

Improving the diagnostic performance of ultrasound in classifying breast lesions: the potential value of S-detect for residents-in-training

Chenyang Zhao

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital
<https://orcid.org/0000-0001-9389-0652>

Mengsu Xiao

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital

Yuxin Jiang

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital

He Liu

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital

Ming Wang

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital

Hongyan Wang

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital

Jing Zhang

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital

Qingli Zhu (✉ zqlpumch@126.com)

Chinese Academy of Medical Sciences and Peking Union Medical College Hospital
<https://orcid.org/0000-0002-0618-2381>

Research article

Keywords: Breast Neoplasms; Breast Ultrasonography; Computer-Assisted Diagnosis

Posted Date: June 20th, 2019

DOI: <https://doi.org/10.21203/rs.2.10429/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background To explore the potential value of S-Detect™, a high-end computer-assisted diagnosis (CAD) software system for residents-in-training. **Methods** Routine breast ultrasound (US) examinations were conducted and assessed by an experienced radiologist. Archived images of the lesions (including grayscale, color Doppler flow and elastography images) were retrospectively assessed by each of five in-training residents who were blinded to the histopathological findings and any other US diagnosis. The diagnostic performances of S-Detect™ and the five residents were measured and compared. Afterwards, category 4a lesions assessed by the residents were downgraded when classified as possibly benign by S-Detect™. The diagnostic performance of the integrated results was compared with the original results of the residents. **Results** A total of 195 focal breast lesions were consecutively enrolled, including 82 malignant lesions and 113 benign lesions. S-Detect™ presented higher specificity and area under the curve (AUC) than the residents. After combination with S-Detect™ in category 4a lesions, the specificity and AUC of the five residents were significantly improved. The intraclass correlation coefficient (ICC) of the five residents also increased after integration. **Conclusions** With the help of the CAD software, the specificity, overall diagnostic performances and interobserver agreements of the residents greatly improved. S-Detect™ can be utilized as an assistant tool for residents-in-training in classifying breast lesions.

Background

On account of the increasing incidence rate of breast cancer in the past decade, it has become a growing public health concern worldwide¹. Early detection of breast cancer can largely improve patient prognosis^{2,3}. As an important adjunctive tool to mammography, ultrasound (US) has shown great potential for diagnosing breast masses, especially in dense breast tissue, allowing identification of the masses that are occult on mammography⁴. In consideration of the accessibility and cost-effectiveness, US has become the most popular imaging method for breast screening in China, which has also been proved to perform superior or not inferior to mammography.⁵

Nevertheless, low specificity and high interobserver variability remain problematic disadvantages for US, especially for residents who have only received short-term training in breast US⁶⁻⁹. Although the Breast Imaging Report and Data System (BI-RADS) lexicon was put forward by the American College of Radiology¹⁰⁻¹², residents-in-training are still inclined to have relatively poor diagnostic performance when assessing breast lesions¹³. BI-RADS subcategory 4a lesions, presenting a few suspicious features but mainly benign characteristics, can make those inexperienced residents very confused in classifying, and making wrong decisions, leading to subsequent overtreatment.

Computer-aided diagnosis (CAD) systems have played a growing part in many fields of medical imaging, including breast US¹⁴⁻¹⁸. S-Detect™ for Breast is a cutting-edge CAD system that acts as an assistant tool for US imaging diagnosis of breast lesions. The diagnostic efficacy of the CAD software for classifying

breast lesions has been validated by several studies¹⁹⁻²¹. Furthermore, S-Detect™ has been proven to be of value in increasing the diagnostic performance of the in-training residents^{19,22}. BI-RADS 4a lesions posed a potential challenge for breast US. And as far as we know, the feasibility of S-Detect™ in improving the diagnostic accuracy of the residents-in-training in BI-RADS 4a lesions has not been investigated by previous studies.

In this study, we evaluated the diagnostic performance of S-Detect™ and five residents-in-training for classifying breast lesions. The results of the residents were reevaluated after some of the category 4a lesions were downgraded by CAD. The aim of the study was to further explore the potential role of S-Detect™ to aid the in-training readers and determine how this system can help to improve diagnostic performance, especially for BI-RADS category 4a lesions.

Methods

This study was a cross-sectional observational study. The ethics approval of the study has been acquired from the Institutional Review Board of Peking Union Medical College Hospital. Written informed consent was also obtained from the adult patients of the study. And for the patients under 18 year-old, the written informed consent was signed by their guardians who accompanied them to receive US examination.

2.1 Patients and imaging

A total of 195 focal breast lesions from the patients aged between 15-82 years, with a mean age of 45.7 years and a median of 45.0 years, were enrolled consecutively in this study.

The inclusion criteria for the study were as follows:

- (1) palpable masses verified by breast imaging; and
- (2) nonpalpable masses found by breast imaging, with or without other symptoms;

The exclusion criteria were as follows:

- biopsy of the breast lesions performed before US examinations;
- pregnancy, or in lactation;
- neoadjuvant treatment;
- only simple cysts visible on US images;
- no evident focal breast lesions suitable for CAD evaluation

The patients underwent US examinations before they received further treatment. All lesions were biopsied and had a final pathological diagnosis. The pathological results were deemed the gold standard for the study.

The patients received standard bilateral breast US scans performed by an experienced radiologist. A commercial US unit (RS85, Samsung Medison Co., Ltd., Korea) equipped with a L3-12A high-frequency linear probe (3-12 MHz) and the CAD software S-Detect™ for Breast was utilized.

2.2 Study protocol

2.2.1 Image assessment of S-Detect™ for Breast and the five in-training residents

A single grayscale US image demonstrating the lesion with the maximum size was manually selected for S-Detect™ for Breast analysis. First, the radiologist clicked the center of the target mass, and the contour of the lesion was segmented by S-Detect™ automatically. The outline of the lesion was adjusted manually by the radiologist when necessary. Then, the classification of each lesion in a dichotomic form (possibly benign and possibly malignant) was provided by S-Detect™. US descriptors extracted by S-Detect™ were also displayed, including shape, orientation, margins, pattern and posterior acoustic features.

