Appendix 1

1. Artificial intelligence algorithm

We designed a 3D Deep Learning algorithm, SSNet, with 13 3D convolutional layers, 5 max pooing layers, and 2 fully connected layers (*Figure 1*). The input images were 3D shaped data cropped from the CT scan with a volume of size 32 mm × 48 mm × 48 mm at the mass center of a ROI with a histological label. The output of the proposed algorithm was probabilities for different categories. The artificial intelligence algorithm was trained from scratch for three differentiation tasks: (I) aggressive (IA) or indolent (AAH, AIS, MIA); (II) categories of different invasiveness, pre-invasive (AAH, AIS), minimally invasive (MIA), invasive (IA); (III) categories of four histological subtypes.

2. Algorithm training and interpretation

The training of the algorithm was performed on a computer with an NVIDA GTX 1080 (NVIDIA, Santa Clara, Calif) graphics processing unit (GPU) and used the TensorFlow deep learning framework (Google, Mountain View, CA). Momentum optimizer was used to minimize the Softmax cross-entropy between the outputs and reference labels with a batch size of 64 and initial learning rate of 0.01, decayed every 300 iterations using an exponential rate of 0.99. We augmented the samples by randomly rotating each patch to 0, 90, 180, and 270 degrees along the Z axis, and randomly flipping them in the X, Y, and Z directions. To prevent overfitting, we used L2 regularization during training. Our training loss converged after 3,000 iterations. The model with the lowest validation loss was selected eventually. To increase the understandability and dependability of the proposed SSNet, we adopted class activation mapping method to generate heat maps to indicate invasiveness in input images by using the feature map extracted from the developed network. The heat mapping was done with the "Matplotlib" module and all programming was conducted in Python version 3.6.4.

3. Interpretation by a feature-based machine learning method

To exploit the potential difference from traditional feature-based AI technique in interpretation of nodule aggressiveness, our previously published radiomic signature was utilized (10), and analysis was performed with extracted radiomic features. Tumor segmentation, feature extraction, and inter-/intra-observer variability was reported previously. The malignancy risk was computed according to the input features and classified the nodules into IA and non-IA (binary classification).

4. Receiver operating characteristic curves analysis

Instead of a continuous value describing invasiveness, only a binary label was provided by doctors. Thus, the receiver operating characteristic (ROC) curves were estimated for six practicing doctors as a group, radiomic signature, and AI model using partial least-squares regression with constrained splines as previously described to warrant a fair comparison (18). Then linear interpolation and the composite trapezoidal rule were applied to estimate the area under ROC curve (AUC) for three approaches. At last, the confidential intervals (CI) of AUCs were obtained through 10,000 bootstrap replicates drawn from test set, on which three approaches were measured using the same replicate. The difference between AUCs was calculated on these same replicates by the stringent Bonferroni-corrected CIs of 1–0.05/k (k stands for number of classes). There is evidence of difference when 0 was not included in the interval. Similar way for AUC calculation was introduced by Rajpurkar *et al.* previously (18).

