

Effects of 1 year of training on the performance of ultrasonographic image interpretation: A preliminary evaluation using images of Sjögren syndrome patients

Yoshitaka Kise^{1,*}, Anne Møystad², Tore Bjørnland³, Mayumi Shimizu⁴, Yoshiko Ariji¹,
Chiaki Kuwada¹, Masako Nishiyama¹, Takuma Funakoshi¹, Kazunori Yoshiura⁵, Eiichiro Ariji¹

¹Department of Oral and Maxillofacial Radiology, Aichi Gakuin University School of Dentistry, Nagoya, Japan

²Institute of Clinical Dentistry, Faculty of Dentistry, University of Oslo, Oslo, Norway

³Department of Oral Surgery and Oral Medicine, Institute of Clinical Dentistry, Faculty of Dentistry, University of Oslo, Oslo, Norway

⁴Department of Oral and Maxillofacial Radiology, Kyushu University Hospital, Fukuoka, Japan

⁵Department of Oral and Maxillofacial Radiology, Faculty of Dental Science, Kyushu University, Fukuoka, Japan

ABSTRACT

Purpose: This study investigated the effects of 1 year of training on imaging diagnosis, using static ultrasonography (US) salivary gland images of Sjögren syndrome patients.

Materials and Methods: This study involved 3 inexperienced radiologists with different levels of experience, who received training 1 or 2 days a week under the supervision of experienced radiologists. The training program included collecting patient histories and performing physical and imaging examinations for various maxillofacial diseases. The 3 radiologists (observers A, B, and C) evaluated 400 static US images of salivary glands twice at a 1-year interval. To compare their performance, 2 experienced radiologists evaluated the same images. Diagnostic performance was compared between the 2 evaluations using the area under the receiver operating characteristic curve (AUC).

Results: Observer A, who was participating in the training program for the second year, exhibited no significant difference in AUC between the first and second evaluations, with results consistently comparable to those of experienced radiologists. After 1 year of training, observer B showed significantly higher AUCs than before training. The diagnostic performance of observer B reached the level of experienced radiologists for parotid gland assessment, but differed for submandibular gland assessment. For observer C, who did not complete the training, there was no significant difference in the AUC between the first and second evaluations, both of which showed significant differences from those of the experienced radiologists.

Conclusion: These preliminary results suggest that the training program effectively helped inexperienced radiologists reach the level of experienced radiologists for US examinations. (*Imaging Sci Dent* 2021; 51: 129-36)

KEY WORDS: Education; Ultrasonography; Sjogren's Syndrome

Introduction

Ultrasonography (US) has a broad spectrum of indications in diagnosing different conditions, because it is a low-cost modality that does not involve radiation exposure. In the field of oral and maxillofacial radiology, the indications for US include the diagnosis of soft tissue lesions, such as

cervical lymph node metastases from malignant tumors.¹ However, US image interpretation is sometimes difficult, and the diagnostic accuracy usually depends on the examiner's experience. Previous studies of US have reported that the diagnostic accuracy of inexperienced radiologists is lower than that of experienced radiologists in the fields of breast, thyroid, and salivary gland imaging.²⁻⁶ In a recent study, inexperienced radiologists were reported to show a lower diagnostic performance for classifying US images of Sjögren syndrome (SjS) than a deep learning artificial intelligence system.⁷

Received October 27, 2020; Revised November 26, 2020; Accepted December 5, 2020
*Correspondence to : Dr. Yoshitaka Kise
Department of Oral and Maxillofacial Radiology Aichi Gakuin University School of Dentistry, 2-11 Suemori-dori, Chikusa-ku, Nagoya 464-8651, Japan
Tel) 81-527592165, E-mail) kise@dpc.agu.ac.jp

In the actual use of US devices in clinics, diagnoses should ideally be determined by the US examiners, because the examiners can create appropriate sectional images for diagnosis by setting appropriate US beam directions. In some cases, however, this is impractical and, in such instances, the examiner saves a typical static image for other interpreters. Without knowing the exact position of the probe in relation to the anatomical structures on a static US image, the assessment is difficult for an interpreter who did not actually examine the patient. This is especially the case for inexperienced interpreters. Therefore, an effective training or diagnostic support system, using computer-aided diagnosis (CAD) techniques, may be beneficial. However, 2 previous reports on US diagnosis have emphasized that although CAD systems can support inexperienced radiologists, their performance quality did not reach the level of experienced radiologists.^{3,4} Therefore, CAD support alone may be insufficient, indicating a need for training in the interpretation of US images.