Five in-training residents with 1-3 years of working experience were invited to assess the US lesions independently. All images of the lesions (including grayscale, color Doppler flow and elastography images) were retrospectively reviewed by five in-training residents, and they were asked to classify the lesions based on BI-RADS lexicon. The residents were blinded to S-Detect™ and pathology results. R1-5 was used to represent the five residents. R1, R2 and R3 were third-year residents, and each had one-year of experience with breast US. R4 and R5 were second-year residents, each with six months of experience with breast US. The five residents had all received a standard training program for breast US, and have passed the exams of basic US organized by our medical center.

A cutoff value was set at category 4 to transform the residents' results into a dichotomic form. Category 2 and 3 lesions were deemed as possibly benign, and category 4 and 5 were considered possibly malignant. The diagnostic performances of S-Detect™ and the five residents were evaluated, and comparisons were made between S-Detect™ and the residents.

2.2.2 Integration of the result of the five residents and S-Detect™ for Breast

To evaluate the potential of S-Detect™ to help improve the diagnostic accuracy of residents, the results of the five in-training residents were integrated with those of S-Detect™ in category 4a lesions. We compared the results of S-Detect™ and those of the residents for each lesion. If the lesion was diagnosed as category 4a by the residents but possibly benign by S-Detect™, the decision of S-Detect™ was adopted, thus downgrading category 4a lesions to the possibly benign group. Due to the high sensitivity of the residents presented in the preliminary experiments, we did not change the category 3 lesions when they were classified as possibly malignant by S-Detect™. In addition, the rest of the classifications made by the residents remained unchanged.

Diagnostic performances of the integrated results were calculated, and compared with the original results of the residents without S-Detect™. Interrater variability before and after integration with S-Detect™ was assessed using intraclass correlation coefficients (ICCs).

2.3 Statistical analysis

The diagnostic performances of the residents, S-Detect™ and the integrated results of the residents and S-Detect™ for category 4a lesions were evaluated using the sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV), receiver operating characteristic (ROC) curve and area under the receiver operator characteristics curve (AUC). In addition, 2×2 contingency tables were delineated to measure these indicators. We made comparisons of sensitivity and specificity between residents using the chi-square test. The AUC values were compared using the Z test.

ICC with 95% confidence intervals was calculated to evaluate the interrater variability of multiple raters. In this study, each subject was rated by the same raters, and ICC was deemed the absolute agreement of the raters, as the systematic differences among the raters were relevant. ICC value was interpreted as follows.

Poor agreement: $ICC < 0$

Slight agreement: $0 < ICC < 0.20$

Fair agreement: $0.20 < ICC < 0.40$

Moderate agreement: $0.40 < ICC < 0.60$

Substantial agreement: $0.60 < ICC < 0.80$

Perfect agreement: $0.80 < ICC < 1$

Statistical significance was considered when the p-value was less than 0.05. SPSS software (IBM, SPSS 21.0) and Medcalc (MedCalc software, version 15, Ghent, Belgium) were utilized in the study.

Results

A total of 195 focal breast lesions including 82 malignant lesions and 113 benign lesions from 195 consecutive patients (mean age, 45.7 years; median age, 45.0 (15-82) years) who were referred to the medical center were consecutively enrolled.

The diagnostic performances of S-Detect™ and the five residents, and the Comparisons of the sensitivity, specificity and AUC between S-Detect™ and the residents are listed in Table 1. Table 1 highlights that the residents had an incline in presenting a high sensitivity, but an evidently low specificity. All the residents

showed a relatively high sensitivity (92.68-100.00%). While, the specificity of S-Detect™ (77.88%) was higher than that of R2-5 (19.47-48.67%), with a p value of < 0.05. The AUC value of S-Detect™ (0.82) was significantly higher than those of the five residents (0.62-0.74), with a p value of <0.05 for all the residents, as shown in Table 1. In this study, S-Detect™ had overall better diagnostic performance than the residents-in-training with limited breast US experiences.

The number of downgraded lesions that were classified as category 4a lesions by the residents but possibly benign by S-Detect™ are listed in Table 2. The sensitivity of the integrated results still remained at a relatively high level (92.68-100.00%). The specificities of all the residents were significantly improved after using the results of S-Detect™ (46.02-76.11%), with a p value of <0.001 for all the residents.

The ROC curves of the five residents, S-Detect™ and the residents combined with S-Detect™ are presented in Fig 1-5. From the ROC curves of the residents, we could determine that the curve was elevated at the top left after combination with S-Detect™. Additionally, the AUC value of the residents with S-Detect™ had an evident increase (0.71-0.85), with statistical significance ($p < 0.001$), indicating the improvement of the overall diagnostic performances of the five residents (Table 1).

To evaluate the interobserver variability of the five residents, we calculated the ICC value of the integrate results and original results. Systematic differences among the five raters were found to be relevant after ANOVA ($p < 0.05$), and the ICC was regarded as a measure of absolute agreement. The single measure of ICC of the five residents increased from 0.480 (0.415-0.549) to 0.643(0.586-0.700) after integration with the results of S-Detect™, indicating that the agreement level increased from moderate to substantial.

Discussion

US is one of the most commonly used modalities in breast imaging. As a convenient and cost-effective imaging method, US has played an essential role in the detection and evaluation of breast lesions in many countries, as well as in China²³. However, despite the promotion of BI-RADS lexicon, operator dependence and interobserver variability are still the major flaws of US⁶⁻⁹. The performance of the BI-RADS lexicon can be largely affected by the clinical experiences of the operators. The specificity of a resident-in-training has been reported to be significantly inferior to that of a high-level radiologist, when using the BI-RADS lexicon in the assessment of breast lesions⁸. As a result, methods to enhance diagnostic efficiency of inexperienced readers and to decrease the interobserver variability for breast US findings are in demand.

