Supplementary materials

Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study

Supplementary Methods

1. Treatment protocols and clinical information

In our study cohorts, patients underwent laryngeal surgery and END based on the expert consensus guidelines established in China.[1] In summary, the recommended approach for neck dissection was as follows: (1) Elective neck dissection should encompass at least neck level II-III. (2) For supraglottic carcinoma, ipsilateral neck dissection was recommended if the primary tumor did not extend beyond the midline of the larynx. (3) Bilateral neck dissection was indicated when primary lesions extended beyond the midline in supraglottic carcinoma cases. (4) In early-stage glottic cancer (T1-2), prophylactic neck dissection was generally not necessary. (5) For cases of T3-4 glottic cancer, the decision to perform bilateral or unilateral dissection was based on the midline invasion.

We collected demographic and clinical characteristics, such as gender, age, primary site, clinical T stage, tumor grade, and pathological lymph node status. While the radiomics and deep learning approaches are valuable, it may overlook macroscopic alterations present throughout the entire tumor region and information adjacent to the tumor boundary. To address this limitation, we developed a set of radiological features for comprehensive tumor evaluation. These features include paraglottic/preepiglottic spaces invasion, cartilage invasion, maximum tumor diameter (cm), and tumor enhancement pattern. Figure S7 provides visual representations of these features.

2. CT parameters of different Centres

Qilu centre:

CT scancer: multidetector row CT system (Discovery CT750HD, GE Healthcare); Parameters: tube voltage of 120 kVp, tube current of 250–400 mAs, and slice thickness of 0.625 mm. Contrast medium: The patients were injected intravenously with 1.5 ml/kg contrast medium (Ultravist 370, Bayer Schering Pharma) into the antecubital vein at a rate of 3.0 ml/s via an automated power injector.

Huaxi centre：

CT sanner: Somatom Definition AS+ and Somatom Definition Flash; Parameters: tube voltage of 120 kV, tube current of 200-210 mA, and slice thickness of 2.0 mm; Contrast medium: An anionic contrast medium (Iohexol 350 mg/ml; GE Healthcare, Chicago, IL), administered at a dose of 1.5 mL/kg body weight, was injected intravenously using a power injector at a rate of 3 mL/s.

PKUPH centre:

CT scancer: Philips iCT 256; Philips Medical Systems, Best, Netherlands; Parameters: tube voltage of 120 kVp, tube current of 100–370 mAs, slice thickness = 1.0 mm. Contrast medium: The patients were injected intravenously with 1.5 ml/kg contrast medium (Ultravist 370, Bayer Schering Pharma) into the antecubital vein at a rate of 2.5 ml/s via a power injector (Missouri XD2001, Ulrich GmbH&Co, Buchbrunnenweg, Ulm, Germany).

QFS centre：

CT scancer: multidetector row CT system (Discovery CT750HD, GE Healthcare); Parameters: tube voltage of 120 kVp, automatic tube current, and slice thickness of 1.0 mm. Contrast medium: Intravenous 60−120 mL (1.5 mL/kg) contrast agent (iopamidol injection, 37g/100ml; Shanghai Bracco Sine Pharmaceutical Industry, Shanghai, China) was managed with an injection rate of 3.0 mL/s.

3. Radiomics feature extraction

This study utilized the majority of default parameter settings provided by the PyRadiomics package. A fixed bin number of 64 was used for image discretization. The aggregation method for feature extraction can be found in Supplementary Table S3. We applied feature transformation using eight different filters, namely LoG (Laplacian of Gaussian), Wavelet, LBP3D (Local Binary Patterns in 3D), Exponential, Square, SquareRoot, Logarithm, and Gradient. These filters help in extracting additional information from the radiological images.

4. Data normalization of image preprocessing

In our study, we normalized the test data using the training parameters. Specifically, we employed the default ImageNet mean-subtraction method to normalize the images for 2D DCNN. For 3D DCNN, all images were normalized using min-max normalization to ensure accurate and reliable model evaluation. For radiomics model, the radiomics features were normalized using z-score method.

5. Five-fold cross-validation in model construction

The 5-fold cross-validation was performed on the training set (n=300, Qilu Hospital). We utilized the Grid search algorithm to optimize the hyperparameters of the model by analyzing the results obtained from the 5-fold cross-validation. This enabled us to identify the best model parameters. Once the optimal parameters were determined, we applied them to train the model using the entire training set. The trained model was then evaluated on the internal and external test sets. The following outlines the procedures for conducting 5-fold cross-validation.

1) **Data Preparation**: Initially, we collected data from various faculties and institutions. However, for the purpose of 5-fold cross-validation, we only choose the training set (n=300) from Qilu Hospital.
2) **5-Fold Cross-Validation**: We performed 5-fold cross-validation on this unified dataset as follows:
   - We divided the dataset into five equal-sized, non-overlapping subsets (folds).
   - During each fold, four subsets were used for training, and the remaining one subset was used for validation.
   - We repeated this process five times, with each fold serving as the validation set once while the others were used for training.
   - This allowed us to train and validate the model five times, each time with a different partition of the data.
3) **Model Parameter Tuning**: After each fold of cross-validation, we obtained model performance metrics. We utilized these metrics, such as accuracy or other relevant criteria, to fine-tune the

algorithm's hyperparameters using a Grid Search algorithm. This process helped us find the best set of hyperparameters for our model.

