The ROC Diagonal is not Layperson's Chance: a New Baseline Shows Useful Points and Areas

André M. Carrington[1,2,3], Paul W. Fieguth[3], Franz Mayr[4], Nick D. James[2], Andreas Holzinger[5], John W. Pickering[6,7], and Richard I. Aviv[1,2,8]

[1] University of Ottawa, Ottawa, Canada
[2] The Ottawa Hospital, Ottawa, Canada
{acarrington,njames,raviv} at toh.ca
[3] University of Waterloo, Waterloo, Canada
pfieguth at uwaterloo.ca
[4] Universidad ORT Uruguay, Montevideo, Uruguay
mayr at ort
[5] University of Natural Resources and Life Sciences Vienna, Vienna, Austria
andreas.holzinger at human-centered.ai
[6] University of Otago, Christchurch, New Zealand
john.pickering at otago.ac.nz
[7] Christchurch Hospital, New Zealand
[8] Ottawa Hospital Research Institute, Ottawa, Canada

Abstract. In health care and other fields, the performance of a binary diagnostic test or classification model is often illustrated as a curve in a receiver operating characteristic (ROC) plot. In ROC plots, the major diagonal is often labelled 'chance', but in general, it does not represent a layperson's concept of chance for binary outcomes. It represents either a special case, or the ROC curve for a classifier that has the same distribution of scores for both classes, and thus zero information to distinguish positives from negatives. Where a model's ROC curve deviates from the major diagonal there is information, but not all information is 'useful information' relative to chance, including some areas above the diagonal. We define the binary chance baseline to identify points and areas in an ROC plot that are more useful than chance. We conduct experiments and compare the explanations between our new baseline and the major diagonal.

Keywords: ROC · AUC · C-statistic· Chance · Explainable AI · Classification · Diagnostic Tests

1 Introduction

In health care and other fields, the receiver operating characteristic (ROC) plot [16, 20] depicts the performance of a binary diagnostic test or classification model across all possible decision thresholds as a ROC curve (Figure 1, $r(x)$). Each threshold specifies a different classifier.

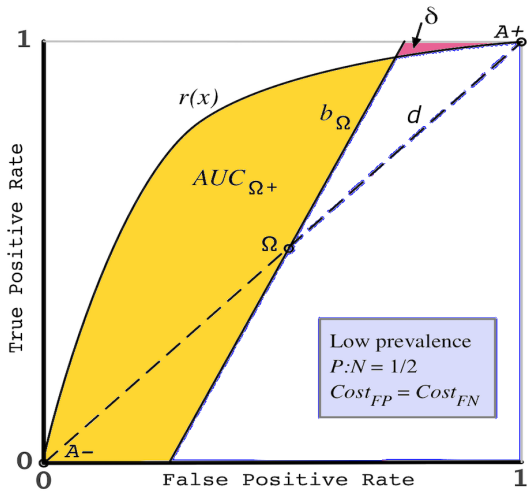

Fig. 1. Layperson's chance or 'binary chance', is a coin toss represented by the point $\Omega = (0.5, 0.5)$. We depict a line of equal cost and prevalence weighted accuracy as b_Ω, the binary chance baseline (the solid straight line). For real-world performance evaluation and explanation, the part of the ROC curve $r(x)$ that is more useful than chance, is above and to the left of b_Ω not the major diagonal d (the dashed line). The area under the ROC curve that is better than binary chance is denoted $AUC_{\Omega+}$ (yellow), while an area of negative utility is denoted δ. For low prevalence and equal costs, the slope of b_Ω is greater than the slope of d.

In the ROC plot, a line drawn from the bottom left to the top right is called the major diagonal or chance diagonal (Figure 1, dashed line) [26, 9] because it is commonly said to represent chance [26, 17], while others describe it as representing a random classifier [6].

