Applied Medical Informatics
Original Research Vol. 45, No. 3 /2023, pp: 103-117

# Min-Max, Min-Max-Median, and Min-Max-IQR in Deciding Optimal Diagnostic Thresholds: Performances of a Logistic Regression Approach on Simulated and Real Data

**Ilie-Andrei CONDURACHE and Sorana D. BOLBOACĂ***

Department of Biostatistics and Medical Bioinformatics, "Iuliu Haţieganu" University of Medicine and Pharmacy Cluj-Napoca, Louis Pasteur Str., No. 6, 400349 Cluj-Napoca, Romania.
E-mails: condurache.ilie.andrei@elearn.umfcluj.ro; sbolboaca@umfcluj.ro

* Author to whom correspondence should be addressed;

**Received: September 23, 2023/Accepted: September 29, 2023/ Published online: 30 September 2023**

**Abstract**
Combining biomarkers and their statistics is used to increase the prediction performance of a diagnosis, but no gold standard method exists. We introduced and evaluated an approach using linear combinations of summary-based statistics tested in logistic regression models with 10-fold repeated cross-validation. We used AUC (area under the ROC- receiver operating characteristic curve), the value of the Youden index, sensitivity (Se), specificity (Sp), diagnostic odds ratio (DOR), Efficiency Index (EI) and Inefficiency Index (InI) as performance metrics on the real-data set. We tested the approaches in multivariate normal distribution simulations with 4, 10, and 100 biomarkers and on real data. The results show that the summary-based models, especially minimum-maximum-median regression model (LR(MMM)) and minimum-maximum-interquartile range model (LR(MMIQR)), have similar performances or slightly better performances than the classical LR model regardless of the imposed mean of biomarkers or covariance matrixes on both simulated and real-data. The differences in AUCs were higher as the number of combined biomarkers increased (LR(MMIQR) model vs. LR model: 0.09 equal or unequal means of four biomarkers, 0.26 equal means, and 0.11 unequal means of 10 biomarkers). In real data, the linear combination of four biomarkers on LR(MMM) and LR(MMIQR) slightly increases the AUCs compared to the LR model. The model's performances were marginally low and without clinical relevance. The linear combination of summary-based statistics, specifically LR(MMM) and LR(MMIQR), exhibits similar performances as the classical LR model when biomarkers are linearly combined to increase diagnostic accuracy. Although the models perform on simulation data-sets, no clinical relevance of the combination is observed in the applied real-data.

**Keywords:** ROC (receiver operating characteristic curve); Logistic regression; Youden; Biomarker; Diagnostic accuracy

## Introduction

The patient's clinical and paraclinical data are used to identify a possible medical condition or differentiate between different diseases. Accuracy, defined as the ability of a test to determine those with and without the disease of interest correctly, is essential in medical diagnosis [1]. High accuracy of a diagnostic test is beneficial for the patients and the medical providers by shortening the time-to-treatment initiation [2] and increasing the diagnostic cost-effectiveness [3].

Biomarkers are measured biological characteristics used to identify the existence of a specific disease [4], but since they are sensitive to particular aspects of the disease, one biomarker exhibits limited performances [5]. An increase in diagnostic performance by identification of new biomarkers [4] or by combining different biomarkers is of interest and has been reported [6-9]. The methods used to combine biomarkers available in the scientific literature are linear [10-12], non-linear [14], or flexible [15]. However, most methods are linear because it is easier to implement in clinical settings and to interpret.

The ROC (receiver operating characteristics) method is used to determine the diagnostic performance of a biomarker by plotting sensitivity (Se, the ability of the test to identify subjects with the disease) against (1-specificity) (Sp, the ability of a test to determine the subjects without the disease). The performance of a diagnostic test can be estimated by the area under the curve (AUC) along with 95% confidence intervals in the case of a measured biomarker. A higher value for the AUC indicates a higher diagnostic performance (AUC=1 belongs to a test able to discriminate subjects with and without the disease of interest perfectly) [16,17]. The cut-off values from the ROC curve can be obtained for the biomarker (each value having a diagnostic Se and Sp). The Youden (J) index is used to estimate the cut-off value (max(Se+Sp-1)) that correctly classifies most subjects [18]. Other methods have been introduced to define the cut-off values: min(Euclidean distance between the ROC curve and the (0,1) point) [19], max($\chi^2$) in 2×2 contingency table [20], max(Se*Sp) [21], min(IU) (IU is Index of Union, where UI = (|Se-AUC|+|Sp-AUC|) [22], or other [23,24]. The AUC (≥0.9 indicate an excellent diagnostic test) and the J index measure are most frequently used to evaluate the diagnostic performance of a biomarker [25,26].

Logistic regression (LR) models are frequently used to combine biomarkers, where the predicted probability of the disease is plotted in ROC curves against the observed probability. Several approaches have been developed and evaluated considering the multivariate normality assumption [27], the Mann-Whitney statistic [28], step-wise methods for outcome diagnostic as an ordinal variable (three outcomes) [29], empirical likelihood ratios [30], the minimum and maximum of the biomarkers [31] or the minimum, maximum, median and interquartile range of the biomarkers [32]. The proposed approaches aimed to maximize the AUC [28, 30] or pAUC (partial AUC) [31,33]

Aznar-Gimeno et al. [32] extended the Liu et al. method [31] by adding the median or interquartile (IQR) range to the min-max values of biomarkers applied to maximize the J index. They considered any theoretical distribution of the continuous biomarkers, used a step-wise approach to select the best linear combination of summary statistics, and applied a simple cross-validation approach to validate the method [32]. The Aznar-Gimeno et al. [32] method proved better than the Liu et al. min-max approach [31], with higher Se and Sp on real-data.

We proposed a new approach that incorporates the summary statistics (Min-Max, Min-Max-Median, and Min-Max-IQR) in logistic regression models to increase diagnostic accuracy in combining normally distributed biomarkers.

## Material and Method

The code used in our study was written using the R software (R Project for Statistical Computing – version 4.1.1) under the RStudio GUI build 576.

### *Simulated Data-Sets*

The data-sets were created using the method described by Aznar-Gimeno et al. [32]. The simulated data was made using the *MASS* R package (under the *mvrnorm* command). We generate 100 random samples of 30/100/500 virtual subjects per group (disease-free and disease group) for each scenario. The generated values of biomarkers followed a theoretical normal distribution (4/10/100 biomarkers) by group (disease and disease-free groups). We conducted the simulations for null means of biomarkers (m1=0) in the disease-free population. In the disease population, we considered equal means (m2=1.0) or different means (Figure 1). The variance for each biomarker was set to 1, and the covariance matrixes were set to be equal ($\Sigma 1 = \Sigma 2$) or different between populations ($\Sigma 1 \neq \Sigma 2$, see

Figure 1). We evaluated the same covariance matrixes (Σ1 = Σ2) as independence (Σ1 = Σ2 = *I*), medium correlation (Σ1 = Σ2 = 0.5**I* + 0.5**J*), and negative correlation (Σ1 = Σ2 = -0.1*I),where *I* is the identity matrix, *J* is a matrix of all ones. In the scenarios with different covariance matrixes, we evaluated high correlation (Σ1=0.3**I*+0.7**J*) and low correlation (Σ2=0.7**I*+0.3**J*).

