# Quantifying Surgical Complexity in Advanced Vestibular Schwannomas Using a Novel Video-Based Grading System

Yunjia Ni, MD[1]; Shawn M. Stevens, MD[2]; Kaith Almefty MD[3]; Randall W. Porter, MD[3]; Julio Juarez Alvarez, BFA[4]; Marie E. Clarkson, BFA[4]; Jonathan Tangsrivimol, MD[5]; Kivanc Yangi, MD[5]; Takuma Maeda, MD, PhD[6]; Hirotaka Sato, MD, PhD[6]; Anthony M. Asher, MD[3]; Mark C. Preul, MD[5]; Richard D. Dortch, PhD[1]; Michael T. Lawton, MD[3]

[1]Barrow Neuroimaging Innovation Center, Department of Translational Neuroscience, Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ; [2] Department of Otolaryngology and Skull Base Surgery, Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ; [3]Department of Neurosurgery, Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ; [4] Neuroscience Publications, Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ; [5] The Loyal & Edith Davis Neurosurgical Research Laboratory, Department of Neurosurgery, Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ; [6] Barrow Aneurysm & AVM Research Center, Department of Translational Neuroscience, Barrow Neurological Institute, St Joseph's Hospital and Medical Center, Phoenix, AZ

## Introduction

- Large tumors (Koos Grade III–IV) are surgically challenging, but operative difficulty varies widely even within the same Koos grade
- Koos grading reflects tumor size and brainstem contact, not intraoperative surgical complexity
- Surgeons therefore rely on experience and intraoperative judgement to anticipate operative difficulty and risk
- Although MRI may contain predictive information, no validated intraoperative grading system exists to establish surgical ground truth

**Goal:** Develop and validate a video-based Vestibular Schwannoma Grading System (VSGS) to objectively quantify key drivers of surgical complexity: adhesion, vascularity, facial nerve configuration, and tumor texture.

## Methods

- Analyzed 108 intraoperative video clips (≤5 min) from 7 Koos Grade IV vestibular schwannoma resections
- Clips sampled three domains: tumor–nerve interface, tumor–brain interface, and tumor-specific features
- Five blinded neurosurgeons (PGY 5–37) independently scored each clip after standardized VSGS training
- Four surgical complexity features were graded: adhesion, vascularity, facial nerve configuration, and tumor texture
- Interrater reliability assessed using Gwet's AC2
- Criterion validity evaluated using weighted Cohen's κ against senior surgeon ratings

## Results

- Overall clip-to-clip scoring variability was low to moderate (SD ≤ 0.60)
- Substantial agreement for vascularity (AC2 = 0.69)
- Moderate agreement for adhesion (0.58) and facial nerve configuration (0.47)
- Tumor texture demonstrated the lowest reliability
- Validity against senior surgeon ratings paralleled reliability

**Key finding:** Vascularity and adhesion are the most reproducible drivers of surgical complexity, while tumor texture requires refinement.