The major diagonal is commonly used to interpret results in two ways. First, for the classifiers where a model's ROC curve is higher than the major diagonal, the model is said to be informative. Second, a model is thought to be better than chance where it is higher than the major diagonal—but we show that is not true for the most intuitive concept of chance for binary outcomes, a fair coin toss. It is useful to compare any classifier, with balanced or imbalance data to chance (a fair coin toss) to understand what the classifier offers above chance as an alternative mechanism as a black box classifier (albeit not a good one). For example, the point (0,0) on the major diagonal at the bottom-left of an ROC plot (Figure 1, $A-$) classifies all inputs as negative. For data with low prevalence, i.e., few positives, its predictions are mostly correct—more than the 50% correct predictions one obtains with a fair coin toss. Yet, $A-$ is on the major diagonal which is said to represent chance. Clearly it does not perform the same as coin toss—the layperson's concept of chance.

Conversely, the point (1,1) on the major diagonal at the top-right of an ROC plot (Figure 1, $A+$), is an all-positive classifier. It has the lowest possible threshold and predicts that all instances are positive. Predictions from $A+$ are mostly wrong for low prevalence data, yet $A+$ is also on the major diagonal which is also said to represent chance. Clearly a coin toss performs better, being half right and half wrong, instead of mostly wrong.

The major diagonal is commonly treated as a performance baseline, as in the examples above, but when it is used that way, we explain that what actually captures the user's intention and expectation is a line with performance equal to a fair coin toss.

In the sections that follow we discuss: binary chance, the binary chance baseline and useful area under the ROC curve, the major diagonal and information theory, implications, related work, experiments, results and conclusions.

2 Binary Chance

We refer to "chance" in this manuscript as something which happens by chance, i.e., "luck" or "without any known cause or reason" [1]. Randomness is a synonym, and for binary outcomes, a fair coin toss produces a random outcome.

We introduce the term "binary chance" for a fair coin toss. This is different from continuous chance as in teh major diagonal (Section 4) and it is also distinguished from "a chance" of rain which implies a non-zero probability without fairness. .

In this paper, we consider a coin toss, all-positive decisions and all-negative decisions to be classifiers with a hidden mechanism. The mechanism may consider the inputs or ignore them altogether when producing a corresponding output. Such classifiers are not good or useful by themselves—but they act as a baseline against which other classifiers are compared to determine if they are useful or not.

A fair coin toss (binary chance) is represented by the centre point $\Omega = (0.5, 0.5)$ [22] in a ROC plot (Figure 2) because on average, the rate of heads (events) is 0.5 and the rate of tails (non-events) is 0.5.

To make comparison easy and visible between chance and points and areas in an ROC plot, we can draw a line through Ω that has performance equivalent to binary chance as established in the literature [13].

3 The Binary Chance Baseline

To find points with performance equivalent to binary chance, we draw an iso-performance line [13, 6, 7], denoted b_Ω through Ω (Figure 2) and call it the binary chance baseline. It can represent accuracy, cost-weighted accuracy, or balanced accuracy, equivalent to chance.

If prevalence is considered but costs are ignored then it represents **accuracy** (or iso-accuracy) equivalent to chance. However, more often than not, prevalence and costs have counteracting effects (explained further in Section 8).

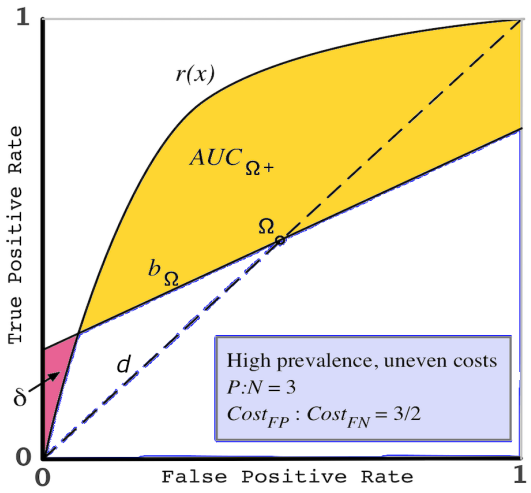

Fig. 2. Layperson's chance or 'binary chance' represented by the point $\Omega = (0.5, 0.5)$ has a line of equal cost and prevalence weighted accuracy b_Ω (solid straight line), which may have a slope lesser than the diagonal d (dashed line), as in this case of high prevalence data with uneven costs. The ROC curve $r(x)$ has a useful area $AUC_{\Omega+}$ (yellow) relative to binary chance Ω, and an area of negative utility denoted δ.