4) **Final Model Training**: Once the hyperparameters were optimized using cross-validation, we applied the selected model configuration, including the optimized hyperparameters, to train the model on the entire training set.

6. The interpretation of parameters in convolutional neural networks

Learning Rate: In a convolutional neural network (CNN), the learning rate is a hyperparameter that determines the step size at which the model's parameters are updated during the training process. It controls the speed and magnitude of the parameter updates, influencing how quickly the model adapts to the training data.

L2 Regularization: L2 regularization, also known as weight decay, is a technique used in CNNs to prevent overfitting. It involves adding a regularization term to the loss function that penalizes large parameter values. This regularization term encourages the model to have smaller parameter values, effectively simplifying the model and reducing its sensitivity to individual data points. By including L2 regularization, the CNN becomes more robust and generalizes better to unseen data.

Batch Size: In CNNs, the training data is divided into smaller batches during the training process. The batch size refers to the number of samples that are processed together before updating the model's parameters. When training with a larger batch size, more samples are processed simultaneously, which can lead to faster training as it takes advantage of parallelization. However, larger batch sizes may require more memory and can result in slower convergence or poorer generalization. Conversely, smaller batch sizes may take longer to train but can result in better convergence and generalization as they offer more frequent parameter updates and increased model variability.

7. Configuration file for Pyradiomics

```
imageType:
  Original: {}
  LoG:
    sigma: [1.0, 2.0, 3.0]    # If you include sigma values >5, remember to also increase the padDistance.
  Wavelet: {}
  LBP3D: {}
  Exponential: {}
  Square: {}
  SquareRoot: {}
  Logarithm: {}
  Gradient: {}

featureClass:
  shape:
  firstorder:
  glcm:    # Disable SumAverage by specifying all other GLCM features available
    - 'Autocorrelation'
    - 'JointAverage'
    - 'ClusterProminence'
    - 'ClusterShade'
    - 'ClusterTendency'
    - 'Contrast'
    - 'Correlation'
    - 'DifferenceAverage'
    - 'DifferenceEntropy'
    - 'DifferenceVariance'
    - 'JointEnergy'
    - 'JointEntropy'
    - 'Imc1'
    - 'Imc2'
    - 'Idm'
    - 'Idmn'
    - 'Id'
    - 'Idn'
    - 'InverseVariance'
    - 'MaximumProbability'
    - 'SumEntropy'
    - 'SumSquares'
  glrlm:
  glszm:
  gldm:
  ngtdm:
setting:
  # Normalization:
  # most likely not needed, CT gray values reflect absolute world values (HU) and should be comparable between scanners.
  # If analyzing using different scanners / vendors, check if the extracted features are correlated to the scanner used.
  # If so, consider enabling normalization by uncommenting settings below:
  #normalize: true
  #normalizeScale: 500    # This allows you to use more or less the same bin width.

  # Resampling:
  # Usual spacing for CT is often close to 1 or 2 mm, if very large slice thickness is used,
  # increase the resampled spacing.
  # On a side note: increasing the resampled spacing forces PyRadiomics to look at more coarse textures, which may or
  # may not increase accuracy and stability of your extracted features.
  interpolator: 'sitkBSpline'
  resampledPixelSpacing: [1, 1, 1]
  padDistance: 10    # Extra padding for large sigma valued LoG filtered images

  # Mask validation:
  # correctMask and geometryTolerance are not needed, as both image and mask are resampled, if you expect very small
  # masks, consider to enable a size constraint by uncommenting settings below:
  #minimumROIDimensions: 2
  #minimumROISize: 50

  # Image discretization:
```

```
  # The ideal number of bins is somewhere in the order of 16-128 bins. A possible way to define a good binwidt is to
  # extract firstorder:Range from the dataset to analyse, and choose a binwidth so, that range/binwidth remains approximately
  # in this range of bins.
  binCount: 64

  # first order specific settings:
  # This amount is added to the gray level intensity in features Energy, Total Energy and RMS, this is to prevent negative values.
  # If using CT data, or data normalized with mean 0, consider setting this parameter to a fixed value (e.g. 2000) that ensures non-
negative numbers in the image.
  voxelArrayShift: 1000

  # Misc:
  # default label value. Labels can also be defined in the call to featureextractor.execute, as a commandline argument,
  # or in a column "Label" in the input csv (batchprocessing)
  label: 1
```