## Conclusions

- VSGS reliably quantified vascularity, adhesion, and CN VII shape features in Koos IV VS. The tumor texture rubric was less reliable.
- Ongoing work will include refining the rubrics based on these preliminary data and to further assess reliability and validity according to experience with VS surgery.
- Ultimately, a fully validated VSGS would be used to correlate with preoperative imaging (predictive capacity to anticipate intraoperative findings) and postoperative outcomes.
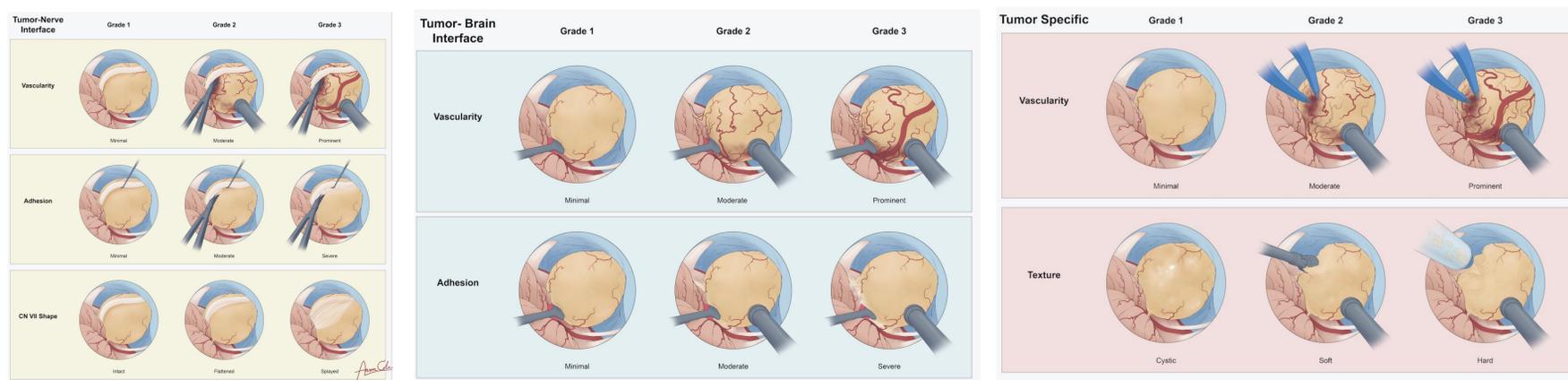


**Figure 1.** There are three domains of the VSGS: tumor-nerve interface, tumor-brain interface, and tumor-specific features. Under each domain, different surgical complexity features would be evaluated. We created an ordinal scale ranging from 0-1 for each surgical complexity feature.
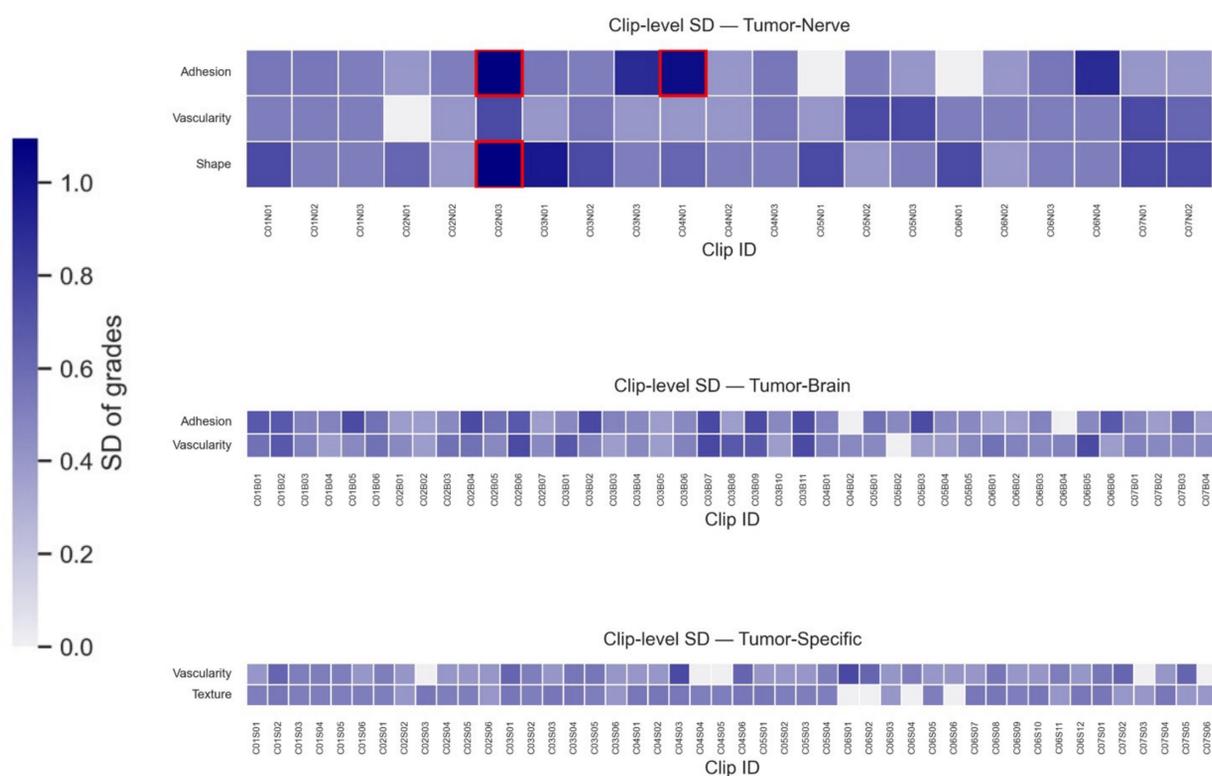


