Supplementary

Table S1 Basic characteristics of enrolled participants in the training set

Dataset	Male	Female	Total
Training Set (discharge diagnosed as otosclerosis)			
Number of patients(retrieved)	298	696	994
Age (mean ± SD)	39.11±10.97	41.49±10.62	40.76±10.78
Number of CT scans	238	556	794
Labels assigned to ears (stapedial)	288	702	990
Total number of CT slices	-	-	101,446
Cropped ear slices	-	-	2,458
Training Set (discharge diagnosed as external auditory canal tumor/new organism)			
Number of patients(chosen)	151	149	300
Age (mean ± SD)	48.89±19.82	44.95±18.52	46.93±19.29
Number of CT scans	151	149	300
Labels assigned to ears	173	168	341
Total number of CT slices	-	-	33,128
Cropped ear slices	-	-	994

Table S2 Basic characteristics of enrolled participants in the retrospective clinical test set

Dataset	Male	Female	Total
Testing Set (bilateral stapedial otosclerosis)			
Number of patients	13	29	42
Age (mean ± SD)	32±14.31	40.71±10.42	38.14±12.36
Number of CT scans	13	29	42
Labels assigned to ears	26	58	84
Total number of CT slices	-	-	9,160
Testing Set (unilateral stapedial otosclerosis)			
Number of patients	0	2	2
Number of CT scans	0	2	2
Labels assigned to ears	0	4	4
Testing Set (bilateral normal)			
Number of patients	37	63	100
Age (mean ± SD)	39.11±14.37	48.71±15.58	44.98±15.91
Number of CT scans	37	63	100
Labels assigned to ears	74	126	200
Total number of CT slices	-	-	22614