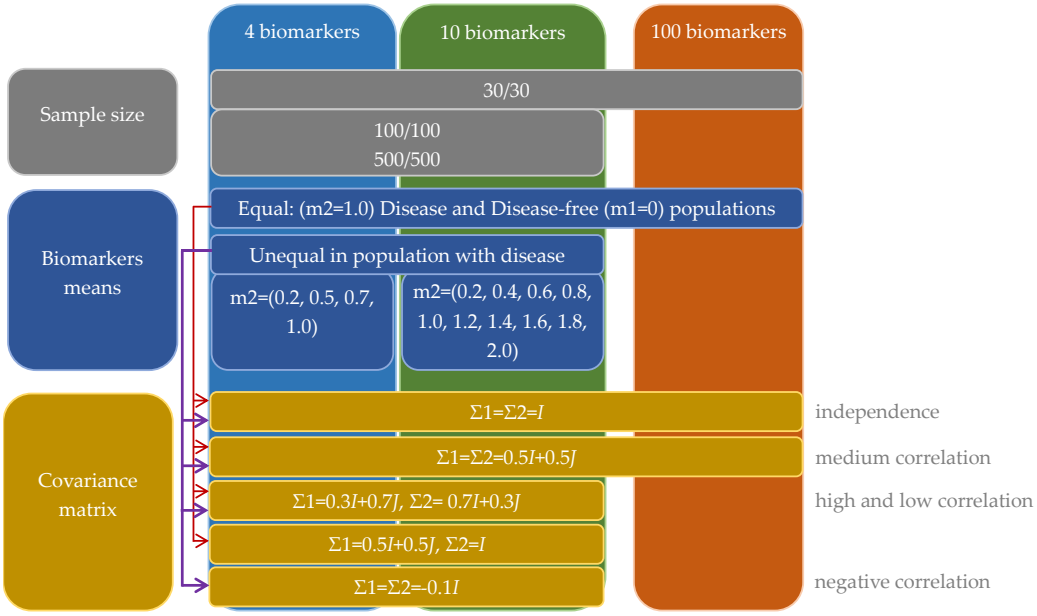


**Figure 1.** Simulated populations used in the study based on the number of biomarkers, covariance matrixes, and biomarker means in the population with the disease (equal/different). *I* = identity matrix, *J* = matrix of all ones.

In each simulated data-set, for each population (disease-free and disease), we calculated five parameters: X_max – the maximum value of the biomarker, X_min – the minimum value of the biomarker, X_q50 – the median value of the biomarker X_q25 – the 25th percentile of the biomarker, and X_q75 – the 75th percentile of the biomarker.

*Proposed Approach*

One-hundred random sets were simulated for each scenario. Logistic regression models were obtained using the resubtitution method and a repeated 10-fold cross-validation approach for each set. We used the package *caret* under the *train* command, R. Four models were created (Table 1).

**Table 1.** Used models and covariates.

| Model | Abb | Covariates |
|---|---|---|
| Classic logistic regression | LR | X1+X2+…+Xk* |
| Minimum-maximum | LR(MM) | X_max+X_min |
| Minimum-maximum-median | LR(MMM) | X_max+X_min+X_q50 |
| Minimum-maximum-interquartile range | LR(MMIQR) | X_max+X_min+X_q50+X_q25+X_q75 |

* *k* was 4/10/100 according to the number of biomarkers; X_max = the highest value; X_min = the smallest value; X_q50 = the value of median; X_q25 = the value of the 25th percentile; X_q75 = the value of the 75th percentile

A ROC curve was created for each binormal model and scenario using the predicted probability of disease against the observed probability. We used the *pROC* package under the roc command, R software, for this task. The AUC and the Youden index were calculated for each set, and the average and the standard deviation were reported.

*R Code by Example*

Examples of written lines of code are presented to exemplify the applied approach. The author's comments are highlighted with "#" symbol before the text. In this example, a repeated 10-fold cross-validation method is used in the scenario where values for four biomarkers are simulated using 30 subjects in each group, the means of biomarkers in the disease group are different (m2=(0.2, 0.5, 0.7, 1.0)) as well as the covariance matrixes are between groups ($\Sigma 1=0.3I+0.7J$, $\Sigma 2= 0.7I+0.3J$):

```
rm(list = ls())
options(warn=-1) #for ignoring any errors in the output
library(pROC)
library(MASS)
library(caret)
mu1 <- c(0,0,0,0) #the mean matrix for the disease-free population
stddev <-c(1,1,1,1) #the variance was 1 for each biomarker
mu2 <- c(0.2,0.5,1,0.7) #the mean matrix for the disease group, with different means of biomarkers
ident <- diag(4) #the identity matrix
mat1 <- matrix(1,4,4) #the matrix of all ones
matf1 <- 0.3*ident+0.7*mat1 #base for the covariance matrix for the disease-free population
matf2 <- 0.7*ident+0.3*mat1 # base for the covariance matrix for the population with the disease
covMat1 <- stddev%*% t(stddev) * matf1 #creating the covariance matrix for the disease-free population
covMat2 <- stddev%*% t(stddev) * matf2 #creating the covariance matrix for the population with the disease
youden1 <- as.numeric()
auc_roc1 <- as.numeric()
youden2 <- as.numeric()
auc_roc2 <- as.numeric()
youden3 <- as.numeric()
auc_roc3 <- as.numeric()
youden4 <- as.numeric()
auc_roc4 <- as.numeric() #defining the Youden index and the AUC value for each model as numeric values
data_ctrl <- trainControl(method="repeatedcv", repeats= 10, savePredictions=TRUE, classProbs=TRUE, number=10, p=0.9, summaryFunction = twoClassSummary, returnResamp = "all") #the cross-validation instruction, for a 10-fold repeated cross-validation
for (i in 1:100) #the command to repeat the same process 100 times to obtain 100 different sets of samples and outputs
{
set.seed(i)
dat1 <- mvrnorm(n=30, mu=mu1, Sigma=covMat1, empirical=TRUE)
dat2 <- mvrnorm(n=30, mu=mu2, Sigma=covMat2, empirical=TRUE) #using the TRUE criteria, the mean values and the covariance matrixes obtained will be exactly the ones specified
condition <- rep(x="normal", times=30)
df1 <- data.frame(condition, dat1)
condition <- rep(x="diseased", times=30)
df2 <- data.frame(condition, dat2) #creating a virtual database by combining the simulated values and a column in which criteria for the disease-free/disease population is specified
df1$X_max <- apply(df1[, 2:5], 1, max)
df1$X_min <- apply(df1[, 2:5], 1, min)
df1$X_q50 <- apply(df1[, 2:5], 1, quantile, probs=0.5)
df1$X_q25 <- apply(df1[, 2:5], 1, quantile, probs=0.25)
df1$X_q75 <- apply(df1[, 2:5], 1, quantile, probs=0.75)
df2$X_max <- apply(df2[, 2:5], 1, max)
```

```
df2$X_min <- apply(df2[, 2:5], 1, min)
df2$X_q50 <- apply(df2[, 2:5], 1, quantile, probs=0.5)
df2$X_q25 <- apply(df2[, 2:5], 1, quantile, probs=0.25)
df2$X_q75 <- apply(df2[, 2:5], 1, quantile, probs=0.75) #calculation of minimum, maximum,median, and interquartile range for each population group
datafinal <- rbind(df1, df2) #obtaining the final database by combining the first two
datafinal <- within(datafinal, {condition <- as.factor(condition)})
model1 <- train(condition~X1+X2+X3+X4, data=datafinal, method="glm", family=binomial(link=logit),trControl=data_ctrl, metric = "ROC", na.action = na.pass) #creating a logistic regression model with repeated 10-fold cross-validation, the model in this example is the classic one (LR)
roc1 <-roc(model1$pred$obs, model1$pred$diseased) #generating the ROC curve
a1 <- coords(roc1, "best", ret=c("tpr", "tnr", "youden")) #calculating the performance of the ROC curve (here along with the Youden index are also specified the Se and the Sp)
youden1[i] <- a1$youden[1] - 1
auc_roc1[i] <- roc1$auc #specifying the values of the parameters (Youden index and AUC value) for each ROC curve
model2 <- train(condition~X_max+X_min, data=datafinal, method="glm", family=binomial(link=logit),trControl=data_ctrl, metric = "ROC", na.action = na.pass)
roc2 <-roc(model2$pred$obs, model2$pred$diseased)
a2 <- coords(roc2, "best", ret=c("tpr", "tnr", "youden"))
youden2[i] <- a2$youden[1] - 1
auc_roc2[i] <- roc2$auc
model3 <- train(condition~X_max+X_min+X_q50, data=datafinal, method="glm", family=binomial(link=logit),trControl=data_ctrl, metric = "ROC", na.action = na.pass)
roc3 <-roc(model3$pred$obs, model3$pred$diseased)
a3 <- coords(roc3, "best", ret=c("tpr", "tnr", "youden"))
youden3[i] <- a3$youden[1] - 1
auc_roc3[i] <- roc3$auc
model4 <- train(condition~X_max+X_min+X_q50+X_q25+X_q75, data=datafinal, method="glm", family=binomial(link=logit),trControl=data_ctrl, metric = "ROC", na.action = na.pass)
roc4 <-roc(model4$pred$obs, model4$pred$diseased)
a4 <- coords(roc4, "best", ret=c("tpr", "tnr", "youden"))
youden4[i] <- a4$youden[1] - 1
auc_roc4[i] <- roc4$auc
}
mean(youden1)
sd(youden1)
mean(auc_roc1)
sd(auc_roc1)
mean(youden2)
sd(youden2)
mean(auc_roc2)
sd(auc_roc2)
mean(youden3)
sd(youden3)
mean(auc_roc3)
sd(auc_roc3)
mean(youden4)
sd(youden4)
mean(auc_roc4)
sd(auc_roc4) #obtaining the averages and the standard deviations of the Youden indexes and the AUC values for each of the four logistic models
```

