Supplementary Online Content

Kehl KL, Elmarakeby H, Nishino M, et al. Assessment of deep natural language processing in ascertaining oncologic outcomes from radiology reports. *JAMA Oncol.* Published online July 25, 2019. doi:10.1001/jamaoncol.2019.1800

This supplementary material has been provided by the authors to give readers additional information about their work.

.

eTable 1A: Human curator training, abstraction process, and interrater reliability

The curation team began with three curators (two with bachelor's degrees, one with a Master of Public Health degree) who had one year of experience reviewing thoracic oncology medical records without the use of the PRISSMM framework. After initial development of the PRISSMM framework, 164 patients with five different types of cancer (breast, N=20; colorectal, N=20; renal, N=21; lung, N=85; and pancreatic cancer, N=18) each underwent curation per PRISSMM by two of the three curators. An epidemiologist with master's-level training and twenty years of oncology data experience (EL) performed quality assurance (QA) by reviewing each record completely against the source medical record documentation.

In preparation for the current analysis, ten patients with lung cancer (of the initial 85 patients) underwent curation by all three curators, supervised by EL, and any differences in curation of imaging reports were adjudicated to create a 'gold standard' set of examples to train subsequent curators. During this adjudication process, differences were resolved with the assistance of two medical oncologists (DS and KLK).

A curation team lead with masters'-level training was then hired to expand the curation team. Additional curators (N=8) were then hired to review records for the remainder of the lung cancer patients in our cohort; seven of the curators performed the imaging report abstraction. During their initial onboarding process, their curations underwent manual quality assurance by the curation team lead. Among the patients in our training subset, 10% then underwent dual curation to calculate interrater reliability. Interrater reliability statistics are provided below:

Outcome	N of dual curated reports	% agreement	kappa
Any cancer	783	90.0%	0.80
Decrease/response	783	96.7%	0.78
Progression/growth	783	89.7%	0.71

The imaging report abstraction process involved the questions below, answered using a REDCap web form (ETable 1b). Curators were asked to review cross-sectional imaging scans (CT, PET, MRI, PET-CT, and other nuclear medicine scans). They did not review ultrasounds and plain films, given the limited role of those treatment modalities in assessing disease status for non-small cell lung cancer. Imaging reports were abstracted from the date of lung cancer diagnosis through the date on which abstraction was conducted. Reports were abstracted for imaging studies conducted either at our institution or at others, if reports from other institutions had been scanned into our medical record. Since large-scale text corresponding to reports from other institutions would require regulatory approval to use those reports as well as optical character recognition (to read scanned PDF files), the current analysis was restricted to imaging reports for studies done at our institution. Abstractors were asked to review each scan report and identify the type of scan, body parts imaged (eg abdomen, pelvis, brain, neck), the date the study was interpreted, and the date of the reference scan for comparison. Next, abstractors were asked to review the 'Impression' section of each report to identify the presence of any cancer; if cancer was present, they were asked to note how it was changing, as well as what anatomical sites were noted to contain cancer. Curators did not incorporate additional EHR data, such as clinician progress notes, into imaging report abstraction.

eTable 1B: Data collection instruments for human curation of cancer status within each radiology report

Question for human curator	Response Options	Notes
Is there any evidence of cancer on this imaging report? *Use only the Impression section of the imaging report to complete this field.*	Yes, the report states there is evidence of cancer	If selected, the outcome of any cancer was coded as positive in a deep learning model for the current imaging report.
	No, the report states there is no evidence of cancer	
	The report mentions cancer but is uncertain, indeterminate, or equivocal	
	Yes, the report states or implies there is evidence of cancer	If selected, the outcome of any cancer was coded as positive in a deep learning model for the current imaging report.
	No, the report states or implies there is no evidence of cancer	
	The report is uncertain, indeterminate, or equivocal	
	The report does not mention cancer	
Which of the following best describes the radiologist's overall interpretation of the patient's cancer status? *Use only the Impression section of the imaging report to complete this field.*	Improving/Responding	If selected, the outcome of "response" was coded as positive in a deep learning model for the current imaging report.
	Stable/No change	
	Mixed	
	Progressing/Worsening/Enlarging	If selected, the outcome of "progression" was coded as positive in a deep learning model for the current imaging report.
	Not stated/Indeterminate	
Select all of the sites thought to be involved with cancer. *Use only the Impression section of the imaging report to complete this field.*	Adrenal gland	If selected, the outcome of "disease in adrenal" was coded as positive in a deep learning model for the current imaging report.
	Bone	If selected, the outcome of "disease in bone" was coded as positive in a deep learning model for the current imaging report.

	Brain or spine	If selected, the outcome of “disease in brain/spine” was coded as positive in a deep learning model for the current imaging report.
	Liver	If selected, the outcome of “disease in liver” was coded as positive in a deep learning model for the current imaging report.
	Lung	
	Lymph nodes (loco/regional)	If selected, the outcome of “disease in lymph nodes” was coded as positive in a deep learning model for the current imaging report.
	Lymph nodes (Distant metastatic)	If selected, the outcome of “disease in lymph nodes” was coded as positive in a deep learning model for the current imaging report.
	Lymph Nodes – NOS	If selected, the outcome of “disease in lymph nodes” was coded as positive in a deep learning model for the current imaging report.
	Peritoneum or peritoneal fluid	
	Pleura or pleural fluid	
	Skin	
	Other Abdomen	
	Other Chest	
	Other Extremity	
	Other Head/Neck	
	Other Pelvis	

