PharmaSUG China 2023 - Paper CC-158

AFROC Analysis in SAS®:
AFROC curves, AFROC-AUC, and average AFROC curves

Pan Zhemin, Tong Ji University, Shanghai;
Qin Yingyi, Navy Medical University, Shanghai;
Cao Ruiting, Siemens Healthineers, Shanghai;
Wang Lei, Siemens Healthineers, Shanghai
He Jia, Tong Ji University & Navy Medical University, Shanghai;

ABSTRACT

Alternative free-response receiver operating characteristic (AFROC) curves have become increasingly popular in the field of medical imaging due to their ability to illustrate the uncertainty in lesion location as well as allowing for multiple detections per case. However, up till now, AFROC curves and area under the AFROC curve (AFROC AUC) are mostly be obtained through R package or other applications. In this paper, we present a novel approach in SAS® that can plot AFROC curves, calculate AFROC AUC, and generate mean AFROC curves.

The proposed approach is demonstrated using a real dataset from a multi-reader multi-case (MRMC) design study, and the results are compared with those obtained using R. The proposed SAS code provides an alternative and convenient way to generate AFROC curves, and can benefit researchers and practitioners in the field of medical imaging by enabling a more comprehensive and realistic evaluation of the performance of medical imaging systems.

INTRODUCTION

Medical imaging is a critical component of modern healthcare, providing clinicians with non-invasive tools to diagnose and monitor diseases. The accuracy of the medical imaging system is a key factor of their effectiveness, and several statistical methods have been developed to evaluate their performance. One such method is the receiver operating characteristic (ROC) curve. In traditional ROC methods, readers assign confidence ratings to each case and determine whether or not it is diseased, without the need to locate the lesion. However, for imaging diagnostic tests which involving the detection and localization of multiple lesions, using the ROC method may result in potential bias. For instance, if patients are diagnosed as diseased but the locations of the lesions are inaccurately identified, it can possibly result in an over-estimation of sensitivity.

To address this limitation, extended methods of ROC have been proposed, including localization ROC (LROC) methods, region-of-interest (ROI) methods, and free-response ROC (FROC) methods. Among them, FROC studies are more flexible, as they consider lesions as the basic diagnostic unit and allow readers to freely mark all suspicious regions. The AFROC method is an improvement on FROC analysis, by replacing the vertical axis from mean number of false positive markers per case to 1-specificity, the AFROC AUC were then limited within 0 to 1, and this is beneficial for comparing multiple diagnostic tests.

Currently, the analysis of AFROC curves and AFROC AUC in clinical practice and research is limited to R packages such as RJafroc. This limitation might hinder the widespread adoption of the AFROC method in medical imaging area. Therefore, this article aims to provide a comprehensive overview of the AFROC method and proposes a novel approach in SAS® that enables the generation of AFROC curves, AFROC AUC, and mean AFROC curves through simple data step, PROC LOGISTIC, and PROC SGPLOT. The proposed approach is demonstrated using a real dataset from a two-modality five reader MRMC design study that aims to evaluate the diagnostic accuracy of dual-view digital mammography (DM) and single-view breast tomosynthesis (BT) (Svahn et al 2011). Furthermore, we recommend a standardized dataset format for studies evaluating imaging diagnostic tests, including MRMC design studies. This standardized format can easily be derived to calculate important metrics such as AFROC, ROC, sensitivity, specificity, etc.

COMPONENTS OF AFROC CURVE

FROC PARADIGM

In contrast to the ROC paradigm, which evaluates the performance at patient level, the FROC paradigm considers all lesions detected. It requires the reader to not only judge whether there is any abnormality in the case, but also to specify and locate an arbitrary number of defects on an image along with their confidence rating for each defect. Thus, under FROC paradigm the data collected were lesion level information, which consists of two parts: lesion localization (LL) - those correctly localized markers of lesions, and non-lesion localization (NL) - those that do not correspond to any lesion.

ROC, FROC AND AFROC

Table 1 summarizes the comparison between ROC, FROC, and AFROC curves in terms of their vertical and horizontal axes, while Figure 1 displays examples of each curve type.