https://dx.doi.org/10.21037/tlcr-21-971

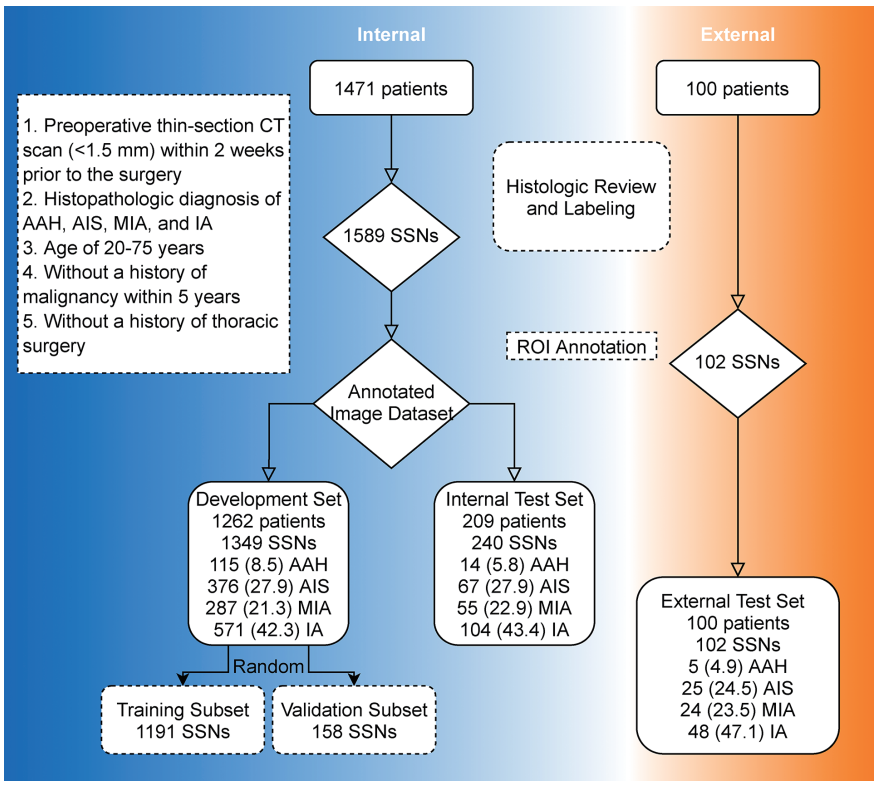

Figure S1 Flowchart of patient allocation in the retrospective dataset and external dataset. Number in parentheses of the left panel represents the percentage of each histological subtype for SSNs. SSN, subsolid nodule; AIS, adenocarcinoma *in situ*; IA, invasive adenocarcinoma; CT, computed tomography MIA, minimally invasive adenocarcinoma; ROI, region of interest.

Table S1 Summary statistics of patients in the Shanghai cohort (training dataset and test dataset) and Ningbo cohort

Characteristics	Development dataset (n=1,262)	Testing dataset (n=209)	External dataset (n=100)	P_1	P_2
Age (years)				0.209	0.692
<65	1,008 (79.9)	159 (76.1)	74 (74.0)		
≥65	254 (20.1)	50 (23.9)	26 (26.0)		
Sex				0.174	0.952
Male	435 (34.5)	62 (29.7)	30 (30.0)		
Female	827 (65.5)	147 (70.3)	70 (70.0)		
Nodule count				0.003	0.956
Solitary	1,168 (92.6)	205 (98.1)	98 (98.0)		
Multiple	94 (7.4)	4 (1.9)	2 (2.0)		

P_1 value, training dataset compared with testing dataset; P_2 value, training dataset compared with external dataset.

https://dx.doi.org/10.21037/tlcr-21-971

Table S2 Comparison of SSNet and practicing doctors to differentiate AAH/AIS, MIA, and IA