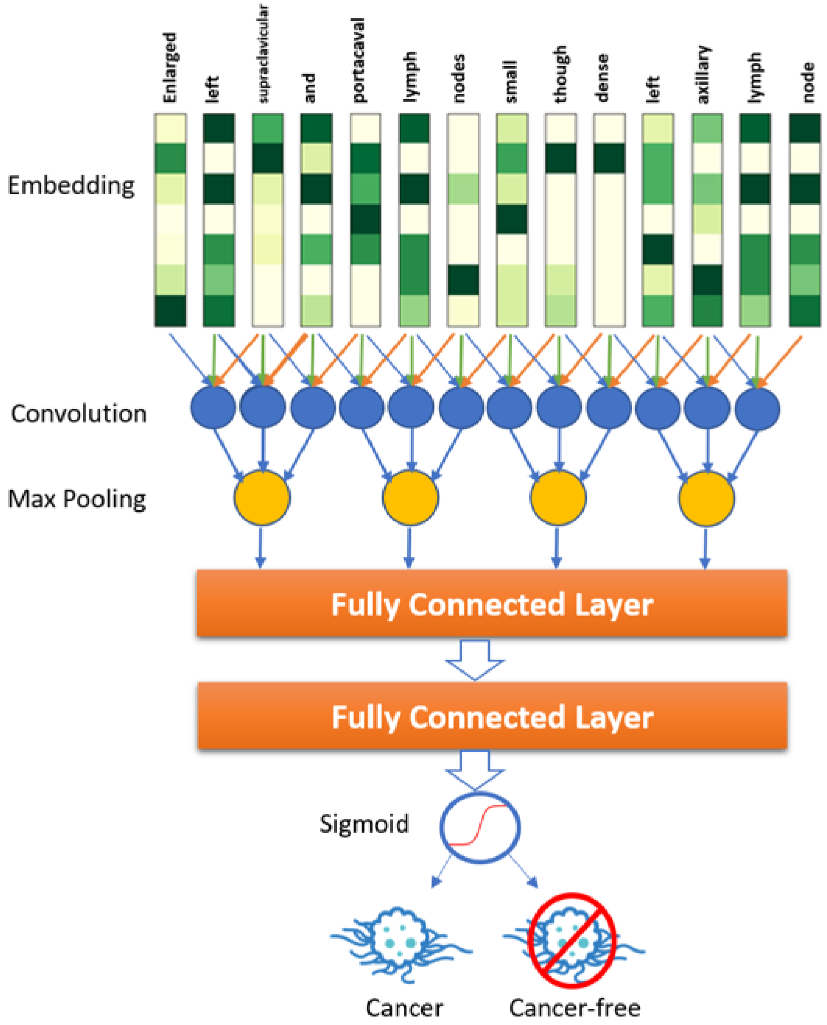
eTable 2: Characteristics of radiology reports

	Curation set (N=14,230 reports)			Reserve set (N=15,000 reports)
	% of training subset (N=11182)	% of validation subset (N=1545)	% of test subset (N=1503)	% of total
All imaging reports curated	100.0	100.0	100.0	100.0
Type of imaging				
CT	74.9	72.8	72.6	67.3
MRI brain/spine	15.3	18.0	17.8	21.0
PET/CT	8.8	8.2	8.6	9.7
MRI body	0.7	0.8	0.8	1.3
Bone scan	0.3	0.3	0.2	0.7
By human curation, did imaging report indicate:				
Any cancer	62.3	57.9	61.2	*
Cancer worsening/progression	25.6	22.5	20.7	*
Cancer improvement/response	11.3	11.0	13.2	*
Cancer in liver	8.8	5.6	5.3	*
Cancer in bone	18.3	15.8	11.6	*
Cancer in brain/spine	8.7	7.9	6.1	*
Cancer in lymph nodes	13.9	13.5	9.1	*
Cancer in adrenal gland	5.3	2.6	2.4	*
Characteristics of patients associated with imaging reports				
Histology				
Adenocarcinoma	78.3	68.7	80.2	77.9
Squamous cell carcinoma	11.4	12.9	10.0	8.5
Small cell lung cancer	4.5	2.8	4.5	3.8
Other/mixed	5.8	15.5	5.3	9.9
Disease extent at original diagnosis				
Early (stage I-III)	54.5	59.6	60.1	*
Metastatic (stage IV)	45.5	40.4	39.9	*

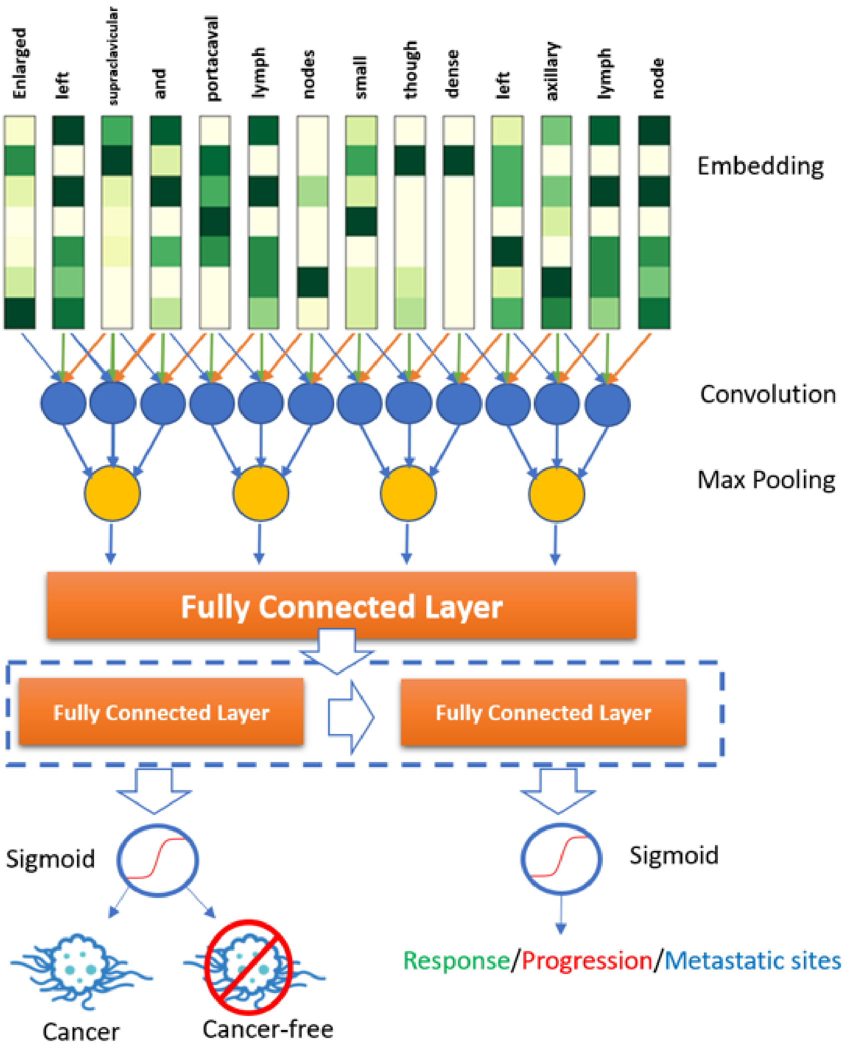
Age at sequencing				
< 50	7.7	6.2	10.7	9.4
50-60	23.4	23.4	20.5	19.8
60-70	36.5	38.1	35.9	36.7
70-80	26.3	27.4	25.9	26.3
> 80	6.1	5.0	7.0	7.8
Gender				
Female	62.0	59.7	61.7	59.5
Male	38.0	40.3	38.3	40.5
Self-reported race				
White	89.8	85.2	88.0	89.8
Asian	4.1	7.5	5.1	4.4
Black/African-American	3.2	0.5	3.9	3.6
Other/unknown	3.0	6.8	3.0	2.2