If prevalence and costs are considered, then it represents **cost-weighted accuracy** equivalent to chance (also called utility or net benefit).

If prevalence and costs are ignored, then the iso-performance line represents **balanced accuracy** and it coincides with the major diagonal–but it **does not explain** real-life performance and its impact in a ROC plot (as the motivation for this paper). Prevalence is rarely 50% and costs of error usually differ. In that context the major diagonal is misleading for explanations.

That said, balanced accuracy, and the AUC as another balanced measure, are useful for other purposes: to view performance without the majority class dominating when data are imbalanced. If and when prevalence and costs are not included in optimization loss functions, optimizing the AUC is a good but misleading way to mitigate the dominance of the majority class.

However, ignoring prevalence and costs, does not negate their presence in real-life and how they affect outcomes and our explanations about those outcomes. There is a choice between two different goals: theoretic performance with balanced measures to temporarily put aside the effect of class imbalance, versus realistic performance measures for explanations and cost optimal decision-making.

Historically and for specific purposes, the ROC plot, the AUC and the major diagonal ignore prevalence and costs. However, artificial intelligence (AI) is being increasingly applied to situations that affect people in everyday life. Hence, just as explainable AI is important, we also must be able to properly explain results.

We can better interpret how a model behaves in ROC plots when we include prevalence and cost information (as we demonstrate in Section 10).

For explanations, the major diagonal only represents binary chance in the special case of classes and costs being balanced, together, or both separately. That is, the latter case occurs when the prevalence is 50%, when the cost of a false positive equals the cost of a false negative, and when the cost of a true positive equals the cost of a true negative.

Let x represent the false positive rate or x-axis of an ROC plot and y represent the true positive rate or y-axis. Binary chance, or a fair coin flip, is on average 50% heads and 50% tails, represented by the point $\Omega = (\Omega_x, \Omega_y) = (0.5, 0.5)$ [22] in an ROC plot.

We define the binary chance baseline $b_\Omega(x)$ for data $\mathcal{X}$ as an iso-performance line passing through Ω with slope $m_\mathcal{X}$ as follows:

$$b_\Omega(x) := m_\mathcal{X}\,(x - \Omega_x) + \Omega_y \tag{1}$$

where the literature defines the slope $m_\mathcal{X}$ (or skew) of iso-performance lines [18, 7, 13] as a function of: P positives and N negatives (or prevalence π); and costs $C_{(.)}$ of false positives (FP), false negatives (FN), treatment for true positives (TP) and non-treatment for true negatives (TN) as follows.

$$m_\mathcal{X} := \frac{N}{P} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} \tag{2}$$

$$= \frac{(1-\pi)}{\pi} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} \tag{3}$$

Which results in the equation:

$$b_\Omega(x) := \frac{(1-\pi)}{\pi} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}}(x - 0.5) + 0.5 \tag{4}$$

The baseline above represents **cost-weighted accuracy** equivalent to (binary) chance. Or, if we ignore costs by setting $\frac{C_{FP}-C_{TN}}{C_{FN}-C_{TP}}$ to 1, then it represents **accuracy** equivalent to (binary) chance.

Costs are fixed or constant in the literature for the iso-performance approach. This appears to be reasonable except where constraints apply or where variable costs are known.

The binary chance baseline is meaningful for explanation because a model starts to become useful only when it performs better than chance—otherwise, a tossed coin performs just as well. While there are other baselines and other ideas about when a model is useful, the focus of this paper is on binary chance as a proper alternative to the ROC major diagonal.

4 The Major Diagonal, Continuous Chance, Divergence and Dissimilarity

The literature on ROC plots sometimes refers to the ROC major diagonal as chance [26, 17]—an ROC curve produced by classification scores (or probabili-

ties) drawn from *the same distribution* for events and non-events. The distribution does not need to be specified, as long as it is the same. We refer to this as "continuous chance", because there is an equal chance that the test or classifier producing each continuous score for both classes.