Supplementary Figures

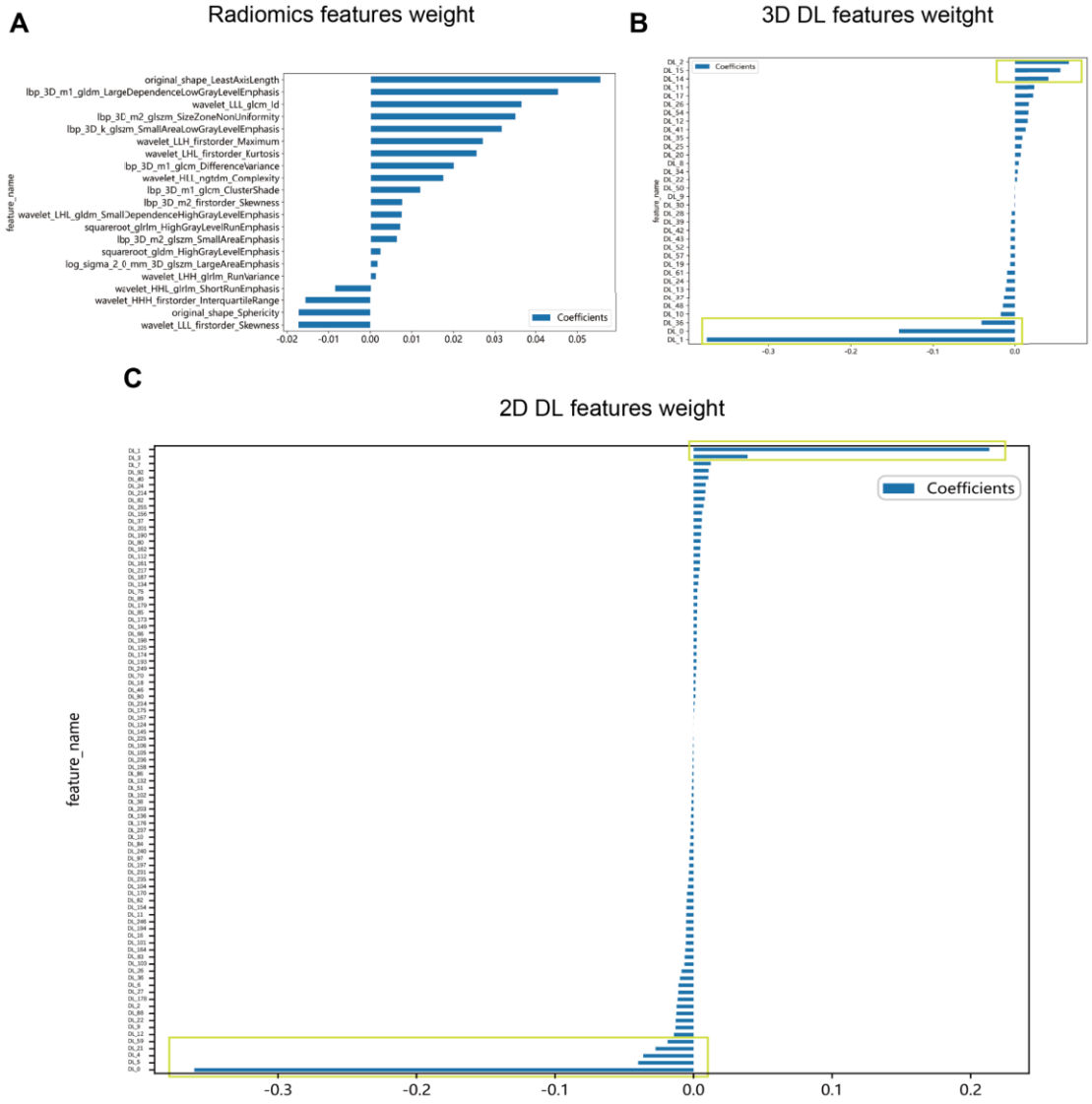

Figure S1. Key radiomics and deep learning (DL) features after LASSO regression. Weight visualization of radiomics features (A), three-dimensional (3D) DL features (B), and two-dimensional (2D) DL features (C) after LASSO selection. Green boxes indicate the top 2D and 3D DL features.