*Evaluation of the Approach on Real Data*

The algorithm was evaluated on real data using the raw data of Ciocan et al. [34,35]. The objective was to compare the performance of the combined biomarkers models with univariate ROC curves for the prediction of colorectal cancer metastasis using lymphocyte (Lim), monocytes (Mon), total proteins (TProt), and albumin (Alb). Limited performances of univariate models were previously reported (lymphocyte-to-monocyte ratio-LMR=(Absolute Lymphocyte Count)/(Absolute Monocyte Count) and Prognostic Nutritional Index - PNI = 10 × Serum Albumin (g/dL) + 0.005 × Lymphocyte Count (per mm3))) [35]. The data-set used contained information from 1688 patients, with the outcome as metastasis (present/absent as 418/1270) and numerical value for biomarkers [34].

The biomarkers (Lim, Mon, TProt, and Alb) were normalized in each group (with and without metastasis) using the STANDARDIZE function in Microsoft Office Excel (Microsoft Office 365). The z-score for each value was used in the analysis. Maximum, minimum, median, and interquartile ranges were computed, and logistic regression models were applied using the resubstitution and the 10-fold cross-validation methods. The AUC values with 95% confidence intervals along with the Se, Sp, diagnostic odds ratio (DOR) [36], Efficiency Index (EI) [37], and Inefficiency Index (InI) [38] were calculated and are reported. We used the clinical utility index [39] to classify the performance of the diagnostic test for case-finding (*rule-in*) or screening (*rule-out*).

## Results

*Simulated Data*

The mean values of AUCs and Youden index varied with the sample size, method, number of biomarkers, and covariance matrix, regardless of the mean of biomarkers in the populations (Figures 2 to 6; SupplementaryMaterial – Tables S1-S5). As expected, the highest dispersion is observed on small sample size and cross-validation (Figures 2 to 6; Supplementary material – Tables S1-S5).

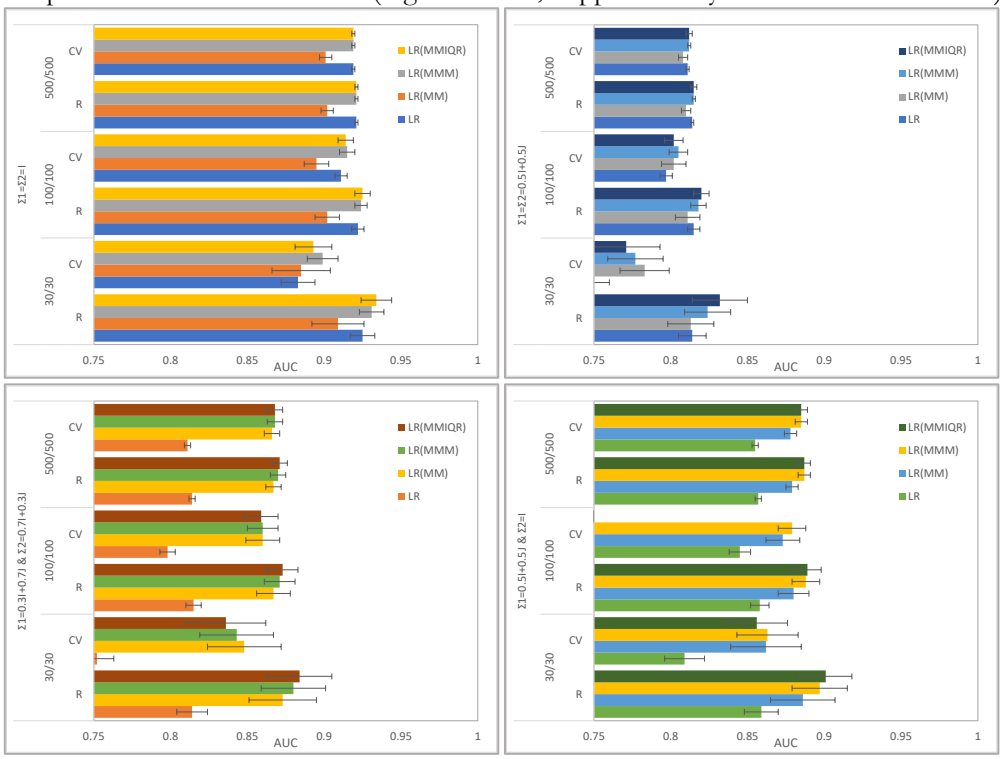


**Figure 2.** Distribution of AUCs for 100 random samples with normally distributed biomarkers and equal means: four biomarkers. (R = resubstitution method, CV = cross-validation method. The bar is the value of the mean and the whiskers are the values of the standard deviation. I = identity matrix, J = matrix of all ones; LR = logistic regression; M=min; M=max; M=median; IQR = interquartile range; AUC = area under the curve)