Several studies have aimed to create technical improvements in US training, including the acquisition of high-quality images or the detection of target lesions using various training methods, such as lectures, hands-on sessions, and apprenticeships with an experienced radiologist.⁸⁻¹³ However, no studies have addressed the problem of how to improve the evaluation of static US images, especially when the examiner and the interpreter are not the same person.

The authors' future goal is to determine the duration and type of training that an inexperienced radiologist needs to become an expert in US diagnosis. To make progress towards this goal, this preliminary study investigated the effects of a 1-year training program that was routinely performed at the authors' department for imaging diagnosis, using static US images of salivary glands in SjS patients.

Materials and Methods

The design of this preliminary study was approved by the Ethics Committee of Kyushu University, Japan, and written informed consent was obtained from all the patients (IRB serial number: 29-006).

Subjects

This study involved 3 inexperienced radiologists (observers A, B and C) who agreed to participate in this study and had different levels of experience in imaging diagnosis. These 3 inexperienced radiologists evaluated a total of 400 static US images of the parotid gland (PG) and the submandibular gland (SMG) twice at a 1-year interval, in April

2019 (the first evaluation) and April 2020 (the second evaluation). These US images were acquired at a different institution from the study location. The acquisition procedure is described below in detail. At the authors' department, where the study was carried out, novices receive 2 years of training in diagnostic imaging, including US examinations. Observer A had already completed the first-year training and had just started the second year when the first evaluation was performed. Observers B and C were beginning their first-year training at the time of the first evaluation.

To compare the observers' performance, 2 experienced radiologists with more than 10 years of experience evaluated the same images. In April 2019, the 3 inexperienced radiologists each independently evaluated 100 static US images of the PG obtained from patients with verified SjS and those without SjS (non-SjS). They also evaluated a total of 200 US images of the SMG in SjS patients (100 images) and non-SjS patients (100 images). All personal information and examination dates included in the obtained images were removed, and the images were randomly displayed in PNG format using Microsoft PowerPoint (Microsoft Corp., Redmond, WA, USA). Typical images are shown in Figures 1A-F. Figures 1A and 1D show US images of the PG and SMG of a non-SjS patient. The parenchyma was homogeneous and well-defined. Figures 1B, 1C, 1E, and 1F are US images of the PG and SMG of a patient with SjS. Figures 1B and 1E are typical examples characterized by multiple hypoechoic areas, and Figures 1C and 1F are typical examples characterized by multiple hyperechoic lines. The evaluations were classified with a 4-point scale of SjS probability: 1, definitely SjS; 2, probably SjS; 3, probably non-SjS; and 4, definitely non-SjS. After finishing the 1-year training, the same 3 radiologists evaluated the same images using the same method in April 2020. Their diagnostic performance at the first and second evaluations was compared in terms of the area under the receiver operating characteristic curve (AUC).

Acquisition of US images

All patients complained of dry mouth and received SjS examinations between 2011 and 2018. Clinical examinations to diagnose SjS were performed in accordance with both Japanese criteria¹⁴ and American-European Consensus Group criteria¹⁵, and US examinations of PG and SMG were also performed. US examinations were performed using a diagnostic unit (Logiq 7: GE Healthcare, Tokyo, Japan) with a center frequency of 12 MHz. The PG was scanned with the coronal plane parallel to the posterior border of the ramus of the mandible. The SMG was scanned with the