Performance metrics	SSNet	Practicing doctors													
		Unassisted							Assisted						
		Junior		Middle		Senior		Micro average	Junior		Middle		Senior		Micro average
		1	2	1	2	1	2		1	2	1	2	1	2	
Sensitivity															
Class 1	0.803	0.790	0.740	0.914	0.815	0.802	0.802	0.811	0.852	0.840	0.790	0.901	0.827	0.728	0.823
Class 2	0.309	0.327	0.418	0.218	0.382	0.382	0.691	0.403	0.345	0.273	0.309	0.273	0.273	0.618	0.348
Class 3	0.933	0.894	0.702	0.750	0.933	0.933	0.923	0.856	0.798	0.731	0.885	0.885	1.000	0.769	0.845
Micro average	0.746	0.782	0.749	0.751	0.808	0.805	0.870	0.734	0.771	0.757	0.777	0.816	0.814	0.797	0.727
Specificity															
Class 1	0.887	0.887	0.818	0.730	0.887	0.893	0.962	0.863	0.774	0.730	0.887	0.836	0.918	0.855	0.833
Class 2	0.919	0.897	0.746	0.924	0.919	0.919	0.881	0.881	0.886	0.838	0.870	0.941	0.946	0.800	0.880
Class 3	0.794	0.794	0.934	0.860	0.831	0.816	0.904	0.857	0.912	0.941	0.816	0.831	0.772	0.949	0.870
Micro average	0.873	0.871	0.818	0.853	0.886	0.884	0.889	0.867	0.865	0.823	0.868	0.858	0.889	0.847	0.863
PPV															
Class 1	0.783	0.674	0.793	0.786	0.780	0.915	0.786	0.750	0.657	0.613	0.780	0.737	0.838	0.720	0.716
Class 2	0.531	0.329	0.583	0.583	0.486	0.633	0.583	0.502	0.475	0.333	0.415	0.577	0.600	0.479	0.464
Class 3	0.776	0.890	0.795	0.808	0.769	0.881	0.808	0.820	0.874	0.905	0.786	0.800	0.770	0.920	0.833
Micro average	0.746	0.705	0.803	0.806	0.780	0.821	0.806	0.747	0.769	0.715	0.775	0.771	0.811	0.752	0.736
NPV															
Class 1	0.898	0.861	0.899	0.904	0.892	0.905	0.904	0.899	0.911	0.899	0.892	0.943	0.913	0.861	0.902
Class 2	0.817	0.812	0.833	0.833	0.818	0.906	0.833	0.832	0.820	0.795	0.809	0.813	0.814	0.876	0.820
Class 3	0.939	0.804	0.941	0.942	0.908	0.939	0.942	0.886	0.855	0.821	0.902	0.904	1.000	0.843	0.880
Micro average	0.873	0.848	0.886	0.888	0.872	0.921	0.888	0.867	0.866	0.853	0.869	0.889	0.891	0.877	0.864
F1 score															
Class 1	0.793	0.706	0.798	0.800	0.785	0.855	0.800	0.779	0.742	0.708	0.785	0.811	0.832	0.724	0.766
Class 2	0.391	0.368	0.462	0.462	0.391	0.661	0.462	0.447	0.400	0.300	0.354	0.370	0.375	0.540	0.398
Class 3	0.847	0.785	0.858	0.866	0.827	0.901	0.866	0.838	0.834	0.809	0.833	0.840	0.870	0.838	0.839
Micro average	0.746	0.726	0.804	0.807	0.781	0.845	0.807	0.779	0.770	0.735	0.776	0.793	0.812	0.774	0.766
Accuracy															
Class 1	0.858	0.884	0.926	0.926	0.921	0.952	0.926	0.916	0.889	0.868	0.921	0.924	0.940	0.897	0.907
Class 2	0.779	0.803	0.886	0.886	0.868	0.912	0.886	0.871	0.865	0.829	0.852	0.881	0.884	0.863	0.863
Class 3	0.854	0.909	0.929	0.933	0.912	0.954	0.933	0.923	0.926	0.919	0.916	0.921	0.931	0.931	0.924
Micro average	0.831	0.884	0.922	0.923	0.945	0.937	0.923	0.924	0.907	0.888	0.910	0.915	0.926	0.906	0.921
AUPRC															
Macro average	0.685	0.668	0.620	0.606	0.709	0.706	0.806		0.659	0.606	0.657	0.674	0.692	0.701	
Micro average	0.750	0.729	0.650	0.683	0.767	0.763	0.829		0.713	0.663	0.721	0.750	0.775	0.721	
Fleiss' kappa					0.601							0.596			

1, 2 represents doctors 1 and 2; class 1 represents AAH/AIS, class 2 represents MIA, and class 3 represents IA. AAH, atypical adenomatous hyperplasia. AIS, adenocarcinoma *in situ*; AUPRC, area under precision-recall curve; IA, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma; NPV, negative predictive value; PPV, positive predictive value.