* Data not available for the reserve set, which has not undergone manual medical record curation

eFigure 1: Deep learning model architectures for imaging report interpretation*



eFigure 1A: Deep learning architecture for ascertaining the presence of cancer

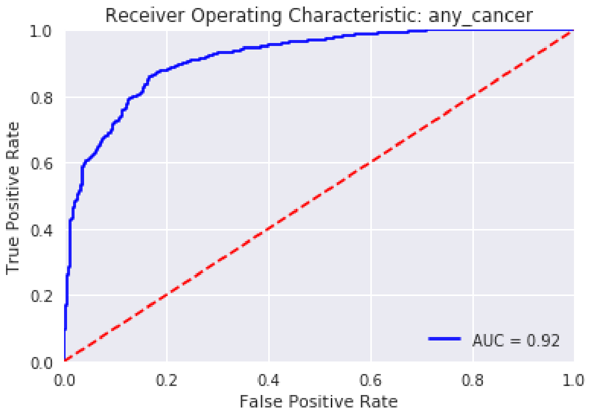
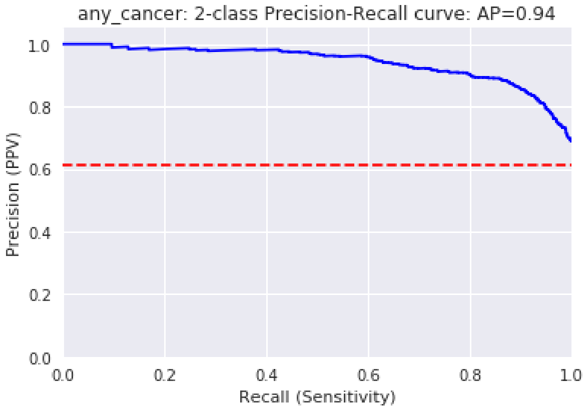
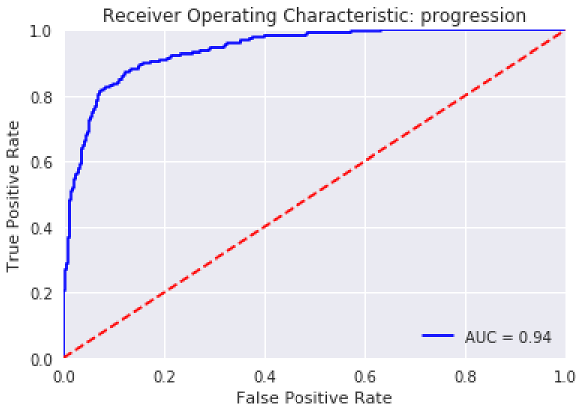
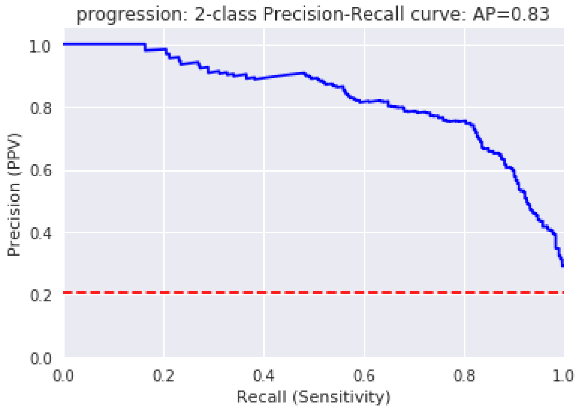


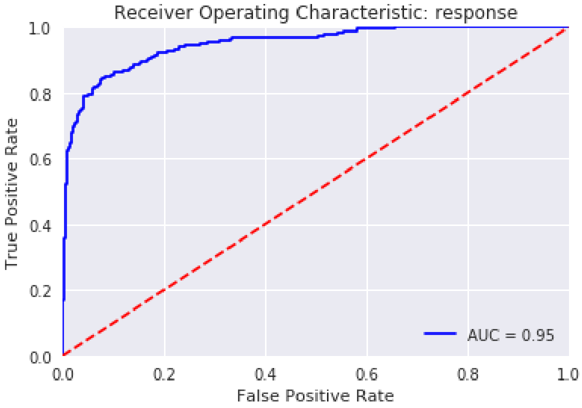
eFigure 1B: Deep learning architecture for ascertaining additional specific outcomes, including the presence of cancer, cancer progression, and cancer response, and specific sites of disease

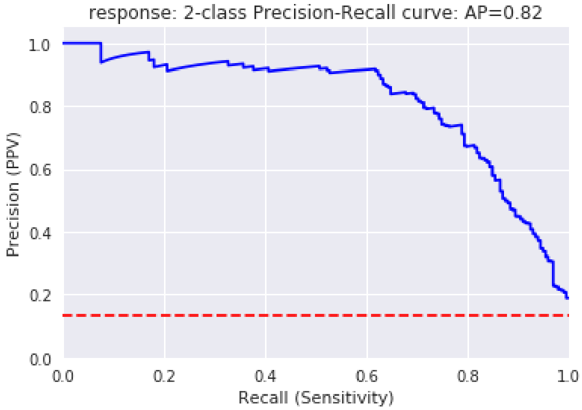
Legend to eFigure 1:

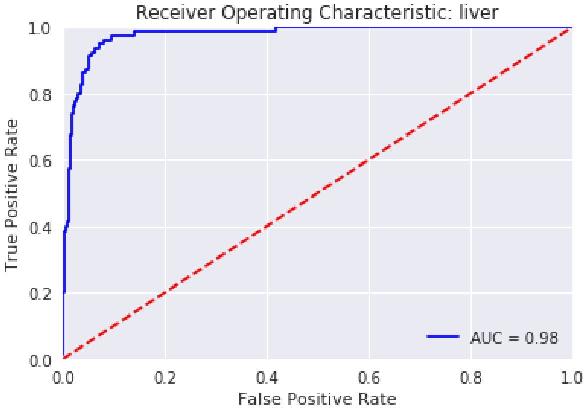
* Specific outcomes in addition to any cancer included response to therapy, progression of disease, and presence of metastases in liver, bone, brain/spine, lymph nodes, and adrenal gland. For each outcome, five separate cross-validation models were constructed and trained on the training dataset. The mean of the predictions from these five models was then calculated to generate an ensemble model prediction for evaluation on the validation and test datasets.