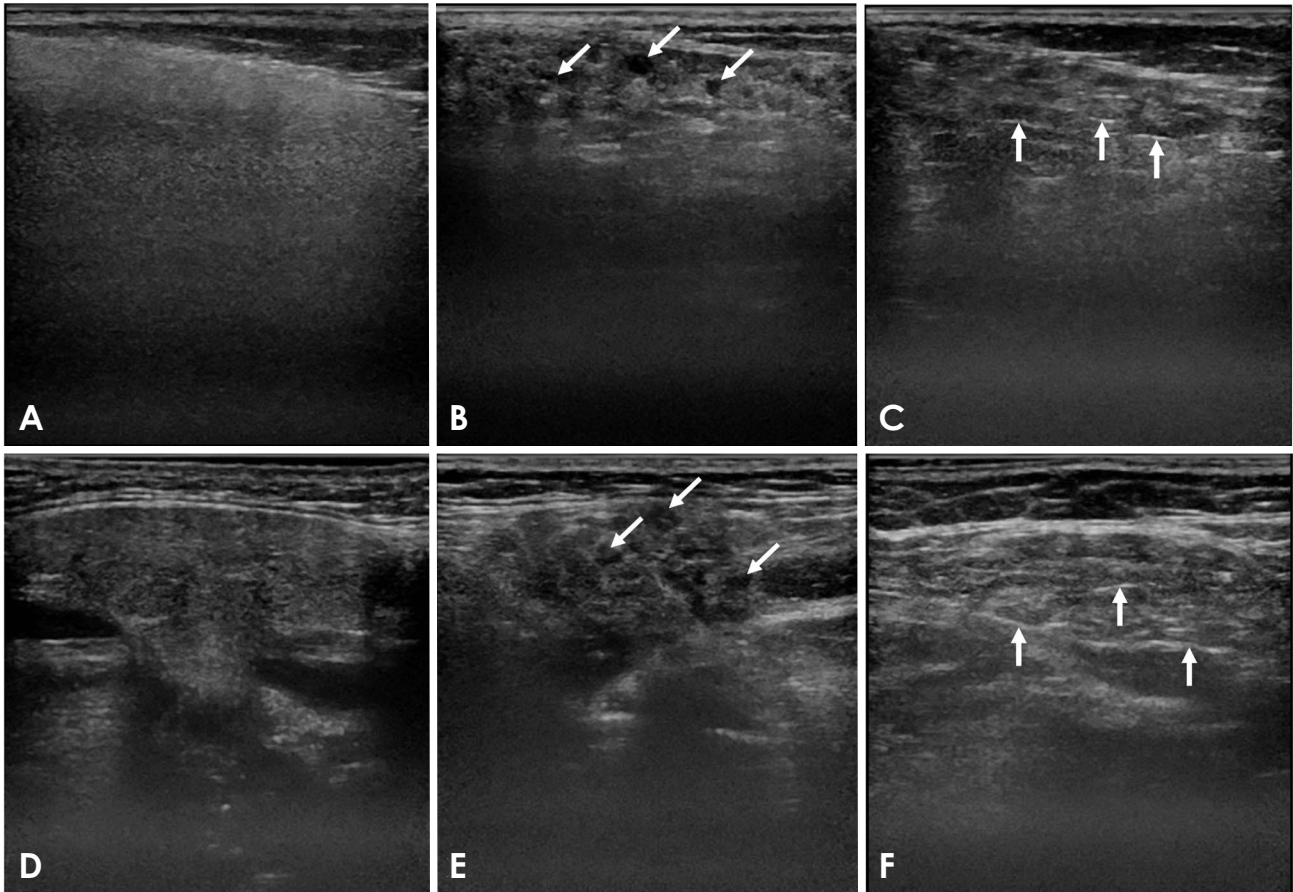


Fig. 1. Examples of ultrasonographic (US) images. A. Parotid gland (PG) of a patient without Sjögren syndrome (SjS). The parenchyma is characterized as homogeneous and well-defined. B and C. PG of a patient with SjS. B. The US image shows typical examples characterized by multiple hypoechoic areas. The glandular parenchyma is heterogeneous with multiple hypoechoic areas found within the parenchyma (white arrows). C. The US image shows typical examples characterized by multiple hyperechoic lines. The glandular parenchyma is heterogeneous and multiple hyperechoic lines are found within the parenchyma (white arrows). D. Submandibular gland (SMG) of a patient without SjS. The parenchyma is characterized as homogeneous and well-defined. E and F. SMG of a patient with SjS. E. The US image shows typical examples characterized by multiple hypoechoic areas. The glandular parenchyma is heterogeneous with multiple hypoechoic areas found within the parenchyma (white arrows). F. The US image shows typical examples characterized by multiple hyperechoic lines. The glandular parenchyma was heterogeneous and multiple hyperechoic lines are found within the parenchyma (white arrows).

sagittal plane parallel to the inferior border of the mandible.

An experienced radiologist (MS) with more than 20 years of experience in US examinations selected the static US images used for evaluation. The selected images were downloaded from the hospital imaging database in Digital Imaging and Communications in Medicine (DICOM) format. Another radiologist (YK) then converted the images from DICOM format to PNG format. Consequently, 100 SjS images (13 men and 87 women with an average age of 56.6 years) and 100 non-SjS images (21 men and 79 women with an average age of 61.1 years) were available for this study.