	Vertical Axis	Horizontal Axis
ROC	True Positive Fraction (TPF) probability of actual positive cases be called positive $\frac{n(diseased\ cases\ rated \geq threshold\ \xi)}{n(all\ diseased\ cases(truth))}$	False Positive Fraction (FPF) probability of actual negative cases be called positive $\frac{n(non\ diseased\ cases\ rated\ \geq threshold\ \xi)}{n(all\ non\ diseased\ cases(truth))}$
FROC	Lesion localization Fraction (LLF) fraction of true positive decisions with correct localization $\frac{n(LLs\ rated \geq threshold\ \xi)}{n(all\ lesions(truth))}$	Non-Lesion localization Fraction (NLF) mean number of false positive markers per case $\frac{n(NLs\ rated \geq threshold\ \xi)}{n(all\ cases - truth)}$
AFROC	Lesion localization Fraction (LLF) fraction of true positive decisions with correct localization $\frac{n(LLs\ rated \geq threshold\ \xi)}{n(all\ lesions(truth))}$	(inferred) False Positive Fraction (FPF) probability of actual negative cases be called positive based on the inferred case rating $\frac{n(non\ diseased\ cases\ rated\ \geq threshold\ \xi)}{n(all\ non\ diseased\ cases(truth))}$

Table 1 The Comparison of ROC, FROC, AFROC in Terms of Vertical Axis and Horizontal Axis

n=the total number of objects that satisfy the given condition;

LL=lesion localization; NL=non-lesion localization.

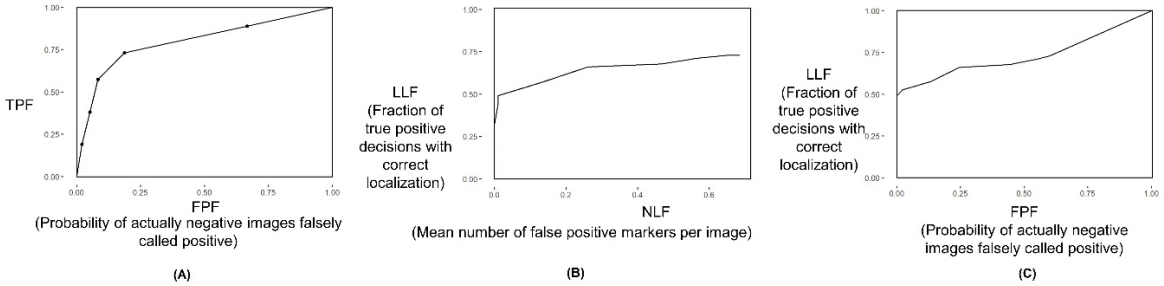

Figure 1 an example of ROC Curve, FROC curve and AFROC Curve

(A)ROC Curve, (B) FROC Curve, (C) AFROC Curve

Back to the topic of this paper, the AFROC curve, is defined as the plot of LLF along the ordinate vs FPF along the abscissa for each decision threshold.

In comparison to FROC, the AFROC plot has identical ordinates LLF, but a different horizontal axis that change from NLF to FPF. This change constrains the AFROC endpoint within the unit square and allows for easier comparison of different modalities.

When comparing AFROC and ROC curves, the horizontal axis is identical, with the only slight difference being that for AFROC curves, FPF is an inferred value based on the maximum rating of a normal case to represent the case-level rating. Meanwhile, the vertical axis of AFROC and ROC plots are different but analogous. As illustrated in Figure 2, if we replace the diseased case information (case level) in the source data with lesion information (lesion level), then the calculation of TPF returns the same value as LLF. This modification transforms the ROC curves into AFROC curves.

Based on this, in this paper, we proposed a standardized data format suitable for imaging diagnostic tests evaluation studies using the FROC paradigm. And building on the analogy between ROC and AFROC, we developed an approach for analyzing AFROC in SAS® by adopting the idea of ROC analysis.