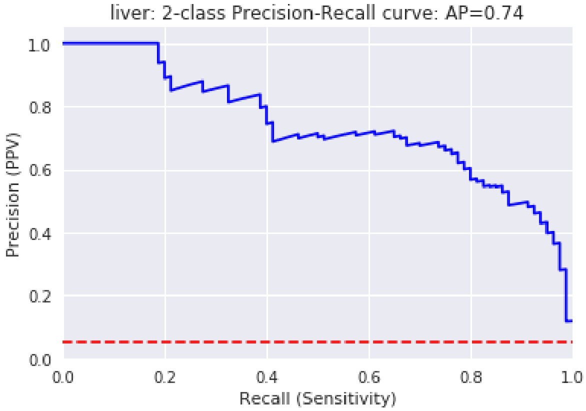
eFigure 2: Graphical depictions of model performance

Outcome	Area under ROC curve*	Area under precision-recall curve†	NPV at F1-optimal threshold‡
Any cancer			81.6%
Disease worsening/progression			95.5%

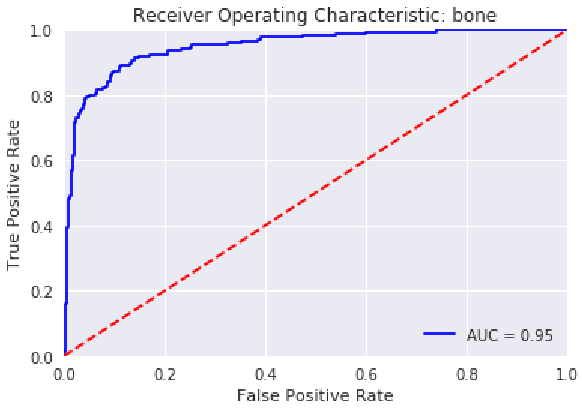
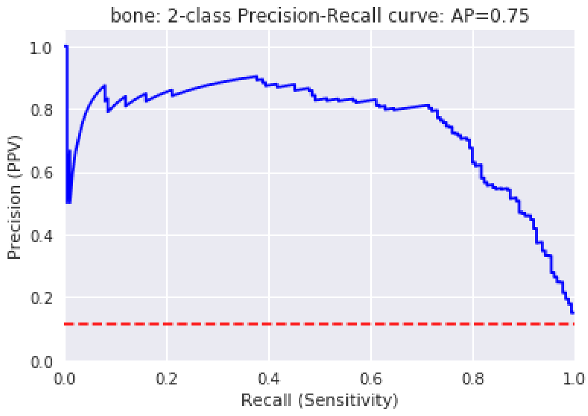
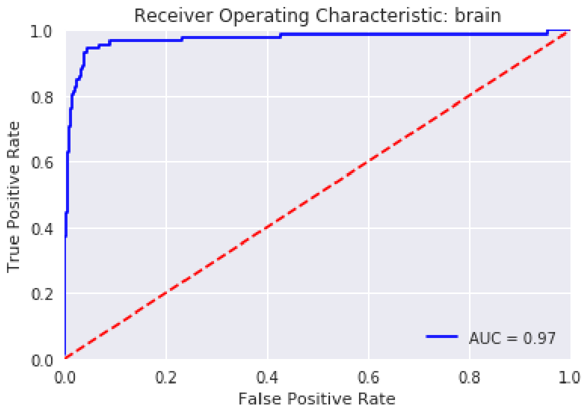
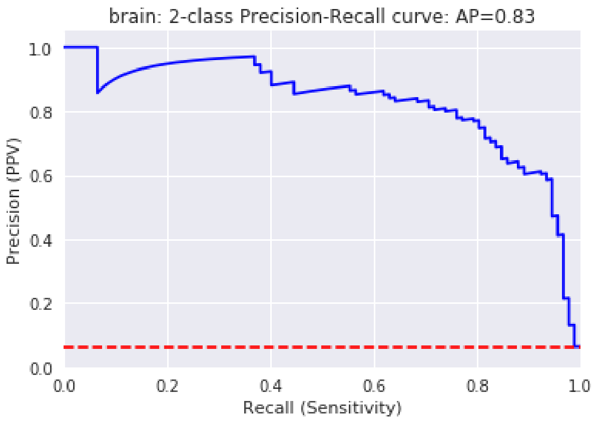
<table>
<tr>
<td><strong>Disease improvement/response</strong></td>
<td>


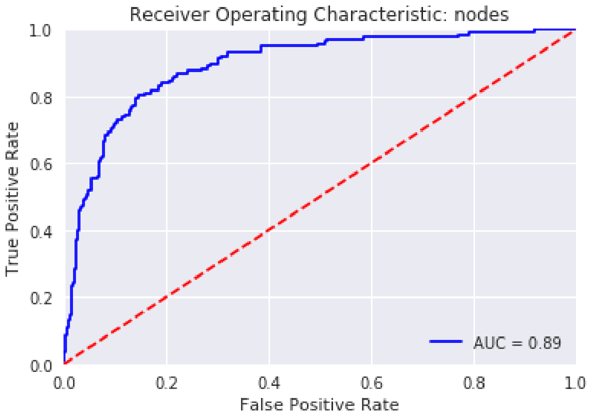
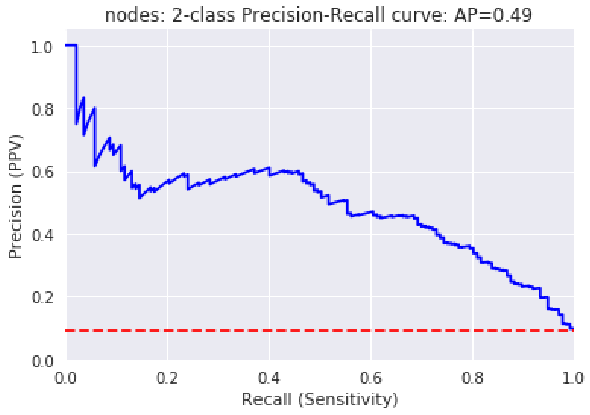
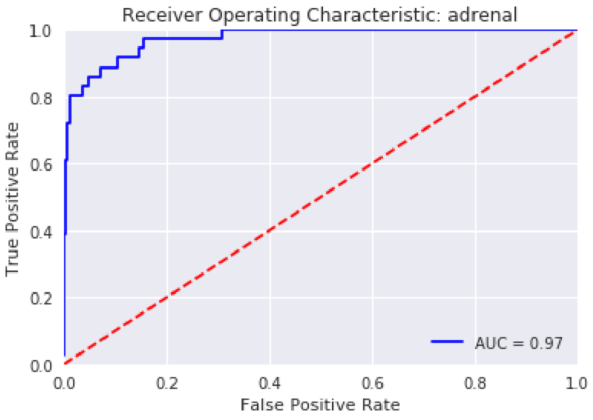
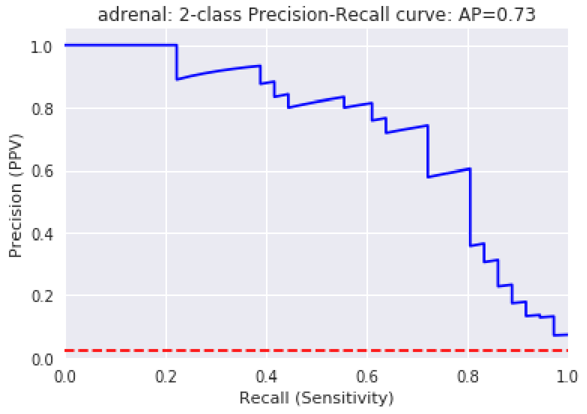
</td>
<td>