http://dx.doi.org/10.21037/atm-21-1171

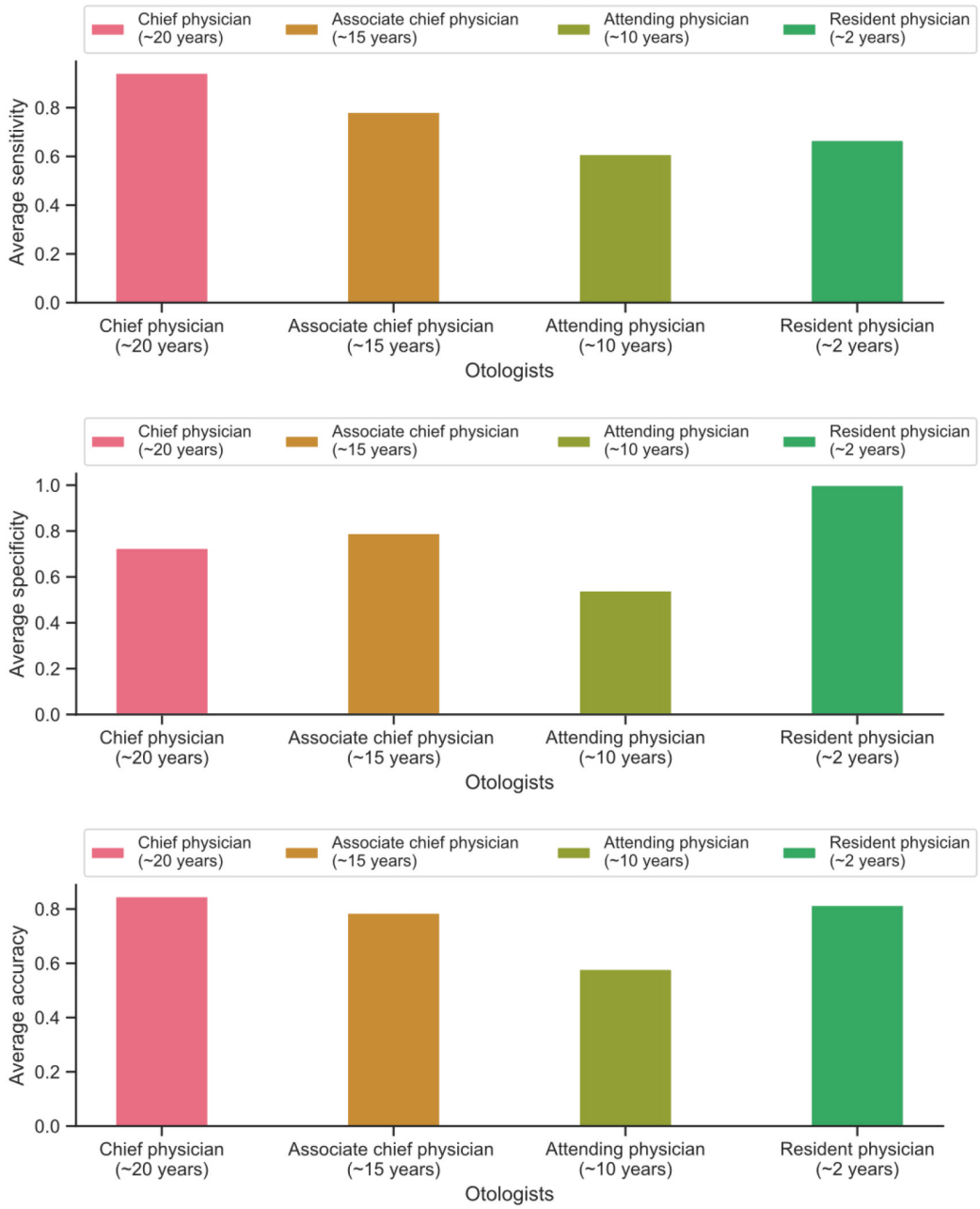

Figure S1 Demonstration of the study results of the ability of otolaryngologists to clinically diagnose otosclerosis. We showed the comparison of average diagnostic performance between chief physician, associate chief physician, attending physician, and resident physician in terms of sensitivity, specificity, and accuracy.

http://dx.doi.org/10.21037/atm-21-1171

Table S3 Comparison of our otosclerosis-LNN model with seven otologists in terms of sensitivity and specificity on the prospectively collected clinical test set, which contains a total of 140 ears (otosclerosis =78, normal =62)

Otologist	Diagnosis time (seconds per ear)	Total number of ears, n=140 (otosclerosis =78, normal =62)					
		True positives	False negatives	True negatives	False positives	Sensitivity	Specificity
Otosclerosis-LNN	0.06	75	3	62	0	96.15%	100%
Chief physician A (~20 years)	34.3	76	2	28	34	97.44%	45%
Chief physician B (~20 years)	32.2	71	7	62	0	91.03%	100%
Associate chief physician A (~15 years)	30	50	28	57	5	64.10%	92%
Associate chief physician B (~15 years)	42.9	72	6	41	21	92.31%	66%
Associate chief physician C (~10 years)	77.1	42	36	28	34	53.85%	45%
Attending physician (~10 years)	20	53	25	39	23	67.95%	63%
Resident (~2 years)	30	52	26	62	0	66.67%	100%