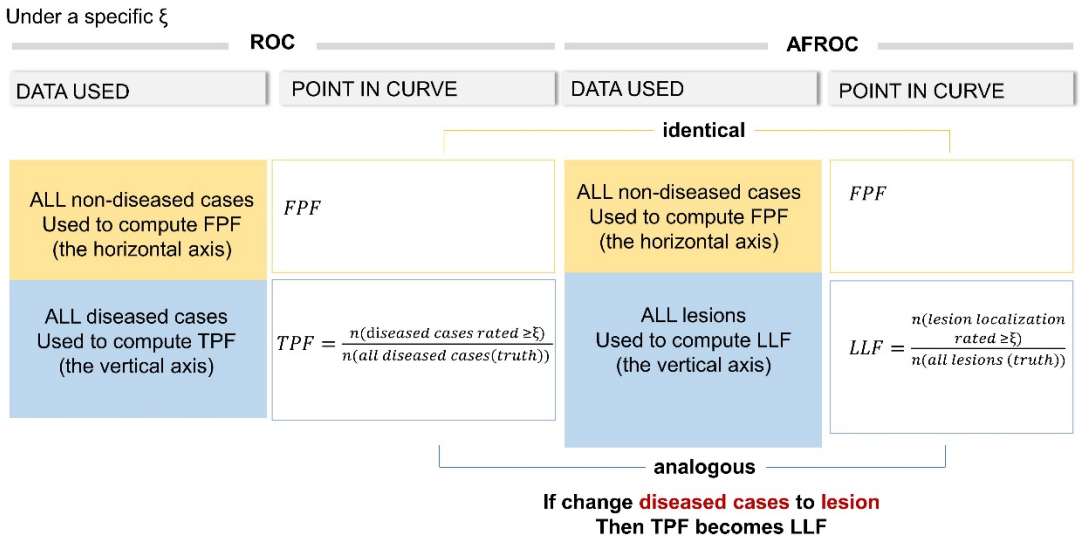

Figure 2 the difference ROC and AFROC under a specific threshold ξ

AFROC ANALYSIS IN SAS®

DATA SOURCE

To demonstrate the codes more effectively, we utilized a real dataset from a MRMC design diagnostic accuracy study (Svahn et al 2011). The primary objective of this study was to evaluate the accuracy of breast tomosynthesis (BT, modality 1) and digital mammography (DM, modality 2). The study involved 185 cases, among which 89 were identified as diseased. Five experienced radiologists assessed all cases using the FROC paradigm. In this paradigm, all suspicious lesions were identified and marked, along with a confidence rating for each mark ranging from 1 (representing most likely non-diseased) to 10 (representing most likely diseased). The raw data for our analysis was obtained from the RJafroc package (version 2.1.1, dataset01) (Dev Chakraborty et al., 2022).

DATA PREPARATION – STANDARDIZED DATA FORMAT

Prior to conducting the analysis, we implemented a meticulous data preparation procedure. This involved aligning the reading results of each radiologist with the corresponding truth (golden standard) data. Additionally, we derived case-level and lesion-level flags, which might be used to calculate the sensitivity and specificity. We suggest organizing data in the this standardized format (Table 2) when analysising a FROC paradigm study to ensure the accuracy and completeness of data analysis, considering it allows for

all relevant information to be presented together and can easily be derived to facilitates the calculation of important metrics such as AUC, AFROC, sensitivity, and specificity.

To be specific, this data format presented readers' reading results on the left side and lesion-matched truth information on the right side. It included all markers identified by readers and corresponding gold standard, and for non-diseased cases with zero markers under certain reader-modality pairs, a standalone line is derived to indicate the case-level true negative information, and the confidence rating of this line would be derived to 0.

Lesion-level flags were generated based on each individual lesion. If both the reader (ReaderYN=1) and the gold standard (TruthYN=1) identified a lesion, it was labeled as a true positive marker at the lesion level. On the other hand, if the reader identified a lesion (ReaderYN=1) but the gold standard did not (TruthYN=0), it was considered a false positive at the lesion level. Similarly, if the reader did not identify a lesion (ReaderYN=0) but the gold standard did (TruthYN=1), it was categorized as a lesion level false negative and the confidence rating be derived to 0.

Case-level flags were determined for cases under each reader-modality combination. For diseased cases, if there was at least one true positive lesion identified, the case was marked as a true positive under this reader-modality pair. Otherwise, if no true positive lesions were found, it was labeled as a case level false positive. For non-diseased cases, if there were no false positive lesion markers, the case was marked as a true negative. However, if any false positive lesion markers were present, it was categorized as a false negative at the case level.

Detail data steps for generating the proposed data format please refer to Appendix 3.

	Readers Reading Results					Lesion Matched Truth			Derived Variable	
	ReaderID	ModalityID	CaseID	Rating	ReaderYN	LesionID	TruthYN	GROUP (1=nondiseased, 2=diseased)	CaseFL (derived)	LesionFL (derived)
Non diseased	1	1	11	4	1	.	0	1	CaseFP	LesionFP
	1	1	11	3	1	.	0	1	CaseFP	LesionFP
	.....									
	1	1	34	0(derived)	0		0	1	CaseTN	
	.....									
Diseased	1	1	175	0(derived)	0	1	1	2	CaseFN	LesionFN
	.....									
	1	1	177	3	1	.	0	2	CaseTP	LesionFP
	1	1	177	3	1	1	1	2	CaseTP	LesionTP
	1	1	177	3	1	2	1	2	CaseTP	LesionTP
	.....									

Table 2 Proposed Standardized Data Format

FP=false positive; TP=true positive; FN=false negative; FN=false negative.

ADD AFROC FLAG FOR FURTHER ANALYSIS

To conduct the AFROC analysis, it is necessary to include all lesions (truth) to compute the LLF portion of the AFROC curve.