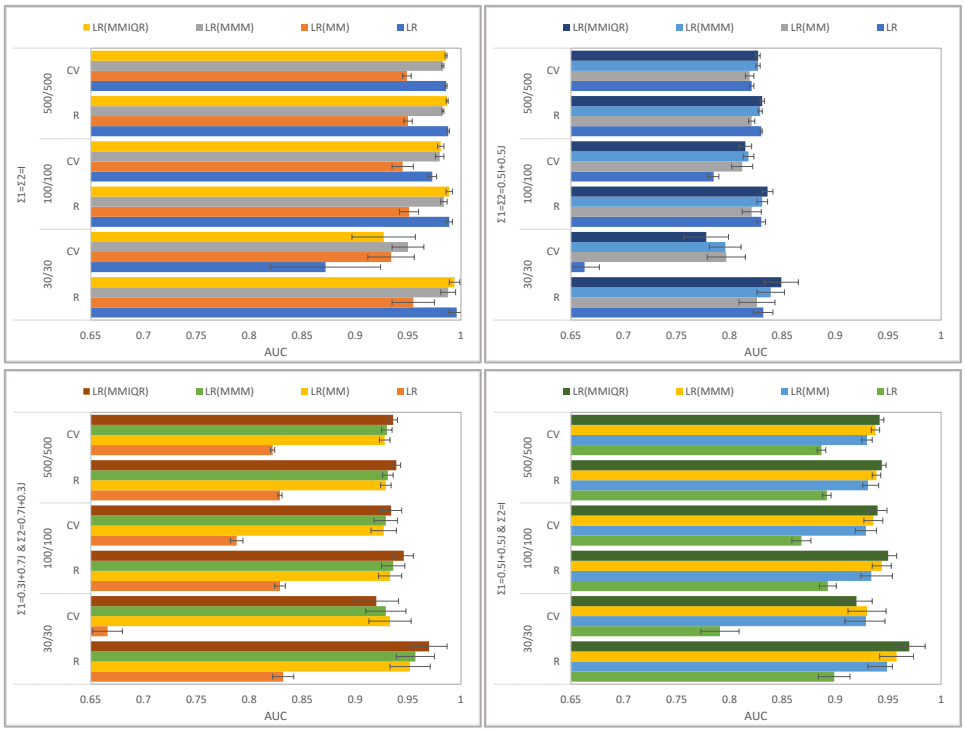


**Figure 3.** Distribution of AUCs for 100 random samples with normally distributed biomarkers and equal means: ten biomarkers. (R = resubstitution method, CV = cross-validation method. The bar is the value of the mean and the whiskers are the values of the standard deviation. I = identity matrix, J = matrix of all ones; LR = logistic regression; M=min; M=max; M=median; IQR = interquartile range; AUC = area under the curve)

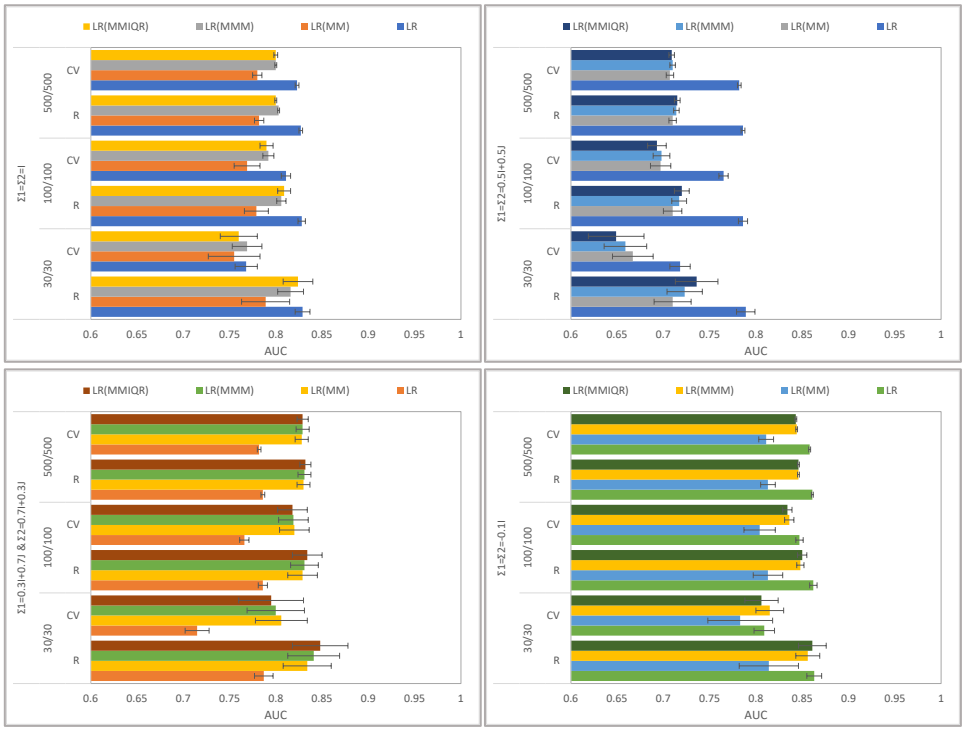


**Figure 4.** Distribution of AUCs for 100 random samples of normally distributed biomarkers with different means: four biomarkers. (R = resubstitution method, CV = cross-validation method. The bar is the value of the mean and the whiskers are the values of the standard deviation. I = identity matrix, J = matrix of all ones; LR = logistic regression; M=min; M=max; M=median; IQR = interquartile range; AUC = area under the curve)

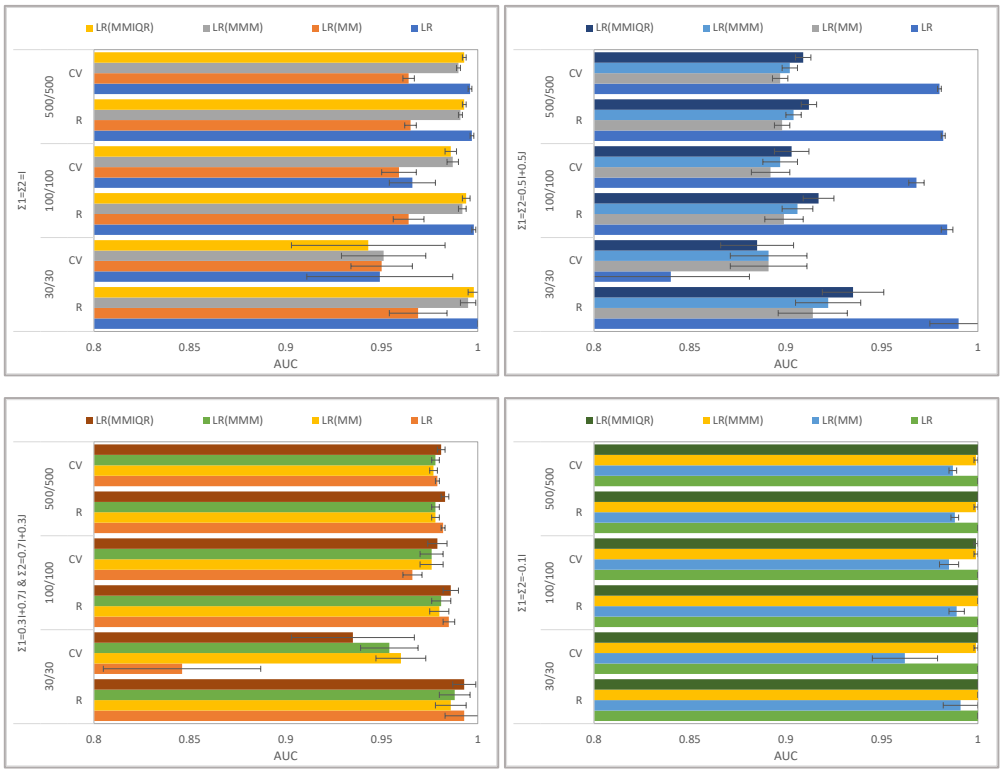


**Figure 5.** Distribution of AUCs for 100 random samples of normally distributed biomarkers with different means: ten biomarkers. (R = resubstitution method, CV = cross-validation method. The bar is the value of the mean and the whiskers are the values of the standard deviation. I = identity matrix, J = matrix of all ones; LR = logistic regression; M=min; M=max; M=median; IQR = interquartile range; AUC = area under the curve)

The increase of the number of normally distributed biomarkers at 100 (equal means of biomarkers), led, as expected, to overfitting when sample sizes are small (30/30), excepting LR and LR(MM) in case of resubstitution method (Figure 6). In the cross-validation, in most of the cases, the LR(MM), LR(MMM), LR(MMIQR) models outperform the LR model.

