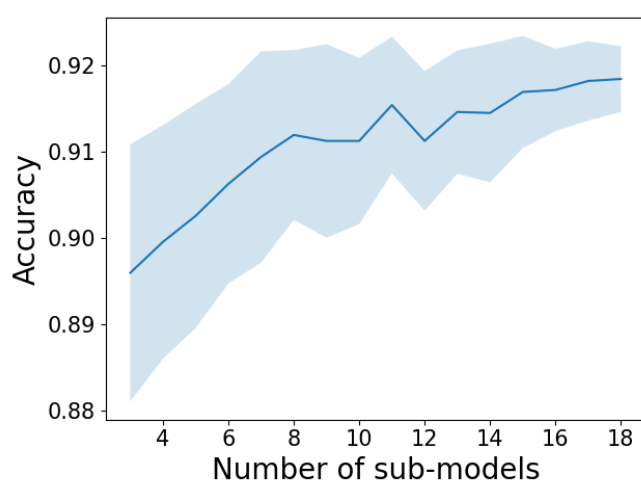
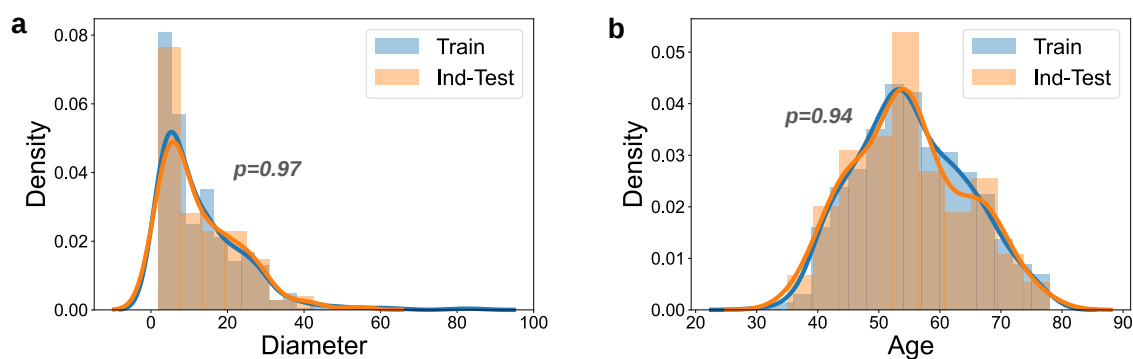


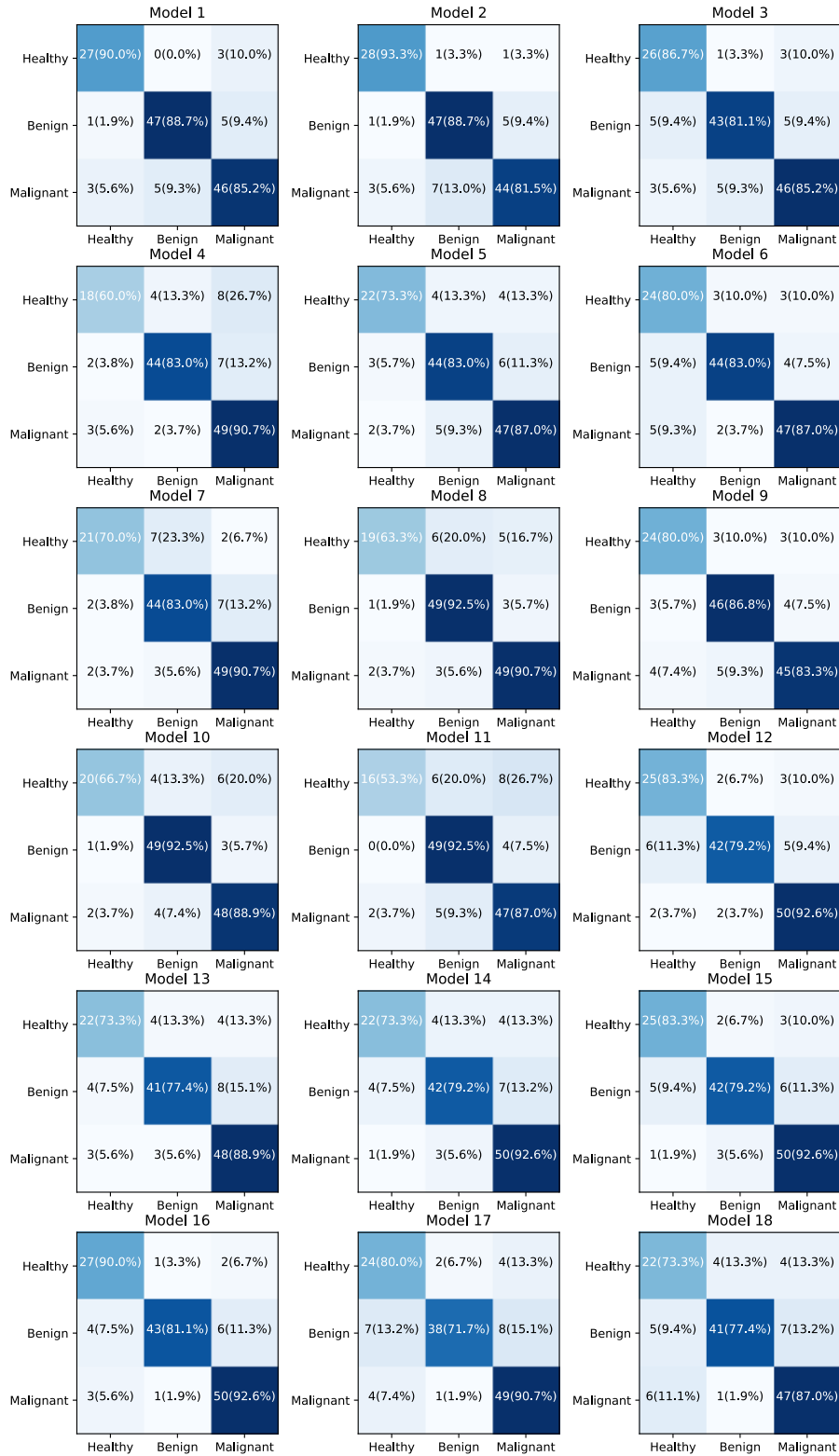
Supplementary Fig. 1 | Comparison of traditional methods and end-to-end methods. The running time was estimated based on the discovery dataset (n=686).