While for the computing of FPF portion, the non-diseased cases need to be included as one record per case with the corresponding maximum lesion rating as the case rating.

By this process the dataset for analysis will be one record per case for non-diseased cases and one record per lesion (truth) for diseased cases under each reader-modality pair.

	ReaderID	ModalityID	CaseID	Rating	ReaderYN	LesionID	TruthYN	GROUP (1=nondiseased, 2=diseased)	CaseFL (derived)	LesionFL (derived)	AFROCFL
Non diseased	1	1	11	4	1	.	0	1	CaseFP	LesionFP	Y
	1	1	11	3	1	.	0	1	CaseFP	LesionFP	
	.....										
	1	1	34	0(derived)	0		0	1	CaseTN		Y
	.....										
	1	1	175	0(derived)	0	1	1	2	CaseFN	LesionFN	Y
	.....										
Diseased	1	1	177	3	1	.	0	2	CaseTP	LesionFP	
	1	1	177	3	1	1	1	2	CaseTP	LesionTP	Y
	1	1	177	3	1	2	1	2	CaseTP	LesionTP	Y
	.....										

Table 3 Derived Data For AFROC Analysis (with AFROCFL="Y")

AFROC ANALYSIS

Utilizing the dataset generated in previous step, the AFROC AUC can be computed through the implementation of proc logistic. The process of calculating the AFROC AUC is analogous to that of ROC AUC, with the exception of a modification in the ordinates and abscissa.

For illustrative purposes, let us consider the calculation of the AFROC AUC specifically for reader 1 under modality 1:

```
*calculate AFROC AUC(ReaderID=1 and ModalityID=1);

ods graphics on;

proc logistic data=TONY_AFROC(where=(AFROCFL="Y" and ReaderID=1 and
ModalityID=1)) plots(only)=roc (id=prob);

        model TruthYN(event="1") = Rating;

        ods output Association = Association;

        run;
```

	Label1	cValue1	nValue1	Label2	cValue2	nValue2
1	Percent Concordant	68.6	68.574561	Somers' D	0.518	0.518311
2	Percent Discordant	16.7	16.743421	Gamma	0.608	0.607505
3	Percent Tied	14.7	14.682018	Tau-a	0.261	0.260513
4	Pairs	9120	9120.000000	c	0.759	0.759156

Figure 3 the Association dataset

The Association dataset contains the AFROC AUC value for ReaderID=1 and ModalityID=1, denoted as "c". By changing the ReaderID and ModalityID in above codes, the AFROC AUC can be obtained for each reader and modality combination.

ITEM	AFROC AUC SAS Result	AFROC AUC RJafroc Result
ReaderID=1, ModalityID=1	0.7591557018	0.7591557
ReaderID=2, ModalityID=1	0.8375000000	0.8375000
ReaderID=3, ModalityID=1	0.8131030702	0.8131031
ReaderID=4, ModalityID=1	0.8071271930	0.8071272
ReaderID=5, ModalityID=1	0.8277960526	0.8277961

ITEM	AFROC AUC SAS Result	AFROC AUC RJafroc Result
ReaderID=1, ModalityID=2	0.6406798246	0.6406798
ReaderID=2, ModalityID=2	0.7034539474	0.7034539
ReaderID=3, ModalityID=2	0.7359100877	0.7359101
ReaderID=4, ModalityID=2	0.7694078947	0.7694079
ReaderID=5, ModalityID=2	0.6805921053	0.6805921

Table 4 Results from SAS vs Results from RJafroc Package

As a summary, in the first part "COMPONENTS OF AFROC CURVE" we introduced the theoretical feasibility of converting ROC curves into AFROC curves by replacing the diseased case information in the source data with lesion information. In this part, we used real FROC paradigm data and standardized it into a format where non-diseased cases kept one record, while diseased cases kept one record per lesion (truth) in the dataset (by using the AFROC Flag). Then we leveraged proc logistic which is commonly used for ROC analysis to conduct AFROC analysis.

Our results demonstrated that the AFROC analysis using the proposed approach in SAS produced yielded identical results to those conducted in RJafroc package (version 2.1.1) by R. This consistency verifies the feasibility and validity of our proposed approach.

PLOT AFROC CURVES AND MEAN AFROC CURVES

PLOT AFROC CURVES

By adding `plots(only)=roc (id=prob)` to above codes, the raw AFROC curve for each reader under each modality can be acquired. While the coordinate labels were still indicated for ROC curves by default.