Training program

The diagnostic imaging training was conducted 1 or 2 days

a week under the instruction of an experienced radiologist at the dental hospital where the study was conducted. Four experienced radiologists with more than 10 years of experience were randomly paired with each inexperienced radiologist. The training program included participation during clinical examinations, collecting patient histories, physical examinations including palpation, and imaging examinations of the patients (Fig. 2). This program included not only US, but also computed tomography (CT), cone-beam CT for dental use (CBCT) and magnetic resonance imaging (MRI) examinations for various maxillofacial diseases, such as tumors, cysts, and inflammation. At the hospital, approximately 10 to 15 patients per day required these examinations, including 8-10 CT or CBCT examinations, 1-2 US

examinations, and 1-2 MRI examinations. The inexperienced radiologists performed these examinations, including actual operation of the US machine, with support from an experienced radiologist. Approximately 10 diagnostic reports a week were generated based on discussions with an experienced radiologist, and 2 or 3 US reports were created in a week. Additionally, the participants attended a weekly conference involving all departmental staff, and presented and discussed important cases during the week. A total of 172 US examinations were performed and 97 cases were handled at the weekly conference during the 1-year program.

Statistical analysis

The differences between the AUC values were tested using chi-square analysis in the JMP statistical software package (version 13.0.0; SAS Institute, Cary, NC, USA). The significance level was set to $P < 0.05$.

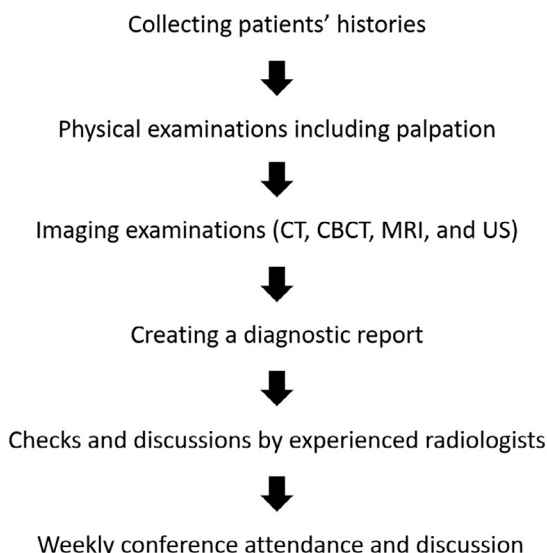


Fig. 2. Schematic illustration of the training program for the inexperienced radiologists at the dental hospital. CT: computed tomography, CBCT: cone-beam computed tomography, MRI: magnetic resonance imaging, US: ultrasonography.

Results

Observers A and B had completed the 1-year training at the time of the second evaluation. Therefore, the former completed 2 years of training and the latter finished 1 year of training. Observer C interrupted the training to take maternity leave for 8 months between August 2019 and March 2020.

A summary of the AUCs of the observers for PG and SMG is shown in Table 1. For observer A, who completed 2 years of training, no difference was found in the AUC for either PG or SMG between the first and second evaluations, and the values were similar to those of the experienced radiologists (Figs. 3A and 4A).

For observer B, who completed 1 year of training, the AUCs were significantly higher at the second evaluation than at the first evaluation for both PG and SMG ($P = 0.0463$ and $P = 0.0005$, respectively). The AUC reached the level of the experienced radiologists for PG assessment, but still showed a difference for SMG classification ($P = 0.0249$) (Figs. 3B and 4B).

For observer C, who did not complete the training, the AUCs were not significantly different between the first and second evaluations for both PG and SMG, and were significantly different from those of the experienced radiologists ($P < 0.0001$, $P < 0.0001$) (Figs. 3C and 4C). This might indicate that a 4-month training period was insufficient.