CAD systems have emerged as powerful tools for medical imaging with the dramatic advancement of artificial intelligence technology¹⁴. The feasibility of using CAD systems¹⁴ to aid in the diagnosis of breast lesions has been verified by previous studies^{24,25}. S-Detect™, which is a dedicated CAD software integrated on a high-end US unit, is constructed based on deep learning algorithms and trained by large-scale clinical databases. The diagnostic process of S-Detect™ is free from the interference of man-identified features. The potential use of S-Detect™ to assist doctors in improving diagnostic performance,

especially for those who lack experience, has been elucidated in previous studies. Ji-Hye Choi and Eun Cho verified that the diagnostic performance of inexperienced readers could be improved with the help of S-Detect™^{26,27}. Mattia Di Segni et al. also suggested that S-Detect™ could act as a teaching tool for in-training residents to improve the accuracy of diagnosing breast lesions¹⁹.

According to the results of our study, S-Detect™ was distinguished by its high specificity, compared with that of the five in-training residents with limited US experience, who presented a remarkable sensitivity but a low specificity. Therefore, we speculated that S-Detect™ could help in improving the residents' specificity. Breast lesions classified into BI-RADS 4a were defined as having a low suspicion of malignancy. In the clinical settings, category 4a is a relatively complicated subgroup of the BI-RADS classifications, of which the malignant rate is 3-10%, and the PPV value is 6%²⁸. In this study, the ratio of malignancy in 4a lesions classified by the five residents were 9.38%, 8.11%, 15.38%, 10.25%, 9.09%, respectively, most of which were within the range defined by the guidelines. Most category 4a lesions are benign, but may undergo unnecessary biopsies. To better address the tradeoff of 4a lesions, new modalities, such as elastography, have been put into clinical use to lower the false-positive rate^{29,30}. And statistically significant improvement in the specificity and AUC was acquired for the residents after using S-Detect™ for category 4a lesions, suggesting that a dedicated CAD system might also provide additional diagnostic information. A CAD system could also be an effective method to downgrade benign category 4a lesions, and reduce unnecessary biopsies. It is noteworthy that the malignant rate of the CAD-downgraded 4a lesions presented in the current study was not satisfying. They were 0%, 9.68%, 7.89%, 6.25%, 4.76% respectively. It implied that the further improvement of S-detect is necessity for the clinical applications.

ICC of the five residents improved after integration with S-Detect™ from a moderate level of agreement to a good level. This result verified that S-Detect™ could also be effective in decreasing interobserver variability in breast US for inexperienced raters.

In the clinical practice, residents are required to undergo systematic training programs before entering clinical work. S-Detect™ can act as a powerful assistant tool to audit the diagnoses made by inexperienced US readers. Notably, the workflow of S-Detect™ is less time-consuming than that of the double reading process. In addition, the US features extracted by S-Detect™ are displayed for readers, providing a useful reference for residents to learn the images case by case, thus S-Detect™ may possess potential value in the training of inexperienced US readers.

There were several limitations in this study. First, the underestimation of the performance of the residents should be mentioned. In a regular US examination flow, radiologists often evaluate a breast lesion based on overall diagnostic information. Apart from dynamic real-time US images, medical history and mammography results, are taken into consideration. While in this retrospective study, only static images were provided for classification to the residents. Moreover, the good performance of S-Detect™ was guaranteed by the high-quality US images used for classification, which were collected by an experienced radiologist who participated in the study. This condition might not be realized in real clinical

settings when utilized in different medical centers in other regions, which may impair the diagnostic performance of S-Detect™.

Conclusions

In this study, S-Detect™ had better diagnostic performance for classifying breast lesions than the five residents. After category 4a lesions were reclassified by S-Detect™, the diagnostic performances of the residents were significantly enhanced, with higher specificity without sacrificing the sensitivity significantly. It is promising for S-Detect™ to improve the specificity of inexperienced readers and avoid unnecessary biopsies of category 4a lesions. S-Detect™ can also help to decrease interobserver variability among different readers.

Abbreviations

Computer-assisted diagnosis (CAD)

ultrasound (US)

Breast Imaging Report and Data System (BI-RADS)

region of interest (ROI)

sensitivity (SE)

specificity (SP)

positive likelihood ratio (PLR)

negative likelihood ratio (NLR)

positive predictive value (PPV)

negative predictive value (NPV)

area under the curve (AUC)

intraclass correlation coefficient (ICC)

Declarations

Ethics approval and consent to participate

The ethics approval of the study has been acquired from the Institutional Review Board of Peking Union Medical College Hospital. Written informed consent was also obtained from the recruited patients or the guardians of the patients under 16 years old.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests

The authors declare that they have no competing interests.

Fundings

This work was funded by CAMS Innovation Fund for Medical Sciences (2017-I2M-1-006), the 2016 Peking Union Medical College education and teaching reform project (2016zlgc0113), the Fundamental Research Funds for the Central Universities (2017320002).