https://dx.doi.org/10.21037/tlcr-21-971

Table S3 Comparison of SSNet and practicing doctors to differentiate AAH, AIS, MIA, and IA

Performance metrics	SSNet	Practicing doctors													
		Unassisted							Assisted						
		Junior		Middle		Senior		Micro average	Junior		Middle		Senior		Micro average
		1	2	1	2	1	2		1	2	1	2	1	2	
Sensitivity															
Class 1	0.286	0.429	0.286	0.429	0.071	0.286	0.571	0.345	0.429	0.429	0.071	0.286	0.143	0.571	0.321
Class 2	0.761	0.358	0.597	0.791	0.687	0.791	0.701	0.654	0.687	0.552	0.761	0.836	0.731	0.642	0.701
Class 3	0.309	0.327	0.418	0.218	0.382	0.382	0.691	0.403	0.345	0.273	0.309	0.273	0.618	0.273	0.348
Class 4	0.933	0.894	0.702	0.750	0.933	0.933	0.923	0.856	0.798	0.731	0.885	0.885	0.769	1.000	0.845
Micro average	0.704	0.588	0.583	0.621	0.688	0.729	0.788	0.641	0.642	0.559	0.671	0.696	0.688	0.708	0.640
Specificity															
Class 1	0.991	0.836	0.929	0.956	0.965	1.000	0.960	0.941	0.942	0.858	0.991	0.982	0.996	0.934	0.951
Class 2	0.850	0.913	0.838	0.723	0.832	0.855	0.960	0.854	0.769	0.792	0.838	0.798	0.827	0.919	0.824
Class 3	0.919	0.897	0.746	0.924	0.919	0.919	0.881	0.881	0.886	0.838	0.870	0.941	0.800	0.946	0.880
Class 4	0.794	0.794	0.934	0.860	0.831	0.816	0.904	0.857	0.912	0.941	0.816	0.831	0.949	0.772	0.870
Micro average	0.901	0.863	0.861	0.874	0.896	0.91	0.925	0.908	0.881	0.853	0.890	0.899	0.896	0.903	0.913
PPV															
Class 1	0.667	0.140	0.200	0.375	0.111	1.000	0.471	0.266	0.316	0.158	0.333	0.500	0.667	0.348	0.287
Class 2	0.662	0.615	0.588	0.525	0.613	0.679	0.870	0.634	0.535	0.507	0.646	0.615	0.620	0.754	0.606
Class 3	0.531	0.486	0.329	0.462	0.583	0.583	0.633	0.502	0.475	0.333	0.415	0.577	0.479	0.600	0.464
Class 4	0.776	0.769	0.890	0.804	0.808	0.795	0.881	0.820	0.874	0.905	0.786	0.800	0.920	0.770	0.833
Micro average	0.704	0.588	0.583	0.621	0.688	0.729	0.788	0.588	0.642	0.558	0.671	0.696	0.688	0.708	0.587
NPV															
Class 1	0.957	0.959	0.955	0.964	0.944	0.958	0.973	0.959	0.964	0.960	0.945	0.957	0.949	0.972	0.958
Class 2	0.902	0.786	0.843	0.899	0.873	0.914	0.892	0.864	0.864	0.820	0.901	0.926	0.888	0.869	0.877
Class 3	0.817	0.818	0.812	0.799	0.833	0.833	0.906	0.832	0.820	0.795	0.809	0.813	0.876	0.814	0.820
Class 4	0.939	0.908	0.804	0.818	0.942	0.941	0.939	0.886	0.855	0.821	0.902	0.904	0.843	1.000	0.880
Micro average	0.901	0.863	0.861	0.874	0.896	0.910	0.929	0.888	0.881	0.853	0.890	0.899	0.896	0.903	0.888
F1 score															
Class 1	0.400	0.235	0.444	0.087	0.211	0.516	0.301	0.400	0.364	0.231	0.118	0.364	0.235	0.432	0.303
Class 2	0.708	0.593	0.731	0.648	0.453	0.777	0.644	0.631	0.601	0.529	0.699	0.709	0.671	0.694	0.651
Class 3	0.391	0.368	0.462	0.462	0.391	0.661	0.447	0.296	0.400	0.300	0.354	0.370	0.540	0.375	0.398
Class 4	0.847	0.785	0.858	0.866	0.827	0.901	0.838	0.776	0.834	0.809	0.833	0.840	0.838	0.870	0.839
Micro average	0.704	0.583	0.729	0.688	0.588	0.788	0.643	0.621	0.642	0.558	0.671	0.696	0.688	0.708	0.644
Accuracy															
Class 1	0.950	0.943	0.979	0.952	0.897	0.968	0.874	0.961	0.954	0.909	0.968	0.970	0.972	0.954	0.879
Class 2	0.825	0.871	0.912	0.884	0.863	0.940	0.812	0.852	0.854	0.841	0.899	0.894	0.889	0.914	0.807
Class 3	0.779	0.803	0.886	0.886	0.868	0.912	0.796	0.865	0.865	0.829	0.852	0.881	0.863	0.884	0.788
Class 4	0.854	0.909	0.929	0.933	0.912	0.954	0.847	0.897	0.926	0.919	0.916	0.921	0.931	0.931	0.848
Micro average	0.852	0.884	0.927	0.915	0.885	0.944	0.951	0.895	0.902	0.876	0.910	0.918	0.915	0.921	0.955
AUPRC															
Macro average	0.559	0.471	0.495	0.526	0.516	0.624	0.714		0.550	0.467	0.501	0.571	0.571	0.593	
Micro average	0.667	0.588	0.583	0.621	0.688	0.729	0.788		0.642	0.558	0.671	0.696	0.688	0.675	
Fleiss' kappa					0.480							0.496			