To plot the AFROC curves of all reader-modality pairs in one figure as well as to modify the label of ordinates and abscissa, below codes can be used.

```
data TONY_AFROC2;set TONY_AFROC;if AFROCFL="Y";run;
proc sort data= TONY_AFROC2 out=TONY_AFROC_MEAN;
    by ModalityID ReaderID;
    run;
ods graphics on;
proc logistic data=TONY_AFROC_MEAN plots(only)=roc (id=prob);
   by ModalityID ReaderID;
   model TruthYN(event="1") = Rating;
    ods output roccurve=ROCdata;
    run;
proc sort data=rocdata out=rocdata; by ReaderID;run;
proc sgplot data=ROCdata aspect=1;
   xaxis label="FPF" values=(0 to 1 by 0.25)
   grid offsetmin=.05 offsetmax=.05;
   yaxis label="LLF" values=(0 to 1 by 0.25)
   grid offsetmin=.05 offsetmax=.05;
   lineparm x=0 y=0 slope=1 / transparency=0.3 lineattrs=(color=black);
```

```
series x=_1mspec_ y=_sensit_ / group= ModalityID ;
run;
```

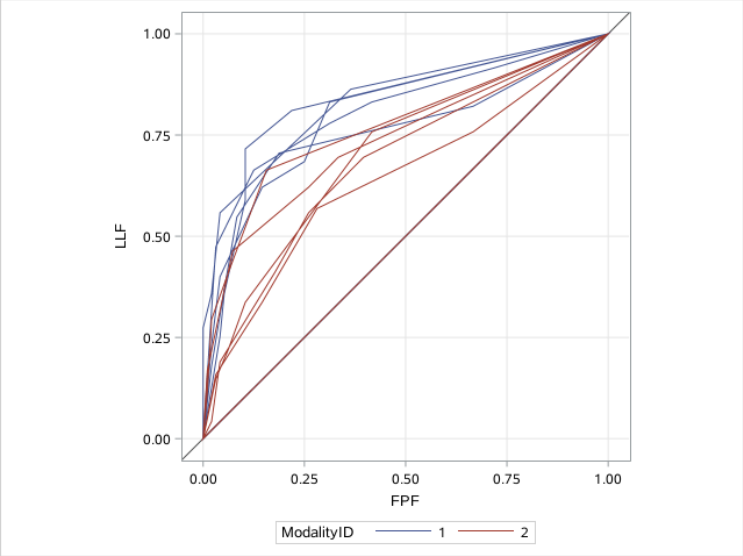

Figure 4 AFROC Curves

PLOT MEAN AFROC CURVES

In certain medical scenarios, it may be advantageous to plot mean AFROC curves over readers in order to better illustrate the differences between modalities. To facilitate the creation of such curves, we adopted the idea from RJafroc package (Dev Chakraborty et al., 2022): Dividing each AFROC curve into 1001 equally spaced points along the FPF axis, spanning the range from 0.000 to 1.000, and then calculating the mean LLF values for each point by averaging the LLF values of all readers under each modality. By connecting these 1001 points the mean AFROC curves for each modality can be obtained.

To be specific, as the original AFROC curves are empirical, we utilized the coordinates of each two adjacent points on the original AFROC curve to fit a line and obtain the corresponding LLF value for each FPF point (from 0.000 to 1.000). In some cases, one FPF value may correspond to multiple LLF values. To ensure a conservative result, the smallest LLF value would be selected in this situation. Finally, we added the coordinate origin (LLF=0, FPF=0) to both curves and plotted the mean AFROC curves for each modality.

