

Shape-Based Inductive Bias for Glioma Grading from Tumor Contours

Puneet Velidi

Department of Mathematics and Statistics
University of Victoria

puneet.velidi@gmail.com

Farouk Nathoo

Department of Mathematics and Statistics
University of Victoria

nathoo@uvic.ca

Michelle F. Miranda

Department of Mathematics and Statistics
University of Victoria

michellemiranda@uvic.ca

Ashery Mbilinyi

Department of Computer Science
University of Victoria

ashery@uvic.ca

Cédric Beaulac

Département de mathématiques
Université du Québec à Montréal

beaulac.cedric@uqam.ca

Abstract

Glioma grading from tumor contours is often treated as a pixel problem even when the signal of interest is shape. We align closed contours with a functional shape-alignment framework, separate global deformation from residual Fourier shape, and organize these quantities as frequency-ordered tokens. In five-fold patient-disjoint cross-validation on BraTS 2020 tumor contours, with model selection performed using grouped inner validation, a compact multi-layer perceptron (MLP) achieves the highest mean balanced accuracy at 71.5%, compared with 65.9% for ResNet-18 and 63.3% for ViT-Tiny. It also gives the highest mean low-grade glioma F1 at 54.9%. Its pooled out-of-fold balanced accuracy is 72.4% (patient-bootstrap 95% CI: 66.4–77.8%). The selected MLPs use 2.9k–117.3k parameters across folds, at least 46 times fewer than the pixel baselines. In a controlled noise-free simulation, shape-based models reach 56.3–71.5% balanced accuracy while the pixel models remain at 50.0–52.5%. This work demonstrates how incorporating a shape-based inductive bias at the representation level can improve interpretability and scalability while enabling substantial dimensionality reduction.

1. Introduction

Object shape is often the signal of interest in scientific imaging. Tumor margins, cell outlines, organ silhouettes, crystal boundaries, and particle contours are all examples where

labels depend more on geometry than on texture. The default deep-learning formulation typically rasterizes the object and asks pixel-based image models to indirectly learn morphological features while also learning invariance to translation, rotation, scale, and starting point. In small labelled cohorts, this spends statistical power on variation that a user may already have prior information about.

Classical shape analysis has traditionally addressed this issue by first aligning objects and then classifying or comparing them in a coordinate system where nuisance transformations have been removed, using tools such as Fourier descriptors, Procrustes alignment, and square-root velocity function (SRVF) representations [13, 20, 33, 34, 41]. The machine-learning analogue is a shape-based inductive bias, in which curve geometry supplies a quotient space, the space of shapes with translation, rotation, scale, and starting point factored out, before supervised learning begins, in the same spirit as geometric and equivariant learning methods that build symmetry structure into the representation [7, 9].

We study this representation question in glioma grading from BraTS 2020 tumor masks [2, 25], an application in which radiologists and radiomics pipelines have long treated tumor morphology as clinically meaningful [1, 16, 21]. High-grade gliomas (HGGs) often present with larger and more irregular margins, while low-grade gliomas (LGGs) tend to be smaller and more compact [31].

The technical basis for our paper is the functional-shape alignment framework introduced by Moindjié *et al.* [26]. A closed contour is treated as a deformation of a latent shape under translation, scale, rotation, and cyclic reparameteri-

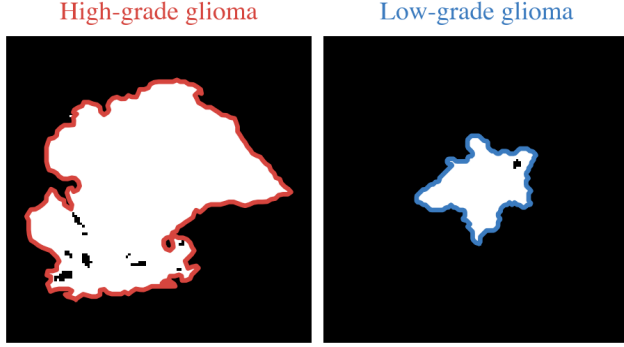


Figure 1. Tumor contours extracted from BraTS 2020 segmentation masks. High-grade gliomas (left) tend to be larger with more irregular boundaries; low-grade tumors (right) are typically smaller and more compact. These examples motivate a contour-level representation for morphology classification.

zation. In a Fourier basis, the starting-point shift becomes a sparse orthogonal block transform, making joint phase-and-rotation alignment a finite-dimensional Procrustes problem. We retain the estimated deformation variables as model inputs alongside the residual Fourier coefficients described in Section 4.2.

We use the term shape token for either the deformation vector or one harmonic residual vector. Lower harmonics encode coarse geometry such as elongation and asymmetry, while higher harmonics encode lobes, indentations, and local boundary roughness. This representation allows us to compare a Shape Transformer, which updates each token using the other deformation and harmonic tokens through self-attention, with a compact MLP that operates on the flattened token representation. In five-fold patient-disjoint cross-validation, the MLP attains the highest mean balanced accuracy and LGG F1. Its selected configurations use 2.9k–117.3k parameters, compared with 0.80M–4.75M for the Shape Transformer and 5.43M–11.2M for the pixel baselines.

This minority-class result suggests greater statistical efficiency. In the imbalanced high-grade glioma (HGG)/LGG setting, overall accuracy can be dominated by HGG cases even when LGG tumors are poorly identified. The compact Fourier representation allows a small MLP to remain competitive on LGG detection without a high-capacity classifier.

Contributions.

1. We formulate aligned Fourier contours as *shape tokens*, with one six-value deformation token and one six-value residual shape token per harmonic. With $K = 32$, the representation has $(K + 1) \times 6 = 198$ scalar inputs.
2. We show on BraTS 2020 that compact MLPs over at most 198 scalar inputs attain the highest mean balanced accuracy and LGG F1 under patient-disjoint cross-

validation. The selected MLPs use 2.9k–117.3k parameters, at least 46 times fewer than the pixel baselines.