Discussion

In this preliminary study, the authors evaluated the effects of a 1-year training program in diagnostic imaging using US images of SjS patients, and found that trainees' diagnostic performance significantly improved after the completion of training. This may have been because inexperienced radiologists paired up with experienced radiologists for examinations, diagnostic imaging, and discussions of cases

Table 1. The area under the curve of the first and second evaluations of ultrasound images for inexperienced radiologists (A, B, C) and experienced radiologists (D, E)

Observers	Parotid gland		Submandibular gland	
	First evaluation	Second evaluation	First evaluation	Second evaluation
A	0.835 (0.774-0.882)	0.826 (0.766-0.872)	0.859 (0.801-0.903)	0.877 (0.820-0.918)
B	0.810 [†] (0.748-0.860)	0.859* (0.800-0.902)	0.704 [†] (0.630-0.768)	0.821* [†] (0.756-0.871)
C	0.789 [†] (0.721-0.843)	0.785 [†] (0.717-0.841)	0.795 [†] (0.730-0.848)	0.786 [†] (0.719-0.840)
D, E		0.865 (0.825-0.896)		0.857 (0.816-0.891)

The 95% CI is shown in parentheses. *: $p < 0.05$ between the first and second evaluation, [†]: $p < 0.05$ compared with the experienced observers

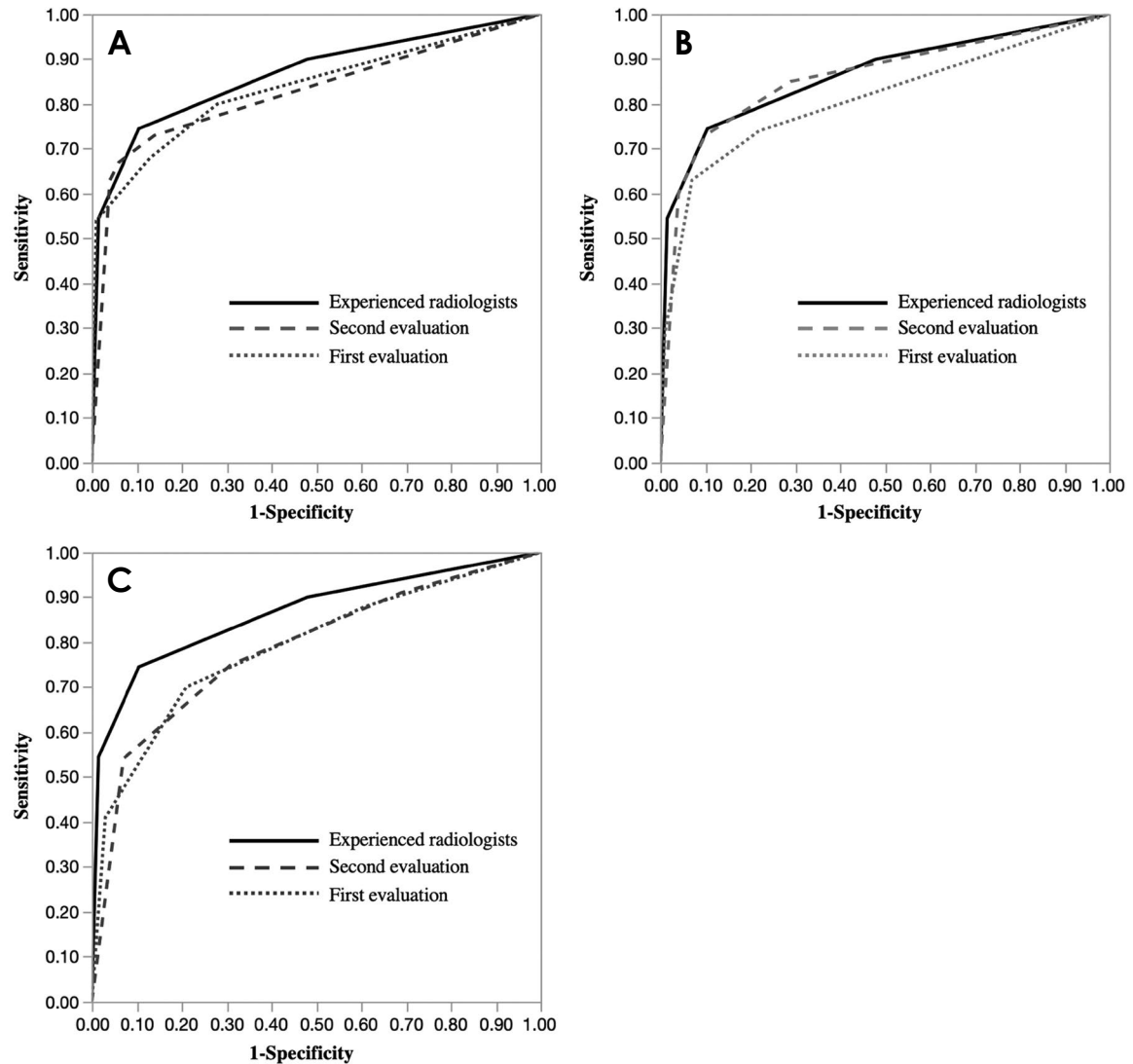


Fig. 3. Receiver operating characteristic curves for the parotid gland at the first and second evaluations by the inexperienced radiologists (A. Two years of experience, B. One year of experience, C. Four months of experience) compared with the experienced radiologists.