This process allowed us to more clearly compare the performance of different modalities in detecting lesions, thereby enabling more informed decision-making in medical practice.

```
***to gain the points for Empirical AFROC curves;
data TONY_AFROC_MEAN;set TONY_AFROC;if AFROCFL="Y";run;
proc sort data=TONY_AFROC_MEAN out=TONY_AFROC_MEAN;
    by ModalityID ReaderID;run;
ods graphics on;
proc logistic data=TONY_AFROC_MEAN plots(only)=roc (id=prob);
   by ModalityID ReaderID;
         model TruthYN(event="1") = Rating;
         ods output roccurve=ROCdata_old;
   run;
proc sort data=ROCdata_old;
```

```
    by ModalityID ReaderID  _1MSPEC_ _SENSIT_;
    run;
data ROCdata_old;
   set ROCdata_old;
   by  ModalityID ReaderID _1MSPEC_;
   if first.ModalityID or first.ReaderID or first._1MSPEC_  then output;
    *Considering sometimes one _1MSPEC_ may corresponding to several
    _SENSIT_, thus we here limited the data by "if first._1MSPEC_ then
    output", in order to utilize the smallest _SENSIT_ if several _SENSIT_
    owns the same _SENSIT_;
    run;
***get the correspond _SENSIT_;
data ROCdata;
   set ROCdata_old;
         _SENSIT_1=lag(_SENSIT_);
         _1MSPEC_1=lag(_1MSPEC_);
    run;
data ROCdata2;
   set ROCdata;
   by ModalityID ReaderID;
         if first.ModalityID or first.ReaderID then delete;
    run;
data roc_new;
   set ROCdata2;
   do i=0 to 1000;
         x=i/1000;
         if x>_1MSPEC_1 and x<=_1MSPEC_ then output;
   end;
    run;
data roc_new;set roc_new;
   y=_SENSIT_1+(_SENSIT_-_SENSIT_1)*(x-_1MSPEC_1)/(_1MSPEC_-
   _1MSPEC_1);run;
proc sort data=roc_new;by ModalityID  x;run;
proc means data=roc_new noprint;
   var y;
   by ModalityID x notsorted;
   output out=roc_final mean=y_mean;run;
***add point x=0, y=0;
data _a;
```

```
   ModalityID=1 ;x=0.000; y_mean=0.000;run;
data _b;
   ModalityID=2 ;x=0.000; y_mean=0.000;run;

data roc_final;set _a _b roc_final;
   attrib ModalityID label="Modality";
         if ModalityID=1 then Modality="BT";
         if ModalityID=2 then Modality="DM";run;
***plot mean AFROC curves;
proc sgplot data=roc_final aspect=1;
    xaxis label="FPF" values=(0 to 1 by 0.25)
    grid offsetmin=.05 offsetmax=.05;
    yaxis label="LLF" values=(0 to 1 by 0.25)
    grid offsetmin=.05 offsetmax=.05;
     lineparm x=0 y=0 slope=1 / transparency=0.3 lineattrs=(color=black);
     series x=x y=y_mean / group=Modality ;
     run;
```

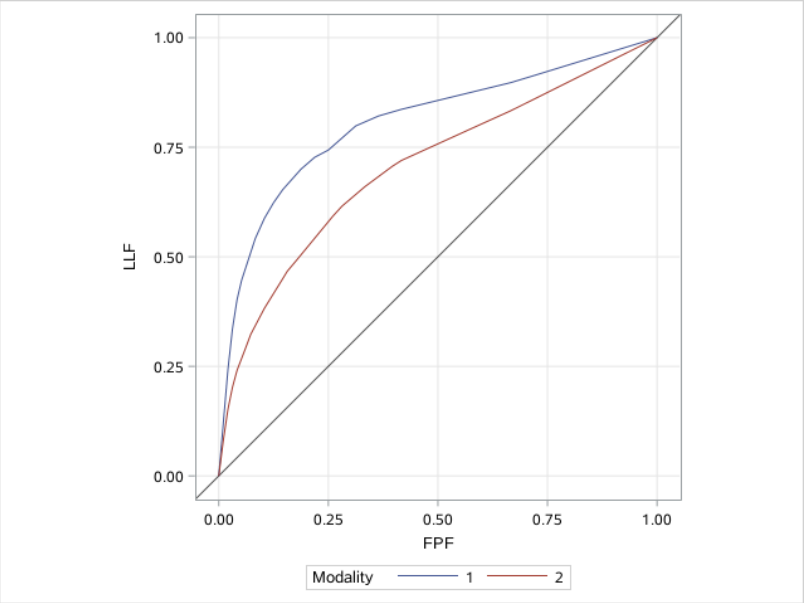

Figure 5 Mean AFROC Curves

CONCLUSION

This paper has introduced the components of AFROC curves and provided SAS code that enables the analysis of AFROC, including the plotting of curves, the calculation of AFROC AUC, and the generation of mean AFROC curves. Additionally, we recommended a standardized data format for the analysis of imaging diagnostic tests evaluation studies under FROC paradigm.

The proposed approach has the potential to facilitate the use of AFROC curves in clinical practice. And by providing a standardized data format and SAS code, we established a feasible and efficient workflow to

evaluate lesion-detection performances by AFROC in SAS. We hope this work would encourage the more widespread use of the AFROC and enable more accurate comparisons of imaging modalities.

In the meantime, in the evaluation of medical diagnostic test performance the calculation of AFROC AUC is the first step, and based on this further statistical inference including interval estimation and hypothesis testing are necessary. Recent years, MRMC design are encouraged for imaging diagnostic tests evaluation studies due to its ability to account for the variability of both readers and cases, and corresponding statistical inference methods like OR and DBM methods have been developed. However, the statistical inference of AFROC AUC based on SAS®, especially under the MRMC design, is an area that still requires further exploration. Our team has developed a SAS macro special for the statistical inference of MRMC design, and are currently under internal validation. Hope to share these works with fellow professionals in the field in near future.