3. We use a controlled simulation to examine how global deformation and mid-frequency shape information affect shape-based and pixel-based classifiers applied to the same generated objects.

2. Related Work

Radiomics and medical shape. Radiomics encodes clinical images through quantitative features and has repeatedly found morphology to be informative for tumor phenotype and outcome [1, 16, 21]. Prior glioma-grading studies have combined shape with intensity or texture features [8, 40], while CNN pipelines have used the largest axial slice or full 3D MRI [39, 45]. Modern segmentation networks such as U-Net [30], nnU-Net [19], and UNETR [17] produce high-quality masks. Our analysis begins once a mask has been obtained and uses the largest-area axial slice, as described in Section 3. This paper studies the value of converting that mask into a shape coordinate system before classification.

Classical shape representations. Classical contour representations include Fourier descriptors [20, 28, 41], chain codes [14], shape contexts [3, 4], active shape models [10], and Procrustes or elastic shape analysis [6, 13, 29, 33, 34]. Contemporary shape analysis often uses the SRVF framework, which represents curves through velocity-normalized functions and compares shapes after quotienting rotation and reparameterization [33, 34]. These methods make invariance explicit by aligning shapes before comparing them. In statistical shape analysis, downstream classifiers are often intentionally simple. Aligned shapes are projected to tangent, principal-component, or basis coordinates and then passed to functional linear or component-based classifiers. Moindjié, Beaulac, and Descary develop a functional view of curve alignment and latent-shape estimation [26], with an application to functional classification of multivariate planar X-ray curves [27]. Inspired by that alignment framework, we ask whether a compact nonlinear classifier can learn from the resulting coefficients.

Functional neural representations. A small but growing body of work has integrated neural networks into functional data analysis (FDA), including deep-learning models with functional inputs [35], neural networks with functional outputs [37], and autoencoder-based approaches to smoothing and representation learning [38]. This literature motivates a middle ground between classical functional linear models and large pixel-based networks, where nonlinear decision rules can be learned without giving up a functional coordinate system whose axes retain geometric meaning.

Tokenized learning. Transformers [36] and BERT-style classification [11] motivate learned pooling over token sequences. Vision Transformers [12] tokenize images into patches, leaving geometric invariance largely to the data and model capacity. Transformer-based models have also been used when the input is naturally a collection of elements, including set-structured data and point clouds [22, 43]. In our case, the tokens are frequency-ordered coordinates of an aligned contour. Because harmonic bands correspond to different scales of morphology, self-attention could plausibly help the classifier learn interactions among deformation variables, low-frequency elongation and asymmetry, and higher-frequency boundary detail.

2.1. Shape Inductive Bias

We introduce a shape-based inductive bias at the representation level through a functional basis representation of tumour contours. This contrasts with CNNs, which are often biased toward classification using texture rather than shape [15, 42] and are not necessarily translation invariant [5]. Increasing the amount of shape information available to a model has also been shown to improve generalizability and robustness to perturbations and some forms of distribution shift. Geirhos *et al.* [15], for example, augment the training data to induce a stronger shape bias by suppressing texture while preserving object shape. Our approach can be viewed as a more direct version of this idea, since we in effect entirely suppress texture.

3. Tasks

3.1. BraTS 2020

We use BraTS 2020 expert segmentation masks [2, 25] to construct a morphology-derived binary target from the tumor subregions present in each extracted component. Components containing the complete triad of necrotic or non-enhancing tumor core, peritumoral edema, and enhancing tumor (labels 1, 2, and 4) are assigned HGG. Their co-occurrence reflects three complementary features of aggressive disease which are tissue necrosis, disruption of the blood–brain barrier with contrast enhancement, and an extensive infiltrative or edematous response. The remaining observed combinations are assigned LGG because they lack this heterogeneous signature and represent comparatively lower-grade-like morphology. Thus, the target is an operational imaging-phenotype classification rather than a histopathologic diagnosis [31], which suffices for our demonstration of shape-based classification. The export contains 413 contours from 368 patients, comprising 318 HGG and 95 LGG contours. Of these patients, 331 contribute one contour and 37 contribute multiple contours. We focus on the minority LGG class because accurate detection must exceed majority-class recognition and may iden-

tify disease before progression to less treatable HGG.

From 3D volumes to 2D contours. For each patient we select the axial cross-section with the largest total tumor area and binarize the whole-tumor mask. Each connected component with at least 50 pixels is treated as a separate tumor and converted to a level-set contour with the marching-squares implementation in scikit-image [24, 32]. The largest closed boundary within each component is kept. Marching squares produces contours with different numbers and spacings of vertices, so we resample each boundary to $P=256$ points uniformly along arc length before Fourier projection. The coefficient-space formulation no longer depends on the original vertex count after this common projection [26]. Finite resampling and truncation can still introduce aliasing, which is one reason we use uniform arc-length sampling before retaining $K \leq 32$ harmonics. In the simulation, noise is added to the 256 generated vertices and both tokenization and rasterization use those same noisy vertices without another resampling step.

Splits and metric. The reported BraTS numbers use five-fold outer cross-validation grouped by patient and stratified by the component-level binary label. The outer folds function as test sets: each contour appears in exactly one outer test fold, and no patient appears in both outer training and test data within a fold. Within each outer-training fold, the first split from a four-fold patient-grouped partition provides inner training and validation subsets. Validation balanced accuracy selects token bandwidth, architecture, hyperparameters, and training epoch; the selected setup is then retrained from scratch on the complete outer-training fold and evaluated once on the outer test fold. All models use the same outer and inner partitions. We report fold means and sample standard deviations and additionally pool all 413 out-of-fold predictions. Confidence intervals are obtained by resampling patients together with all of their contours.