When the distribution of scores is the same for events and non-events, the model is not informative, i.e., it has zero information to distinguish the two classes, resulting in an ROC curve along the major diagonal [10, Fig. 2]. The absence of information is demonstrated by divergence and distance measures which are zero in this case: the Jensen-Shannon (J-S) divergence [14, 11, 12] is zero, the Kulback-Liebler divergence [5] in either direction is zero, and the Hellinger distance [3] is zero.

The Hellinger distance fits the classification context nicely since it achieves a value of 1 when the two distributions are non-overlapping, or when two unimodel distributions are linearly separable—a situation where many classifiers can perform perfectly with an AUC of 1 [10, Fig. 2]. That is, the [0, 1] range of the Hellinger distance correlates to the [0.5, 1] range of AUC achievable by most classifiers.

5 The Major Diagonal as an ROC Curve

Points (and curves) on the major diagonal are said to be non-informative and of no predictive value [16, 6, 15]—however, that is not quite accurate. A model may not have any information along the diagonal, but if a threshold is chosen wisely based on prevalence as prior knowledge, then the resulting classifier may be useful at that threshold (informed by the threshold choice, not the model's own intelligence).

In all situations except those in which costs and prevalence exhibit a rare and unusual equilibrium, some points on the major diagonal if chosen with prior knowledge of prevalence, are more predictive than (binary) chance.

As explained in the introduction, for data with low prevalence the all-negative classifier $A-$ at the bottom-left of the ROC plot (Figure 1) and classifiers at nearby points along the diagonal, perform better than a fair coin toss. $A-$ and nearby points are above the binary chance baseline. They have better cost-weighted accuracy (and better accuracy) in Figure 1.

Although some points on the diagonal chosen with prior knowledge, are useful, on the whole, the diagonal as an ROC curve, lacks information by itself. The JS divergence is measured for the whole distribution (or all samples along the ROC curve as the sample distribution).

The predictive ability and utility of the endpoints for imbalanced data are not seen nor included by the diagonal baselines for AUC nor balanced accuracy, because they ignore the prevalence (and hence class ratio) and they ignore costs.

6 Explaining ROC Plots: Useful Areas

To explain performance in an ROC plot, we must consider both the binary chance baseline and the major diagonal.

The area under the ROC curve but above the binary chance baseline (the yellow area denoted $AUC_{\Omega+}$ in Figure 2) tells us where a model is useful (more than chance).

The area under the ROC curve but above the major diagonal (the status quo approach) tells us where a model has information to distinguish positives from negatives, but some of the associated area and ROC points perform worse than chance.

Underneath the major diagonal and the binary chance baseline, the areas are both 0.5, but the location of those areas differ. Similarly, the areas between the curve and each baseline over the whole plot, only differ by δ, but the locations of the two areas differ.

Hence, the differences stand out more in a range or region of interest (Figures 3a and b). Also, there are some points on an ROC curve that perform worse than chance—those which border δ (Figure 2).

7 The Status Quo is Limiting and May Cause Errors

The status quo approach to ROC plots, historically but unnecessarily, ignores prevalence and costs, limiting the use of ROC plots to abstract interpretations for initial model development, instead of interpreting a model in the context of real-life applications [8]. Halligan *et al.* [8] imply that this is an inherent limitation of ROC plots, but our work demonstrates that sometimes it is not. Others have also suggested additions to status quo ROC plots for better explanations [2].

The useful areas and points on ROC curves identified by the binary chance baseline in ROC plots, provide explanations that are complimentary to decision curve analysis. Furthermore, ROC plots can relate a variety of pre-test and post-test measures to each other, including predictive values and likelihood ratios.

We have shown errors in explanations about what is useful in an ROC plot with the major diagonal as the status quo approach. Those errors occur because ignoring prevalence and costs is equivalent to implicitly and erroneously assuming that both are balanced within or between themselves, which is seldom the case.

8 Including Only One of Prevalence or Costs May Cause Errors: They Counteract Each Other

The previous section explains that ignoring prevalence and costs may cause errors in interpretation and choosing the best classifier. However, we may also incur errors by including prevalence while ignoring costs, or including costs while ignoring prevalence, because they often have counteracting effects in (2) as we explain in the following example.