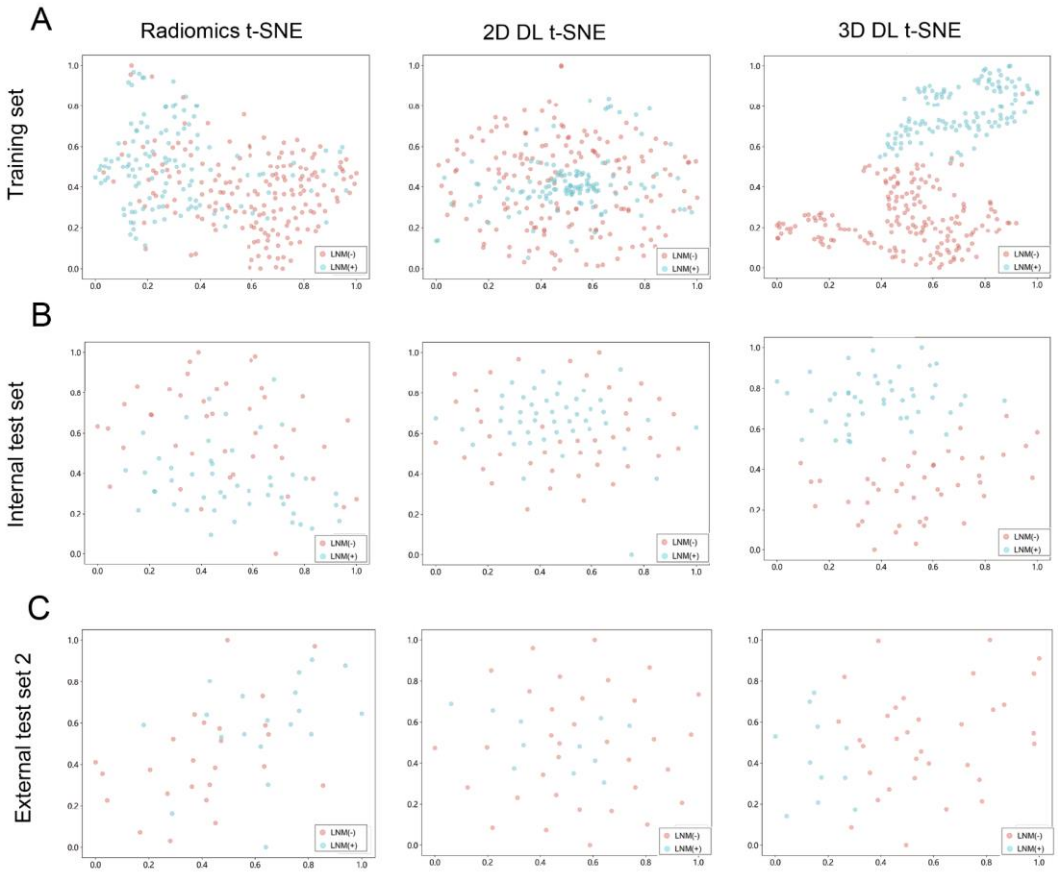

Figure S2. Discriminative ability analysis for radiomics and deep learning (DL) features. T-distributed stochastic neighbour embedding (t-SNE) visualizations for the radiomics features, two-dimensional (2D) and three-dimensional (3D) DL features in the training set (A), internal test set (B), and the external test 2 (C). Each dot represents a patient. Blue dots indicate patients with lymph node metastasis (LNM), and red dots indicate patients without LNM.