**Figure 2.** Clip level variance of rater's grading was low-to-moderate (SD ≤ 0.60) with two clips as outliers (SD > 0.6).
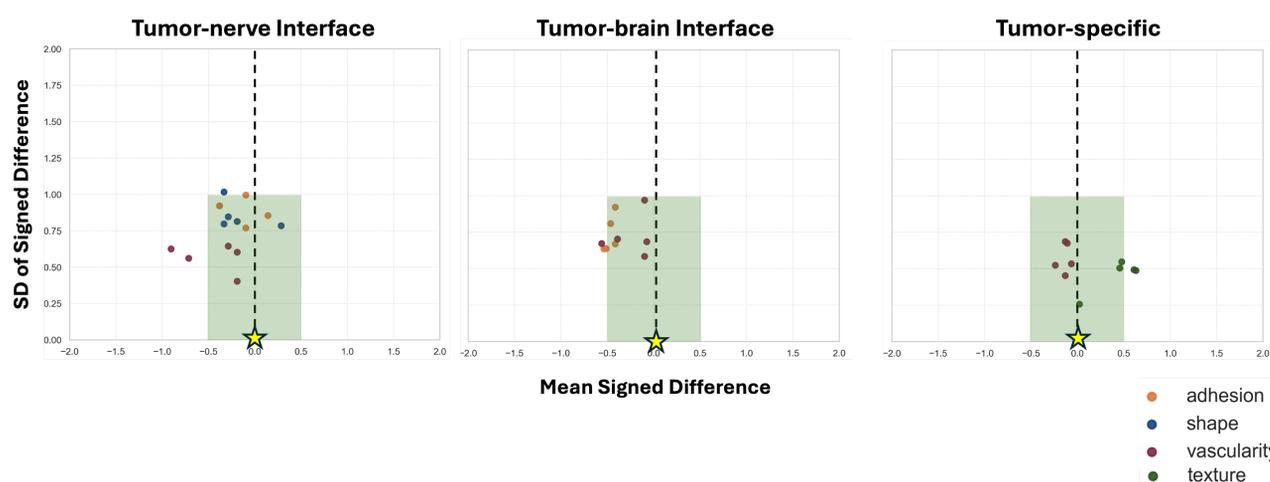


**Figure 3.** Bias-precision plot showing deviation of different raters' ratings from the operating surgeon's (gold standard) grading. We observed that most ratings had a mean signed difference of 0 and SD ≤1.
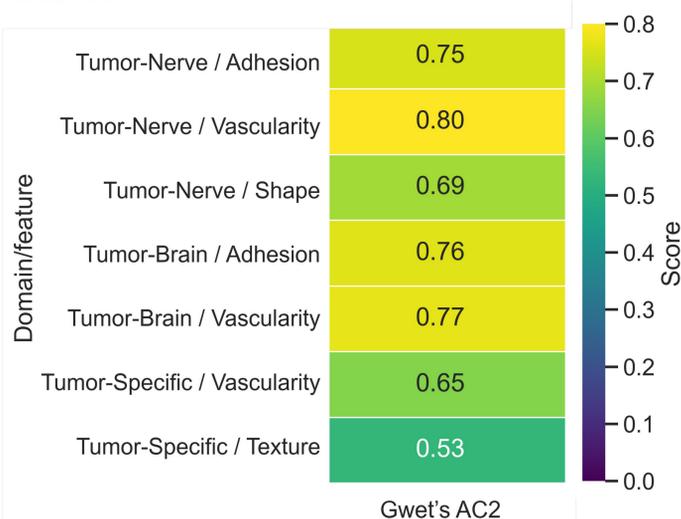