Authors' contributions

QL Zhu and YX Jiang designed the study of the comparison between S-detect and doctors. MS Xiao, H liu, M Wang, HY Wang, J Zhang, QL Zhu and CY Zhao collected the patients' data and performed ultrasound examination and S-detect software. CY Zhao and QL Zhu summarize and analyze all the data and figures of the manuscript, and wrote the original version of the manuscript. All the authors took part in discussing and revising the manuscript. The authors had also read and approved the final manuscript,

Acknowledgements

Not applicable.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA: a cancer journal for clinicians*. 2015;65(1):5-29.
2. Dubey AK, Gupta U, Jain S. Breast cancer statistics and prediction methodology: a systematic review and analysis. *Asian Pacific journal of cancer prevention : APJCP*. 2015;16(10):4237-4245.
3. Tabar L, Vitak B, Chen HH, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiologic clinics of North America*. 2000;38(4):625-651.
4. Brem RF, Lenihan MJ, Lieberman J, Torrente J. Screening breast ultrasound: past, present, and future. *AJR American journal of roentgenology*. 2015;204(2):234-240.
5. Shen S, Zhou Y, Xu Y, et al. A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. *British journal of cancer*. 2015;112(6):998-1004.
6. Abdullah N, Mesurolle B, El-Khoury M, Kao E. Breast imaging reporting and data system lexicon for US: interobserver agreement for assessment of breast masses. *Radiology*. 2009;252(3):665-672.
7. Elverici E, Zengin B, Nurdan Barca A, Didem Yilmaz P, Alimli A, Araz L. Interobserver and Intraobserver Agreement of Sonographic BIRADS Lexicon in the Assessment of Breast Masses. *Iranian journal of radiology : a quarterly journal published by the Iranian Radiological Society*. 2013;10(3):122-127.
8. Lee YJ, Choi SY, Kim KS, Yang PS. Variability in Observer Performance Between Faculty Members and Residents Using Breast Imaging Reporting and Data System (BI-RADS)-Ultrasound, Fifth Edition (2013). *Iranian journal of radiology : a quarterly journal published by the Iranian Radiological Society*. 2016;13(3):e28281.
9. Park CS, Kim SH, Jung NY, Choi JJ, Kang BJ, Jung HS. Interobserver variability of ultrasound elastography and the ultrasound BI-RADS lexicon of breast lesions. *Breast cancer (Tokyo, Japan)*. 2015;22(2):153-160.
10. Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H. A Pictorial Review of Changes in the BI-RADS Fifth Edition. *Radiographics : a review publication of the Radiological Society of North America, Inc*. 2016;36(3):623-639.
11. Radiology ACO. The American College of Radiology Breast Imaging Reporting and Data System (BI-RADS). 2003.
12. Baert AL. Breast Imaging Reporting and Data System (BI-RADS). 2013.
13. Youk JH, Jung I, Yoon JH, et al. Comparison of Inter-Observer Variability and Diagnostic Performance of the Fifth Edition of BI-RADS for Breast Ultrasound of Static versus Video Images. *Ultrasound in medicine & biology*. 2016;42(9):2083-2088.
14. Dromain C, Boyer B, Ferre R, Canale S, Delaloge S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *European journal of radiology*. 2013;82(3):417-423.
15. Chang RF, Wu WJ, Moon WK, Chen DR. Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis. *Ultrasound in medicine & biology*. 2003;29(5):679-686.

16. Chen CM, Chou YH, Han KC, et al. Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology*. 2003;226(2):504-514.
17. Yap MH, Pons G, Marti J, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE journal of biomedical and health informatics*. 2018;22(4):1218-1226.
18. Becker AS, Mueller M, Stoffel E, Marcon M, Ghafoor S, Boss A. Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. *The British journal of radiology*. 2018;91(1083):20170576.
19. Di Segni M, de Soccio V, Cantisani V, et al. Automated classification of focal breast lesions according to S-detect: validation and role as a clinical and teaching tool. *Journal of ultrasound*. 2018;21(2):105-118.
20. Kim K, Song MK, Kim EK, Yoon JH. Clinical application of S-Detect to breast masses on ultrasonography: a study evaluating the diagnostic performance and agreement with a dedicated breast radiologist. *Ultrasonography (Seoul, Korea)*. 2017;36(1):3-9.
21. Choi JH, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic performance according to reader experience. *Ultrasonography (Seoul, Korea)*. 2017.
22. Bartolotta TV, Orlando A, Cantisani V, et al. Focal breast lesion characterization according to the BI-RADS US lexicon: role of a computer-aided decision-making support. *La Radiologia medica*. 2018;123(7):498-506.
23. Guo R, Lu G, Qin B, Fei B. Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review. *Ultrasound in medicine & biology*. 2018;44(1):37-70.
24. Drukker K, Giger ML, Metz CE. Robustness of computerized lesion detection and classification scheme across different breast US platforms. *Radiology*. 2005;237(3):834-840.
25. Drukker K, Giger ML, Vyborny CJ, Mendelson EB. Computerized detection and classification of cancer on breast ultrasound. *Academic radiology*. 2004;11(5):526-535.
26. Cho E, Kim EK, Song MK, Yoon JH. Application of Computer-Aided Diagnosis on Breast Ultrasonography: Evaluation of Diagnostic Performances and Agreement of Radiologists According to Different Levels of Experience. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine*. 2018;37(1):209-216.
27. Choi JH, Kang BJ, Baek JE, Lee HS, Kim SH. Application of computer-aided diagnosis in breast ultrasound interpretation: improvements in diagnostic performance according to reader experience. *Ultrasonography (Seoul, Korea)*. 2018;37(3):217-225.
28. Lazarus E, Mainiero MB, Schepps B, Koelliker SL, Livingston LS. BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value. *Radiology*. 2006;239(2):385-391.
29. Koh J, Kim EK, Kim MJ, Yoon JH, Park VY, Moon HJ. Role of elastography for downgrading BI-RADS category 4a breast lesions according to risk factors. *Acta radiologica (Stockholm, Sweden : 1987)*. 2018:284185118780901.

30. Au FW, Ghai S, Moshonov H, et al. Diagnostic performance of quantitative shear wave elastography in the evaluation of solid breast masses: determination of the most discriminatory parameter. *AJR American journal of roentgenology*. 2014;203(3):W328-336.

