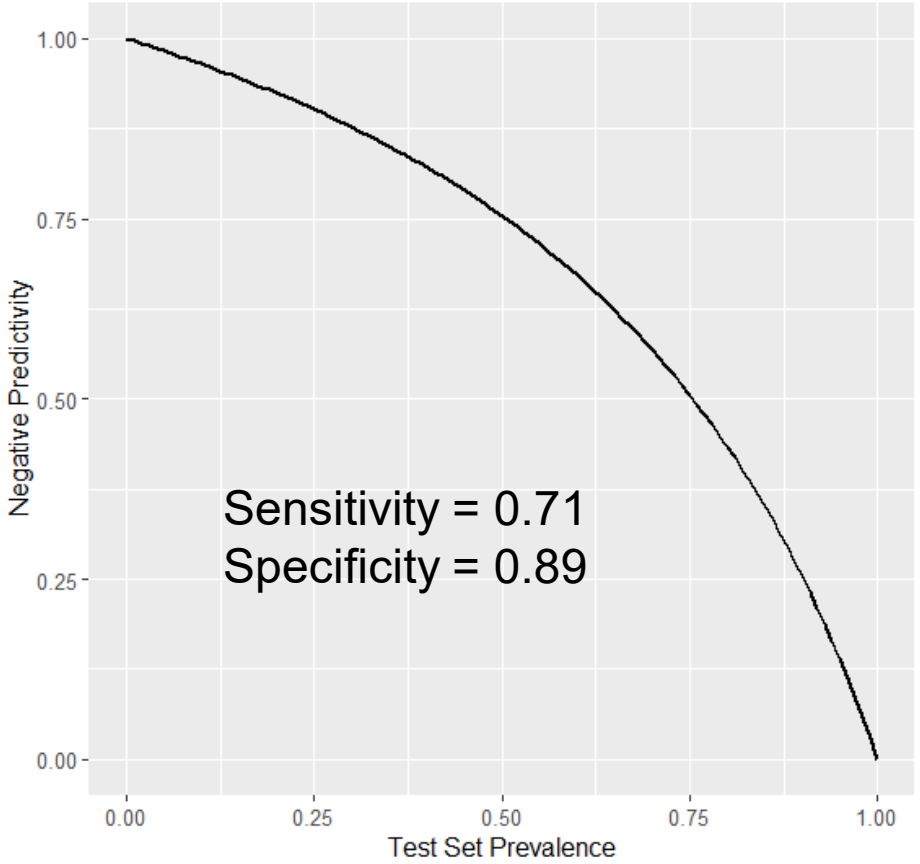
Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

$$NPV = \frac{TN}{TN + FN}$$

$$\Rightarrow NPV = \frac{Spe \cdot (1 - Pre^{+})}{Spe \cdot (1 - Pre^{+}) + (1 - Sen) \cdot Pre^{+}}$$

From NPV to balanced NPV

$$NPV = \frac{Spe \cdot (1 - Pre^{+})}{Spe \cdot (1 - Pre^{+}) + (1 - Sen) \cdot Pre^{+}}$$



$$when\ Pre^{+} = 0.5\ then$$

$$NPV^{0.5} = \frac{Spe}{Spe + 1 - Sen} = Bal\ NPV$$

Lhasa Limited

Kappa

Actual \ Predicted	+	-
+	$Sen \cdot Pre^{+} \cdot N$	$(1 - Sen) \cdot Pre^{+} \cdot N$
-	$(1 - Spe) \cdot (1 - Pre^{+}) \cdot N$	$Spe \cdot (1 - Pre^{+}) \cdot N$

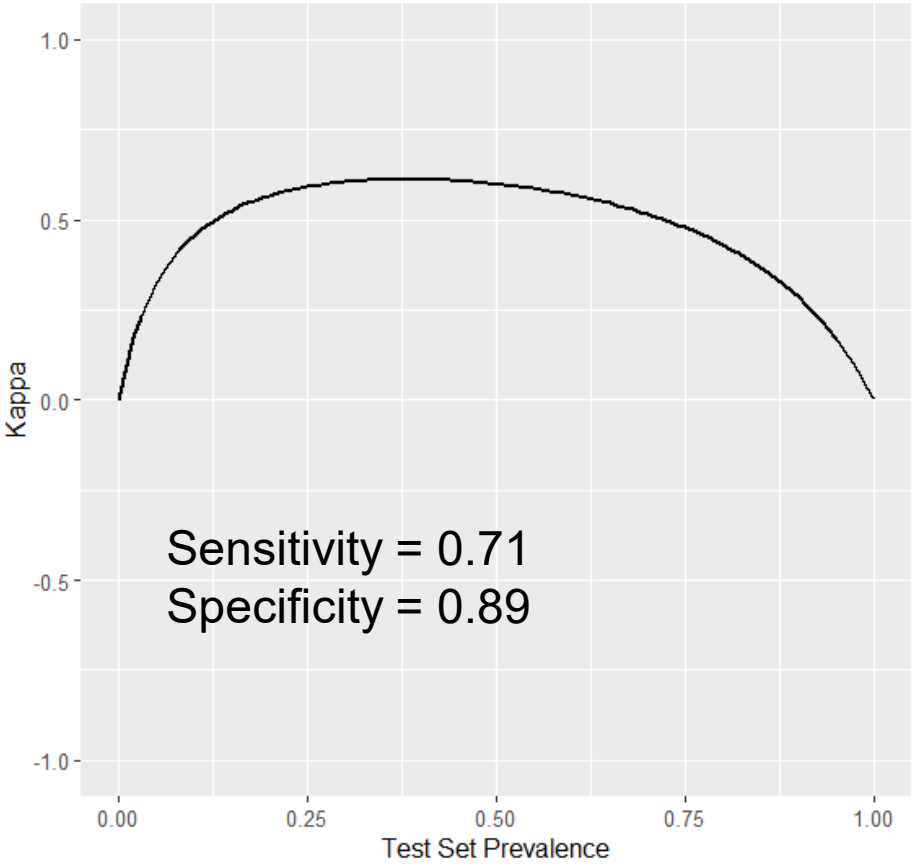
$$K = \frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (TN + FP) + (TP + FN) \cdot (TN + FN)}$$

$$\Rightarrow K = \frac{2 \cdot (Sen + Spe - 1)}{\left[Sen + (1 - Spe)\frac{(1 - Pre^{+})}{Pre^{+}}\right] + \left[Spe + (1 - Sen)\frac{Pre^{+}}{(1 - Pre^{+})}\right]}$$

Lhasa Limited

From kappa to balanced kappa

$$K = \frac{2 \cdot (Sen + Spe - 1)}{\left[Sen + (1 - Spe)\frac{(1 - Pre^{+})}{Pre^{+}}\right] + \left[Spe + (1 - Sen)\frac{Pre^{+}}{(1 - Pre^{+})}\right]}$$



$when\ Pre^{+} = 0.5\ then$

$$K^{0.5} = Sen + Spe - 1 = Bal\ K$$

Lhasa Limited