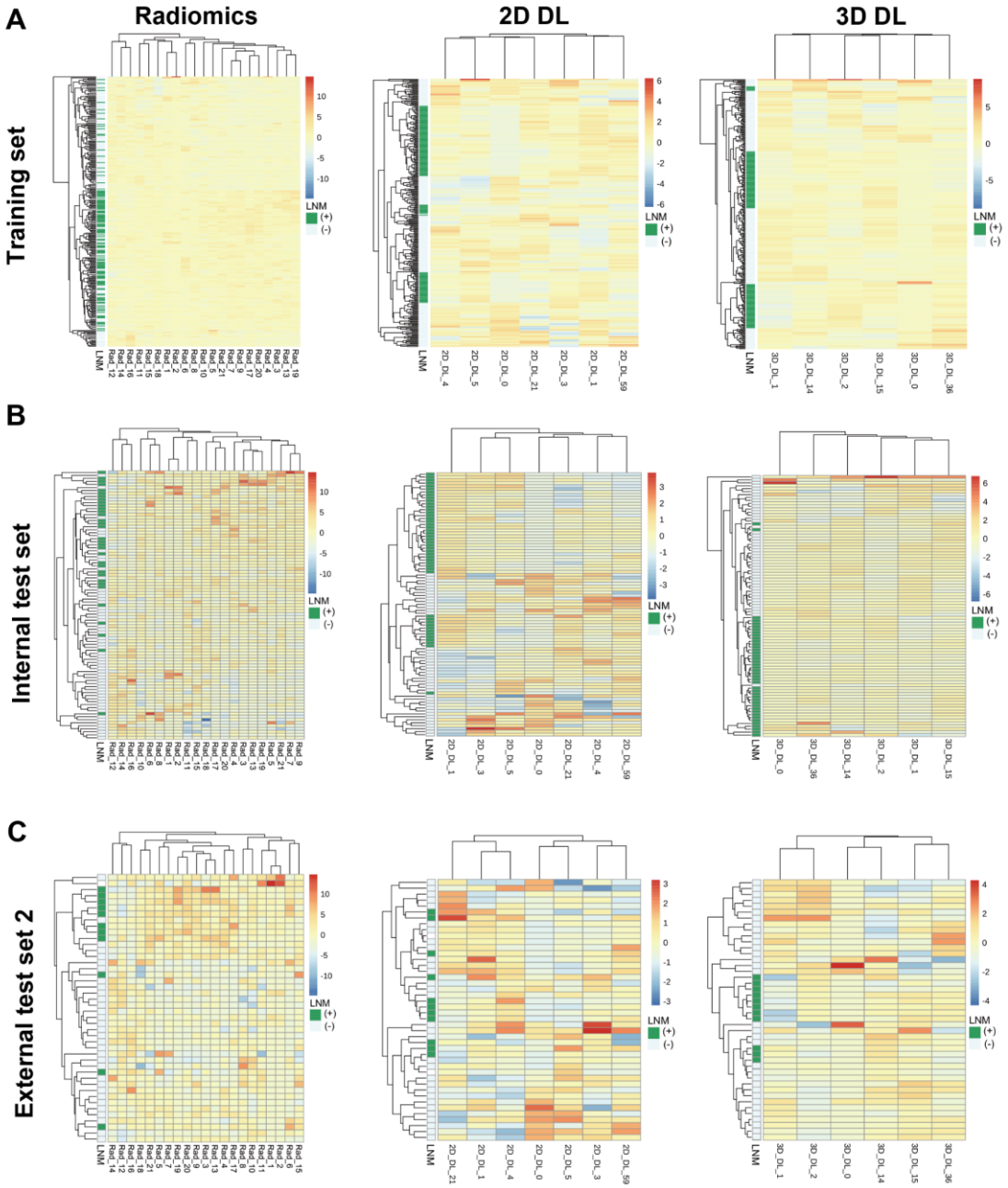

Figure S3. Clustering analysis for key radiomics and deep learning (DL) features. Hierarchical clustering heatmap for key radiomics features, top two-dimensional (2D) DL features, and top three-dimensional (3D) DL features in the training set (A), internal test set (B), and external test set 2 (C). The x-axis represents the IDs of radiomics and DL features, and the y-axis represents patients. Patients belong to the same cluster (adjacent rows) share similar features in the Euclidean space. The status of lymph node metastasis (LNM) is displayed on the white-green bar located on the left side next to the y-axis.

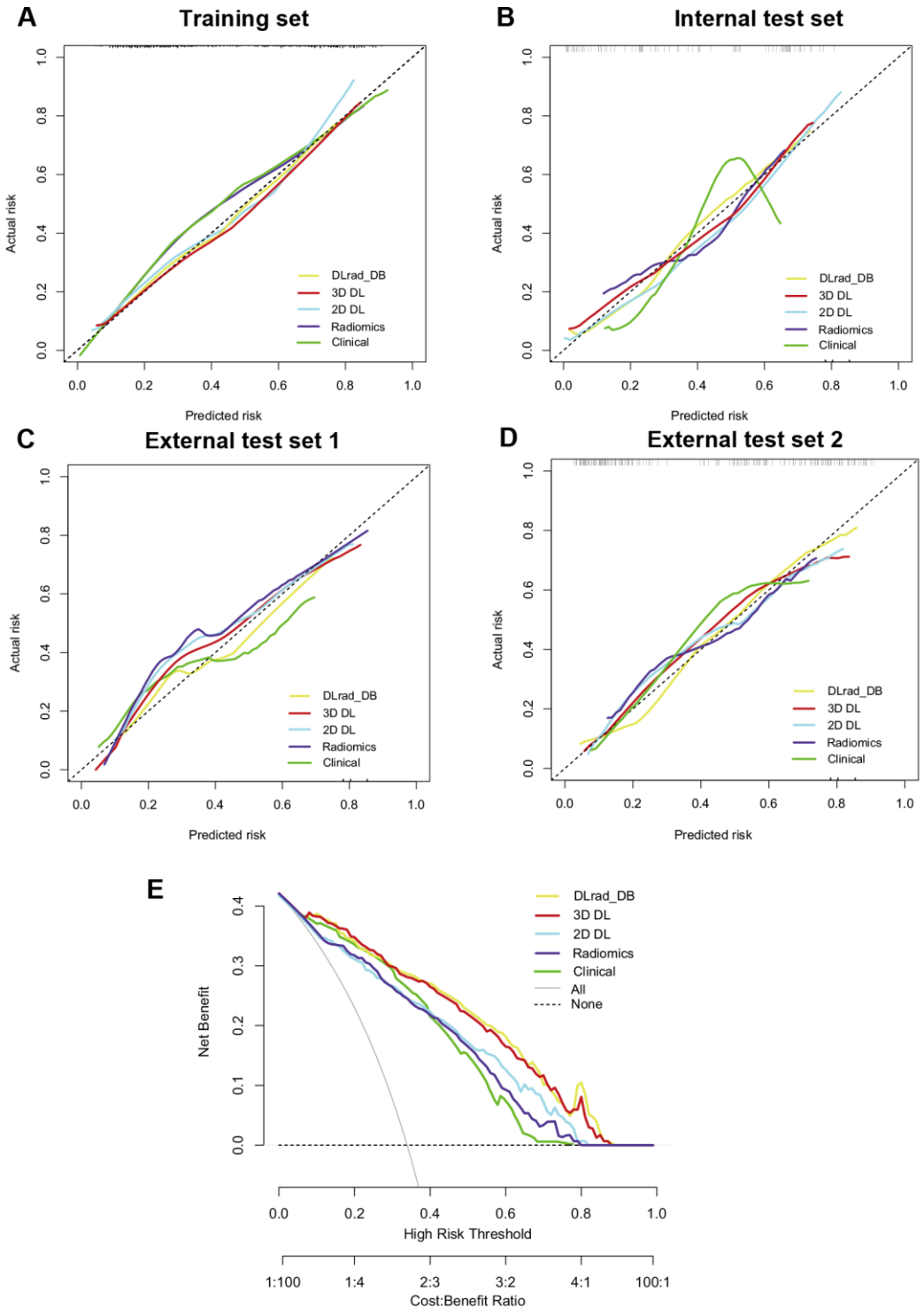

Figure S4. Performances for occult lymph node metastasis (LNM) prediction. Calibration curves for the DLRad_DB model, three-dimensional (3D) deep learning (DL) model, two-dimensional (2D) DL model, radiomics model, and clinical model in the training set (A), internal test set (B), external test set 1 (C), and the external test set 2 (D). (E) Decision curve analysis for each model in the external test set 1.