Tables

	SE (%)	SP (%)	PLR	NLR	PPV (%)	NPV (%)	AUC
	95%CI	95%CI	95%CI	95%CI	95%CI	95%CI	95%CI
S-detect	85.37	77.88	3.86	0.19	73.68	88	0.82
	75.83-92.20	69.10-85.14	2.70-5.52	0.11-0.32	63.65-82.19	79.98-93.64	0.75-0.87
R1	100	48.67	1.95	0	58.57	100	0.74
	95.60-100	39.16-58.26	1.63-2.33	0	49.95-66.83	87.66-100.00	0.68-0.80
R2	100	24.78	1.33	0	49.1	100	0.62
	95.60-100	17.14-33.78	1.20-1.48	0	41.30-56.94	87.66-100.00	0.55-0.69
R3	96.34	45.13	1.76	0.08	56.03	94.44	0.71
	89.68-99.24	35.75-54.77	1.48-2.09	0.03-0.25	47.43-64.37	84.61-98.84	0.64-0.77
R4	98.78	19.47	1.23	0.06	47.09	95.65	0.59
	93.39-99.97	12.62-27.98	1.12-1.35	0.01-0.46	39.45-54.84	78.05-99.89	0.52-0.66
R5	97.56	37.17	1.55	0.07	52.98	95.45	0.67
	92.47-99.70	28.26-46.76	1.34-1.80	0.02-0.26	44.70-61.14	84.53-99.44	0.60-0.74
R1+S	100	69.91*	3.32	0	70.69	100	0.85*
	95.60-100	60.57-78.18	2.51-4.40	0	61.52-78.77	95.44-100	0.79-0.90
R2+S	96.34*	49.56*	1.91	0.07	58.09	94.92	0.73*
	89.68-99.24	40.02-59.12	1.58-2.30	0.02-0.23	49.33-66.49	85.85-98.94	0.66-0.79
R3+S	92.68*	76.11*	3.88	0.1	73.79	93.48	0.84*
	84.75-97.27	67.17-83.63	2.78-5.42	0.04-0.21	64.20-81.96	86.34-97.57	0.79-0.89
R4+S	96.34**	46.02*	1.78	0.08	56.43	94.55	0.71*
	89.69-99.24	36.60-55.65	1.50-2.13	0.03-0.25	47.80-64.78	84.88-98.86	0.64-0.77
R5+S	95.12*	72.57*	3.47	0.07	71.56	95.35	0.84*
	87.98-98.66	63.37-80.54	2.56-4.70	0.03-0.18	62.12-79.79	88.52-98.72	0.78-0.89

Table 1 The diagnostic performance of S-detect, the five residents and the integrated results

*: The integrated result of the residents and S-detect was significantly different with the original ones, with a p value <0.001.

** : The integrated result of the residents and S-detect was significantly different with the original ones, with a p value <0.05

	Total number of 4a lesions	Downgraded lesions		Histological malignant	Histological benign
R1	32	24	S-detect malignant	3	5
			S-detect benign	0	24
				Histological malignant	Histological benign
R2	37	31	S-detect malignant	0	6
			S-detect benign	3	28
				Histological malignant	Histological benign
R3	52	38	S-detect malignant	5	9
			S-detect benign	3	35
				Histological malignant	Histological benign
R4	39	32	S-detect malignant	2	5
			S-detect benign	2	30
				Histological malignant	Histological benign
R5	55	42	S-detect malignant	3	10
			S-detect benign	2	40