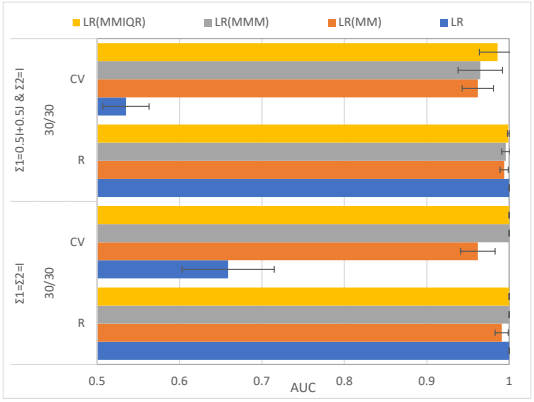


**Figure 6.** Distribution of AUCs for 100 random samples with equal means of 100 normally distributed biomarkers with equal means. (R = resubstitution method, CV = cross-validation method. The bar is the value of the mean and the whiskers are the values of the standard deviation. I = identity matrix, J = matrix of all ones; LR = logistic regression; M=min; M=max; M=median; IQR = interquartile range; AUC = area under the curve)

*Real Data*

The univariate ROC curves obtained after normalizing the biomarker's values show that the lymphocyte counts (Lim) have the highest performance (Table 2). The resubstitution and the cross-

validation method had similar results when combining the biomarkers (Table 2), with the highest Youden index in the LP model (0.194 for resubstitution method and 0.173 for cross-validation method). The lowest performance is observed in the LR(MM) model in both methods, while the LR model has the best prediction performances (Table 2).

**Table 2.** Metric of diagnostic performance by methods and models obtained in real-data analysis

| Method | Biomarker /Model | AUC [95% CI] | Se (%) | Sp (%) | DOR | EI | InI |
|---|---|---|---|---|---|---|---|
| **Univariable ROC curve** | **Lim** | 0.600 [0.567-0.633] | 45.45 | 72.44 | 2.2 | 1.9 | 0.5 |
| | **Mon** | 0.523 [0.490-0.556] | 30.14 | 75.98 | 1.4 | 1.8 | 0.5 |
| | **TProt** | 0.524 [0.492-0.555] | 51.43 | 53.85 | 1.2 | 1.1 | 0.9 |
| | **Alb** | 0.532 [0.501-0.563] | 74.16 | 31.96 | 1.3 | 0.7 | 1.4 |
| **Combined biomarkers Resubstitution** | **LR** | 0.613 [0.581-0.646] | 49.52 | 69.92 | 2.3 | 1.8 | 0.5 |
| | **LR(MM)** | 0.560 [0.529-0.592] | 54.78 | 56.53 | 1.6 | 1.3 | 0.8 |
| | **LR(MMM)** | 0.566 [0.534-0.597] | 50.95 | 59.52 | 1.5 | 1.3 | 0.7 |
| | **LR(MMIQR)** | 0.566 [0.534-0.597] | 48.08 | 62.04 | 1.5 | 1.4 | 0.7 |
| **Combined biomarkers Cross-validation** | **LR** | 0.606 [0.595-0.616] | 46.86 | 70.50 | 2.1 | 1.8 | 0.5 |
| | **LR(MM)** | 0.554 [0.544-0.564] | 58.61 | 51.42 | 1.5 | 1.1 | 0.9 |
| | **LR(MMM)** | 0.557 [0.547-0.567] | 49.40 | 59.52 | 1.4 | 1.3 | 0.8 |
| | **LR(MMIQR)** | 0.553 [0.543-0.563] | 42.44 | 65.88 | 1.4 | 1.5 | 0.7 |

Lim = lymphocyte; Mon = monocytes; Tprot = total proteins; Alb = albumin; LR = regression model; MM = X_min, X_max; MMM = X_min, X_max, X_q50; MMIQR = X_min, X_max, X_Q25, X_75; AUC = area under the ROC curve; CI = confidence interval; Se = Sensitivity; Sp = Specificity; DOR = diagnostic odds ratio; EI = Efficiency Index; InI = Inefficiency Index

In both methods applied to combine the biomarkers, only the LR model performs similarly to the Lim. The LR(MMIQR) method in cross-validation exhibits similar performances in ruling-in and ruling-out as the LR mode (Figure 7). However, poor ability in case finding (rule-in) and fair performances in screening (rule-out) are observed, showing limited clinical relevance.

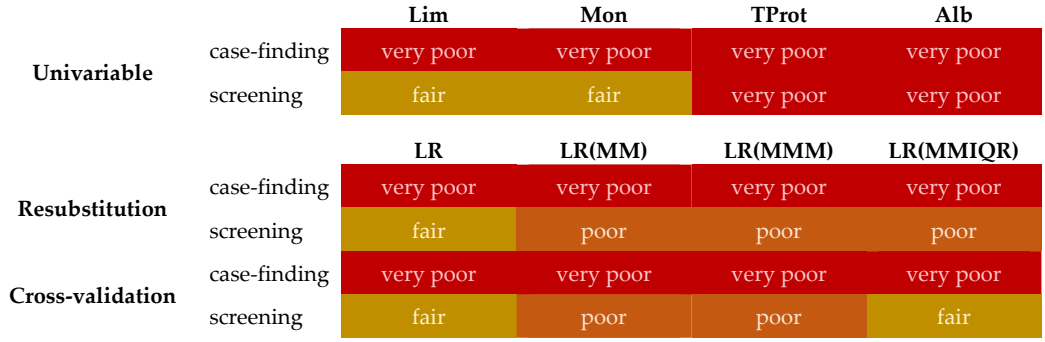


**Figure 7.** Performances based on clinical utility of the biomarker(s) for case-finding (rule-in) and for screening (rule-out). (Lim = lymphocyte; Mon = monocytes; Tprot = total proteins; Alb = albumin; LR = regression model; MM = X_min, X_max; MMM = X_min, X_max, X_q50; MMIQR = X_min, X_max, X_Q25, X_75)

## Discussion

Our simulation study shows different performances according to the imposed scenario (covariance matrixes, method- substitution vs. cross-validation, and sample size), generally with high performances of LR and LR(MMIQR) models. As expected, including more biomarkers in the model leads to an increase in model performances. On real-data, the LR models showed performances outside the clinical relevance, with the outperformance of the LR model, in both resubstitution and cross-validation, all the proposed approaches (LR(MM), LR(MMM) and LR(MMIQR) models), and with best performances obtained by the LR(MMIQ) models.

We implemented a method similar to Aznar-Gimeno et al. [32] to increase the diagnostic accuracy of biomarkers based on their statistics (minimum, maximum, median, and interquartile range) compared to the classical logistic regression model. We incorporated the summary statistics in a logistic regression model, opposite to a step-wise approach [32], and we used 10-fold repeated cross-validation instead of simple cross-validation [32]. The use of smaller random samples in our study could explain our slightly lower performances compared to Aznar-Gimeno et al. [32].