**Figure 4.** We used Gwet's AC2 to benchmark multi-rater reliability assessment. There was an observed intermediate to excellent agreement across raters except for the tumor-specific texture.
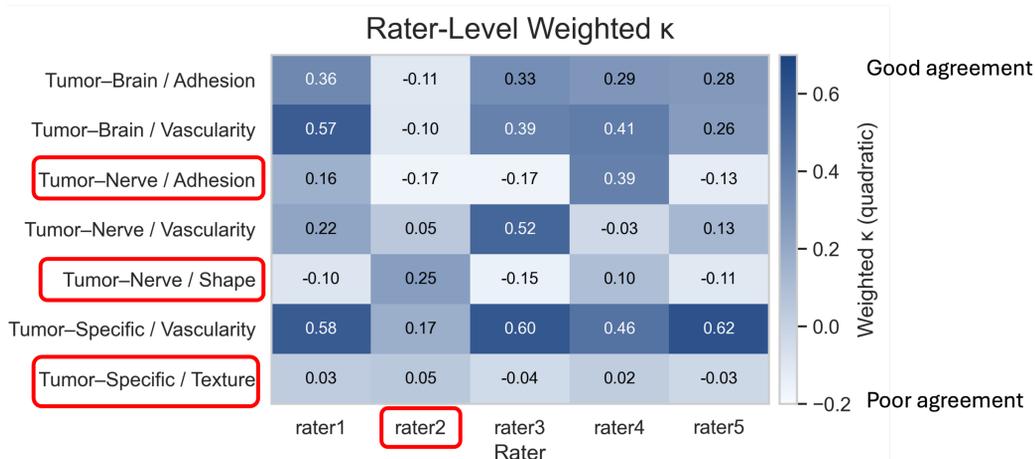
| Domain/feature | Gwet's AC2 |
|---|---|
| Tumor-Nerve / Adhesion | 0.75 |
| Tumor-Nerve / Vascularity | 0.80 |
| Tumor-Nerve / Shape | 0.69 |
| Tumor-Brain / Adhesion | 0.76 |
| Tumor-Brain / Vascularity | 0.77 |
| Tumor-Specific / Vascularity | 0.65 |
| Tumor-Specific / Texture | 0.53 |

### Rater-Level Weighted κ

| | rater1 | rater2 | rater3 | rater4 | rater5 |
|---|---|---|---|---|---|
| Tumor–Brain / Adhesion | 0.36 | -0.11 | 0.33 | 0.29 | 0.28 |
| Tumor–Brain / Vascularity | 0.57 | -0.10 | 0.39 | 0.41 | 0.26 |
| Tumor–Nerve / Adhesion | 0.16 | -0.17 | -0.17 | 0.39 | -0.13 |
| Tumor–Nerve / Vascularity | 0.22 | 0.05 | 0.52 | -0.03 | 0.13 |
| Tumor–Nerve / Shape | -0.10 | 0.25 | -0.15 | 0.10 | -0.11 |
| Tumor–Specific / Vascularity | 0.58 | 0.17 | 0.60 | 0.46 | 0.62 |
| Tumor–Specific / Texture | 0.03 | 0.05 | -0.04 | 0.02 | -0.03 |

**Figure 5.** Using weighted kappa, we observed scattered disagreement across some domains, such as tumor-nerve adhesion, tumor-nerve shape, and tumor-specific texture. This warrants the definitions of the gradings system. We also noticed that some raters tend to disagree with the gold standard, which might be due to limited clinical experience.

## Acknowledgements