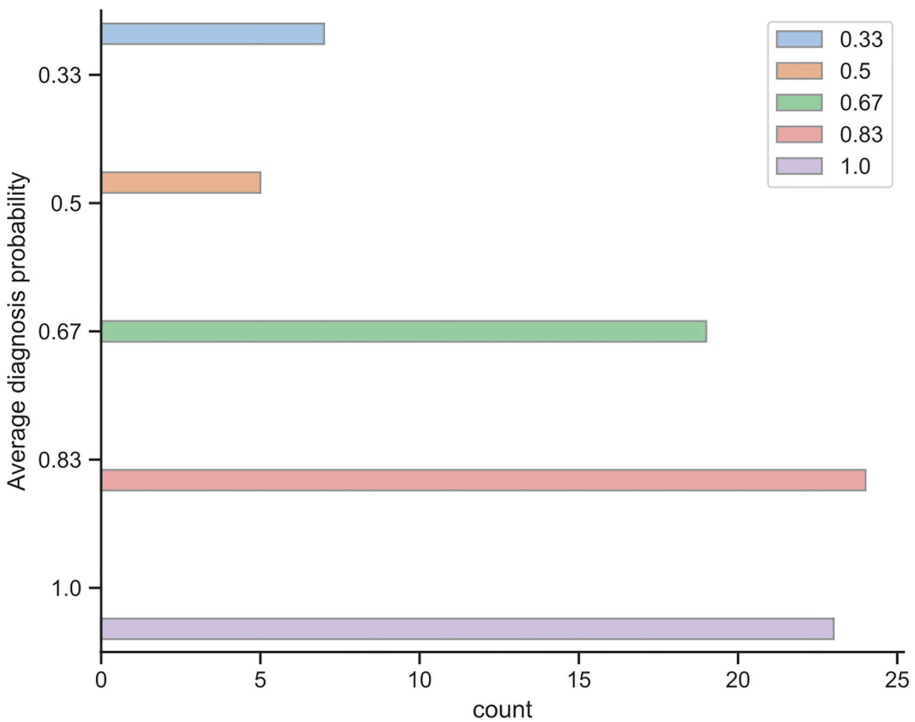

Figure S2 Histogram of the average diagnosis accuracy of otolaryngologists of each ear in the clinical test set.