In simulated data, two similarities were observed when equal and unequal means of the biomarkers were investigated (Figures 2-5). First, as expected, increasing the number of biomarkers increases the performances, especially for the proposed models (LR(MM), LR(MMM), and LR(MMIQR)). Second, in terms of performances defined as high AUC and high Youden index, the worst model is LR(MM), and the best performing model is LR model (Figures 2-5, Tables S1-S4).

Specific behavior could be observed when the same mean of biomarkers was imposed (Figures 2 and 3). The best performant model is the LR(MMIQR) when the sample size is small (30/30), and the resubstitution method is used. As the sample size increased, the LR / LR(MMM) or LR(MMIQR) showed equivalent performances. In the cross-validation, the LR(MM) / LR(MMM) / LR(MMIQR) models perform slightly better than LR; in general, the best performant model is the LR(MMIQR) but with a different pattern in different populations sizes.

The analysis of the results for different means of the biomarkers (Figures 4 and 5), a situation closer to what we would have found in clinical practice, showed the following: 1) the LR model has highest performances regardless the sample size, the number of biomarkers or the validation method in the independence scenario ($\Sigma1 = \Sigma2 = I$) or the medium correlation scenario ($\Sigma1 = \Sigma2 = 0.5*I + 0.5*J$); 2) the highest performances of the LR(MMIQR), different covariance matrices scenario ($\Sigma1=0.3*I+0.7*J$, $\Sigma2=0.7*I+0.3*J$) for four (resubstitution and cross-validation) and ten biomarkers (cross-validation); 3) the outperformance of the LR model in the negative correlation scenario ($\Sigma1 = \Sigma2 = -0.1*I$), but with comparable performances of the LR(MMIQR) model; 4) a tendency to overfit, especially for the LR model in the independence ($\Sigma1 = \Sigma2 = I$) and the negative correlation scenarios ($\Sigma1 = \Sigma2 = -0.1*I$).

Our results on simulated data (Figures 2-6, Tables S1-S4) are generally similar to those reported by Aznar-Gimeno et al. [32] regarding Youden values and the tendency of the models. The slightly lower performances of the performances reported in our study compared to those reported by Aznar-Gimeno et al. [32], show that the lower number of random samples (from 1,000 to 100 in our study) and the elimination of the step-wise approach in selecting the models did not have an impact on the prediction performances. Our study reports higher performances on the ten biomarkers summary statistics-based models in the different correlation scenarios with different means as compared to Aznar-Gimeno et al. [32]. We also investigated the 10 biomarkers in negative correlation scenarios with different means compared to Aznar-Gimeno et al. [32], the scenario that generally leads to overfitting of the models (Figure 5). Overfitting is also observed when the number of combined biomarkers increased to 100 in the scenario of equal means of biomarkers in the independence and medium correlation scenarios (Figure 6).

Our results on real data (Table 2) demonstrate the superiority of LR model with similar performances of LR(MMM) and LR(MMIQR) models and the lowest performance of the LR(MM) model. Our result is similar to the results reported by Aznar-Gimeno et al. [32] on Duchenne muscular dystrophy data-set and small for gestational age data-set, when four biomarkers are combined, with higher performance of LR model and similar performances of MMM and MMIQR models. The similarity of summary statistics models in our real data set could be explained by the similar performances in univariable models of the investigated sample (Table 2). The classification level based on AUC, DOR (the odds of metastasis is 2.3 times greater for resubstitution and 2.1 for cross-validation – LR model), EI (1.8, where 1 is the inflection point and denotes the same accuracy and inaccuracy [37,38]), and InI (0.5 for combined biomarkers any method, the value closest to zero indicating less inaccuracy – false positive and false negative results [37,38]) shows the absence of clinical relevance (Table 2), a result similar to those reported by Ciocan et al. [35]. The reported results showed that the combination of biomarkers had only fair abilities in ruling-out (LR model, Figure 7), so without clinical relevance and utility.

The domain of diagnostic accuracy is investigated by researchers all over the world due to its relevance in clinical practice. Gerke and Zapf [40] demonstrated using a simulation environment with a heuristic algorithm that the optimal cut-off point is unbiased for a disease prevalence of 0.5, with positively biased for a prevalence < 0.5 and negatively biased when disease prevalence is > 0.5. Gerke and Zapf [40] also reported a convergence of the optimal cut-off to the true values when the sample size exceeds 1,000 subjects. Ciocan et al. [41] also reported that when a ROC model is developed on at least 70% of the total available population ($n$>1,000), the performances of the models are similar to the full-model. Remaley et al. [42] introduced a different method of evaluating diagnostic accuracy (in comparison to the ROC curve) – prevalence-value-accuracy plots (PVA) in which the effect of disease prevalence, false positives, false negatives, and accuracy are incorporated in the diagnostic test performance [42].

The combinations of biomarkers or their summary statistics are not novel, similar to the use of cross-validation and logistic regression. Borowiak and Reed [43] proposed a systematic algorithm to combine two diagnostic tests by minimizing false positive and false negative results. They demonstrated a lower error rate when combining Doppler ultrasound and pneumoplethysmography to diagnose severe carotid stenosis [43]. Esteban et al. [44] proposed a step-by-step method for combining multiple biomarkers in multivariate normal and non-normal scenarios using linear combinations that maximize the AUC value and demonstrate the similarity with the LR model. In a multivariate normal distribution model, Pinksy and Zhu [45] showed that additional biomarkers negatively correlated with the primary marker can increase diagnostic accuracy. In contrast, biomarkers positively correlated with the primary marker have a smaller added value [45].

Liu et al. [31] used a linear combination of minimum and maximum values of biomarkers, to increase AUCs detected in simulated ROC curves. Aznar-Gimeno et al. [32] demonstrated the outperformance of the summary-based statistics models (using min-max-median / min-max-IQR models) compared to a min-max approach, although the differences between the two proposed models were not significant. Biggs et al. [46] combined different rapid diagnostic tests used to identify the dengue virus (IgM- Immunoglobulin M, IgG, and NS1- dengue virus nonstructural protein 1) to increase the diagnosis of immune status using logistic regression. They demonstrated the superiority of biomarkers combination [46].

Several limitations of our study must be highlighted. First, we used restricted criteria on simulated data-sets that are not necessarily seen in clinical practice (e.g., equal means of biomarkers and normal distribution), using mainly the AUCs and Youden index values, inducing an overestimation. Generation of the biomarkers to closely reflect the real-data and known proportion of disease would capture the clinical reality and utility more appropriately. However, such a strategy is not a *holy grail* in true AUCs estimation since disease prevalence affects the test performances [47-49] and varies from one population to another (e.g., metastases in patients with colorectal cancer 24.8% [35], 30.0-31.0% [50], 33.0% [51], from 22.6 to 41.1% of lung metastasis [52], 83.3% liver metastasis on patients ≥ 70 years-n=210 [53]). A solution would be to evaluate the performance of the diagnosis methods using other metrics (e.g., "likelihood to be diagnosed or misdiagnosed" [54], BEI- balanced EI - efficiency index and UEI- unbiased EI) [38,55]), some of them were previously used in our study when we evaluated real data. Second, we investigate only scenarios based on multivariate normal distribution, so the results of the proposed approach must also be assessed on distribution-free scenarios considering that in real life, raw biomarkers data do not necessarily respect the theoretical normal distribution, interactions between different biomarkers, non-linear association between biomarkers and the outcome variable etc. All above-listed factors have an impact on the estimation of regression coefficients, impacting the accuracy of the model. Third, we did not consider the presence of other covariates that could reflect and affect test performances [56-59]. Fourth, using the logistic regression method could be seen as a weakness of our study because the results are not in the biomarkers space but in the probability space. However, this problem could be solved by an implementation of a valid model in an application to assist the medical staff in daily practice.