Table 2 The downgraded lesions by S-detect in category 4a

Figures

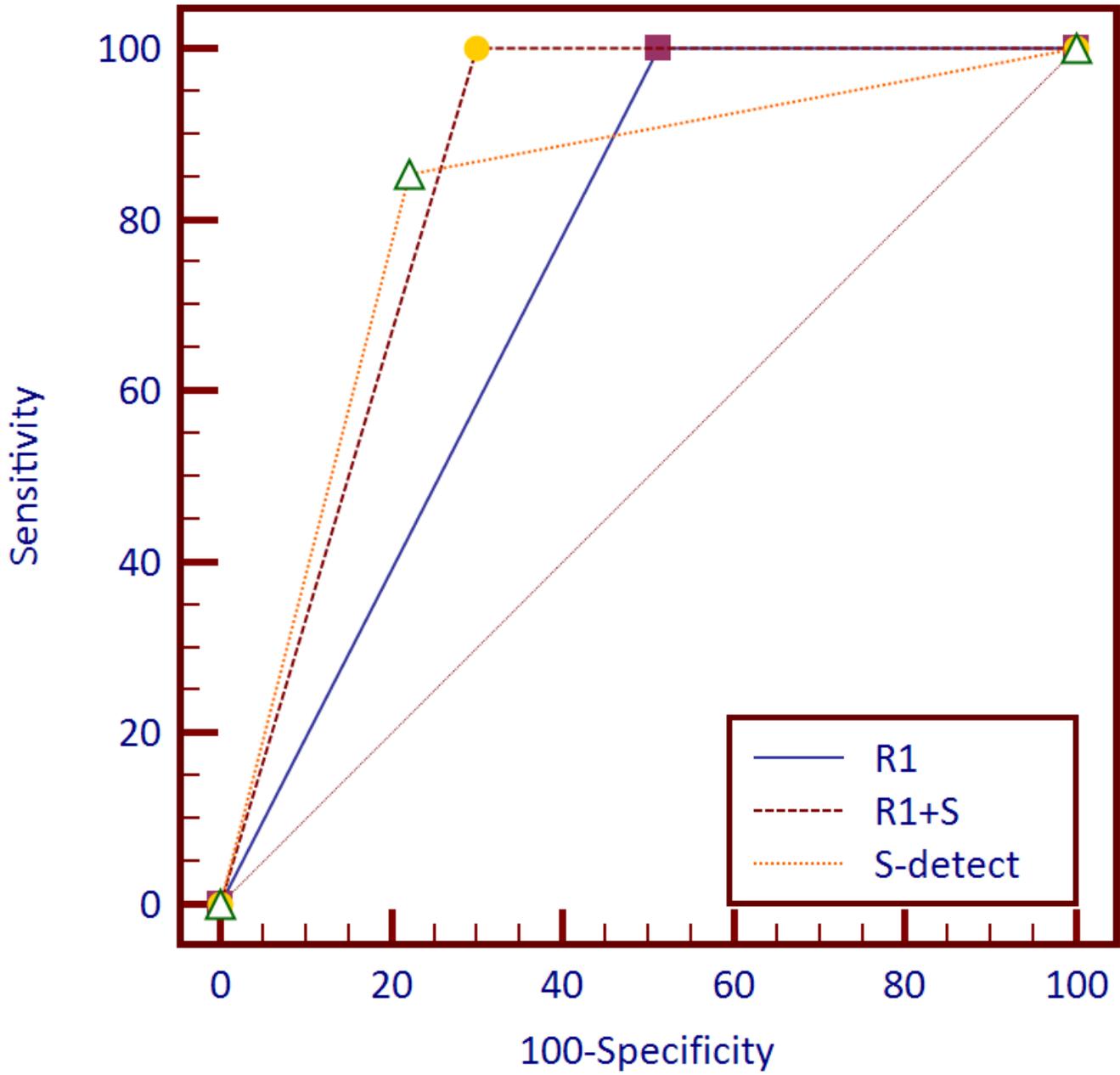


Figure 1

The ROC of R1, S-detect and the integrated results of residents and S-detect

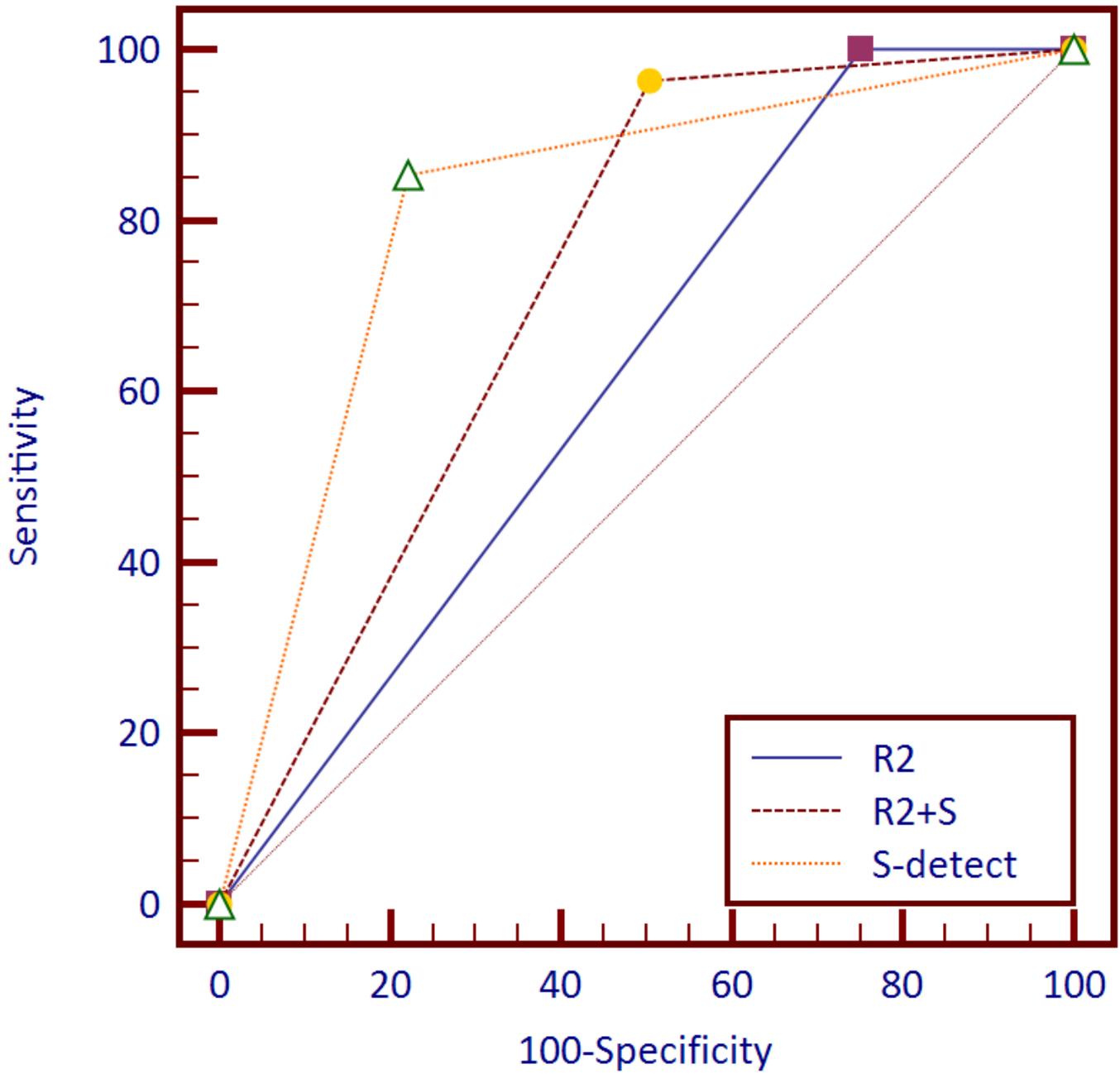


Figure 2