during conferences. It is considered important that the inexperienced radiologists were placed in an environment where they could efficiently absorb the expertise of experienced radiologists.

Delzell et al.¹¹ reported the effect of a comprehensive musculoskeletal US training program to improve accuracy in the diagnosis of rotator cuff tears in relatively inexperienced examiners. The training program consisted of lectures on US application software and scanning techniques for the musculoskeletal system, hands-on imaging sessions, and apprenticeships with an experienced radiologist, until the candidates gradually started performing examinations independently. Weekly ongoing medical education conferences were held to present and discuss all clinical cases from the previous week. In contrast, the training program did not

include lectures or hands-on imaging sessions in advance, although trainees received real-time lectures and advice from experienced radiologists during patient examinations and discussion after the examination. The training program also included attendance at weekly conferences, which probably assisted in learning retention and are recommended for inclusion in the training program.

The training program analyzed in this study included not only US imaging, but also CT and MRI imaging, the collection of patients' medical history, and information about physical examinations. The addition of physical and anatomical information about the target tissue makes it possible to make a comprehensive diagnosis. This broad focus is thought to improve diagnostic performance to a greater extent than US training alone.

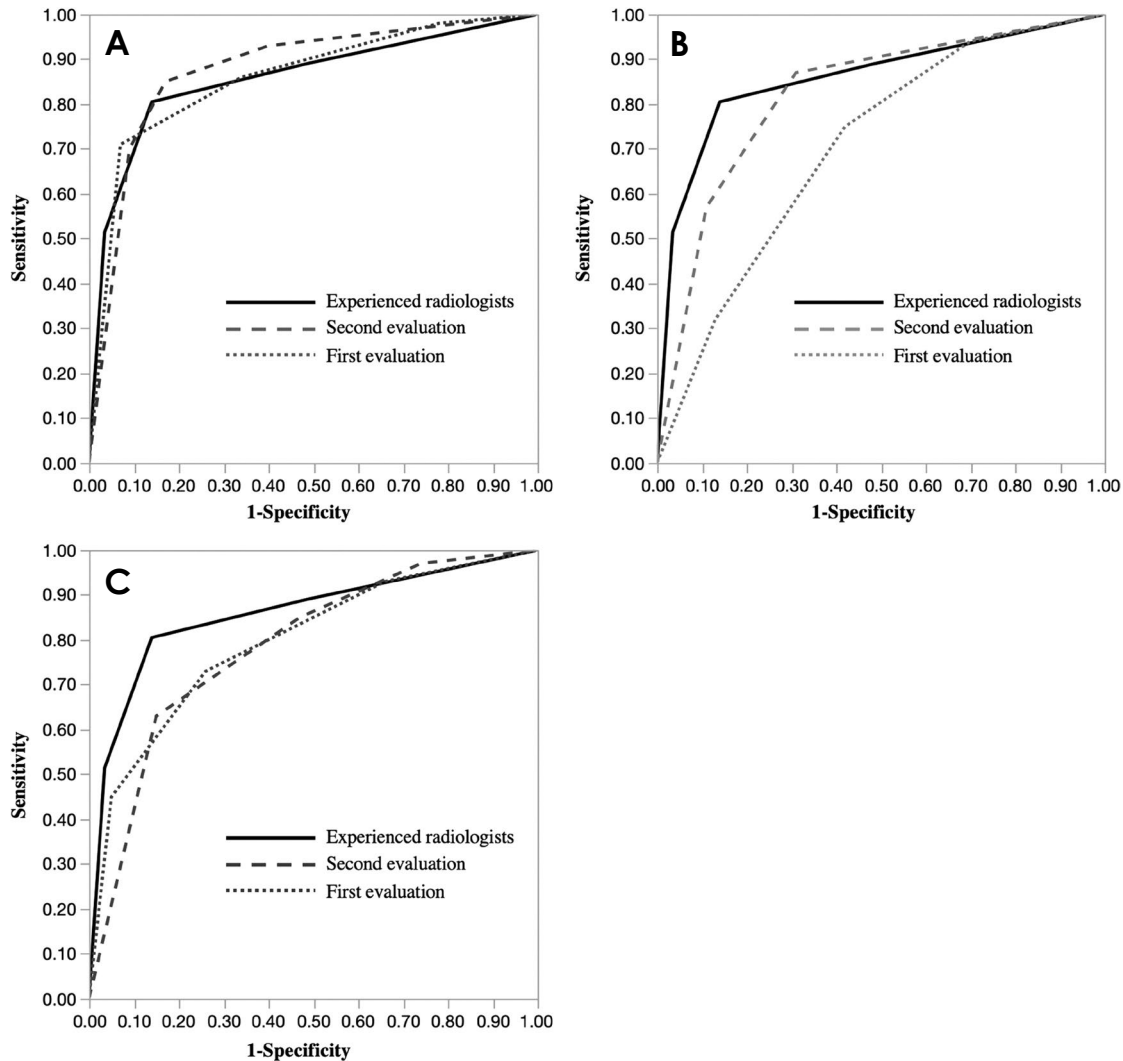


Fig. 4. Receiver operating characteristic curves for the submandibular gland at the first and second evaluations by the inexperienced radiologists (A. Two years of experience, B. One year of experience, C. Four months of experience) compared with the experienced radiologists.

Observer A, who had already completed 1 year of training and was in the second year, reached the level of the experienced radiologists for both the first and second evaluations. For Observer B, who was training for the first time, PG assessment reached the level of the experienced radiologists, while SMG assessment reached a level close to that. However, for observer C, who discontinued training, the AUC values did not improve and remained significantly lower than those of the experienced radiologists. Although this study was limited to still images of SjS, its findings show that 1 year of training may be sufficient. The difference between the results for PG and SMG was presumably because the US images of the PG did not contain other structures (such as muscle, lymph nodes, and blood vessels) within the image, whereas the US images of the SMG often contained such