Although the proposed methods showed performances in simulated data-sets, no clinical relevance was observed in the evaluated real-data set. The evaluation of the proposed approach on other real data-sets could appropriately assess the reliability of LR linear combination of summary statistics in an increase of the diagnosis accuracy. Once reliable and accurate combination of

biomarkers is identified (characterized by an AUC with 95% confidence boundaries exceeding 0.9, DOR as larger as possible, EI as larger as possible, InI as closest to zero as possible, positive likelihood ratio (PLR) higher than 10, negative LR smaller than 0.1, a high value of positive and negative clinical utility index) [25,37,38,39]), the translation towards clinical practice (e.g. model integration into an online application) could transfer the knowledge towards daily medical practice.

## Conclusion

Linear combination of summary-based statistics tested in logistic regression models showed performances in in-creasing the diagnostic accuracy, with higher performances, similar to the LR model, of minimum-maximum-interquartile range model (LR(MMIQR)) and lower performances of minimum-maximum regression model (LR(MM)) on the identification of metastasis in patients with colorectal cancer. The proposed approach needs evaluation on other real-data sets to appropriately assess the clinical relevance of the proposed approach also considering the presence of covariates.

## Data Availability Statement

Data from the Ciocan et al. study can be found at https://doi.org/10.6084/m9.figshare.19398641.v1, accessed on 19 January 2023.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## Abbreviations

| | |
|---|---|
| AUC | Area Under the ROC Curve |
| C.I. | Confidence Interval |
| CV | Cross-Validation |
| IQR | Interquartile Range |
| LR | Logistic Regression |
| MM | Min-Max Approach |
| MMIQR | Min-Max-IQR Approach |
| MMM | Min-Max-Median Approach |
| NLR | Neutrophil-to-Lymphocyte Ratio |
| dNLR | Derived Neutrophil-to-Lymphocyte Ratio |
| PLR | Platelet-to-Lymphocyte Ratio |
| PVA | Prevalence-value-accuracy Plot |
| ROC | Receiver Operating Characteristic |

## References

1. Šimundić, A.M. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009;19(4):203–211.
2. Balogh EP, Miller BT, Ball JR. (Eds.) The Diagnostic Process [Internet]. Improving Diagnosis in Health Care. National Academies Press (US); 2015 [cited 2022 Dec 2]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK338593/

3. Koffijberg H, van Zaane B, Moons KG. From accuracy to patient outcome and cost-effectiveness evaluations of diagnostic tests and biomarkers: an exemplary modelling study. BMC Med Res Methodol. 2013;13(1):12. https://doi.org/10.1186/1471-2288-13-12
4. Califf RM. Biomarker definitions and their applications. Exp Biol Med. 2018;243(3):213–221. https://doi.org/10.1177/1535370217750088
5. Barański K, Schlünssen V. The Accuracy of a Screening Tool in Epidemiological Studies-An Example of Exhaled Nitric Oxide in Paediatric Asthma. Int J Environ Res Public Health. 2022;19(22):14746. https://doi.org/10.3390/ijerph192214746.
6. Xu T, Fang Y, Rong A, Wang J. Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy. BMC Med Res Methodol. 2015;15(1):94. https://doi.org/10.1186/s12874-015-0085-z
7. Mann T, Gupta RK, Reeve BW, Ndlangalavu G, Chandran A, Krishna AP, et al. Blood RNA biomarkers for tuberculosis screening in people living with HIV prior to anti-retroviral therapy initiation: A diagnostic accuracy study. medRxiv [Preprint]. 2023;2023.06.01.23290783. https://doi.org/10.1101/2023.06.01.23290783.
8. Sun L, Tu H, Chen T, Yuan Q, Liu J, Dong N, Yuan Y. Three-dimensional combined biomarkers assay could improve diagnostic accuracy for gastric cancer. Sci Rep. 2017;7:11621. https://doi.org/10.1038/s41598-017-12022-1.
9. Lycke M, Ulfenborg B, Kristjansdottir B, Sundfeldt K. Increased Diagnostic Accuracy of Adnexal Tumors with A Combination of Established Algorithms and Biomarkers. J Clin Med. 2020;9:299. https://doi.org/10.3390/jcm9020299
10. Su J, Liu J. Linear combinations of multiple diagnostic markers. J Am Stat Assoc. 1993;88:1350–1355.
11. Pepe M, Thompson M. Combining diagnostic test results to increase accuracy. Biostatistics. 2000;1:123–140.
12. Yin J, Tian L. Optimal linear combinations of multiple diagnostic biomarkers based on Youden index. Stat Med. 2014;33:1426–1440.
13. Hua J, Tian L. Combining multiple biomarkers to linearly maximize the diagnostic accuracy under ordered multi-class setting. Stat Methods Med Res. 2021;30(4):1101–1118. https://doi.org/10.1177/0962280220987587.
14. Fong Y, Yin S, Huang Y. Combining biomarkers linearly and nonlinearly for classification using the area under the ROC curve. Stat Med. 2016;35(21):3792-809. https://doi.org/10.1002/sim.6956.
15. Xu T, Fang Y, Rong A, Wang J. Flexible combination of multiple diagnostic biomarkers to improve diagnostic accuracy. BMC Med Res Methodol. 2015;15:94. https://doi.org/10.1186/s12874-015-0085-z
16. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry. 1993;39(4):561–577.
17. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J Radiol. 2004;5(1):11–18. https://doi.org/10.3348/kjr.2004.5.1.11
18. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3
19. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol. 2006;163(7):670-675. https://doi.org/10.1093/aje/kwj063.
20. Miller R, Siegmund D. Maximally selected chi square statistics. Biometrics. Journal of the Biometric Society. 1982;38(4):1011–1016. https://doi.org/10.2307/2529881.
21. Liu X. Classification accuracy and cut point selection. Statistics in Medicine. 2012;31(23):2676–2686. https://doi.org/10.1002/sim.4509.
22. Unal I. Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. Comput Math Methods Med. 2017;2017:3762651. https://doi.org/10.1155/2017/3762651.
23. López-Ratón M, Rodríguez-Álvarez MX, Cadarso-Suárez C, Gude-Sampedro F. Optimal Cutpoints: An R package for selecting optimal cutpoints in diagnostic tests. J Stat Softw. 2014;61:1–36. https://doi.org/10.18637/jss.v061.i08
24. Nahm FS. Receiver operating characteristic curve: overview and practical use for clinicians. Korean J Anesthesiol. 2022;75(1):25–36. https://doi.org/10.4097/kja.21209.