3.2. Controlled Simulation

Our simulation is a diagnostic for how the classifiers use deformation and latent-shape information in a controlled geometric setting. We draw a latent shape indicator $\text{cls} \sim \text{Bernoulli}(0.5)$ independently of the deformation variables. This indicator controls the added harmonics but is not the target label y . For $t \in [0, 2\pi)$, the planar base curve is $\mathbf{C}_0(t) = r(t)(\cos t, \sin t)$, where

$$r(t) = 1 + c_3 \cos(3t + \phi_3) + \mathbf{1}\{\text{cls} = 1\} \sum_{k=4}^6 b_k \cos(kt + \phi_k), \quad (1)$$

Here $c_3 \sim U(0.18, 0.32)$ and $\phi_3 \sim U(0, 2\pi)$ randomize the shared three-lobed base shape. For class 1 only, the ampli-

tudes $b_k \sim U(0.020, 0.030)$ add small bumps at harmonics $k = 4, 5, 6$, with independent random phases ϕ_k . The base curve is then transformed by isotropic scale $s = \exp(u)$, $u \sim U(\log 0.3, \log 3.0)$, small shear, arbitrary rotation, and translation $(t_x, t_y) \sim U(-1.5, 1.5)^2$.

Labels deliberately combine information from the deformation and latent-shape components,

$$\begin{aligned} \eta &= \log s + 0.5 \operatorname{sign}(t_x) + 1.2(\text{cls} - 0.5) + \epsilon, \\ y &= \mathbf{1}\{\eta > 0\}, \quad \epsilon \sim \mathcal{N}(0, 0.1). \end{aligned} \quad (2)$$

The terms $\log s$ and $\operatorname{sign}(t_x)$ are available to a shape-based classifier through the deformation token, while cls is encoded by the mid-frequency residual shape tokens. Pixel-based classifiers read a 64×64 filled raster produced after recentering and rescaling the polygon. This preprocessing removes the explicit scale and translation variables used by the label, although quantization and boundary statistics may retain indirect correlates. To simulate noise, independent Gaussian noise was added to each coordinate of every contour vertex, $(\tilde{\mathbf{c}}_i = \mathbf{c}_i + \epsilon_i)$, where $(\epsilon_i \sim \mathcal{N}(\mathbf{0}, \sigma_c^2 \mathbf{I}_2))$ and $(\sigma_c \in 0, 0.10)$. The same perturbed contour was used for Fourier tokenization and rasterization, while the label remained unchanged and no pixel-level noise was added.

Sample sizes, noise, and bandwidth. The reported simulation aggregate uses tokenizer bandwidth $K=32$, selected by the procedure in Section 5, sample sizes $N \in \{50, 100, 250, 1,000\}$, and contour-noise levels $\sigma \in \{0, 0.10\}$. Here “noise-free” means $\sigma = 0$. At $\sigma = 0.10$, independent Gaussian noise is added to contour vertices before either tokenization or rasterization. Architectures are selected once on a separate noise-free validation set and then held fixed for both test conditions. Final estimates use 20 independent training runs and the same fixed 2,000-sample test set within each noise condition. This prevents model selection from adapting to a particular test-noise level, but it also means the noisy condition measures robustness away from the selection distribution.

4. Shape-Based Method

The core representation pipeline is shared by BraTS and the simulation. For efficient batched computation, a segmentation boundary or simulated contour is first represented on a uniform 256-point grid and projected once to a common Fourier basis. Alignment and token construction then operate in coefficient space, and either a compact MLP or a Shape Transformer performs classification.

4.1. Functional-Shape Alignment and Tokenization

Inspired by Moindjié *et al.*’s functional formulation [26], each observed planar contour $\mathbf{C}(t) = (X(t), Y(t))$, $t \in$

$[0, 1]$, is treated as a deformation of a latent shape,

$$\mathbf{C}(t) = \rho \mathbf{O}_\theta \tilde{\mathbf{C}} \circ \gamma_\delta(t) + \mathbf{T}, \quad (3)$$

with scale $\rho > 0$, rotation $\mathbf{O}_\theta$, translation $\mathbf{T}$, and cyclic starting-point shift $\gamma_\delta(t) = \text{mod}(t - \delta, 1)$. We use a training-derived reference template $\boldsymbol{\mu}$. During inner model selection it is the first preshape in the deterministic inner-training order; after selection it is recomputed as the first preshape in the complete outer-training fold. Validation and outer-test contours are aligned to this training reference and never contribute to its estimation. This fixed-reference choice differs from iterative Fréchet-mean estimation in the full functional-shape procedure [26], and sensitivity to the reference choice remains to be tested.

Preshape. Our implementation estimates translation and scale from the uniformly resampled points,

$$\mathbf{T}_i = P^{-1} \sum_{j=1}^P \mathbf{C}_{ij}, \quad \rho_i = \left(P^{-1} \sum_{j=1}^P \|\mathbf{C}_{ij} - \mathbf{T}_i\|_2^2 \right)^{1/2}. \quad (4)$$

The preshape is $\mathbf{C}_i^* = (\mathbf{C}_i - \mathbf{T}_i)/\rho_i$. This point-space calculation was used to share a simple vectorized preprocessing interface across BraTS contours and simulated curves. It is not required by Moindjié *et al.*’s functional approach. We then compute the Fourier coefficients of the uniformly sampled preshape using

$$\begin{aligned} a_k &= \frac{2}{P} \sum_{j=0}^{P-1} x_j \cos(2\pi k j / P), \\ b_k &= \frac{2}{P} \sum_{j=0}^{P-1} x_j \sin(2\pi k j / P), \end{aligned} \quad (5)$$

and analogously for the y coordinate.

Rotation and phase. For each candidate δ on a 101-point grid, the Fourier shift property applies the corresponding block-diagonal rotation to each harmonic. The grid gives a starting-point resolution of about one percent of the contour and was chosen as a practical accuracy and cost tradeoff. It can introduce phase quantization below one grid step. The rotation θ is then solved in closed form by the Procrustes trace objective

$$(\hat{\theta}_i, \hat{\delta}_i) = \arg \min_{\theta, \delta} \|\boldsymbol{\alpha}_i \boldsymbol{\beta}(\delta) - \mathbf{O}_\theta \boldsymbol{\mu}\|_F^2, \quad (6)$$

where $\boldsymbol{\alpha}_i$ stacks the preshape Fourier coefficients and $\boldsymbol{\beta}(\delta)$ is the sparse orthogonal phase-shift matrix [26]. The implementation evaluates the grid, keeps the best objective, and returns $(\hat{\theta}_i, \hat{\delta}_i)$.