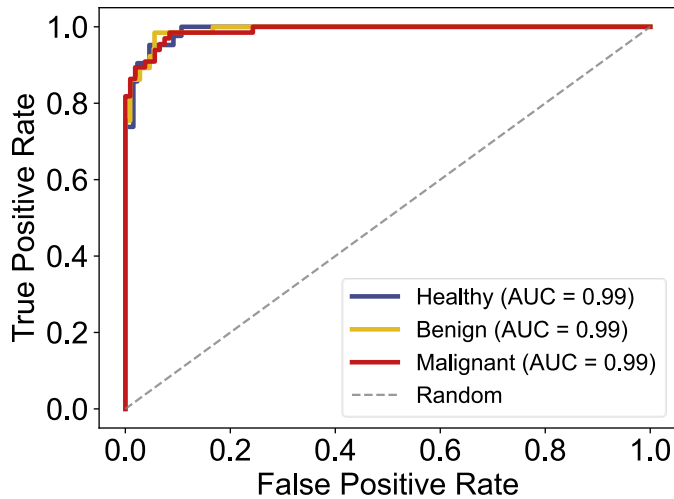
Supplementary Fig. 2 | The impact of the number of sub-models on performance. The shading of solid lines (mean) represents error bars (standard deviation).



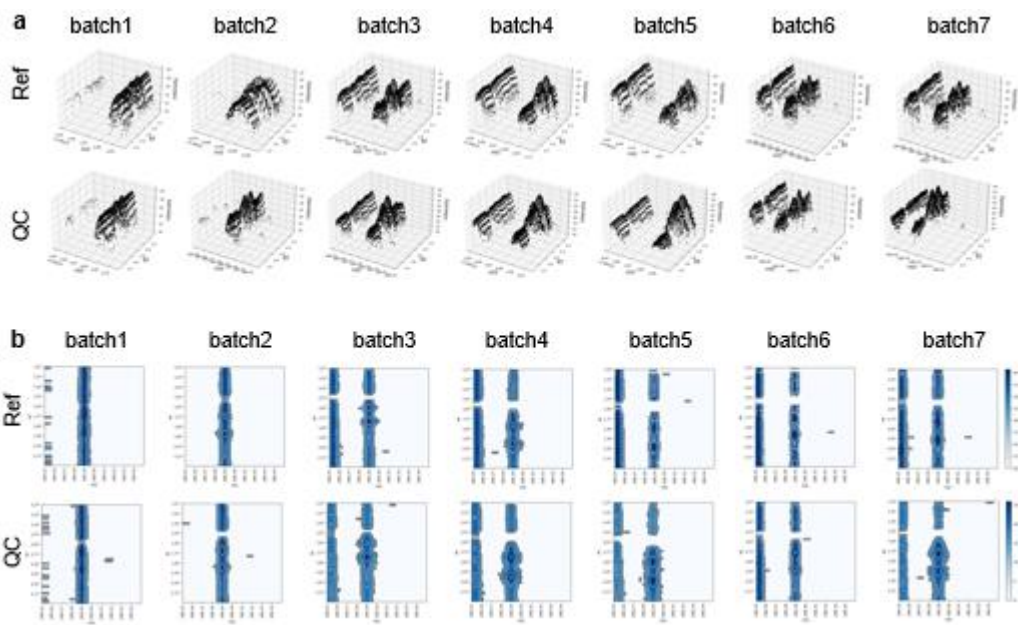
Supplementary Fig. 3 | Distribution of lesion size and age in discovery (Train) dataset (n=686) and independent testing (Ind-Test) dataset (n=173). Independence was calculated by T-test (two-sided).