</td>
<td>95.5%</td>
</tr>
<tr>
<td><strong>Liver involvement</strong></td>
<td>


</td>
<td>


</td>
<td>98.7%</td>
</tr>
</table>

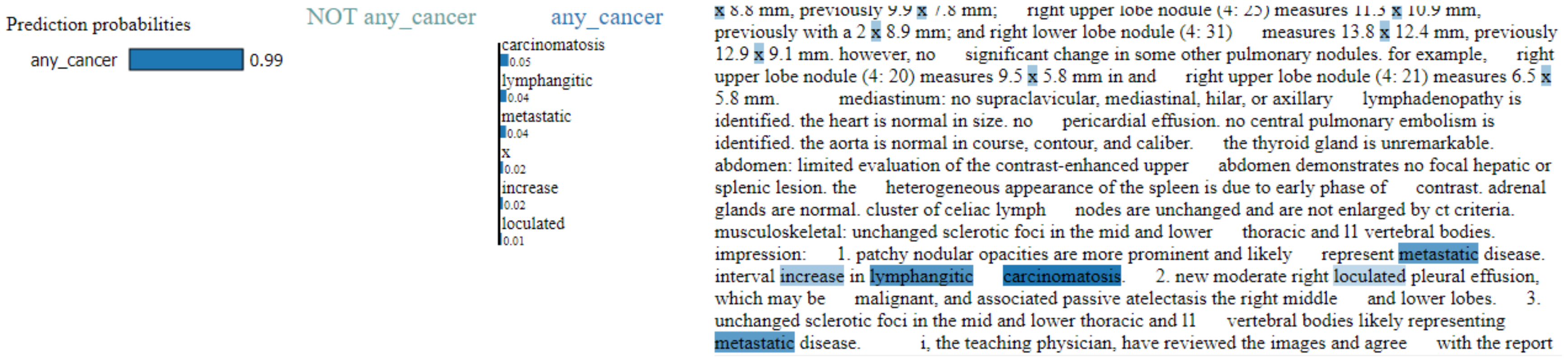
Bone involvement			96.5%
Brain/spine involvement			98.9%

Lymph node involvement			97.0%
Adrenal involvement			99.1%

Legend to eFigure 2:
PPV, positive predictive value
* Area under the receiver-operating characteristic curve. Red diagonal line represents the expected statistic for an uninformative classifier (0.5).
† Area under the precision-recall curve. Red line represents the expected statistic for an uninformative classifier (the proportion of imaging reports that were positive for the outcome of interest).
‡ NPV, negative predictive value. F1 optimal threshold: The threshold probability for defining a 'positive' outcome that maximizes the F1 score, which in turn is the harmonic mean between precision and recall.

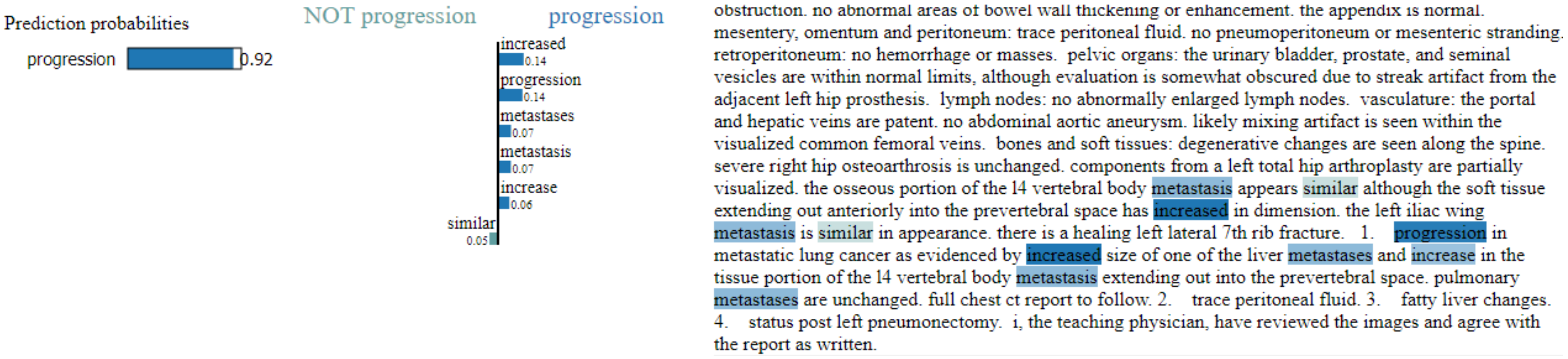
eFigure 3: LIME explanations for individual model predictions

eFigure 3A: Example of prediction regarding any cancer



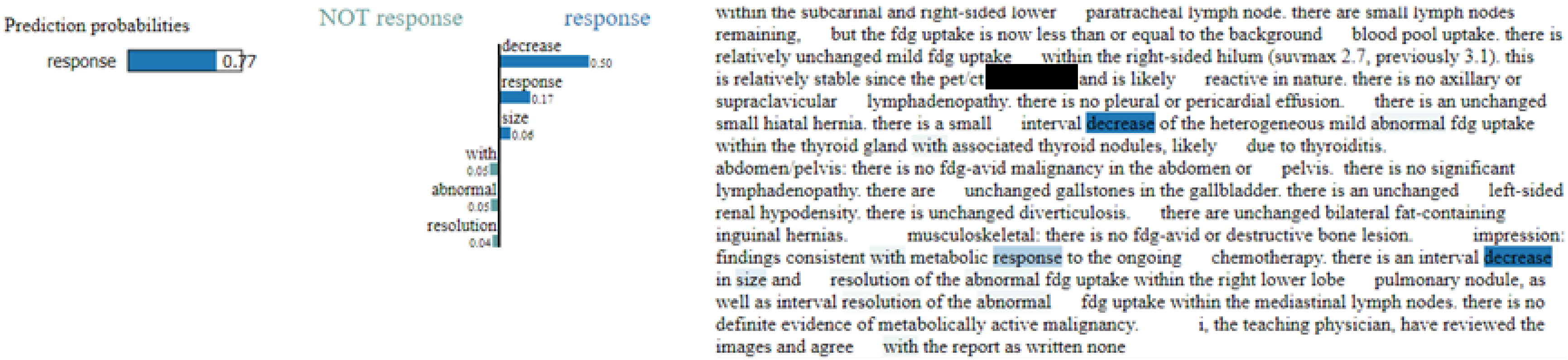
x 8.8 mm, previously 9.9 x 7.8 mm; right upper lobe nodule (4: 25) measures 11.3 x 10.9 mm, previously with a 2 x 8.9 mm; and right lower lobe nodule (4: 31) measures 13.8 x 12.4 mm, previously 12.9 x 9.1 mm. however, no significant change in some other pulmonary nodules. for example, right upper lobe nodule (4: 20) measures 9.5 x 5.8 mm in and right upper lobe nodule (4: 21) measures 6.5 x 5.8 mm. mediastinum: no supraclavicular, mediastinal, hilar, or axillary lymphadenopathy is identified. the heart is normal in size. no pericardial effusion. no central pulmonary embolism is identified. the aorta is normal in course, contour, and caliber. the thyroid gland is unremarkable. abdomen: limited evaluation of the contrast-enhanced upper abdomen demonstrates no focal hepatic or splenic lesion. the heterogeneous appearance of the spleen is due to early phase of contrast. adrenal glands are normal. cluster of celiac lymph nodes are unchanged and are not enlarged by ct criteria. musculoskeletal: unchanged sclerotic foci in the mid and lower thoracic and l1 vertebral bodies. impression: 1. patchy nodular opacities are more prominent and likely represent metastatic disease. interval increase in lymphangitic carcinomatosis. 2. new moderate right loculated pleural effusion, which may be malignant, and associated passive atelectasis the right middle and lower lobes. 3. unchanged sclerotic foci in the mid and lower thoracic and l1 vertebral bodies likely representing metastatic disease. i, the teaching physician, have reviewed the images and agree with the report