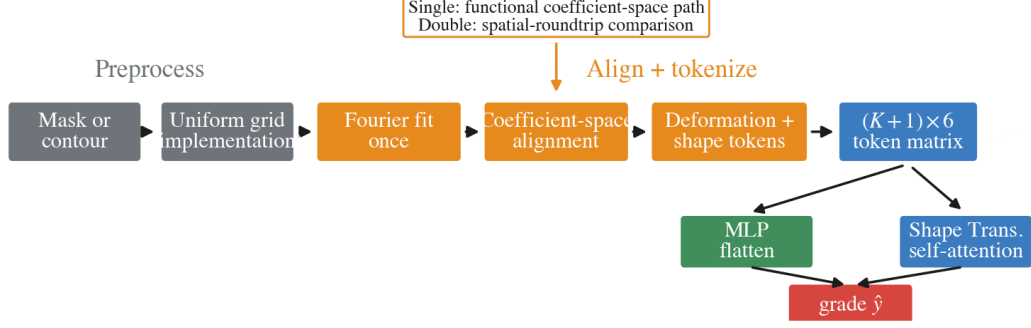


Figure 2. Contour-token pipeline. The contour is converted to a deformation token and frequency-ordered residual shape tokens. Downstream shape-based classifiers receive the same token matrix, the MLP flattens it, while the Shape Transformer mixes harmonic information through self-attention.

Coefficient-space implementation. Following Moindjié *et al.*, phase, inverse rotation, and scale are applied directly to the fitted coefficients through the block structure of the basis [26]. Supplementary Section C compares this implementation with an equivalent spatial-roundtrip approximation and shows identical aggregate metrics in a fixed-split BraTS implementation check.

4.2. Token Layout

The default token matrix is $\mathbf{X}_i \in \mathbb{R}^{(K+1) \times 6}$.

Deformation token. The first token stores the functional-shape deformation variables

$$(T_x, T_y, 1/\rho, \delta, \theta, 0), \quad (7)$$

where the first five values are standardized using means and standard deviations fitted only on the relevant training partition, then applied unchanged to validation or outer-test contours. During final outer-fold training, these statistics are refitted on the complete outer-training fold. The padding value remains zero. Thus, test examples neither influence one another nor contribute distributional information to pre-processing. Unlike a learned [CLS] token, the deformation token carries physical contour information and also serves as the position-0 readout for the shape-based models.

Shape tokens. Each shape token stores the residual coefficients and amplitudes for one harmonic,

$$\mathbf{v}_{ik} = (a_{xk}, b_{xk}, a_{yk}, b_{yk}, A_{xk}, A_{yk})_i, \quad (8)$$

$$\mathbf{z}_{ik} = \text{ZSCORE}_{\text{rows}}[\mathbf{v}_{ik} - \mathbf{v}_{\mu k}]. \quad (9)$$

where $A_{xk} = \sqrt{a_{xk}^2 + b_{xk}^2}$ and $A_{yk} = \sqrt{a_{yk}^2 + b_{yk}^2}$. These amplitudes make harmonic magnitude directly available while the signed coefficients retain phase information. Low-order tokens encode coarse geometry; higher-order tokens encode boundary lobes and local detail. With

$K=32$, the largest token representation has $33 \times 6 = 198$ scalars. Bandwidth is selected independently within each outer fold for each architecture over $K \in \{8, 16, 24, 32\}$. The selected models can therefore receive different harmonic bandwidths, so their comparison is not a controlled capacity comparison on identical inputs.

4.3. Classifiers

All shape-based classifiers consume the same token matrix for a given K .

MLP. The token matrix is flattened and passed through two GELU hidden layers,

$$\begin{aligned} \mathbf{X} &\xrightarrow{\text{flatten}} \mathbb{R}^{6(K+1)} \xrightarrow{\text{Linear+GELU+drop}} \mathbb{R}^{d_h}, \\ \mathbb{R}^{d_h} &\xrightarrow{\text{Linear+GELU+drop}} \mathbb{R}^{d_h} \xrightarrow{\text{Linear}} \hat{y}. \end{aligned}$$

There is no positional embedding and no parameter sharing across harmonics. Across the five BraTS folds, the selected configurations span $K \in \{8, 16, 32\}$, $d_h \in \{32, 128, 256\}$, and dropout in $\{0, 0.1, 0.2\}$, for **2.9k–117.3k** parameters.

Shape Transformer. The transformer is a pre-norm encoder [36] with a linear token projection, learned positional embeddings, L self-attention blocks with H heads, GELU feed-forward layers of width $\max(64, 4d)$, and position-0 readout. Self-attention updates are

$$\alpha_{ij} = \frac{\exp(\mathbf{q}_i^\top \mathbf{k}_j / \sqrt{d_k})}{\sum_m \exp(\mathbf{q}_i^\top \mathbf{k}_m / \sqrt{d_k})}, \quad \mathbf{h}'_i = \sum_j \alpha_{ij} \mathbf{v}_j. \quad (10)$$

Across the five BraTS folds, the selected configurations span $K \in \{8, 16, 24, 32\}$, $d \in \{128, 192, 256\}$, $H \in \{4, 8\}$, and $L \in \{4, 6\}$, for **0.80M–4.75M** parameters.

Table 1. BraTS 2020 token configurations selected by grouped inner-validation balanced accuracy. One configuration is selected within each outer fold; K values are sorted for compactness. MLP widths span 32–256; transformer widths span 128–256, heads span 4–8, and layers span 4–6.