1, 2 represents doctors 1 and 2; class 1 represents AAH, class 2 represents AIS, class 3 represents MIA, and class 4 represents IA. AAH, atypical adenomatous hyperplasia. AIS, adenocarcinoma *in situ*; AUPRC, area under precision-recall curve; IA, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma; NPV, negative predictive value; PPV, positive predictive value.

https://dx.doi.org/10.21037/tlcr-21-971

Table S4 Performance details of different categories in multiclass differentiation on the participant level

AUC	SSNet	Practicing doctors					
		Unassisted			Assisted		
		Junior	Middle	Senior	Junior	Middle	Senior
Three class							
Class 1	0.879	0.841	0.888	0.921	0.829	0.884	0.870
Class 2	0.696	0.652	0.703	0.829	0.641	0.665	0.768
Class 3	0.914	0.900	0.882	0.928	0.913	0.876	0.946
Four class							
Class 1	0.718	0.703	0.752	0.878	0.751	0.718	0.850
Class 2	0.850	0.776	0.796	0.898	0.736	0.828	0.857
Class 3	0.724	0.652	0.703	0.829	0.641	0.665	0.768
Class 4	0.916	0.900	0.882	0.928	0.913	0.876	0.946

In the 3-class differentiation, class 1 represents AAH/AIS, class 2 represents MIA, and class 3 represents IA. In the 4-class differentiation, class 1 represents AAH, class 2 represents AIS, class 3 represents MIA, and class 4 represents IA. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma *in situ*; AUC, area under receiver operating characteristic curve; IA, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma.

