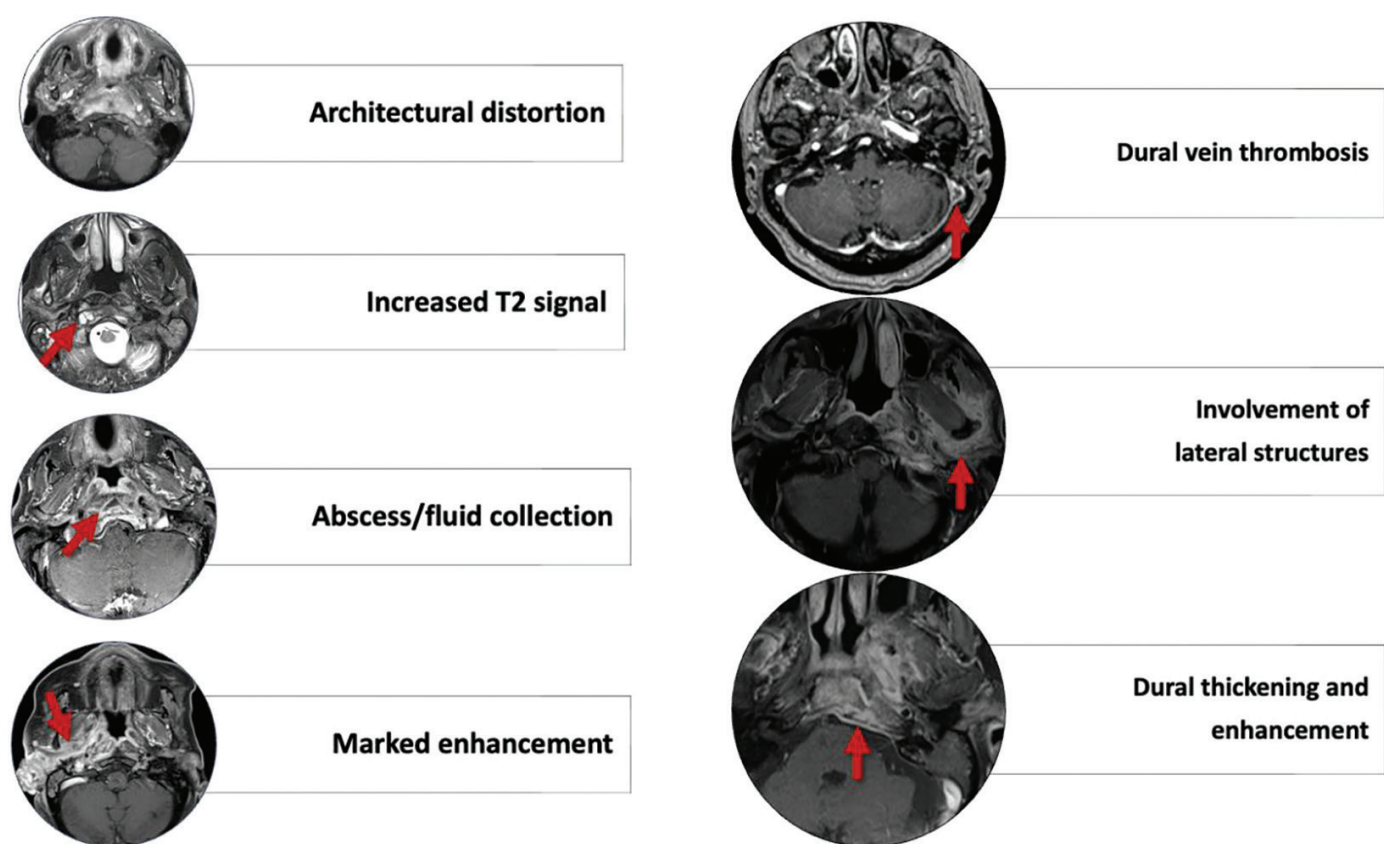


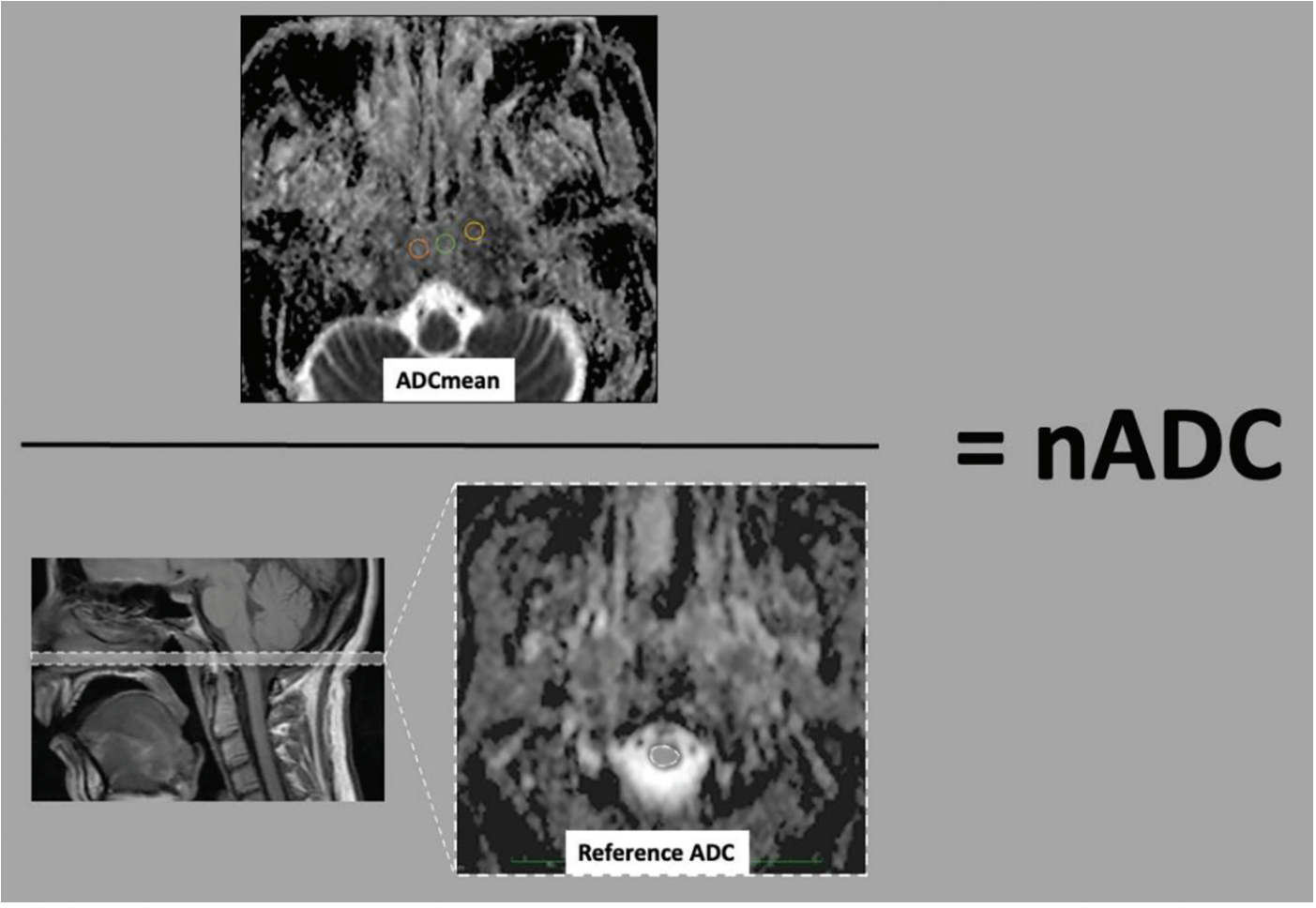
MRI acquisition

Imaging protocols varied due to studies being performed across different centers, timeframes, and scanners. Diffusion-weighted imaging was not available in 6 patients (6.45%). The distribution of magnetic resonance imaging (MRI) vendors was as follows: Siemens (54.8%), Philips (28%), and GE (17.2%). Axial contrast-enhanced T1-weighted images were obtained after the injection of 0.1 mmol/kg of gadolinium-based contrast agents. All but three patients (3.2%) underwent imaging in a 1.5T MRI unit.

The median in-plane resolutions of the T1w, FS-T2w, and T1c series were 0.625 x 0.625 mm², 0.69 x 0.69 mm², and 0.68 x 0.68 mm², respectively (25th-75th percentile: 0.45-0.765 mm, 0.46-0.75 mm, 0.51-0.765 mm pixel width); the median slice thickness for all three series was 4 mm (range: 3-5 mm, 3-5 mm, 1-5 mm). Additionally, the T2 series were obtained using different fat suppression techniques [chemical fat suppression, short tau inversion recovery (STIR), Dixon]. Diffusion-weighted imaging was performed using two b-values, with the highest b-value set at 1000 on Philips scanners and 800 on other scanners.