Model	Selected K	Params
MLP	8, 8, 8, 16, 32	2.9k–117.3k
Shape Trans.	8, 16, 24, 32, 32	0.80M–4.75M

Pixel baselines. BraTS pixel-based baselines are ResNet-18 [18] and ViT-Tiny [12] on one-channel 224×224 rasterized tumor masks. ResNet-18 uses its standard 7×7 , stride-2 stem, and ViT-Tiny uses 16×16 patches. Both are trained from scratch. Simulation pixel-based baselines uses the same architectures on 64×64 filled rasters.

Training. BraTS runs train with AdamW, learning rate 10^{-3} , weight decay 10^{-2} , batch size 16, 40 epochs, and inverse-square-root class-balanced sampling. The simulation token runner uses Adam, learning rate 2×10^{-3} , batch size 64, 60 epochs, and no early stopping. The pixel runner uses the same learning rate and batch size for 30 epochs. These fixed budgets are not compute matched, which limits architecture-level comparisons in the simulation.

5. Model Selection Protocol

The experiments separate model selection from final performance reporting. The outer BraTS folds are the test partitions used to estimate out-of-sample performance, not a final dataset to which cross-validation is subsequently applied. Within each outer-training fold, model selection uses one patient-grouped inner validation split. Simulation selection is performed on a noise-free holdout set. Final simulation tables hold selected configurations fixed across random initializations and noise levels.

BraTS token-model sweep. The BraTS sweep uses the Fourier-space token exporter for $K \in \{8, 16, 24, 32\}$. MLP widths are $d_h \in \{32, 64, 128, 256, 512\}$ with dropout $p \in \{0.0, 0.1, 0.2\}$. The Shape Transformer uses $d \in \{128, 192, 256\}$, $H \in \{4, 8\}$, and $L \in \{4, 6\}$ subject to $H \mid d$. For every candidate, the reference contour and deformation-feature normalization are fitted on the inner-training subset and applied unchanged to its validation subset. After selection, preprocessing is refitted on the full outer-training fold before its test contours are transformed. Table 1 summarizes the configurations selected independently within the five outer folds. Outer-fold results are reported separately in Table 3.

Selection varies materially across folds for both token models, which is expected in this small dataset. The MLP

Table 2. Hybrid simulation on noise-free contours ($\sigma=0$), balanced accuracy (%), mean [95% CI] over 20 independent training runs. Shape-based columns use $K=32$ Fourier harmonics, while pixel-based baselines use the corresponding rasterized shapes.

Shape-based models		
N	MLP	Shape Trans.
50	56.3 [50.9, 61.7]	65.8 [63.4, 68.2]
100	63.4 [58.2, 68.6]	66.4 [64.0, 68.7]
250	69.8 [65.2, 74.5]	67.5 [65.1, 69.9]
1,000	69.6 [64.9, 74.4]	68.5 [63.6, 73.4]
Pixel-based models		
N	CNN	ViT-Tiny
50	50.1 [49.7, 50.6]	50.0 [50.0, 50.0]
100	50.2 [49.8, 50.6]	50.0 [50.0, 50.0]
250	51.1 [50.6, 51.5]	50.0 [50.0, 50.0]
1,000	52.5 [51.9, 53.1]	50.0 [50.0, 50.1]

nevertheless remains compact in every fold, with its largest selected configuration still at least 46 times smaller than ViT-Tiny. Because bandwidth and capacity are selected independently by architecture, the models do not necessarily receive identical token inputs.

Simulation architecture selection. For the simulation, `grid_holdout.py` selects architectures per (K, N, model) on noise-free data. The validation set is a fixed set of 500 independently generated samples held out from training and testing, and each candidate is averaged over three random initializations. Final evaluation uses the selected bandwidth $K=32$, repeats training 20 times per noise level, and uses a fixed 2,000-sample test set so all models are compared on identical examples within each noise condition. These runs were submitted as DRAC batch jobs. Each reported 95% confidence interval is $\bar{x} \pm t_{0.975, 19s} / \sqrt{20}$ across the 20 training runs.

6. Results

6.1. BraTS 2020

BraTS results are means over five patient-disjoint outer folds, which serve as the test partitions. Hyperparameters and training epochs are selected using one grouped inner validation split, after which each model is retrained on the complete outer-training fold and evaluated once on unseen patients. Shape preprocessing is fitted only on the corresponding training subset at both stages. Values in parentheses are sample standard deviations across outer folds.

Observation. The MLP has the highest mean accuracy, balanced accuracy, and LGG F1. Its 71.5% mean balanced accuracy is 5.6 points above ResNet-18 and 8.2 points above ViT-Tiny. The pooled out-of-fold balanced accuracy is 72.4%, with a patient-bootstrap 95% interval of 66.4–

Table 3. BraTS 2020 patient-disjoint cross-validation results, mean (SD) in percent. Parameter ranges reflect fold-specific selection.

Model	Params	Acc.	Bal.
ResNet-18	11.2M	76.8 (11.0)	65.9 (4.6)
ViT-Tiny	5.43M	74.6 (20.3)	63.3 (9.4)
Shape Trans.	0.80–4.75M	76.7 (3.2)	71.0 (3.2)
MLP	2.9–117.3k	81.4 (6.0)	71.5 (6.4)

Table 4. BraTS 2020 patient-disjoint per-class F1, mean (SD) in percent.

Model	LGG F1	HGG F1
ResNet-18	48.4 (11.4)	84.0 (10.1)
ViT-Tiny	42.2 (24.7)	78.5 (25.7)
Shape Trans.	52.8 (9.8)	84.3 (2.7)
MLP	54.9 (12.2)	87.7 (5.6)

Table 5. BraTS 2020 mean final-training wall-clock time and mean balanced accuracy across patient-disjoint folds. Runs use an Apple M3 CPU at batch size 16; selected epoch counts vary by fold and token export is excluded.