Overall, we believe that our approach has the potential to improve the use of AFROC curves in clinical practice and advance the field of medical imaging.

REFERENCES

Svahn T, Andersson I, Chakraborty D, et al.2011. “The diagnostic accuracy of dual-view digital mammography, single-view breast tomosynthesis and a dual-view combination of breast tomosynthesis and digital mammography in a free-response observer performance study ”. Radiat Prot Dosimetry,139(1-3):113-117.

Dev Chakraborty, Peter Phillips, Xuetong Zhai. 2022. “RJafroc: Artificial Intelligence Systems and Observer Performance”. Accessed Oct 7, 2022. https://cran.r-project.org/web/packages/RJafroc/index.html

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Name: Pan Zhemin
Enterprise: Tong Ji University
E-mail: zm_pan@tongji.edu.cn

Name: Qin Yingyi
Enterprise: Navy Medical University
E-mail: yingyi_qin@163.com

Name: Cao Ruiting
Enterprise: Siemens Healthineers
E-mail: RuiTing.Cao@siemens-healthineers.com

Name: Wang Lei
Enterprise: Siemens Healthineers
E-mail: lwwang@siemens-healthineers.com

Name: He Jia
Enterprise: Tong Ji University & Navy Medical University
E-mail: hejia63@yeah.net

ACKNOWLEDGMENTS

The methods are developed during the experiment with Siemens Healthineers on their AI-Rad Companion (Pulmonary) product.

APPENDIX

Appendix 1 AFROC.TONY datasets (contains all LL and NL markers)

	CaseID	LesionID	Rating	ReaderID	ModalityID
1	97	1	4	1	1
2	98	1	4.5	1	1
3	99	1	5	1	1
4	102	1	4	1	1
5	103	1	3.5	1	1
6	104	1	3	1	1
7	106	1	3.5	1	1
8	107	1	3.5	1	1
9	108	1	4	1	1
10	109	1	4.5	1	1


tony.sas7bdat

Appendix 2 AFROC.TRUTH datasets (contains the truth information for each case)

	CaseID	LesionID	Weight	TruthYN	GROUP
1	1	0	0	0	1
2	2	0	0	0	1
3	3	0	0	0	1
4	4	0	0	0	1
5	5	0	0	0	1
6	6	0	0	0	1
7	7	0	0	0	1
8	8	0	0	0	1
9	9	0	0	0	1
10	10	0	0	0	1
11	11	0	0	0	1
12	12	0	0	0	1
13	13	0	0	0	1
14	14	0	0	0	1
15	15	0	0	0	1
16	16	0	0	0	1


truth.sas7bdat

Appendix 3 Data steps to generate the proposed data format (based on source data listed in Appendix 1 and Appendix 2).

- Step 1, we listed all findings of each reader under each modality for each case and derived a binary variable "ReaderYN" to indicate whether the reader identified a lesion in a specific case or not.

```
*add ReaderYN variable;

data AFROC.TONY;
        set AFROC.TONY;
        attrib ReaderYN length=3;
        ReaderYN=1;
*The variable ReaderYN represents whether the reader has identified the
lesion or not(ReaderYN=1:identified, ReaderYN= 0:unidentified).In current
dataset, each observation is identified by a reader, and as a result, all
the values of ReaderYN are currently set to 1;
        run;
```

- Step 2, we processed the truth information to prepare for the further match with readers` reading results.

```
*add variable to truth dataset;

data AFROC.TRUTH;set AFROC.TRUTH;

   attrib TruthYN length=3; *The variable TruthYN represents whether the
lesion was actually detected or not.TruthYN=0(not detected in truth),
TruthYN=1 (detected in truth);
```

```
    if LesionID=0 then TruthYN=0;else TruthYN=1;
   attrib GROUP length=3 ;*GROUP=1 : the case is health, GROUP=2 : the
case is diseased;
    if LesionID=0 then GROUP=1;else GROUP=2;
   run;
*extend the truth dataset to each reader under each modality;
   data NEW_TRUTH;
         set AFROC.TRUTH;
           do ReaderID=1 to 5 ;
               do ModalityID=1 to 2 ;
                     output;
               end;
           end;
         run;
data TRUTH_POS;set NEW_TRUTH;*save the positive truth lesions for
<match1>;
         if LesionID^=0;
         drop GROUP; *GROUP information will be added in <combination2>;
         run;
data TRUTH_NoDup;set NEW_TRUTH;
  keep ReaderID ModalityID CaseID GROUP;
  run;*save the group information for each case, extended to each reader
  and modality for <match 2>;
proc sort data=TRUTH_NoDup out=TRUTH_NoDup nodup;
   by ReaderID ModalityID CaseID;
    run;
```