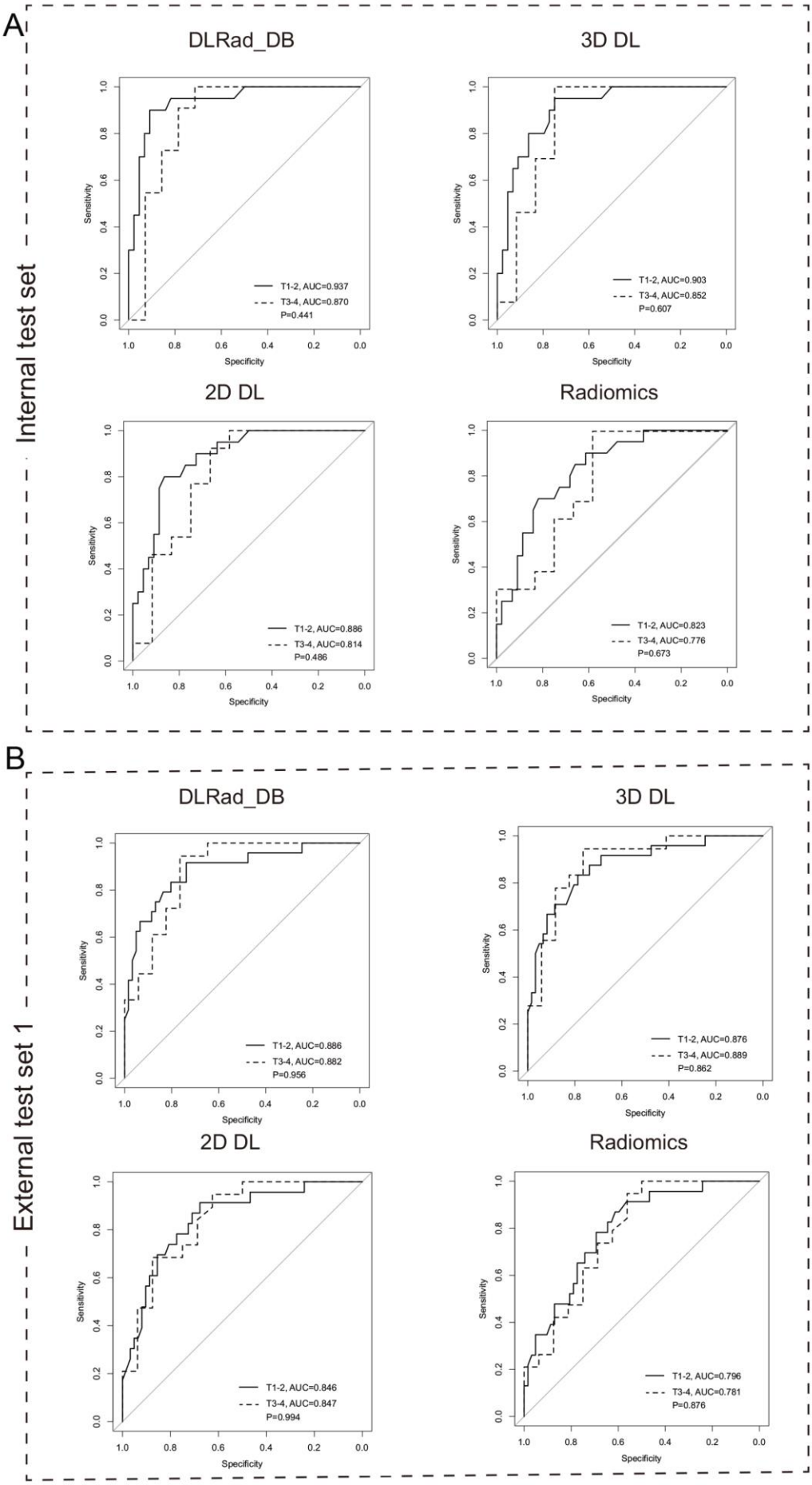
A
Internal test set
DLRad_DB
T1-2, AUC=0.937
T3-4, AUC=0.870
P=0.441
3D DL
T1-2, AUC=0.903
T3-4, AUC=0.852
P=0.607
2D DL
T1-2, AUC=0.886
T3-4, AUC=0.814
P=0.486
Radiomics
T1-2, AUC=0.823
T3-4, AUC=0.776
P=0.673
B
External test set 1
DLRad_DB
T1-2, AUC=0.886
T3-4, AUC=0.882
P=0.956
3D DL
T1-2, AUC=0.876
T3-4, AUC=0.889
P=0.862
2D DL
T1-2, AUC=0.846
T3-4, AUC=0.847
P=0.994
Radiomics
T1-2, AUC=0.796
T3-4, AUC=0.781
P=0.876
Sensitivity
Specificity

Figure S5. Comparisons of the receiver-operating characteristic (ROC) curves of the predictive models in different T stages. AUC indicates area under the curve; DL indicates deep learning; 2D, two-dimensional, and 3D, three-dimensional. P was calculated through the Delong test.