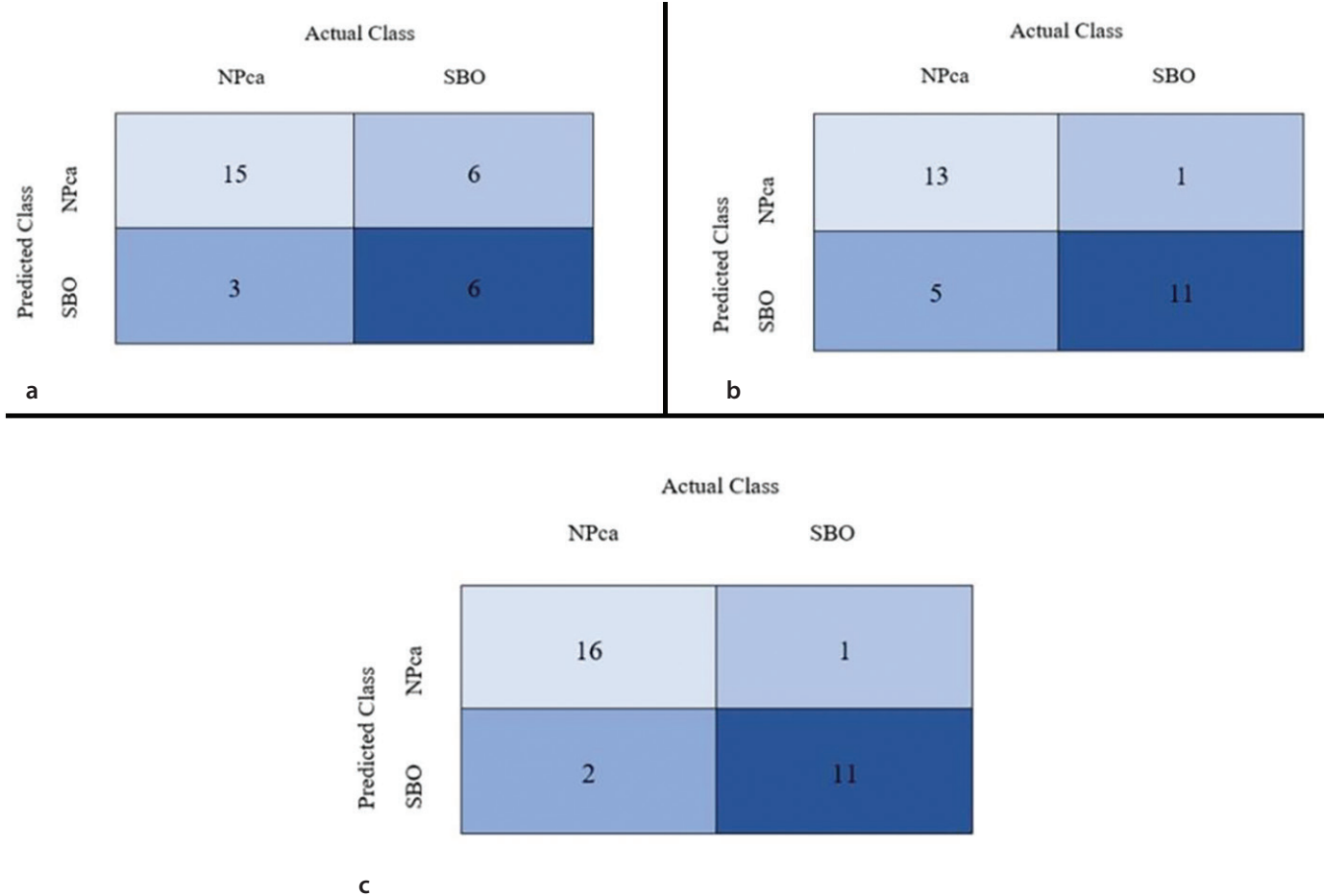
Supplementary Figure 1. Illustration of the evaluated semantic features. The presence of architectural distortion, involvement of lateral structures, marked enhancement, dural thickening, and lymphadenopathy were used for semantic model.