Model	Params	Train	Bal.
ResNet-18	11.2M	1,052 s	65.9
ViT-Tiny	5.43M	377 s	63.3
Shape Trans.	0.80–4.75M	18.1 s	71.0
MLP	2.9–117.3k	0.16 s	71.5

77.8%. The Shape Transformer has higher LGG recall but lower precision and F1; its small balanced-accuracy difference from the MLP is well within fold variability. The efficiency difference is much larger: even the largest selected MLP is at least 46 times smaller than the pixel baselines, and its mean final-training time is 0.16 seconds.

6.2. Controlled Simulation Results

Tables 2 and 9 report balanced accuracy as mean [95% CI] over 20 independent training runs. The main results use Moindjié *et al.*’s direct coefficient-space path, and pixel-based baselines use matched rasters for the same examples. An equivalent spatial-roundtrip implementation gives identical metrics in a fixed-split BraTS check and broadly similar simulation results (Supplementary Sections C–D).

Observation. On noise-free contours, both shape-based classifiers perform above chance at every sample size, whereas the CNN remains between 50.1% and 52.5%. ViT-Tiny remains near 50% with almost no run-to-run variation. A representative rerun at $N=250$ confirmed nonzero gradi-

ents, but the loss converged near $\log 2$ and the model predicted a single class for all 2,000 test examples. This behavior indicates optimization collapse rather than an inactive training loop. We therefore do not interpret the ViT-Tiny result as evidence of a general disadvantage of pixel-based models.

The contrast partly reflects the simulation design. Recentering and rescaling remove the explicit deformation variables from the raster even though those variables contribute to the label, whereas the deformation token retains them. The MLP attains its highest noise-free mean at $N=250$, although its confidence interval is wide. The noise results and attention analysis are reported in the supplement.

The CNN rises from 50.1–52.5% without noise to 59.7–66.5% with noise (Table 9), a large and counterintuitive change. A possible explanation is that vertex-wise contour noise produces raster-level boundary statistics or texture-like cues through its interaction with rasterization, consistent with the discussion in Section 2.1. This mechanism remains a hypothesis and requires direct testing. Overall, the deformation-plus- shape representation is effective in the noise-free diagnostic. Shape-based models achieve the strongest overall results in this controlled setting, although the unequal information retained by the token and raster representations limits broader architecture-level conclusions.

7. Discussion and Limitations

Our main finding from the BraTS analysis is that an explicit shape-based representation supports competitive class-balanced performance with much smaller models. Under patient-disjoint cross-validation, the MLP has the highest mean balanced accuracy at 71.5% and the highest mean LGG F1 at 54.9%. The selected MLPs use 2.9k–117.3k parameters, at least 46 times fewer than the pixel baselines. Its pooled out-of-fold balanced accuracy is 72.4%, with a patient-bootstrap 95% interval of 66.4–77.8%. The stronger result is therefore compactness without sacrificing mean performance. This is precisely the setting in which we expected a shape-based inductive bias to be valuable, as the segmentation mask is already available and the phenotype is plausibly related to morphology.

The class-wise behavior is asymmetric, which is scientifically useful. Pixel-based baselines remain strong for HGG because the majority class is common and high-grade tumors often have large spatial and size cues. Both shape-based models improve mean LGG F1 over the pixel baselines, suggesting that explicit alignment, scale-aware deformation variables, and harmonic residuals can reduce irrelevant variation before the supervised learner sees the data. This suggests asking where in the class structure, and in which frequency bands, shape analysis provides information that pixel-based models dilute or use inefficiently.

Our simulation helps us understand the MLP result, but doesn't establish a normative ranking. When the functional-shape representation is noise-free and the target is recoverable from a small set of deformation and harmonic coordinates, flattening the token matrix is often enough.

The importance of this point for the machine-learning literature is that it changes the role of domain structure. The pipeline treats shape analysis as the step that defines the learning problem itself, with learned nonlinear classifiers operating on inputs organized by transformations and harmonic scales that have a mathematical interpretation before training begins.

The same representation also addresses two practical goals beyond accuracy. It supports mechanistic interpretability where deformation, low-frequency elongation and asymmetry, mid-frequency lobulation, and higher-frequency boundary detail to be inspected separately and, in future work, ablated. It is also scalable for very large imaging collections because each contour becomes a fixed-size matrix of tens or hundreds of numbers, independent of the original image resolution, making storage, CPU training, and inference cheaper than repeated pixel-based model training.

Limitations. We reduce each 3D segmentation to one axial slice and one or more connected-component contours, and a multi-slice or surface-based extension would test whether the representation scales to full 3D tumor geometry. BraTS tumour contours are annotated by experts and while automated segmentation models have approached expert level accuracy, segmentation error could plausibly reduce the performance of similar models. Patient grouping prevents subject overlap between outer training and test folds, but 95 LGG contours remain distributed across only five test folds, producing high variability in minority-class estimates. Model selection uses one patient-grouped validation split per outer fold rather than averaging selection over all possible inner folds. We use the first training preshape as the reference μ , leaving training-fold Fréchet-mean estimation for future work [26]. ResNet-18 and ViT-Tiny are trained from scratch on only a few hundred contours per fold, so their performance may reflect limited-data optimization as well as the input representation, particularly for ViT-Tiny. Binary HGG/LGG grading is simpler than modern molecularly informed glioma classification. Future work can test which harmonic bands remain stable across cohorts and whether attention helps specifically when labels depend on interactions between deformation and local boundary irregularity.

8. Conclusion

We define aligned Fourier shape tokens for glioma grading from tumor boundaries. Functional-shape alignment

converts each contour into one deformation token plus frequency-ordered residual shape tokens, giving a compact representation of morphology. In patient-disjoint cross-validation on BraTS 2020, fold-selected MLPs over these tokens give the highest mean balanced accuracy and LGG F1 while using 2.9k–117.3k parameters, at least 46 times fewer than the pixel classifiers. A fixed-split implementation check shows that the equivalent spatial roundtrip and second projection produce the same aggregate BraTS metrics (Supplementary Section C).

