Appendix 1: Surgical procedures

All patients underwent open surgery, involving concurrent thyroidectomy and cervical lymph node dissection. As part of the standard procedure, central lymph nodes were consistently dissected. Patients displaying signs of lateral cervical LNM through fine-needle aspiration biopsy or preoperative imaging proceeded to lateral cervical dissection. Throughout the procedure, critical structures such as the internal jugular vein, sternocleidomastoid muscle, and spinal accessory nerve were meticulously preserved. The quantification of dissected lymph nodes and identified metastases was diligently performed and independently reviewed by two pathologists.

Appendix 2: Radiomics feature explanation

First-order features characterize the distribution of intensities within the ROIs, including energy, entropy, mean, and median. 2D features describe the ROI's size and shape, such as mesh surface, pixel surface, perimeter, and maximum diameter. Intensity features represent the first-order statistical distribution of voxel intensities within the tumor, while texture features illustrate the patterns or the intensities' second and higher-order spatial distributions. The texture features are extracted using various methods, including the gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM).

Appendix 3: Training parameters

```
params = dict(train=train_f,
              valid=val_f,
              labels_file=labels_f,
              data_pattern=data_pattern,
              j=0,
              max2use=None,
              val_max2use=None,
              batch_balance=False,
              normalize_method='imagenet',
              model_name=get_param_in_cwd('model_name','densenet121'),
              vit_settings = {'patch_size': 64, 'dim': 1,024, 'depth': 6, 'heads': 16, 'mlp_dim': 768},
              gpus=[0],
              batch_size=32,
              epochs=get_param_in_cwd('epoch', 50),
              init_lr=0.01,
              optimizer='sgd',
              retrain=None,
              model_root=get_param_in_cwd('model_root','dl_models'),
              add_date=False,
              iters_start=0,
              iters_verbose=1,
              save_per_epoch=False,
              pretrained=True)
```

Tabel S1 The results of machine learning in five random CV of the internal dataset

CV	Model name	Accuracy	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	F1
CV1	LR	0.603	0.745 (0.666–0.824)	0.81	0.564	0.262	0.939	0.395
	SVM	0.824	0.548 (0.432–0.665)	0.31	0.923	0.433	0.875	0.361
	KNN	0.836	0.634 (0.542–0.727)	0.095	0.977	0.444	0.85	0.157
	RF	0.634	0.724 (0.635–0.812)	0.81	0.6	0.279	0.943	0.415
	ExtraTrees	0.718	0.787 (0.715–0.858)	0.738	0.714	0.33	0.935	0.456
	XGBoost	0.752	0.702 (0.608–0.796)	0.595	0.782	0.342	0.91	0.435
	LightGBM	0.748	0.702 (0.610–0.794)	0.571	0.782	0.333	0.905	0.421
	MLP	0.687	0.752 (0.672–0.832)	0.762	0.673	0.308	0.937	0.438
CV2	LR	0.7	0.768 (0.680–0.857)	0.694	0.7	0.269	0.935	0.388
	SVM	0.806	0.648 (0.531–0.765)	0.5	0.855	0.353	0.915	0.414
	KNN	0.795	0.651 (0.554–0.748)	0.361	0.863	0.295	0.895	0.325
	RF	0.833	0.675 (0.570–0.780)	0.389	0.903	0.389	0.903	0.389
	ExtraTrees	0.772	0.746 (0.652–0.840)	0.583	0.802	0.318	0.924	0.412
	XGBoost	0.837	0.756 (0.643–0.848)	0.583	0.877	0.429	0.93	0.494
	LightGBM	0.715	0.746 (0.643–0.848)	0.722	0.714	0.286	0.942	0.409
	MLP	0.715	0.756 (0.668–0.844)	0.694	0.718	0.281	0.937	0.4
CV3	LR	0.715	0.721 (0.631–0.812)	0.605	0.733	0.277	0.917	0.38
	SVM	0.825	0.594 (0.482–0.706)	0.316	0.911	0.375	0.887	0.343
	KNN	0.852	0.691 (0.598–0.784)	0.158	0.969	0.462	0.872	0.235
	RF	0.673	0.747 (0.655–0.838)	0.737	0.662	0.269	0.937	0.394
	ExtraTrees	0.643	0.746 (0.655–0.836)	0.763	0.622	0.254	0.94	0.382
	XGBoost	0.665	0.734 (0.647–0.822)	0.711	0.658	0.26	0.931	0.38
	LightGBM	0.738	0.726 (0.635–0.817)	0.605	0.76	0.299	0.919	0.4
	MLP	0.711	0.722 (0.629–0.814)	0.632	0.724	0.279	0.921	0.387
CV4	LR	0.63	0.681 (0.583–0.780)	0.667	0.624	0.204	0.929	0.312
	SVM	0.821	0.583 (0.461–0.705)	0.333	0.891	0.306	0.903	0.319
	KNN	0.844	0.658 (0.550–0.758)	0.152	0.943	0.278	0.885	0.196
	RF	0.645	0.675 (0.571–0.779)	0.576	0.655	0.194	0.915	0.29
	ExtraTrees	0.698	0.695 (0.595–0.795)	0.606	0.712	0.233	0.926	0.336
	XGBoost	0.702	0.652 (0.543–0.761)	0.545	0.725	0.222	0.917	0.316
	LightGBM	0.756	0.695 (0.595–0.796)	0.545	0.786	0.269	0.923	0.36
	MLP	0.622	0.675 (0.574–0.776)	0.636	0.62	0.194	0.922	0.298

Tabel S1 (*continued*)

Tabel S1 (*continued*)

CV	Model name	Accuracy	AUC (95% CI)	Sensitivity	Specificity	PPV	NPV	F1
CV5	LR	0.76	0.66 (0.563–0.757)	0.41	0.821	0.286	0.889	0.337
	SVM	0.806	0.611(0.502–0.720)	0.333	0.888	0.342	0.884	0.338
	KNN	0.81	0.646 (0.553–0.739)	0.308	0.897	0.343	0.882	0.324
	RF	0.669	0.666 (0.574–0.759)	0.564	0.687	0.239	0.901	0.336
	ExtraTrees	0.776	0.679 (0.588–0.770)	0.41	0.839	0.308	0.891	0.352
	XGBoost	0.578	0.666 (0.568–0.763)	0.744	0.549	0.223	0.925	0.343
	LightGBM	0.776	0.679 (0.586–0.771)	0.436	0.835	0.315	0.895	0.366
	MLP	0.814	0.659(0.562–0.756)	0.333	0.897	0.361	0.885	0.347

CV, cross-validation; AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; F1, F1-score; LR, logistic regression; SVM, support vector machine; KNN, K-nearest neighbors; RF, random forest; XGBoost, extreme gradient boost; LightGBM, light gradient boosting machine; MLP, multi-layer perceptron.