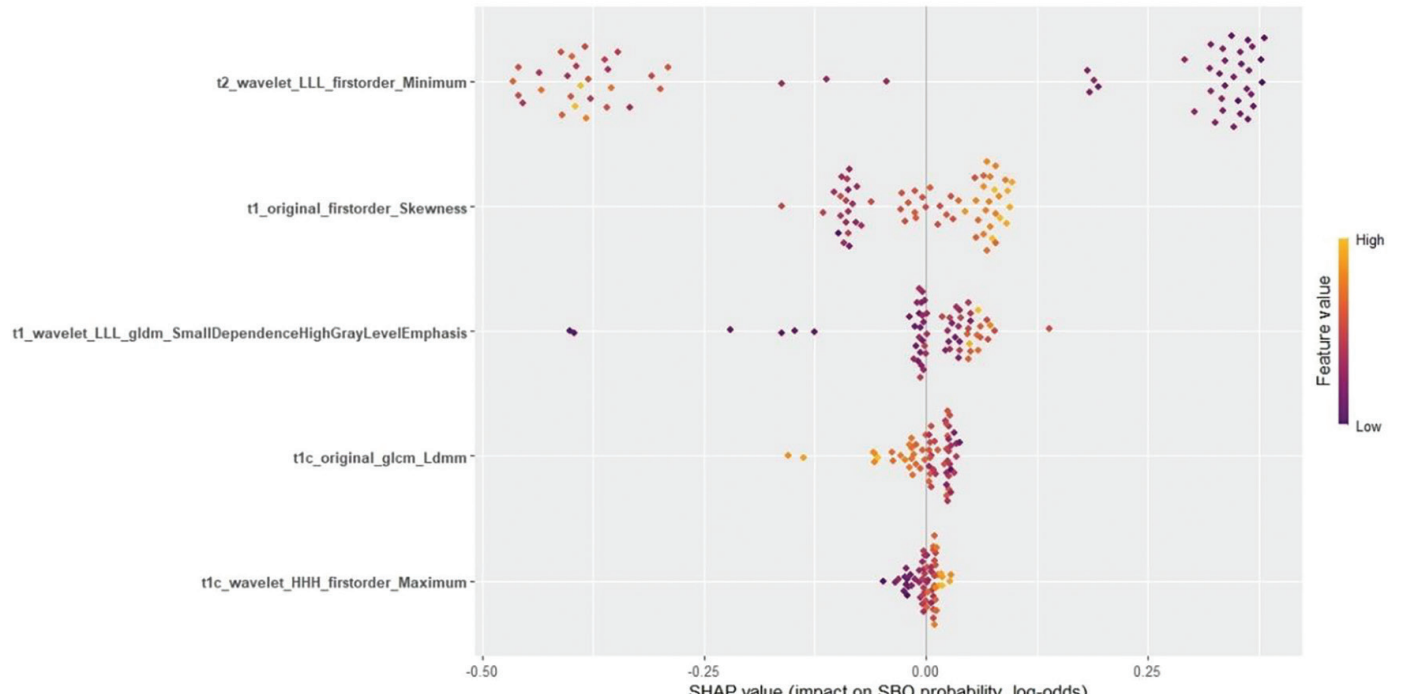
Supplementary Figure 2. Apparent diffusion coefficient (ADC) parameters. For the mean of the ADC values (ADCmean) calculation, three regions of interest (ROIs) of equal size were drawn on the lesion at the level of the skull base selecting areas of higher restricted diffusion visually. Subsequently, the arithmetic mean of these ADC values was calculated. The reference ADC value was calculated as the arithmetic mean of two ADC values which were obtained by the largest possible ROI drawn to cover the majority of the spinal cord on the slice nearest to the obex level and the slice immediately below it. Subsequently, the normalized ADC (nADC) was derived as the ratio of the ADCmean to the reference ADC. This approach was intended to mitigate the impact of center- or device-related factors.

```
1 minimumROIDimensions: 2
2 minimumROISize: null
3 normalize: true
4 normalizeScale: 200
5 removeOutliers: 3
6 resampledPixelSpacing:
7   - 2.0
8   - 2.0
9   - 2.0
10 interpolator: "sitkBSpline"
11 preCrop: false
12 padDistance: 5
13 distances:
14   - 1
15 force2D: false
16 force2Ddimension: 0
17 resegmentRange:
18   - -3
19   - 3
20 label: 1
21 additionalInfo: true
22 binWidth: 5
23 symmetricalGLCM: true
24 correctMask: true
25 resegmentMode: "sigma"
26
```

Supplementary Figure 3. The YAML configuration file used for radiomic feature extraction with PyRadiomics (v3.0.1). This file defines key settings for image preprocessing and feature calculation. Key settings for standardization include isotropic resampling to 2 x 2 x 2 mm voxels, intensity normalization, and discretization with a fixed bin width of 5. The full software citation can be found in the Methods section.



Supplementary Figure 4. Confusion matrices of the (a) ADC model; (b) radiomics model; and (c) semantic model. ADC, apparent diffusion coefficient; SBO, skull base osteomyelitis; NPca, nasopharyngeal carcinoma.



Supplementary Figure 5. SHapley Additive exPlanations (SHAP) summary plot for radiomics model. This plot illustrates the impact of individual feature values on the model's output (SBO probability, log-odds). Each point represents a single instance of a patient, showing its SHAP value for a given feature. The x-axis indicates the SHAP value, where positive values suggest an increase in the predicted probability of SBO, and negative values suggest a decrease. The y-axis lists the top contributing radiomics features. The color of each point represents the actual feature value for that instance, ranging from "Low" (purple) to "High" (yellow), providing insight into how the magnitude of a feature's value correlates with its impact on the prediction. For instance, high values of "t2_wavelet_LLL_firstorder_Minimum" appear to be associated with a positive impact on SBO probability, while low values are associated with a negative impact. SBO, skull base osteomyelitis.