The broader implication of our work is that shape analysis can provide substantial dimensionality reduction while retaining mechanistic meaning. This is relevant beyond medical imaging. Geospatial and remote-sensing data, for example, are often massive, high-dimensional, multiresolution, and spatially structured, and the literature has repeatedly emphasized the computational and statistical challenges of learning directly from these data at scale [23, 44]. Many objects in those settings also have meaningful boundaries, including coastlines, agricultural fields, watersheds, roads, fire perimeters, and urban footprints. For such problems, a shape-based representation could turn very large rasters into compact geometric summaries that can be indexed, audited, and learned from cheaply. Contours may therefore complement full images by providing an initial layer for scalable and interpretable learning whenever morphology is the primary scientific object.

Code. The implementation uses PyTorch 2.8 for model training and a Python implementation inspired by Moindjié *et al.*'s coefficient-space functional-shape alignment. An anonymized repository will accompany double-blind review, and the public URL will be restored after review.

Acknowledgements This research was enabled in part by support provided by BC DRI Group and the Digital Research Alliance of Canada (<https://alliancecan.ca>).

References

- [1] Hugo J. W. L. Aerts, Emmanuel Rios Velazquez, Ralph T. H. Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Daniëlle Rietveld, Frank Hoesbers, Michelle M. Rietbergen, C. René Leemans, Andre Dekker, John Quackenbush, Robert J. Gillies, and Philippe Lambin. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5:4006, 2014. 1, 2
- [2] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the Cancer Genome Atlas glioma MRI collections with expert

- segmentation labels and radiomic features. *Scientific Data*, 4(1):170117, 2017. 1, 3
- [3] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NeurIPS*, pages 831–837, 2000. 2
- [4] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *IEEE TPAMI*, 24(4):509–522, 2002. 2
- [5] Valerio Biscione and Jeffrey S. Bowers. Convolutional neural networks are not invariant to translation, but they can learn to be. *ArXiv*, abs/2110.05861, 2021. 3
- [6] Fred L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991. 2
- [7] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint*, 2021. 1
- [8] Hwan-Ho Cho, Seung-Hak Lee, Jonghoon Kim, and Hyunjin Park. Classification of the glioma grading using radiomics analysis. *PeerJ*, 6:e5982, 2018. 2
- [9] Taco S. Cohen and Max Welling. Group equivariant convolutional networks. In *International Conference on Machine Learning*, pages 2990–2999, 2016. 1
- [10] Timothy F. Cootes, Christopher J. Taylor, David H. Cooper, and Jim Graham. Active shape models—their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995. 2
- [11] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, pages 4171–4186, 2019. 3
- [12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 3, 6
- [13] Ian L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, Chichester, 1998. 1, 2
- [14] Herbert Freeman. On the encoding of arbitrary geometric configurations. *IRE Trans. Electronic Computers*, 10(2):260–268, 1961. 2
- [15] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *ArXiv*, abs/1811.12231, 2018. 3
- [16] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016. 1, 2
- [17] Ali Hatamizadeh, Dong Yang, Holger R. Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1748–1758, 2022. 2
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 6
- [19] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:203–211, 2021. 2
- [20] Frank P. Kuhl and Charles R. Giardina. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 18(3):236–258, 1982. 1, 2
- [21] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G. P. M. van Stiphout, Patrick Granton, Catharina M. L. Zegers, Robert Gillies, Ronald Boellard, Andre Dekker, and Hugo J. W. L. Aerts. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, 2012. 1, 2
- [22] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam R. Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In *International Conference on Machine Learning*, pages 3744–3753, 2019. 3
- [23] Songnian Li, Suzana Dragicevic, Francesc Antón Castro, Monika Sester, Stephan Winter, Arzu Coltekin, Christopher Pettit, Bin Jiang, James Haworth, Alfred Stein, and Tao Cheng. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:119–133, 2016. 8
- [24] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *ACM SIGGRAPH*, pages 163–169, 1987. 3
- [25] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal Brain Tumor image segmentation benchmark (BRATS). *IEEE Trans. Medical Imaging*, 34(10):1993–2024, 2015. 1, 3
- [26] Issam-Ali Moindjié, Cédric Beaulac, and Marie-Hélène Descary. A functional approach to curve alignment and shape analysis. *arXiv preprint*, 2025. 1, 2, 3, 4, 5, 8, 11
- [27] Issam-Ali Moindjié, Marie-Hélène Descary, and Cédric Beaulac. Multivariate planar curves: A statistical framework for shape analysis in images. *arXiv preprint*, 2025. 2
- [28] Eric Persoon and King-Sun Fu. Shape discrimination using fourier descriptors. *IEEE Trans. Systems, Man, and Cybernetics*, 7(3):170–179, 1977. 2
- [29] F. James Rohlf and Dennis Slice. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1):40–59, 1990. 2
- [30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer, 2015. 2
- [31] Max-Ludwig Schäfer, Martin H. Maurer, Michael Synowitz, Joost Wüstefeld, Tim Marnitz, Florian Streitparth, and Edzard Wiener. Low-grade (WHO II) and anaplastic (WHO III) gliomas: differences in morphology and MRI signal intensities. *European Radiology*, 23(10):2846–2853, 2013. 1, 3