eFigure 3B: Example of prediction regarding disease worsening/progression



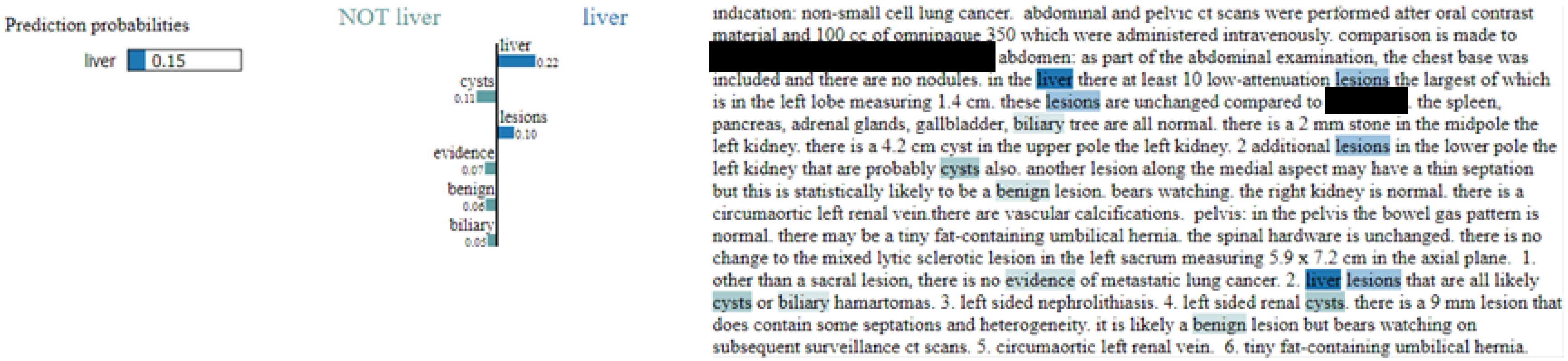
obstruction. no abnormal areas of bowel wall thickening or enhancement. the appendix is normal. mesentery, omentum and peritoneum: trace peritoneal fluid. no pneumoperitoneum or mesenteric stranding. retroperitoneum: no hemorrhage or masses. pelvic organs: the urinary bladder, prostate, and seminal vesicles are within normal limits, although evaluation is somewhat obscured due to streak artifact from the adjacent left hip prosthesis. lymph nodes: no abnormally enlarged lymph nodes. vasculature: the portal and hepatic veins are patent. no abdominal aortic aneurysm. likely mixing artifact is seen within the visualized common femoral veins. bones and soft tissues: degenerative changes are seen along the spine. severe right hip osteoarthrosis is unchanged. components from a left total hip arthroplasty are partially visualized. the osseous portion of the l4 vertebral body metastasis appears similar although the soft tissue extending out anteriorly into the prevertebral space has increased in dimension. the left iliac wing metastasis is similar in appearance. there is a healing left lateral 7th rib fracture. 1. progression in metastatic lung cancer as evidenced by increased size of one of the liver metastases and increase in the tissue portion of the l4 vertebral body metastasis extending out into the prevertebral space. pulmonary metastases are unchanged. full chest ct report to follow. 2. trace peritoneal fluid. 3. fatty liver changes. 4. status post left pneumonectomy. i, the teaching physician, have reviewed the images and agree with the report as written.

eFigure 3C: Explanation of prediction regarding disease improvement/response



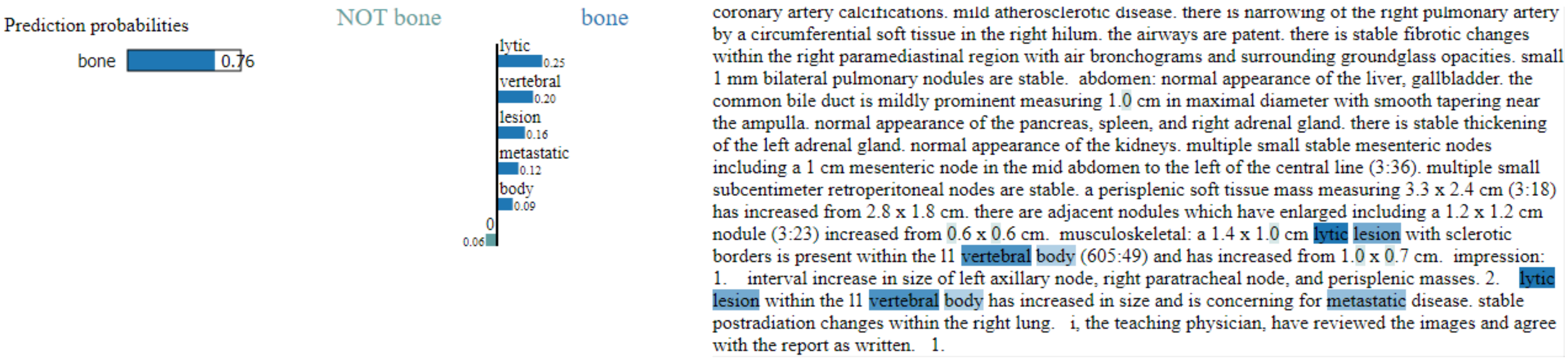
within the subcarinal and right-sided lower paratracheal lymph node. there are small lymph nodes remaining, but the fdg uptake is now less than or equal to the background blood pool uptake. there is relatively unchanged mild fdg uptake within the right-sided hilum (suvmax 2.7, previously 3.1). this is relatively stable since the pet/ct [redacted] and is likely reactive in nature. there is no axillary or supraclavicular lymphadenopathy. there is no pleural or pericardial effusion. there is an unchanged small hiatal hernia. there is a small interval decrease of the heterogeneous mild abnormal fdg uptake within the thyroid gland with associated thyroid nodules, likely due to thyroiditis. abdomen/pelvis: there is no fdg-avid malignancy in the abdomen or pelvis. there is no significant lymphadenopathy. there are unchanged gallstones in the gallbladder. there is an unchanged left-sided renal hypodensity. there is unchanged diverticulosis. there are unchanged bilateral fat-containing inguinal hernias. musculoskeletal: there is no fdg-avid or destructive bone lesion. impression: findings consistent with metabolic response to the ongoing chemotherapy. there is an interval decrease in size and resolution of the abnormal fdg uptake within the right lower lobe pulmonary nodule, as well as interval resolution of the abnormal fdg uptake within the mediastinal lymph nodes. there is no definite evidence of metabolically active malignancy. i, the teaching physician, have reviewed the images and agree with the report as written none