The ROC of R2, S-detect and the integrated results of residents and S-detect

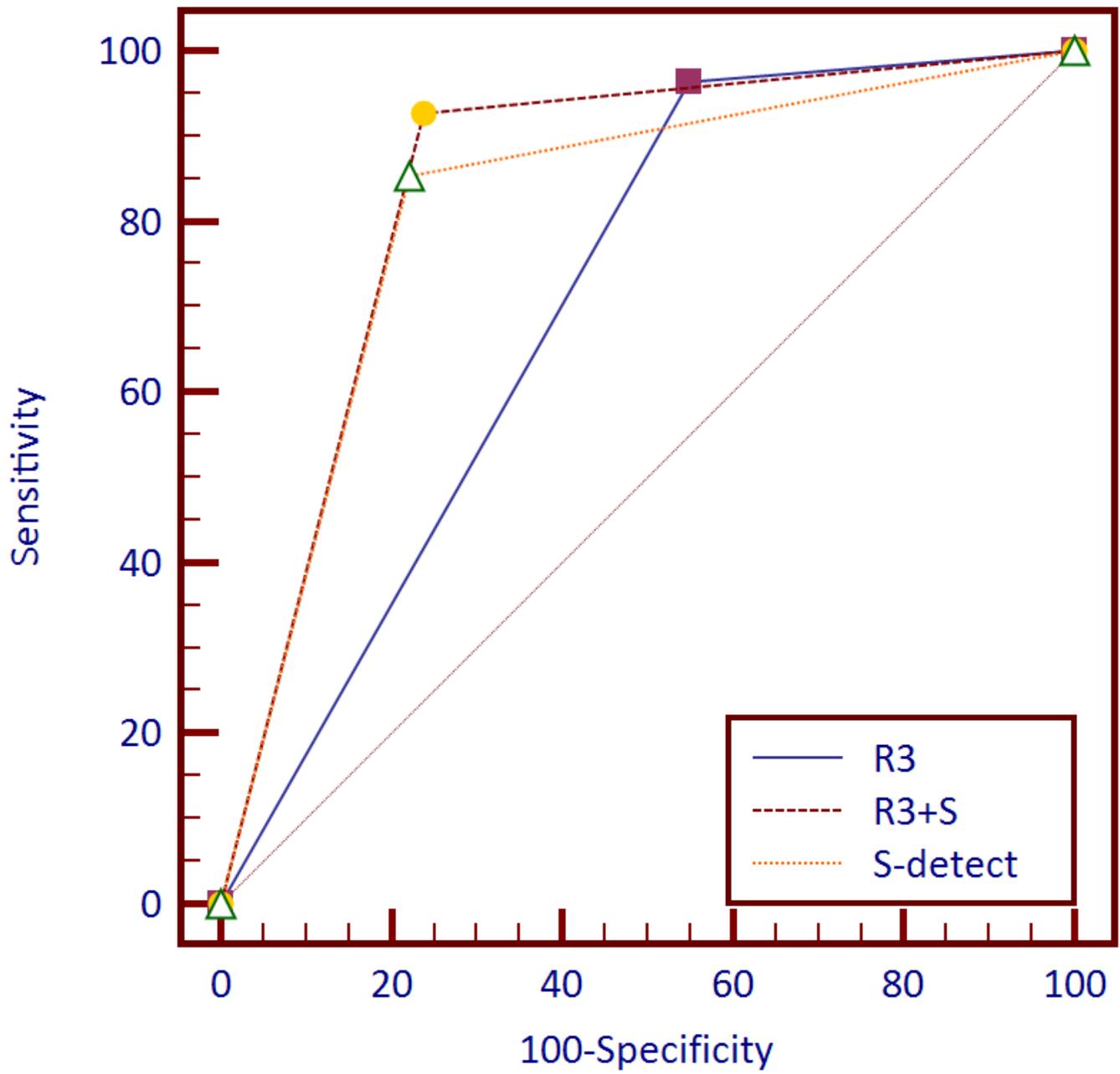


Figure 3

The ROC of R3, S-detect and the integrated results of residents and S-detect

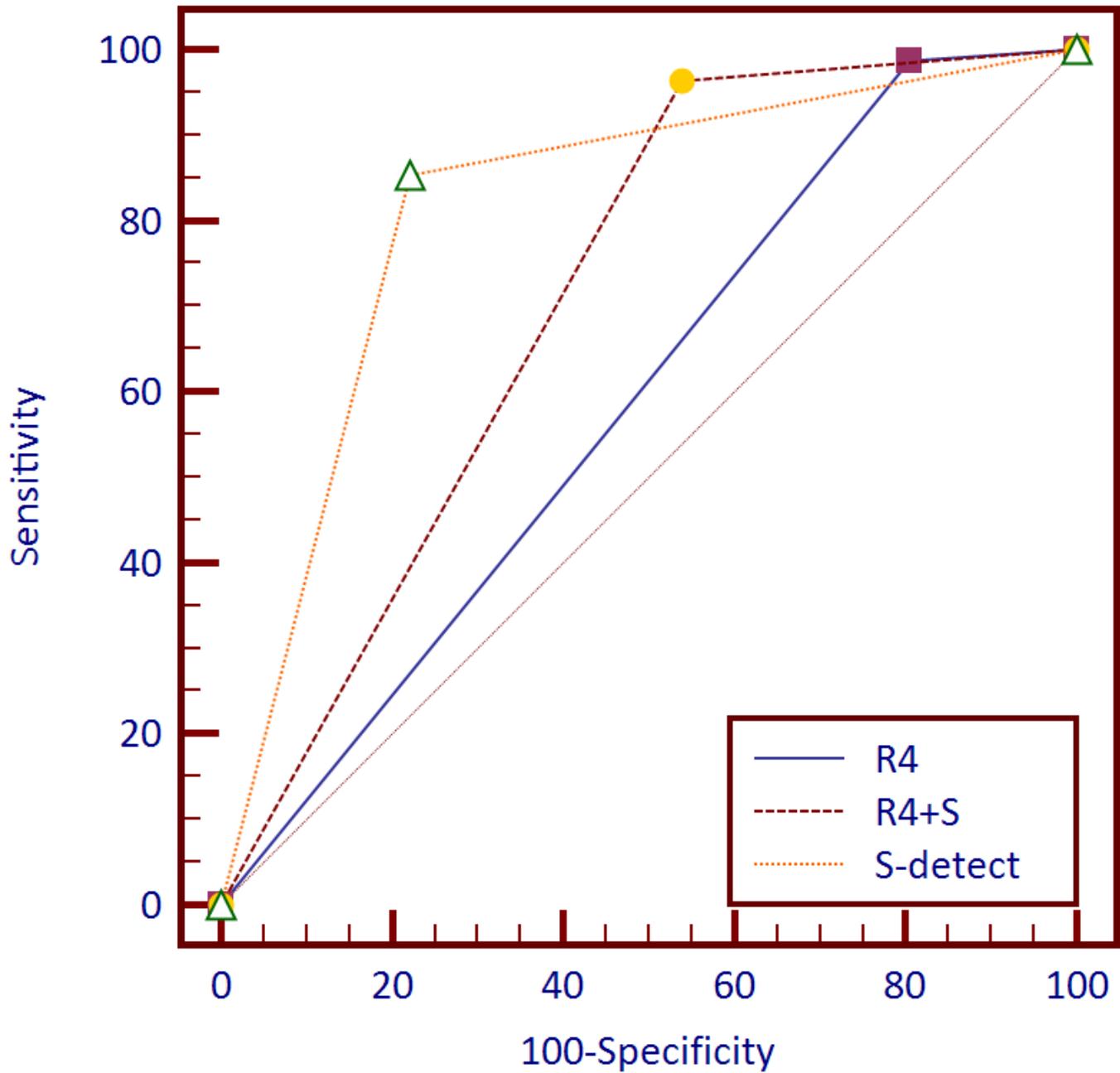


Figure 4