Consider a medical condition, such as colon cancer. The cost C_{FN} of missing the disease in screening, a false negative, is much worse than a mistaken detection C_{FP}, a false positive, for which follow-up tests are conducted with some expense. The term $\frac{C_{FP}-C_{TN}}{C_{FN}-C_{TP}}$ causes a low slope, i.e., less than one.

However, as a condition with low prevalence, the term $\frac{N}{P}$ causes a high slope, i.e., it has an effect that counteracts the costs. These effects do not in general balance each other, and in some cases they act in the same direction. However, more often than not, the minority class tends to be the class of interest, and false negatives tend to be more costly, resulting in counteracting effects.

9 Costs as Rates Square the Effect of the Class Ratio or Prevalence

In the literature on costs for optimal points in ROC analysis [13, 7] or costs in net benefit calculations [23], costs are specified in terms of individuals. For example, the cost of a false negative C_{FN}, an individual, relative to the cost of a false positive C_{FP}, is specified.

However, since the ROC plot axes are rates $TPR = 1 - FNR$ and $FPR = 1 - TNR$, it may be more natural to consider costs as rates. That is, to achieve 1% better TPR and thus reduce FNR by 1%, what percentage increase in FPR do we incur? From Metz's [13] expression (2) for slope of iso-performance lines, we derive the equation for rates, as follows:

$$m_{\mathcal{X}} = \frac{N}{P} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} \tag{5}$$

$$= \frac{N}{P} \cdot \left(\frac{N \cdot P}{P \cdot N}\right) \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} \tag{6}$$

$$= \left(\frac{N \cdot N}{P \cdot P}\right) \cdot \frac{P}{N} \cdot \frac{C_{FP} - C_{TN}}{C_{FN} - C_{TP}} \tag{7}$$

$$= \left(\frac{N}{P}\right)^2 \cdot \left[\frac{C_{FP} - C_{TN}}{N}\right] \cdot \left[\frac{C_{FN} - C_{TP}}{P}\right]^{-1} \tag{8}$$

$$= \left(\frac{N}{P}\right)^2 \cdot [C_{FPR} - C_{TNR}] \cdot [C_{FNR} - C_{TPR}]^{-1} \tag{9}$$

$$= \left(\frac{N}{P}\right)^2 \cdot \frac{C_{FPR} - C_{TNR}}{C_{FNR} - C_{TPR}} \tag{10}$$

When we consider costs as rates, the effect of the class ratio $\frac{N}{P}$ or prevalence π in $\frac{N}{P} = \frac{1-\pi}{p\imath}$ is squared in (10). Hence, ignoring prevalence, or using prevalence without costs, could be more detrimental than expected.

Costs as rates may be useful to ensure the proportionality principle [25] for equitable AI.

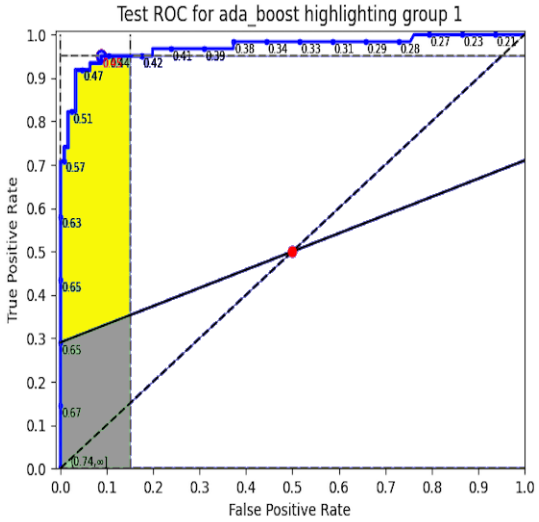

Fig. 3. An ROC plot for Adaboost applied to Wisconsin Breast Cancer recurrence data (size and texture), focused on sensitivity measured by the vertical aspect of the ROC curve in the ROI $FPR = [0, 0.15]$. The area better than binary chance (the solid sloped line passing through the center) is smaller than the information above the major diagonal.