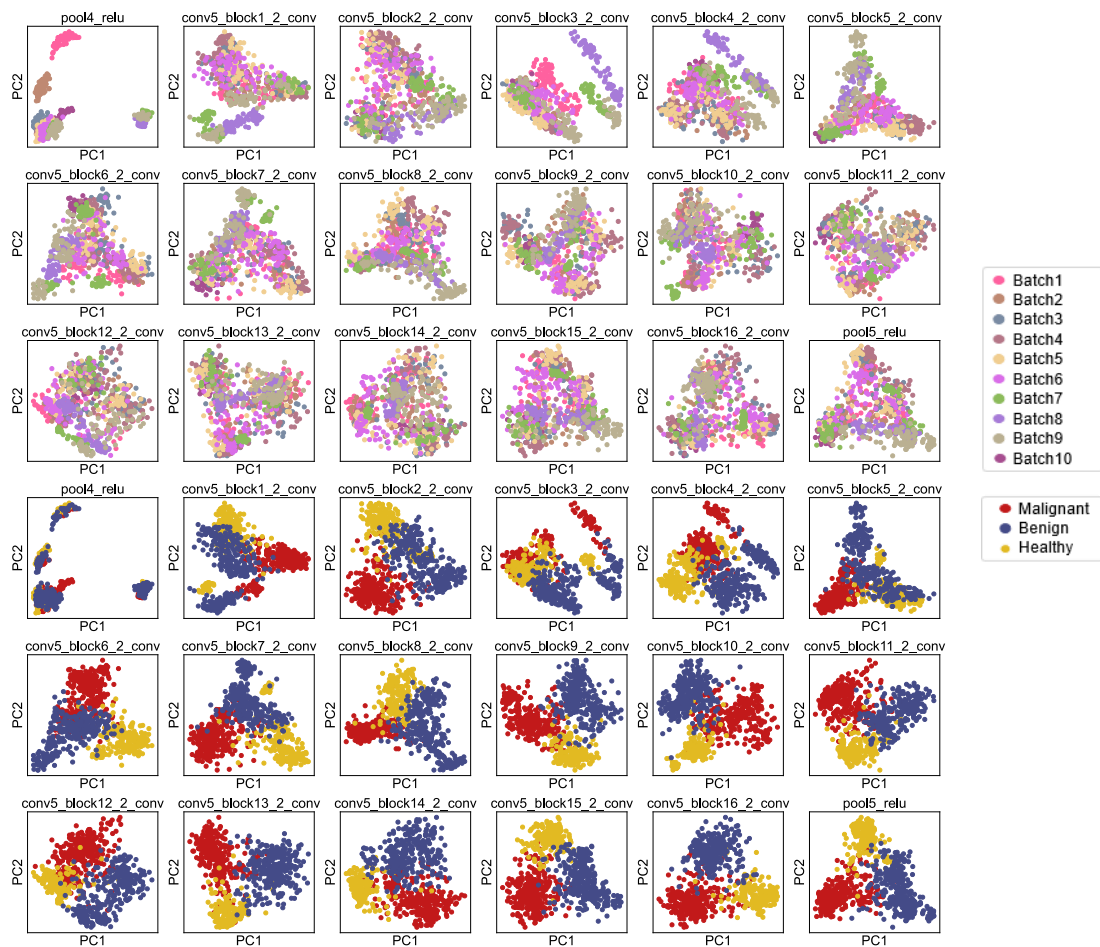
Supplementary Fig. 4 | Confusion matrix of each sub-model in ensemble model. The numbers inside the boxes are the number of matched samples between the true label and the predicted label. The ratio in parentheses is the number of matched samples divided by the number of all samples of the true label.