The ROC of R4, S-detect and the integrated results of residents and S-detect

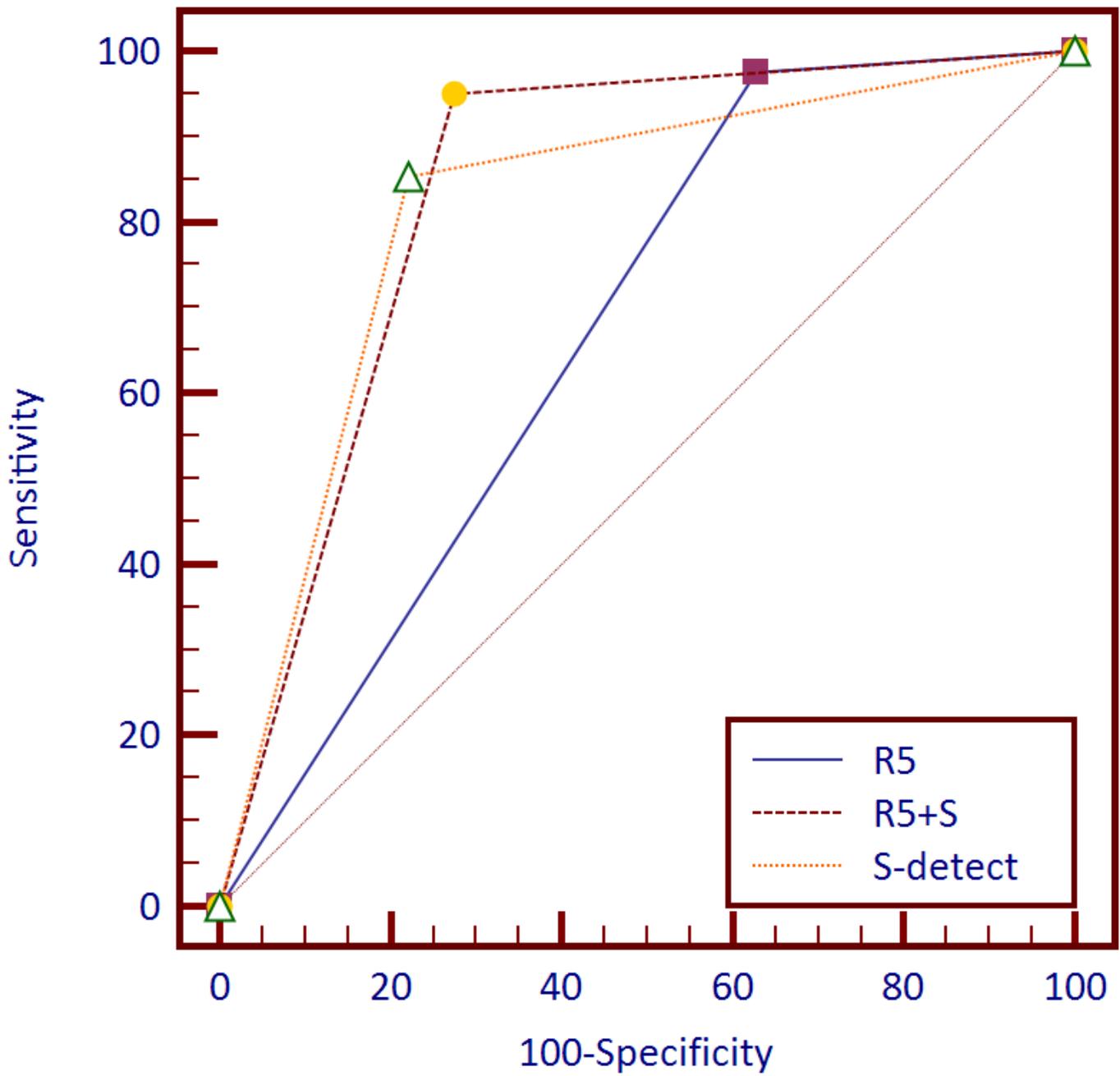


Figure 5

The ROC of R5, S-detect and the integrated results of residents and S-detect

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [STROBEchecklistcohort.doc](#)