structures. Therefore, US interpretation may be more difficult for the SMG than for the PG, possibly implying that inexperienced radiologists need more training in diagnosing diseases in the SMG. After training, the radiologists were able to identify the surrounding anatomical structures, and their diagnostic performance improved significantly.

This study used images of SjS. SjS is an autoimmune disease characterized by dry eyes and dry mouth. Dry mouth results from reduced secretion of saliva caused by deterioration of salivary gland function, and fatty degeneration is a characteristic finding in the parenchyma of affected glands.^{16,17} US images of the PG and SMG in SjS patients are characterized by inhomogeneous parenchyma, multiple hypoechoic areas, and multiple hyperechoic lines.⁶ Cysts and tumors are usually located partially in the PG and SMG

tissue, and their presence can be easily detected using US. However, because SJS causes degeneration of the entire gland, the diagnosis is difficult unless the interpreter has a clear understanding of the appearance of healthy glands and the surrounding anatomy. Therefore, the detection of SJS may be more dependent on experience than the detection of cysts or tumors, which was the reason for choosing SJS images to verify the effects of training.

The present study had some limitations. First, the number of radiologists was small and the study included radiologists with different durations of training experience at the first evaluation. One had 1 year of experience with US, and 2 were inexperienced. The significant effect of the training was confirmed by comparing the diagnostic performance of these 3 inexperienced radiologists. However, a further analysis will be needed with a larger number of radiologists. Second, the effect of training focused on the accuracy of image interpretation, and did not include the technique of image acquisition. In some cases, the examiner does not perform image diagnosis, and another doctor makes a diagnosis based on the image stored on the server. Since remote imaging-based diagnoses have become widespread in recent years, such situations occur increasingly frequently. Most of these images are static images rather than dynamic images due to the large volume of data involved and the storage capacity of the hard disk. It is difficult to interpret static images that are not taken by oneself, but this is a necessary ability for radiologists. Third, the only images evaluated by the radiologists were of salivary glands in patients with or without SJS. Future studies should evaluate other lesions and diseases. Finally, the training in this study is unique to the institution where the study was conducted, and the number of patients and the type of lesions depended on the patients presenting at that time, so bias may have occurred among the evaluators. For future research, multi-institutional training programs should be established and implemented in a unified manner.