eFigure 3D: Explanation of prediction regarding liver involvement



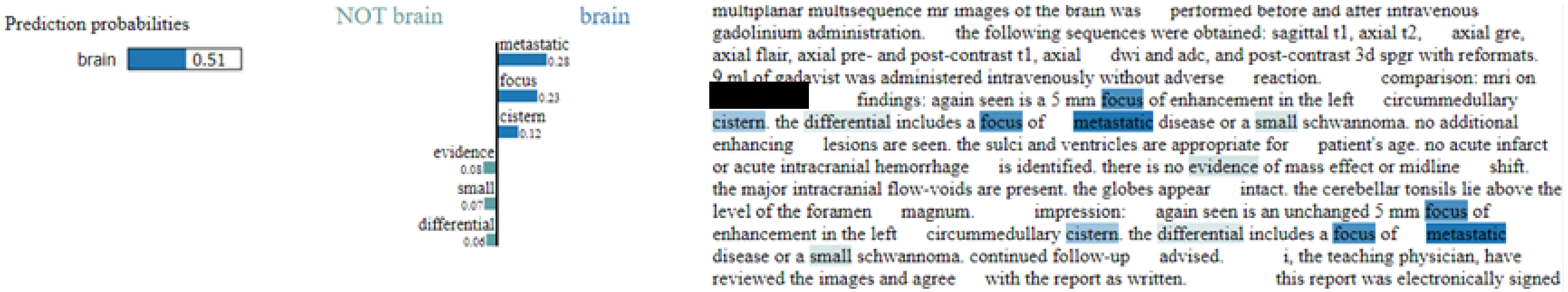
indication: non-small cell lung cancer. abdominal and pelvic ct scans were performed after oral contrast material and 100 cc of omnipaque 350 which were administered intravenously. comparison is made to [redacted] abdomen: as part of the abdominal examination, the chest base was included and there are no nodules. in the liver there at least 10 low-attenuation lesions the largest of which is in the left lobe measuring 1.4 cm. these lesions are unchanged compared to [redacted]. the spleen, pancreas, adrenal glands, gallbladder, biliary tree are all normal. there is a 2 mm stone in the midpole the left kidney. there is a 4.2 cm cyst in the upper pole the left kidney. 2 additional lesions in the lower pole the left kidney that are probably cysts also. another lesion along the medial aspect may have a thin septation but this is statistically likely to be a benign lesion. bears watching. the right kidney is normal. there is a circumaortic left renal vein.there are vascular calcifications. pelvis: in the pelvis the bowel gas pattern is normal. there may be a tiny fat-containing umbilical hernia. the spinal hardware is unchanged. there is no change to the mixed lytic sclerotic lesion in the left sacrum measuring 5.9 x 7.2 cm in the axial plane. 1. other than a sacral lesion, there is no evidence of metastatic lung cancer. 2. liver lesions that are all likely cysts or biliary hamartomas. 3. left sided nephrolithiasis. 4. left sided renal cysts. there is a 9 mm lesion that does contain some septations and heterogeneity. it is likely a benign lesion but bears watching on subsequent surveillance ct scans. 5. circumaortic left renal vein. 6. tiny fat-containing umbilical hernia.

eFigure 3E: Explanation of prediction regarding bone involvement

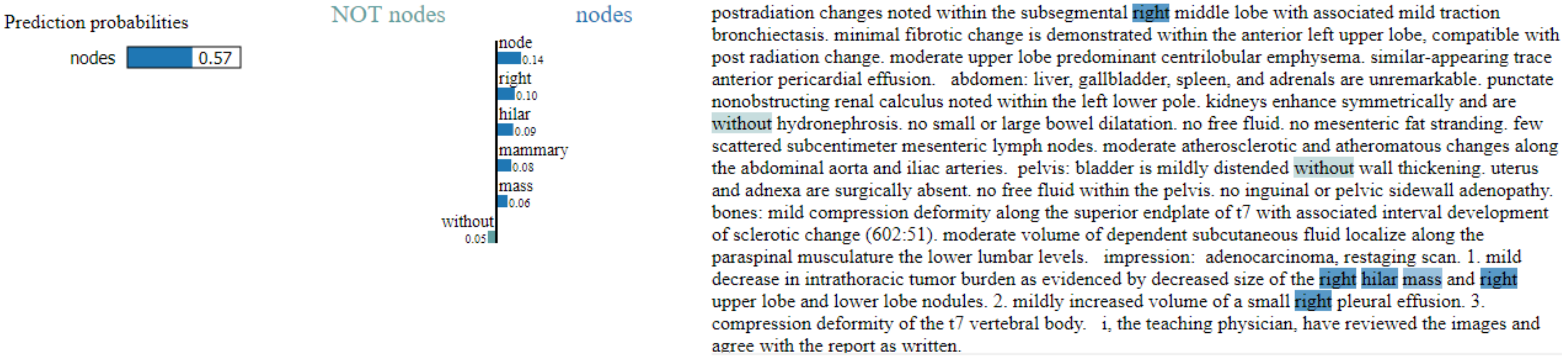


coronary artery calcifications. mild atherosclerotic disease. there is narrowing of the right pulmonary artery by a circumferential soft tissue in the right hilum. the airways are patent. there is stable fibrotic changes within the right paramediastinal region with air bronchograms and surrounding groundglass opacities. small 1 mm bilateral pulmonary nodules are stable. abdomen: normal appearance of the liver, gallbladder. the common bile duct is mildly prominent measuring 1.0 cm in maximal diameter with smooth tapering near the ampulla. normal appearance of the pancreas, spleen, and right adrenal gland. there is stable thickening of the left adrenal gland. normal appearance of the kidneys. multiple small stable mesenteric nodes including a 1 cm mesenteric node in the mid abdomen to the left of the central line (3:36). multiple small subcentimeter retroperitoneal nodes are stable. a perisplenic soft tissue mass measuring 3.3 x 2.4 cm (3:18) has increased from 2.8 x 1.8 cm. there are adjacent nodules which have enlarged including a 1.2 x 1.2 cm nodule (3:23) increased from 0.6 x 0.6 cm. musculoskeletal: a 1.4 x 1.0 cm lytic lesion with sclerotic borders is present within the 11 vertebral body (605:49) and has increased from 1.0 x 0.7 cm. impression: 1. interval increase in size of left axillary node, right paratracheal node, and perisplenic masses. 2. lytic lesion within the 11 vertebral body has increased in size and is concerning for metastatic disease. stable postradiation changes within the right lung. i, the teaching physician, have reviewed the images and agree with the report as written. 1.