10 Experimental Method and Results

We conduct an experiment using the Wisconsin Diagnostic Breast Cancer (WDBC) data set [24]—a data set curated by clinicians for the prediction of whether or not breast cancer recurs by a time endpoint. Our testing examines size, texture and shape features of nuclei sampled by thin needle aspiration. For diagnostic tests the SpPin rule [19] indicates that high specificity, the left side of an ROC plot is paramount. So we define a region of interest (ROI) as 85% specificity or above, $FPR = [0, 0.15]$ (Figure 3). Ideally we want to minimize false negatives as well, and that is achieved by some models (ibid).

If we choose an ROC point (or threshold) that achieves 95% specificity exactly, it may or may not have sufficient sensitivity, and it may or may not be optimal in cost weighted accuracy. Hence, within the ROI we:

- identify the useful ROC points and area with the binary chance baseline,
- identify optimal point(s) in the ROI (hollow blue circles in figures),
- describe how choices and explanations differ from the major diagonal, and
- confirm that points along the major diagonal perform better or worse, as expected, compared to chance $\Sigma = (0.5, 0.5)$.

In Figure 3 about one third of the area in the ROI performs no better than binary chance. Thresholds in the gray region below the binary chance baseline

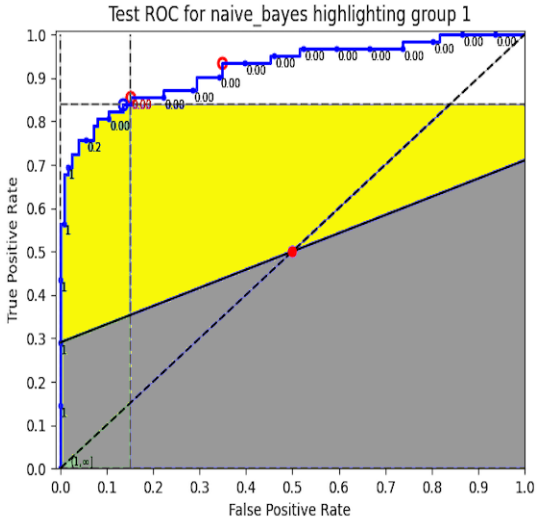

Fig. 4. An ROC plot for Naive Bayes applied to Wisconsin Breast Cancer recurrence data (size and texture), focused on specificity measured by the horizontal aspect of the ROC curve in the ROI $FPR = [0, 0.15]$. There are differences in the interpretation of performance and useful areas or ROC points between the dashed major diagonal and our solid binary chance baseline.

perform worse than chance and should not be used, contrary analysis using the major diagonal. The binary chance baseline has a gradual or low slope, given given 30% prevalence and the cost of false negative specified as five times worse than a false positive (a hypothetical cost). Knowledge of the prevalence is almost always known, or at lease an expected range).

If one were to decide that a more interpretable model than Adaboost was required, they might opt for Naive Bayes (Figure 6) or a Decision Tree (Figure 5), but the latter has almost no points in the ROI better than chance, different from the major diagonal's status quo interpretation.

The binary chance baseline and measurement of areas and cost-weighted accuracy can also be applied to Mean ROC plots. And the prevalence and costs may differ, causing a binary chance baseline with a slope higher than the major diagonal.

For Adaboost we confirmed that the values of cost-weighted accuracy (also called net benefit) are as expecgted (Table 1)

11 Related Work

Zhou *et al.* [26] provide a classic work on the interpretation of ROC curves and plots including discussion of the major diagonal and chance. Numerous other

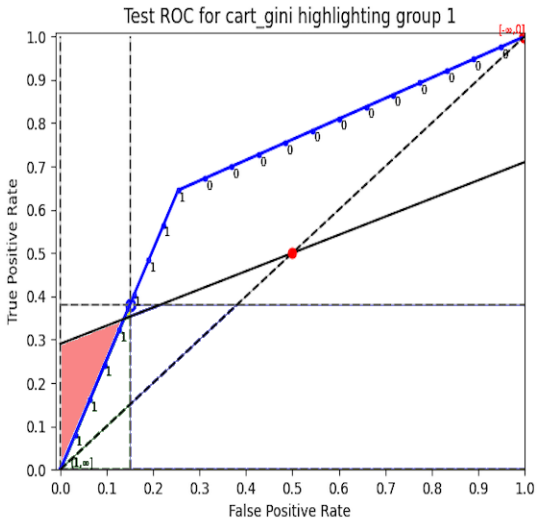

Fig. 5. Using only the texture features in Wisconsin Diagnostic Breast Cancer (WDBC) data set, the decision tree model (cart-gini) performs worse than chance in the region of interest $FPR = (0.5, 0.5)$.