In conclusion, these preliminary results indicate that the training program described herein could be effective in helping inexperienced radiologists to attain the level of experienced radiologists for US examinations.

Acknowledgements

We thank Helen Jeays, BDS Sc AE, from Edanz Group (<https://en-author-services.edanzgroup.com/ac>) for editing a draft of this manuscript.

Conflicts of Interest: None

References

1. Arijji Y, Kimura Y, Hayashi N, Onitsuka T, Yonetsu K, Hayashi K, et al. Power Doppler sonography of cervical lymph nodes in patients with head and neck cancer. *AJNR Am J Neuroradiol* 1998; 19: 303-7.
2. Li T, Jiang Z, Lu M, Zou S, Wu M, Wei T, et al. Computer-aided diagnosis system of thyroid nodules ultrasonography: diagnostic performance difference between computer-aided diagnosis and 111 radiologists. *Medicine (Baltimore)* 2020; 99: e20634.
3. Lee J, Kim S, Kang BJ, Kim SH, Park GE. Evaluation of the effect of computer aided diagnosis system on breast ultrasound for inexperienced radiologists in describing and determining breast lesions. *Med Ultrason* 2019; 21: 239-45.
4. Park HJ, Kim SM, Yun BL, Jang M, Kim B, Jang JY, et al. A computer-aided diagnosis system using artificial intelligence for the diagnosis and characterization of breast masses on ultrasound: added value for the inexperienced breast radiologist. *Medicine (Baltimore)* 2019; 98: e14146.
5. 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. *Br J Radiol* 2018; 91: 20170576.
6. Shimizu M, Okamura K, Yoshiura K, Ohyama Y, Nakamura S, Kinukawa N. Sonographic diagnostic criteria for screening Sjögren's syndrome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2006; 102: 85-93.
7. Kise Y, Shimizu M, Ikeda H, Fujii T, Kuwada C, Nishiyama M, et al. Usefulness of a deep learning system for diagnosing Sjögren's syndrome using ultrasonography images. *Dentomaxillofac Radiol* 2020; 49: 20190348.
8. Chalumeau-Lemoine L, Baudel JL, Das V, Arrivé L, Noblinski B, Guidet B, et al. Results of short-term training of naïve physicians in focused general ultrasonography in an intensive-care unit. *Intensive Care Med* 2009; 35: 1767-71.
9. Syperda VA, Trivedi PN, Melo LC, Freeman ML, Ledermann EJ, Smith TM, et al. Ultrasonography in preclinical education: a pilot study. *J Am Osteopath Assoc* 2008; 108: 601-5.
10. Counselman FL, Sanders A, Slovis CM, Danzl D, Binder LS, Perina DG. The status of bedside ultrasonography training in emergency medicine residency programs. *Acad Emerg Med* 2003; 10: 37-42.
11. Delzell PB, Boyle A, Schneider E. Dedicated training program for shoulder sonography: the results of a quality program reverberate with everyone. *J Ultrasound Med* 2015; 34: 1037-42.
12. Fulton N, Buethe J, Gollamudi J, Robbin M. Simulation-based training may improve resident skill in ultrasound-guided biopsy. *AJR Am J Roentgenol* 2016; 207: 1329-33.
13. Roark AA, Ebuoma LO, Ortiz-Perez T, Sepulveda KA, Severs FJ, Wang T, et al. Impact of simulation-based training on radiology trainee education in ultrasound-guided breast biopsies. *J Am Coll Radiol* 2018; 15: 1458-63.
14. Fujibayashi T, Sugai S, Miyasaka N, Hayashi Y, Tsubota K. Revised Japanese criteria for Sjögren's syndrome (1999): availability and validity. *Mod Rheumatol* 2004; 14: 425-34.
15. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexan-

- der EL, Carsons SE, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* 2002; 61: 554-8.
16. Izumi M, Eguchi K, Nakamura H, Nagataki S, Nakamura T. Premature fat deposition in the salivary glands associated with Sjögren syndrome: MR and CT evidence. *AJNR Am J Neuroradiol* 1997; 18: 951-8.
17. Takagi Y, Sumi M, Sumi T, Ichikawa Y, Nakamura T. MR microscopy of the parotid glands in patients with Sjögren's syndrome: quantitative MR diagnostic criteria. *AJNR Am J Neuroradiol* 2005; 26: 1207-14.