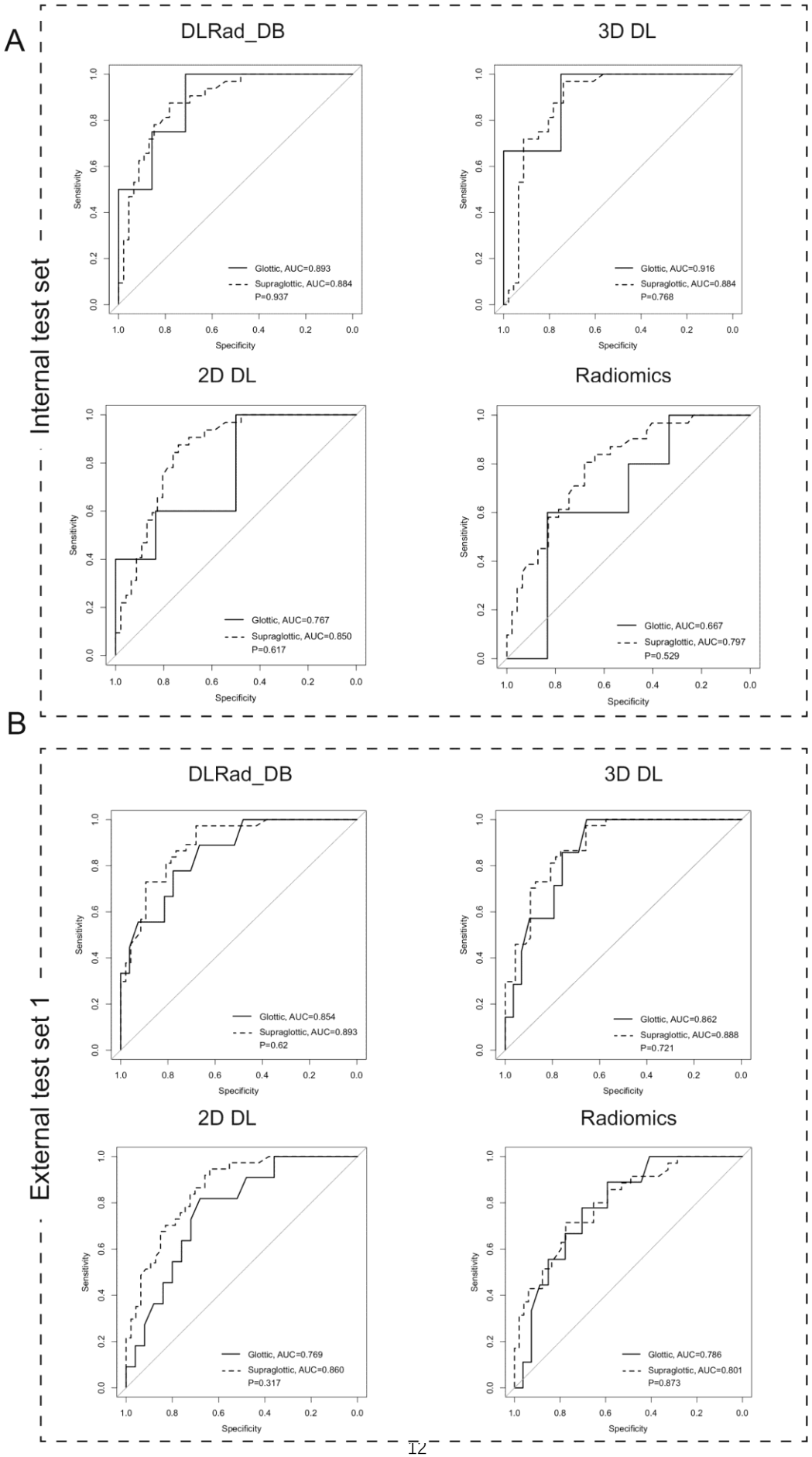
A
Internal test set
DLRad_DB
Sensitivity
Specificity
Glottic, AUC=0.893
Supraglottic, AUC=0.884
P=0.937
3D DL
Glottic, AUC=0.916
Supraglottic, AUC=0.884
P=0.768
2D DL
Glottic, AUC=0.767
Supraglottic, AUC=0.850
P=0.617
Radiomics
Glottic, AUC=0.667
Supraglottic, AUC=0.797
P=0.529
B
External test set 1
DLRad_DB
Glottic, AUC=0.854
Supraglottic, AUC=0.893
P=0.62
3D DL
Glottic, AUC=0.862
Supraglottic, AUC=0.888
P=0.721
2D DL
Glottic, AUC=0.769
Supraglottic, AUC=0.860
P=0.317
Radiomics
Glottic, AUC=0.786
Supraglottic, AUC=0.801
P=0.873

Figure S6. Comparisons of the receiver-operating characteristic (ROC) curves of the predictive models in different primary site. AUC indicates area under the curve; DL indicates deep learning; 2D, two-dimensional, and 3D, three-dimensional. P was calculated through the Delong test.