http://dx.doi.org/10.21037/atm-21-1171

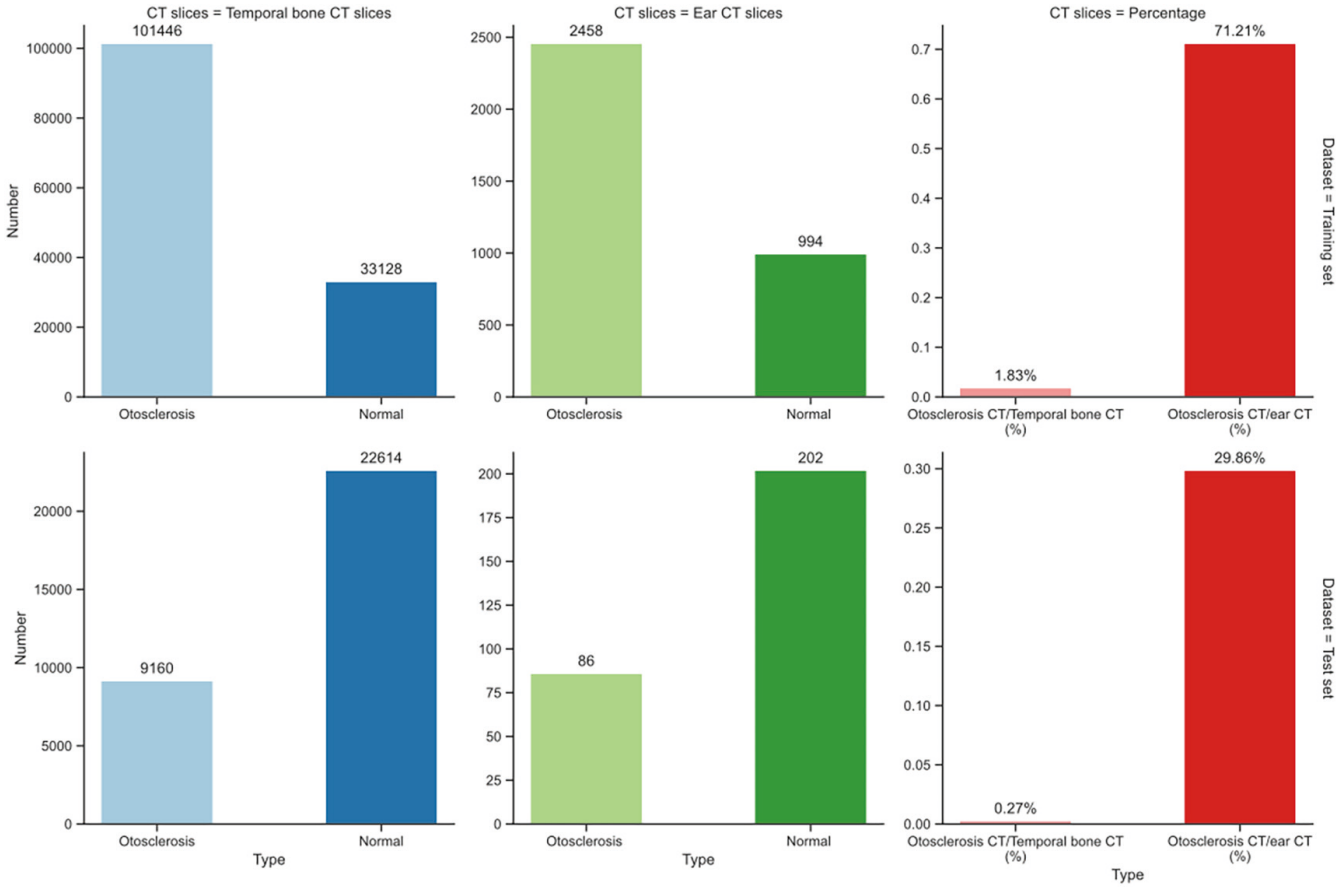

Figure S3 Demonstration of the number of the temporal bone CT slices and ear CT slices in the training and test sets. The rightmost figure shows the ratio of the number of CT slices with otosclerosis to the total number of temporal bone CT slices and the number of ear CT slices, respectively and the proportions are 1.83% and 71.21%, respectively. Therefore, our otosclerosis-LNN model employs an end-to-process strategy instead of end-to-end deep neural networks widely used in existing works to train the deep learning-based otosclerosis detection network. The end-to-process strategy allows us to train the deep learning model using ear CT slices instead of whole temporal bone CT slices, which avoids the extreme class imbalance problem.

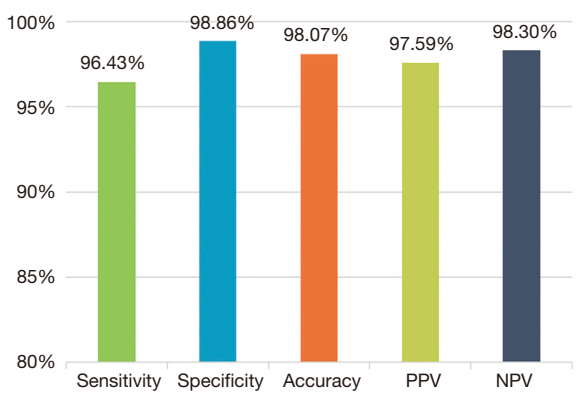

Figure S4 Demonstration of the diagnostic performance of our otosclerosis-LNN model on the external test set.