eFigure 3F: Explanation of prediction regarding brain/spine involvement

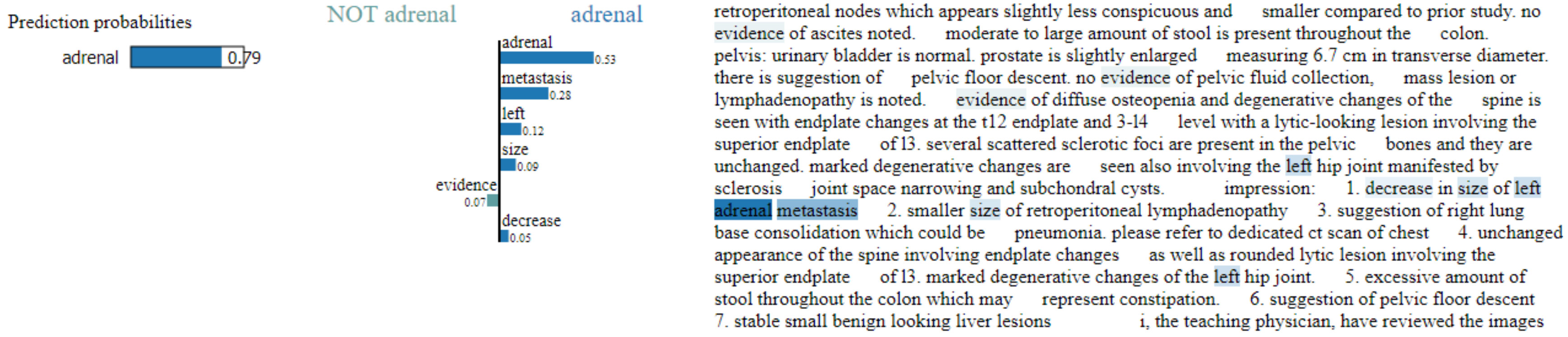


multiplanar multisequence mr images of the brain was performed before and after intravenous gadolinium administration. the following sequences were obtained: sagittal t1, axial t2, axial gre, axial flair, axial pre- and post-contrast t1, axial dwi and adc, and post-contrast 3d spgr with reformats. 9 ml of gadavist was administered intravenously without adverse reaction. comparison: mri on [redacted] findings: again seen is a 5 mm focus of enhancement in the left circummedullary cistern. the differential includes a focus of metastatic disease or a small schwannoma. no additional enhancing lesions are seen. the sulci and ventricles are appropriate for patient's age. no acute infarct or acute intracranial hemorrhage is identified. there is no evidence of mass effect or midline shift. the major intracranial flow-voids are present. the globes appear intact. the cerebellar tonsils lie above the level of the foramen magnum. impression: again seen is an unchanged 5 mm focus of enhancement in the left circummedullary cistern. the differential includes a focus of metastatic disease or a small schwannoma. continued follow-up advised. i, the teaching physician, have reviewed the images and agree with the report as written. this report was electronically signed eFigure 3G: Explanation of prediction regarding lymph node involvement



postradiation changes noted within the subsegmental right middle lobe with associated mild traction bronchiectasis. minimal fibrotic change is demonstrated within the anterior left upper lobe, compatible with post radiation change. moderate upper lobe predominant centrilobular emphysema. similar-appearing trace anterior pericardial effusion. abdomen: liver, gallbladder, spleen, and adrenals are unremarkable. punctate nonobstructing renal calculus noted within the left lower pole. kidneys enhance symmetrically and are without hydronephrosis. no small or large bowel dilatation. no free fluid. no mesenteric fat stranding. few scattered subcentimeter mesenteric lymph nodes. moderate atherosclerotic and atheromatous changes along the abdominal aorta and iliac arteries. pelvis: bladder is mildly distended without wall thickening. uterus and adnexa are surgically absent. no free fluid within the pelvis. no inguinal or pelvic sidewall adenopathy. bones: mild compression deformity along the superior endplate of t7 with associated interval development of sclerotic change (602:51). moderate volume of dependent subcutaneous fluid localize along the paraspinal musculature the lower lumbar levels. impression: adenocarcinoma, restaging scan. 1. mild decrease in intrathoracic tumor burden as evidenced by decreased size of the right hilar mass and right upper lobe and lower lobe nodules. 2. mildly increased volume of a small right pleural effusion. 3. compression deformity of the t7 vertebral body. i, the teaching physician, have reviewed the images and agree with the report as written.

eFigure 3H: Explanation of prediction regarding adrenal involvement



retroperitoneal nodes which appears slightly less conspicuous and smaller compared to prior study. no evidence of ascites noted. moderate to large amount of stool is present throughout the colon. pelvis: urinary bladder is normal. prostate is slightly enlarged measuring 6.7 cm in transverse diameter. there is suggestion of pelvic floor descent. no evidence of pelvic fluid collection, mass lesion or lymphadenopathy is noted. evidence of diffuse osteopenia and degenerative changes of the spine is seen with endplate changes at the t12 endplate and 3-14 level with a lytic-looking lesion involving the superior endplate of l3. several scattered sclerotic foci are present in the pelvic bones and they are unchanged. marked degenerative changes are seen also involving the left hip joint manifested by sclerosis joint space narrowing and subchondral cysts. impression: 1. decrease in size of left adrenal metastasis 2. smaller size of retroperitoneal lymphadenopathy 3. suggestion of right lung base consolidation which could be pneumonia. please refer to dedicated ct scan of chest 4. unchanged appearance of the spine involving endplate changes as well as rounded lytic lesion involving the superior endplate of l3. marked degenerative changes of the left hip joint. 5. excessive amount of stool throughout the colon which may represent constipation. 6. suggestion of pelvic floor descent 7. stable small benign looking liver lesions i, the teaching physician, have reviewed the images