Supplementary Fig. 5 | Predicted receiver operating characteristic (ROC) curves of the DeepMSProfiler in different classes. Healthy: healthy individuals; Benign: benign lung nodules; Malignant: lung adenocarcinoma. Random: performance baseline in a random state.

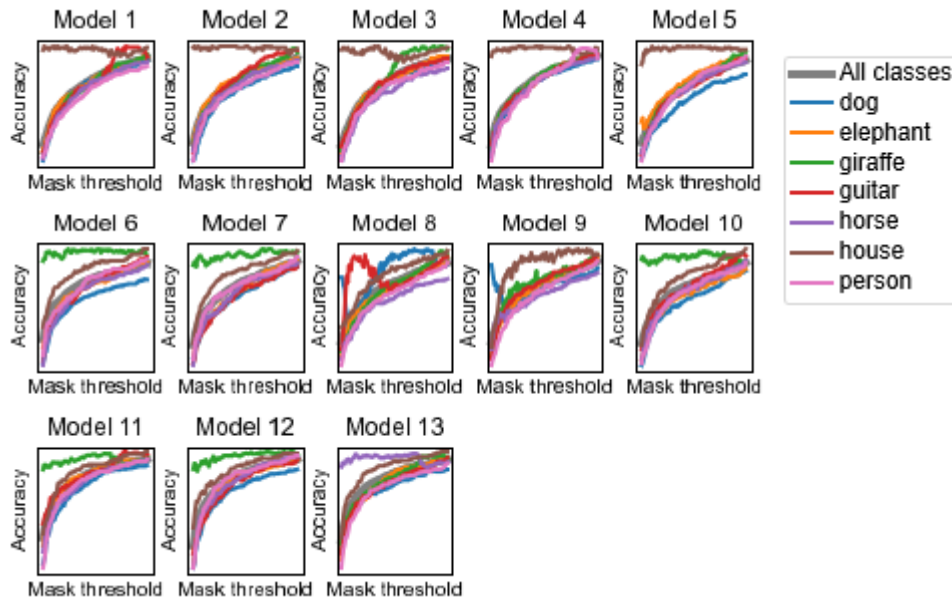


Supplementary Fig. 6 | 3D point array and 2D mapped heatmap of reference samples and QC samples. (a) 3D point array of reference samples and quality control (QC) samples in different batches. (b) 2D mapped heatmap of reference samples and QC samples in different batches. RT: retention time; m/z: mass-to-charge ratio.

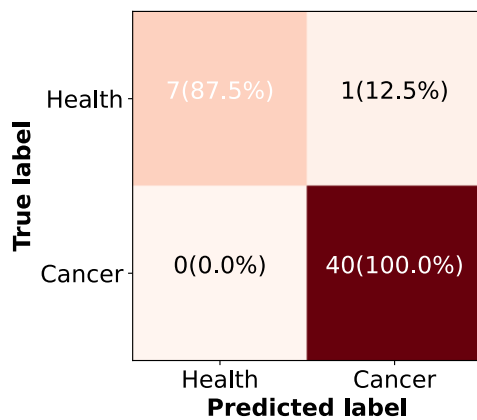


Supplementary Fig. 7 | The visualization of dimensionality reduction for the output data of the hidden

layers of fifth feature extraction module in DeepMSProfiler. Upper: different colours indicate different sample batches; Lower: different colours represent different population classes.



Supplementary Fig. 8 | Feature scoring of different models in Photo-Art-Cartoon-Sketch (PACS) dataset. Lines in different colours represent the scores of different categories, and grey lines represent score across all categories.



Supplementary Fig. 9 | The confusion matrix of the DeepMSProfiler model in colon cancer dataset. The numbers inside the boxes are the number of matched samples between the true label and the predicted label. The ratio in parentheses is the number of matched samples divided by the number of all samples of the true label.