http://dx.doi.org/10.21037/atm-21-1171

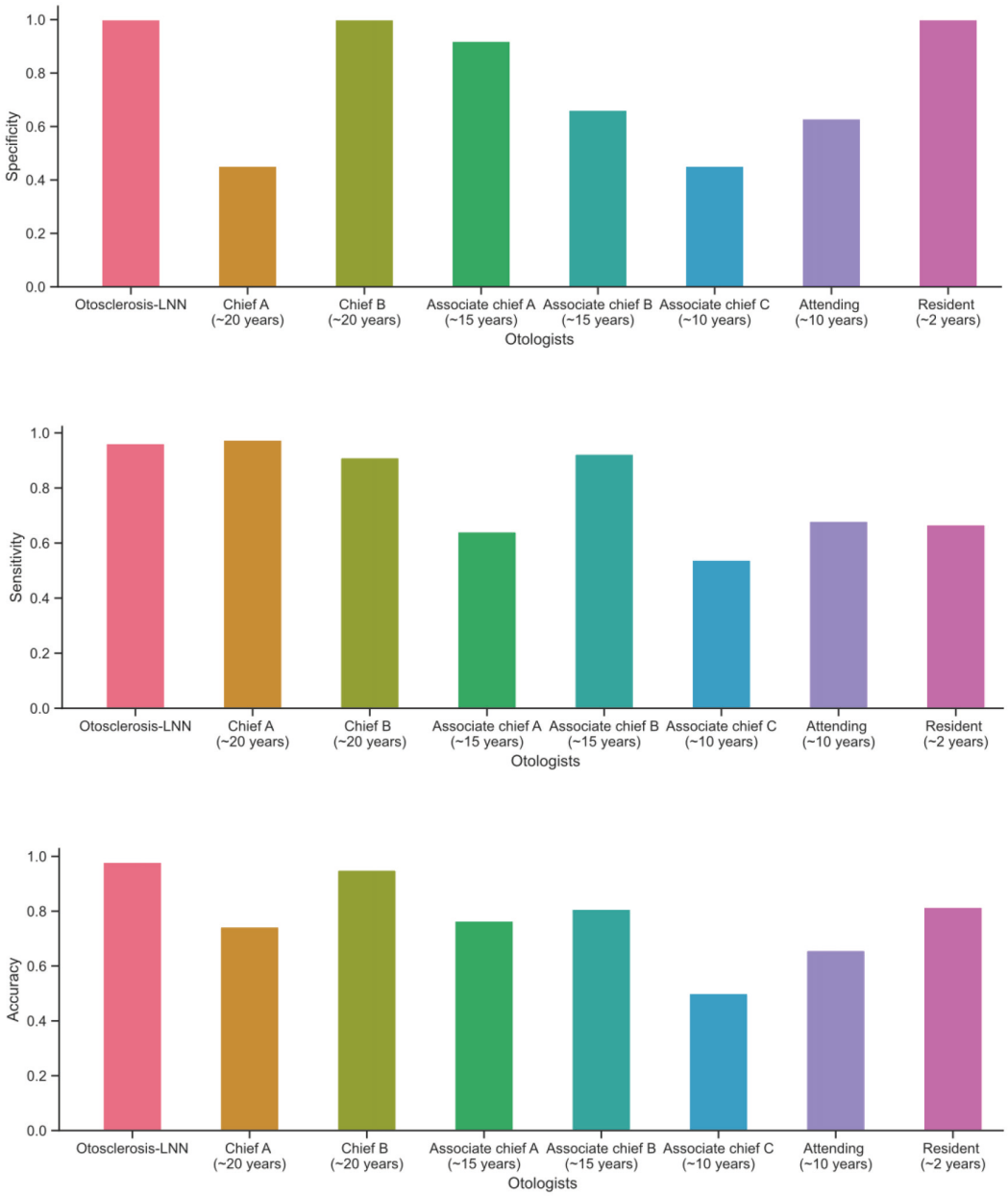

Figure S5 Comparison of our otosclerosis-LNN model with otologists. The otosclerosis-LNN model demonstrated higher screening performance than both two chief physicians (~20 years) enrolled in the comparison study on the prospectively collected test set. The otosclerosis-LNN model achieved sensitivity and specificity of 96.15% and 100%, respectively (Table S3), and the average sensitivity and specificity of the two chief physicians (~20 years) attending the comparison study are 94.20% and 72.55%, respectively. These two important indicators are lower than the otosclerosis-LNN model. In addition, two associate chief physicians achieved average sensitivity and specificity of 70.07% and 67.70%, respectively, which is lower than that of two chief physicians. The attending physician (~10 years) achieved sensitivity and specificity of 67.90% and 62.90%, respectively and the resident (~2 years) achieved sensitivity and specificity of 66.70% and 100%, respectively. Overall, compared with other doctors, the chief physicians demonstrated a higher diagnostic level in the test set, but still lower than the otosclerosis-LNN model.