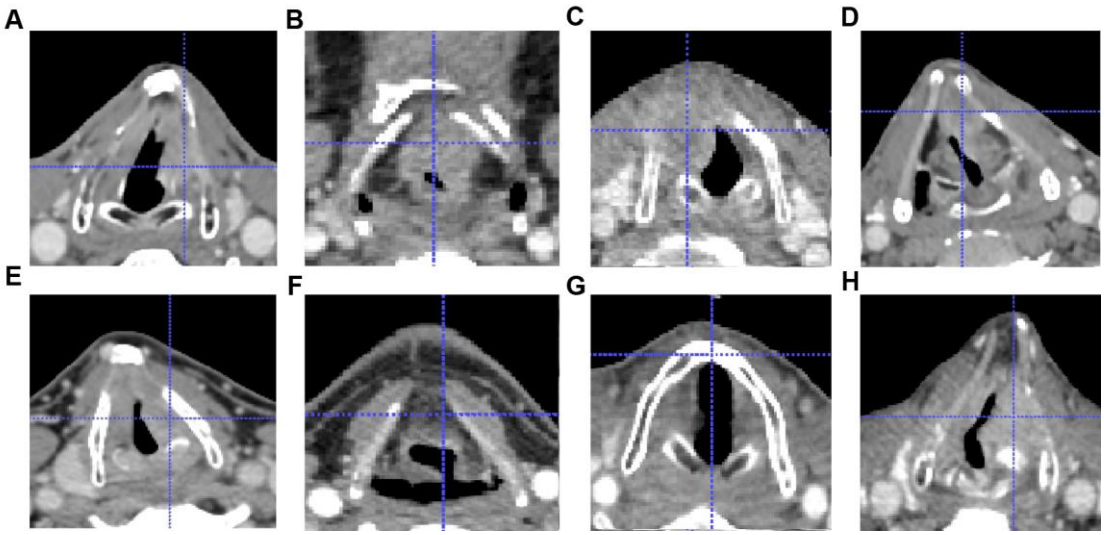

Figure S7. Radiological features of laryngeal squamous cell carcinoma. Paraglottic space invasion (A) or not (E); Preepiglottic space invasion (B) or not (F); cartilage invasion (C) or not (G); enhancement pattern: inhomogeneous enhancement (D) or homogeneous enhancement(H).

Supplementary Tables

Table S1. Regional distribution of metastatic lymph nodes in the training and internal test sets (Qilu Centre) and external test set 1 (Huaxi Centre and PKUPH Centre)

	Laterality and levels	**II**	**II-III**	**II-IV**	**III-IV**
Qilu Centre (pN+, n=157)	Ipsilateral, n	46	68	29	3
	Bilateral, n	0	11	0	0
Huaxi and PKUPH Centres (pN+, n=42)	Ipsilateral, n	11	21	8	0
	Bilateral, n	0	2	0	0

Abbreviations: pN+: pathological lymph node positive; n represents number of patients.

Table S2. Descriptions of the key radiomics features

Feature ID	Feature description
Rad_1	original_shape_LeastAxisLength
Rad_2	lbp_3D_m1_gldm_LargeDependenceLowGrayLevelEmphasis
Rad_3	wavelet_LLL_glcm_ld
Rad_4	lbp_3D_m2_glszm_SizeZoneNonUniformity
Rad_5	lbp_3D_k_glszm_SmallAreaLowGrayLevelEmphasis
Rad_6	wavelet_LLH_firstorder_Maximum
Rad_7	wavelet LHL firstorder Kurtosis
Rad_8	lbp_3D_m1_glcm_DifferenceVariance
Rad_9	wavelet_HLL_ngtdm_Complexity
Rad_10	lbp_3D_m1_glcm_ClusterShade
Rad_11	lbp_3D_m2_firstorder_Skewness
Rad_12	wavelet_LHL_gldm_SmallDependenceHighGrayLevelEmphasis
Rad_13	squareroot_glrlm_HighGrayLevelEmphsis
Rad_14	lbp_3D_m2_glszm_SmallAreaEmphsis
Rad_15	squareroot_gldm_HighGrayLevelEmphasis
Rad_16	log sigma 2.0 mm 3D glszm largeAreaEmphasis
Rad_17	wavelet LHH glm RunVariance
Rad_18	wavelet_HHL_girlm_ShortRunEmphasis
Rad_19	wavelet_HHH_firstorder_InterquartileRange
Rad_20	original_shape_Sphericity
Rad_21	wavelet_LLL_firstorder_Skewness

Table S3. Feature aggregation methods

Feature family	Aggregation methods
IS, IH	3D (calculated over the volume)
GLCM, GLRLM	3D: mrg (merged 3D directions)
GLSZM, NGTDM, NGLDM	3D (calculated from single 3D matrix)

Note: The term definition refers to the IBSI protocol.[2]

Supplementary References

1. [Expert consensus on management of cervical lymph node metastasis of head and neck squamous cell carcinoma]. Zhonghua er bi yan hou tou jing wai ke za zhi = Chinese journal of otorhinolaryngology head and neck surgery 2016; **51**(1): 25-33.

2. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. ArXiv 1612.07003. Cornell University Library: arXiv; 2016.