- Step3, We matched the reading results with the truth based on lesions of each case by each reader under each modality. This matching process was conducted in two sub-steps:

Step 3.1, We matched the reading result with the diseased part of the truth to identify false positive(FP), true positive (TP), and false negative(FN) markers. (Note: For FP, TP and FN markers, one lesion one record)

```
*<match 1>;
*By using this combination, we can identify the ratings of False Positive,
True Positive, and False Negative.;
proc sort data=AFROC.TONY out=TONY;
   by ReaderID ModalityID CaseID LesionID;
   run;
proc sort data=TRUTH_POS;
   by ReaderID ModalityID CaseID LesionID;
```

```
 run;
data TONY1;
 merge  TONY TRUTH_POS;
 by ReaderID ModalityID CaseID LesionID;
 run;
```

Step 3.2, We matched the dataset generated in step 3.1 with the `TRUTH_NoDup` dataset which contains information about whether each case was diseased or not (case level).

During this step, we additionally identified the true negative (TN) cases which is not included in Step 3.1. (note: for true negative cases, one case one record).

This means if ReaderID=1 under modality1 did not find any marker in a non-disease case, originally there would be no records under FROC paradigm. And in this step, we generated a record where ReaderYN=0 and TruthYN=0 to indicate the case level true negative result. (Note: for true negative cases, one case one record).

```
 *<match 2>;
*When combined with TRUTH_NoDup, it includes the rating of True Negative;
   proc sort data=TONY1 out=TONY1;
    by  ReaderID ModalityID CaseID;
    run;
proc sort data=TRUTH_NoDup out=TRUTH_NoDup;
    by  ReaderID ModalityID CaseID;
    run;


data TONY2;
 merge TONY1 TRUTH_NoDup;
 by  ReaderID ModalityID CaseID;
 run;
```

For the unmatched pairs and true negative cases, we derived the corresponding missing Rating, TruthYN and ReaderYN to 0.

```
data TONY2;set TONY2;
      if Rating=. then Rating=0;
      if TruthYN=. then TruthYN=0;
      if ReaderYN=. then ReaderYN=0;
      run;
```

- Step 4, We also added optional case level flags (TP, FP, TN, FN) and lesion level flags (TP, FP, FN) that can be of help for calculating other figures of merit (FOM) besides AFROC.

```
*lesion Level flag;
data  TONY2;set TONY2;
      attrib CaseFL length=$20;
      attrib LesionFL length=$20;
attrib match  length=3;*Intermediate variable that identifies the case
```

```
level flag;
      if  TruthYN=1 and ReaderYN=1 then LesionFL="LesionTP";
      if  TruthYN=1 and ReaderYN=0 then LesionFL="LesionFN";
      if  TruthYN=0 and ReaderYN=1 then LesionFL="LesionFP";
      if LesionFL="LesionTP" then match=1;else match=0;
      run;
*case level flag;
proc sql;
  create table TONY2NEW as
select CaseID, LesionID, Rating, ReaderID, ModalityID,  ReaderYN,
TruthYN, GROUP, CaseFL, LesionFL,sum(match) as matchNEW,sum (ReaderYN)
as ReaderYNSUM from TONY2
      group by ModalityID,ReaderID,CaseID;
      quit;
data TONY2NEW;set TONY2NEW;
      if GROUP=1 and ReaderYNSUM=0 then CASEFL="CaseTN";
      if GROUP=2 and matchNEW>0 then  CASEFL="CaseTP";
      if GROUP=2 and matchNEW=0 then CASEFL="CaseFN";
      if GROUP=1 and ReaderYNSUM^=0 then CASEFL="CaseFP";
      drop matchNEW ReaderYNSUM;
      run;
```

- Step 5, add AFROC Flag for Further Analysis

```
*for the vertical axis(LLF);
data TONY2NEW_1;set TONY2NEW;where GROUP=2;
      if TruthYN=1 then AFROCFL="Y";
 run;
*for the horizontal axis(FPF);
data TONY2NEW_2;set TONY2NEW;where GROUP=1;
      proc sort data=TONY2NEW_2 out=TONY2NEW_2;
 by CaseID ModalityID ReaderID desending Rating ;
 run;
data TONY2NEW_2;set TONY2NEW_2;
   by CaseID ModalityID ReaderID ;
   if first.ReaderID then AFROCFL="Y";
 run;
*Combine;
data TONY_AFROC;set TONY2NEW_1 TONY2NEW_2;run;
```