https://dx.doi.org/10.21037/tlcr-21-971

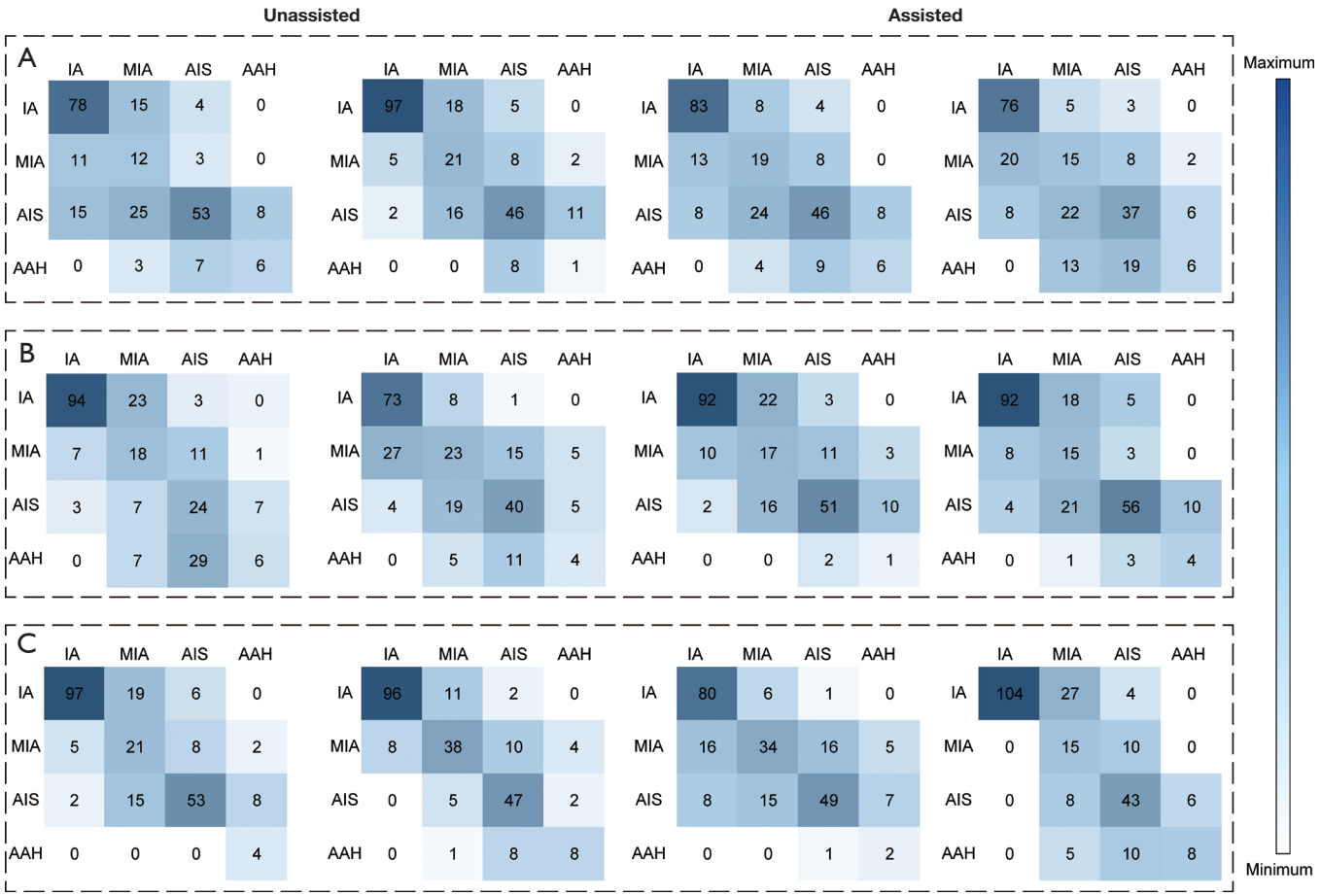

Figure S2 Confusion matrix demonstrating the correlation between prediction (row) and observed (column) labels of subsolid nodules by practicing doctors. (A) Junior rank, (B) middle rank, and (C) senior rank in 4-category classification. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma *in situ*; IA, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma.

https://dx.doi.org/10.21037/tlcr-21-971