25. Bolboacă SD. Medical Diagnostic Tests: A Review of Test Anatomy, Phases, and Statistical Treatment of Data. Comput Math Methods Med. 2019;2019:1891569. https://doi.org/10.1155/2019/1891569.
26. Reibnegger G, Schrabmair W. Optimum binary cut-off threshold of a diagnostic test: comparison of different methods using Monte Carlo technique. BMC Med Inform Decis Mak. 2014;14:99. https://doi.org/10.1186/s12911-014-0099-1.
27. Su JQ, Liu JS. Linear Combinations of Multiple Diagnostic Markers. J Am Stat Assoc. 1993;88(424):1350–1355. https://doi.org/10.2307/2291276.
28. Pepe, M.S, Thompson, M.L. Combining diagnostic test results to increase accuracy. Biostatistics. 2000, 1(2), 123–140. https://doi.org/10.1093/biostatistics/1.2.123.
29. Kang L, Xiong C, Crane P, Tian L. Linear combinations of biomarkers to improve diagnostic accuracy with three ordinal diagnostic categories. Stat Med. 2013;32(4):631–643. https://doi.org/10.1002/sim.5542
30. Chen X, Vexler A, Markatou M. Empirical likelihood ratio confidence interval estimation of best linear combinations of biomarkers. Comput Stat Data Anal. 2015;82:186–198. https://doi.org/10.1016/j.csda.2014.09.010.
31. Liu C, Liu A, Halabi S. A min-max combination of biomarkers to improve diagnostic accuracy. Stat Med. 2011;30(16):2005–2014. https://doi.org/10.1002/sim.4238.
32. Aznar-Gimeno R, Esteban LM, Sanz G, del-Hoyo-Alonso R, Savirón-Cornudella R. Incorporating a New Summary Statistic into the Min–Max Approach: A Min–Max–Median, Min–Max–IQR Combination of Biomarkers for Maximising the Youden Index. Mathematics. 2021;9(19):2497. https://doi.org/10.3390/math9192497.
33. Ma H, Halabi S, Liu A. On the use of min-max combination of biomarkers to maximize the partial area under the ROC curve. J Probab Stat. 2019;2019:8953530. https://doi.org/10.1155/2019/8953530.
34. [dataset] Ciocan A, Bolboacă SD, Ciocan RA, Zaharie VF, Graur F, Puia CI, Al Hajjar N. Colorectal Cancer: Some Variables Commonly Measured in Clinical Practice. figshare. 2022 [cited May 5, 2023] Available from: https://doi.org/10.6084/m9.figshare.19398641.v1
35. Ciocan A, Ciocan RA, Al Hajjar N, Gherman CD, Bolboacă S.D. Abilities of Pre-Treatment Inflammation Ratios as Classification or Prediction Models for Patients with Colorectal Cancer. Diagnostics 2021;11:566. https://doi.org/10.3390/diagnostics11030566.
36. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. Journal of Clinical Epidemiology. 2003;56(11):1129–1135. https://doi.org/10.1016/S0895-4356(03)00177-X
37. Larner AJ. Communicating Risk: Developing an "Efficiency Index" for Dementia Screening Tests. Brain Sci. 2021;11(11):1473. https://doi.org/10.3390/brainsci11111473.
38. Larner AJ. Efficiency Index for Binary Classifiers: Concept, Extension, and Application. Mathematics 2023;11:2435. https://doi.org/10.3390/math11112435
39. Mitchell AJ. Sensitivity × PPV is a recognized test called the clinical utility index (CUI+). Eur J Epidemiol. 2011, 26(3):251–252. https://doi.org/10.1007/s10654-011-9561-x.
40. Gerke O, Zapf A. Convergence Behavior of Optimal Cut-Off Points Derived from Receiver Operating Characteristics Curve Analysis: A Simulation Study. Mathematics. 2022;10(22):4206. https://doi.org/10.3390/math10224206.
41. Ciocan A, Hajjar NA, Graur F, Oprea VC, Ciocan RA, Bolboacă SD. Receiver Operating Characteristic Prediction for Classification: Performances in Cross-Validation by Example. Mathematics 2020;8:1741. https://doi.org/10.3390/math8101741
42. Remaleym AT, Sampson ML, DeLeo JM, Remaley NA, Farsi BD, Zweig MH. Prevalence-value-accuracy plots: a new method for comparing diagnostic tests based on misclassification costs. Clin Chem. 1999;45(7):934–941.
43. Borowiak D, Reed JF. Utility of combining two diagnostic tests. Comput Methods Programs Biomed. 1991;35(3): 171–175. https://doi.org/10.1016/0169-2607(91)90119-E.
44. Esteban LM, Sanz G, Borque A. A step-by-step algorithm for combining diagnostic tests. J Appl Stat. 2011;38(5):899–911. https://doi.org/10.1080/02664761003692373.
45. Pinsky PF, Zhu CS. Building Multi-Marker Algorithms for Disease Prediction–The Role of Correlations among Markers. Biomark Insights. 2011;6:BMI.S7513. https://doi.org/10.4137/BMI.S7513

46. Biggs JR, Sy AK, Ashall J, Santoso MS, Brady OJ, Reyes MAJ, et al. Combining rapid diagnostic tests to estimate primary and post-primary dengue immune status at the point of care. PLoS Negl Trop Dis. 2022;16(5):e0010365. https://doi.org/10.1371/journal.pntd.0010365.
47. Choi BCK. Causal Modeling to Estimate Sensitivity and Specificity of a Test When Prevalence Changes. Epidemiology 1997;8(1):80–86.
48. Willis BH. Empirical evidence that disease prevalence may affect the performance of diagnostic tests with an implicit threshold: a cross-sectional study. BMJ Open 2012;2:e000746. https://doi.org/10.1136/bmjopen-2011-000746
49. Bentley TG, Catanzaro A, Ganiats TG. Implications of the impact of prevalence on test thresholds and outcomes: lessons from tuberculosis. BMC Res Notes 2012;5:563. https://doi.org/10.1186/1756-0500-5-563
50. Riihimäki M, Hemminki A, Sundquist J, Hemminki K. Patterns of metastasis in colon and rectal cancer. Sci Rep. 2016;6:29765. https://doi.org/10.1038/srep29765.
51. Väyrynen V, Wirta E-V, Seppälä T, Sihvo E, Mecklin J-P, Vasala K, Kellokumpu I. Incidence and management of patients with colorectal cancer and synchronous and metachronous colorectal metastases: a population-based study, BJS Open 2020;4(4):685–692. https://doi.org/10.1002/bjs5.50299
52. Jördens MS, Labuhn S, Luedde T, Hoyer L, Kostev K, Loosen SH, Roderburg C. Prevalence of Lung Metastases among 19,321 Metastatic Colorectal Cancer Patients in Eight Countries of Europe and Asia. Curr Oncol. 2021;28(6):5035–5040. https://doi.org/10.3390/curroncol28060423.
53. Nassabein R, Mansour L, Richard C, Vandenbroucke-Menu F, Aubin F, Ayoub J-P, et al. Outcomes of Older Patients with Resectable Colorectal Liver Metastases Cancer (CRLM): Single Center Experience. Curr. Oncol. 2021;28:1899–1908. https://doi.org/10.3390/curroncol28030176
54. Larner AJ. Evaluating cognitive screening instruments with the "likelihood to be diagnosed or misdiagnosed" measure. Int. J. Clin. Pract. 2019;73:e13265. https://doi.org/10.1111/ijcp.13265.
55. Larner AJ. Evaluating binary classifiers: extending the Efficiency Index. Neurodegener. Dis. Manag. 2022;12:185–194. https://doi.org/10.2217/nmt-2022-0006.
56. Janes H, Pepe MS. Adjusting for Covariates in Studies of Diagnostic, Screening, or Prognostic Markers: An Old Concept in a New Setting. American Journal of Epidemiology 2008;168(1):89–97. https://doi.org/10.1093/aje/kwn099.
57. Lewis F, Sanchez-Vazquez MJ, Torgerson PR. Association between covariates and disease occurrence in the presence of diagnostic error. Epidemiology and Infection 2012;140(8):1515–1524.
58. Pardo-Fernandez JC, Rodrıguez-Alvarez MX, Van Keilegom I. A review on ROC curves in the presence of covariates. Statistical Journal 2014;12(1):21–41.
59. Roldán-Nofuentes JA. Simultaneous Comparison of Sensitivities and Specificities of Two Diagnostic Tests Adjusting for Discrete Covariates. Mathematics 2021;9:2029. https://doi.org/10.3390/math9172029