Table 1. Validation of expected cost-weighted accuracy at various points in the ROC plot (Figure **??**).

ROC point	Cost-Weighted Accuracy	Expectation
optimal ROC point	-0.20	is the smallest value, as expected
A- (0, 0)	-1.86	is smaller than chance, as expected
A+ (1, 1)	-0.63	is larger than chance, as expected
optimal point in ROI	-0.20	is a small value, as expected (also the smallest)
chance (0.5, 0.5)	-1.25	
perfect test (0, 1)	0	

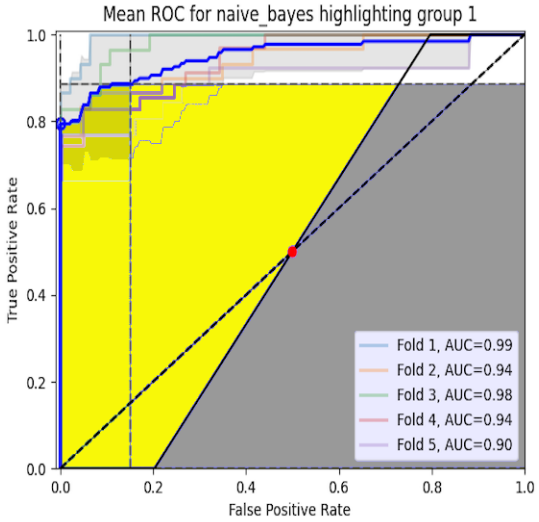

Fig. 6. The binary chance baseline may have a higher slope than the major diagonal and it may be applied to Mean ROC curves and plots too.

sources [26, 9, 17] discuss and describe the major diagonal as chance or a random classifier [6], and the major diagonal has long been used as a point of comparison for performance measures in ROC plots.

Aside from binary chance and continuous chance as concepts we discuss in this paper, or "a chance" of rain, there is also the concept of chance agreement. Kappa [4] describes the amount of agreement beyond chance agreement, between any two models or people that produce scores. Kappa includes prevalence but does not include costs unless modified [**?**]. The unmodified version is more commonly known.

While Kappa and other priors may provide alternative baselines from which to judge utility, our paper focuses on the misunderstanding of the major diagonal as chance, and the clarification of the binary chance baseline as the layperson's concept of chance for binary outcomes. Other baselines may be investigated in other work.

Iso performance lines were introduced by Metz [13], and later used by others, such as Provost and Fawcett [6] for the purpose of identifying an optimal ROC point on the ROC curve. Flach [7] then investigated the geometry of ROC plots with iso performance lines. Subtil and Rabilloud [21] were the first, to apply iso performance lines as a baseline from which to measure performance. They examine performance equivalent to the all-negative and all-positive classifiers we discussed and denoted as $A-$ and $A+$. We apply the same concept, but as applied to binary chance Ω.

12 Conclusions and Future Work

ROC plots that label the major diagonal as chance are misleading for interpretation—chance is conflated with zero information. We clarify and illustrate anew the baseline that represents binary chance as the more intuitive and interpretable layperson's concept for explainable AI. We also explain the major diagonal's relation to divergence and dissimilarity measures, where the latter correlates with expected performance. While ROC plots may have originally been applied to disregard prevalence and costs, that does not serve our needs for explanations. We show that explanations from the major diagonal about several ROC points are faulty, whereas the binary chance baseline is congruent with our expectations and computed values of cost-weighted accuracy. Or if one ignores costs and assumes they are balanced, then our baseline is more congruent with accuracy (which accounts for prevalence). This new perspective should shed new light on prior work.

Contributions

All authors contributed in writing this article. AC conceived the main ideas initially. In consultation with PF, JP, FM, and NJ various ideas were further developed and refined, with AH and RA providing guidance. Experiments were conducted and coded by AC and FM. All authors reviewed and provided edits to the article.