Supplementary Table 1. Sample information statistics

Healthy (N=210)	sex	F	M					
		105	105					
	age	40-49	50-59	60-69	70-79			
65		80	52	13				
Benign (N=323)	sex	F	M					
		155	168					
	age	30-39	40-49	50-59	60-69	70-79		
		14	96	134	69	10		
	diameter (mm)	1-10	11-20	21-30	11-20	21-30	unknown	
289	28	1	1	1	5			
Malignant (N=326)	sex	F	M					
		160	166					
	age	30-39	40-49	50-59	60-69	70-79	unknown	
		4	63	124	85	34	16	
	diameter (mm)	1-10	11-20	21-30	31-40	41-50	>51	unknown
	45	121	94	13	6	2	45	
	stage	I	II	III	IV	unknown		
273		21	10	4	18			
smoke	No	Yes	unknown					
	146	48	132					

Supplementary Table 2. Correlation between sample information and classification labels

	Pearson		Spearman		Kendall		Mutual information
	cor	p value	cor	p value	cor	p value	cor
sex	6.23E-18	1	0	1	0	1	0.000139427
diameter	0.785275	9.36E-82	0.840548	5.32E-104	0.706665	5.90E-61	0.372874274
age	0.072946	0.06750766	0.071043	0.075	0.056084	0.072465	0.017671851
order	-0.05242	0.18919242	-0.052525	0.188308	-0.029977	0.329566	0.291241762
batch	2.25E-17	1	0	1	0	1	0.093945682

Note: The correlation statistical method is two-sided.

Supplementary Table 3. Performance on different datasets

Dataset	Accuracy	AUC score	Precision	Recall	Sensitivity	Specificity
training	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00	1.00±0.00
validation	0.92±0.01	0.99±0.00	0.91±0.01	0.92±0.01	0.92±0.01	0.96±0.00
testing	0.91±0.01	0.99±0.00	0.91±0.01	0.91±0.01	0.91±0.01	0.96±0.00

Supplementary Table 4. Model sensitivity at different specificities

	Sensitivity	
	=0.7	=0.9
Specificity(SVM)	0.95	0.67
Specificity(RF)	0.93	0.67
Specificity(AdaBoost)	0.97	0.79
Specificity(XGBoost)	0.98	0.93
Specificity(Densenet121)	0.99	0.95
Specificity(DeepMSPProfiler)	1.00	0.96

Supplementary Table 5. Metabolites verified with chemical standards

metabolites	mz	rt	score	pathway	robustness_score	confidence_score	annoation	type
Serine	106.0499	9.5115	0.703866551	Glycine, serine and threonine metabolism(0.8)	1	0.996214028	HMDB0003406	Metabolite_Endogenous
Valine	118.0863	8.45325	0.717386295	Valine, leucine and isoleucine degradation(0.61458);Pantothenate and CoA biosynthesis(0.57143);Valine, leucine and isoleucine biosynthesis(0.46154)	0.5	0.983099303	HMDB0000883	Metabolite_Food
Isoleucine	132.1019	8.635428571	0.712766896	Valine, leucine and isoleucine degradation(0.61458);Valine, leucine and isoleucine biosynthesis(0.46154)	1	0.96911901	HMDB0000172	Metabolite_Food
Asparagine	133.0608	12.509	0.718206925	Protein digestion and absorption	1	1	HMDB0000168	Terminal_Metabolite
Glutamic acid	148.0604	13.017	0.710600136	D-Glutamine and D-glutamate metabolism(0.5625)	1	1	HMDB0003339	Metabolite_Food
Glutamine	148.0604	12.071	0.740918756	D-Glutamine and D-glutamate metabolism(0.5625)	1	1	HMDB0003423	Metabolite_Endogenous
Carnitine	162.1125	12.116	0.738626601	Bile secretion(0.027778)	1	0.994365916	HMDB0000062	Terminal_Metabolite
Citrulline	176.103	12.698	0.743663802	Arginine biosynthesis(0.54286)	1	0.951462314	HMDB0000904	Metabolite_Endogenous
Serotonin	177.1022	12.904	0.735578244	Bile secretion(0.027778);Tryptophan metabolism(0.384)	1	1	HMDB0000259	Metabolite_Endogenous
13-cis-Retinoic acid	301.2162	1.486	0.772935613	Retinol metabolism(2.2131)	1	1	HMDB0006219	Metabolite_Endogenous
Docosahexaenoic acid	327.233	1.006	0.942932846	Biosynthesis of unsaturated fatty acids(0.015152)	1	0.96118378	HMDB0002183	Metabolite_Endogenous

Note: The ‘score’ values are derived from the results of PIUMet, while the ‘pathway’, ‘robustness_score’, and ‘confidence_score’ are obtained from the outputs of KEGG pathway enrichment analysis using MetaboAnalyst.