http://dx.doi.org/10.21037/atm-21-1171

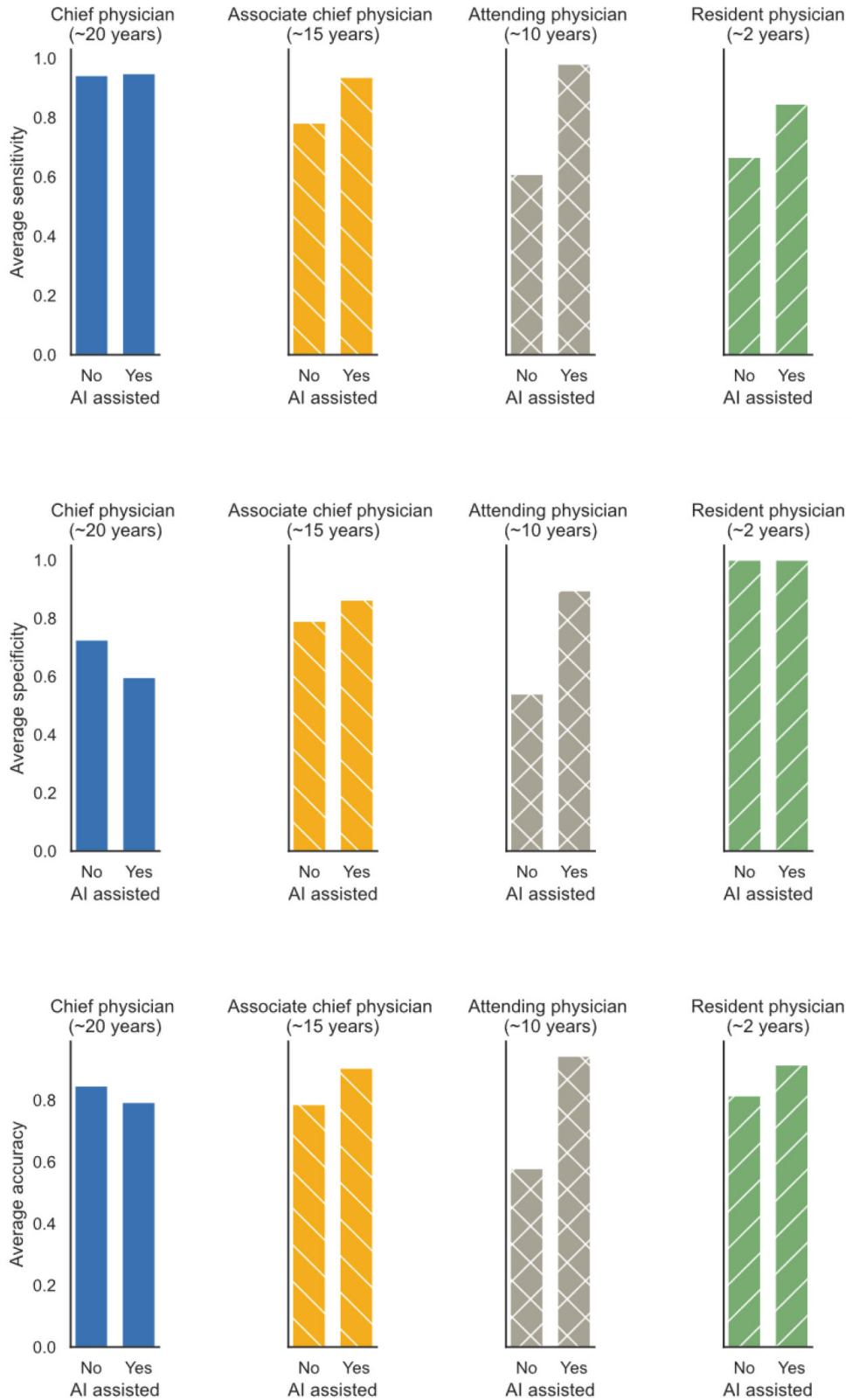

Figure S6 Demonstration of the diagnostic performance of otologists with the assistance of our otosclerosis-LNN model.

http://dx.doi.org/10.21037/atm-21-1171

Table S4 Using the otosclerosis-LNN model to assist otolaryngologists in the diagnosis of stapedial otosclerosis in the temporal bone high-resolution CT images

Otologists	Total number of ears, n=140 (otosclerosis =78, normal =62)					
	True positives	False negatives	True negatives	False positives	Sensitivity	Specificity
Otosclerosis-LNN	75	3	62	0	96.15%	100.00%
Chief physician A (~20 years)	78	0	12	50	100.00%	19.35%
Chief physician B (~20 years)	70	8	62	0	89.74%	100.00%
Associate chief physician A (~15 years)	70	8	58	4	89.74%	93.55%
Associate chief physician B (~15 years)	76	2	49	13	97.44%	79.03%
Associate chief physician C (~10 years)	76	2	55	7	97.44%	88.71%
Attending physician (~10 years)	77	1	56	6	98.72%	90.32%
Resident (~2 years)	66	12	62	0	84.62%	100.00%

http://dx.doi.org/10.21037/atm-21-1171