Acknowledgements Parts of this work has received funding by the Austrian Science Fund (FWF), Project: P-32554 "A reference model for explainable Artificial Intelligence in the medical domain".

References

1. chance, https://dictionary.cambridge.org/dictionary/english/chance
2. Althouse, A.D.: Statistical graphics in action: making better sense of the roc curve. International Journal of Cardiology **100**(215), 9–10 (2016)
3. Beran, R.: Minimum hellinger distance estimates for parametric models. The annals of Statistics pp. 445–463 (1977)
4. Cohen, J.: A coefficient of agreement for nominal scales. Educational and psychological measurement **20**(1), 37–46 (1960)
5. Cover, T.M., Thomas, J.A.: Elements of information theory. John Wiley & Sons (2012)
6. Fawcett, T.: An introduction to ROC analysis. Pattern recognition letters **27**(8), 861–874 (2006)
7. Flach, P.A.: The geometry of ROC space: Understanding machine learning metrics through ROC isometrics. Proceedings of the Twentieth International Conference on Machine Learning (2003)

8. Halligan, S., Altman, D.G., Mallett, S.: Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. European radiology **25**(4), 932–939 (2015)
9. Hand, D.J.: Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine learning **77**(1), 103–123 (2009)
10. Inácio, V., Rodríguez-Álvarez, M.X., Gayoso-Diz, P.: Statistical evaluation of medical tests. Annual Review of Statistics and Its Application **8**, 41–67 (2021)
11. Lin, J.: Divergence measures based on the Shannon entropy. IEEE Transactions on Information Theory **37**(1), 145–151 (Jan 1991). https://doi.org/10.1109/18.61115, http://ieeexplore.ieee.org/document/61115/
12. Menéndez, M., Pardo, J., Pardo, L., Pardo, M.: The jensen-shannon divergence. Journal of the Franklin Institute **334**(2), 307–318 (1997), publisher: Elsevier
13. Metz, C.E.: Basic principles of ROC analysis. In: Seminars in nuclear medicine. vol. 8, pp. 283–298. Elsevier (1978)
14. Nielsen, F.: On a variational definition for the jensen-shannon symmetrization of distances based on the information radius. Entropy **23**(4), 464 (2021)
15. Obuchowski, N.A.: Receiver operating characteristic curves and their use in radiology. Radiology **229**(1), 3–8 (2003). https://doi.org/10.1148/radiol.2291010898
16. Obuchowski, N.A., Bullen, J.A.: Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Physics in Medicine & Biology **63**(7), 07TR01 (2018)
17. Powers, D.M.W.: Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. Tech. Rep. December, Flinders University (2007)
18. Provost, F., Fawcett, T.: Robust classification for imprecise environments. Machine Learning (2001)
19. Sackett, D.L., Straus, S.: On some clinically useful measures of the accuracy of diagnostic tests. BMJ Evidence-Based Medicine **3**(3), 68 (1998)
20. Streiner, D.L., Cairney, J.: What's under the ROC? An introduction to receiver operating characteristics curves. The Canadian Journal of Psychiatry **52**(2), 121–128 (2007)
21. Subtil, F., Rabilloud, M.: An enhancement of ROC curves made them clinically relevant for diagnostic-test comparison and optimal-threshold determination. Journal of clinical epidemiology **68**(7), 752–759 (2015)
22. Van den Hout, W.B.: The area under an ROC curve with limited information. Medical Decision Making **23**(2), 160–166 (2003). https://doi.org/10.1177/0272989X03251246
23. Vickers, A.J., Elkin, E.B.: Decision curve analysis: a novel method for evaluating prediction models. Medical Decision Making **26**(6), 565–574 (2006)
24. Wolberg, W.H., Mangasarian, O.L.: Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proceedings of the national academy of sciences **87**(23), 9193–9196 (1990)
25. Young, H.P.: Equity: in theory and practice. Princeton University Press (1995)
26. Zhou, X.H., McClish, D.K., Obuchowski, N.A.: Statistical methods in diagnostic medicine, vol. 569. John Wiley & Sons (2002)