- [32] scikit-image developers. `skimage.measure.find_contours`. scikit-image API documentation, 2026. Accessed 2026-07-15. [3](#)
- [33] Anuj Srivastava and Eric P. Klassen. *Functional and Shape Data Analysis*. Springer, 2016. [1](#), [2](#)
- [34] Anuj Srivastava, Eric Klassen, Shantanu H. Joshi, and Ian H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE TPAMI*, 33(7):1415–1428, 2011. [1](#), [2](#)
- [35] Barinder Thind, Kevin Multani, and Jiguo Cao. Deep learning with functional inputs. *Journal of Computational and Graphical Statistics*, 32(1):171–180, 2023. [2](#)
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pages 5998–6008, 2017. [3](#), [5](#)
- [37] Sidi Wu, Cédric Beaulac, and Jiguo Cao. Neural networks for scalar input and functional output. *Statistics and Computing*, 33(5):118, 2023. [2](#)
- [38] Sidi Wu, Cédric Beaulac, and Jiguo Cao. Functional autoencoder for smoothing and representation learning. *Statistics and Computing*, 34(6):203, 2024. [2](#)
- [39] Cheng Xu, Yue Peng, Wensheng Zhu, Zhen Chen, Jing Li, Wei Tan, Zhen Zhang, and Xiang Chen. An automated approach for predicting glioma grade and survival of LGG patients using CNN and radiomics. *Frontiers in Oncology*, 12: 969907, 2022. [2](#)
- [40] Evangelia I. Zacharaki, Sumei Wang, Sanjeev Chawla, Dong Soo Yoo, Ronald Wolf, Elias R. Melhem, and Christos Davatzikos. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magnetic Resonance in Medicine*, 62(6):1609–1618, 2009. [2](#)
- [41] Charles T. Zahn and Ralph Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Trans. Computers*, 21(3):269–281, 1972. [1](#), [2](#)
- [42] Yixin Zhang and Maciej A. Mazurowski. Convolutional neural networks rarely learn shape for semantic segmentation. *Pattern Recognit.*, 146:110018, 2023. [3](#)
- [43] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip H. S. Torr, and Vladlen Koltun. Point transformer. In *IEEE/CVF International Conference on Computer Vision*, pages 16259–16268, 2021. [3](#)
- [44] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017. [8](#)
- [45] Ying Zhuge, Holly Ning, Peter Mathen, Jason Y. Cheng, Andra V. Krauze, Kevin Camphausen, and Robert W. Miller. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Medical Physics*, 47(7):3044–3053, 2020. [2](#)

Supplementary Material

A. Geometric Meaning of Individual Harmonics

Each Fourier harmonic controls a qualitatively different aspect of shape geometry. Figure 3 shows the effect of varying a single cosine coefficient a_x at harmonics $k=2$, $k=3$, and $k=5$. Harmonic $k=2$ governs bilateral elongation, $k=3$ introduces triangularity, and $k=5$ adds finer lobes. This frequency-to-geometry correspondence gives our token representation its structure, with each token controlling a specific scale of shape variation.

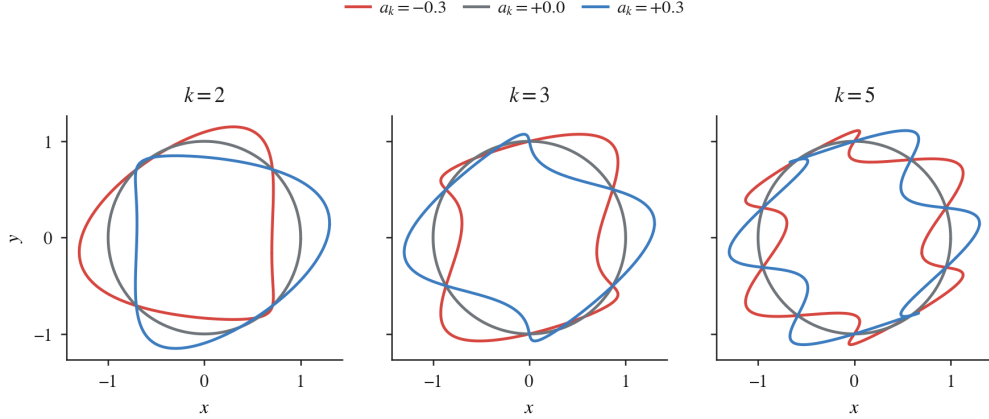


Figure 3. Effect of varying a single Fourier coefficient a_x at different harmonics.

B. Compute Environment and Timing

BraTS timing runs were measured on an Apple M3 chip (CPU only). The simulation sweeps and final repeated evaluations were run on Compute Canada cluster resources, so the simulation tables report accuracy and confidence intervals. Table 6 reports the fixed-split implementation-check timing, inference time, parameter count, and input dimensionality; patient-grouped final-training times appear in Table 5.

Table 6. Fixed-split implementation-check cost on BraTS 2020 (Apple M3 CPU). Training is wall-clock for 40 epochs at batch size 16. Inference is per batch of 16 samples.

Input	Model	Params	Dim.	Train	Infer.
Pixel	ResNet-18	11.2M	224×224	547 s	349.9 ms
Pixel	ViT-Tiny	5.43M	224×224	285 s	180.2 ms
Single	Shape Trans.	3.16M	9×6	15 s	3.66 ms
Single	MLP	7.5k	33×6	0.75 s	0.03 ms
Double	Shape Trans.	3.16M	9×6	11 s	2.97 ms
Double	MLP	7.5k	33×6	0.61 s	0.02 ms

C. Spatial-Roundtrip Robustness Check

Moindjié *et al.*'s formulation remains in functional space after the initial Fourier smoothing. Phase, inverse rotation, and scale are applied directly to the fitted coefficients through the block structure of the basis [26]. We call this the *single* path because it requires only one curve-to-Fourier projection. For comparison, we introduced a spatial-roundtrip approximation, which we call the *double* path. It returns to point space, applies the phase shift by cyclic linear interpolation, rotates the sampled curve, and computes a second Fourier projection. The paths are equivalent in the continuous limit, but the double path introduces interpolation, resolution dependence, and a second truncation that can oversmooth the curve.

On the original fixed contour split, the double path gives the same aggregate BraTS results as the direct coefficient-space path for both shape-based classifiers (Table 7). The per-class precision, recall, and F1 values are also identical across paths. The simulation results are likewise comparable, with overlapping confidence intervals at every sample size (Table 8 and Supplementary Section D). We therefore retain the direct coefficient-space path in the patient-grouped main results and treat the double path as an implementation robustness check.

Table 7. Fixed-split BraTS 2020 comparison of the direct coefficient-space (single) and spatial-roundtrip (double) implementations. LGG/HGG columns report class recall, and G-mean is their geometric mean.

Path	Model	Acc.	Bal.	G-mean	LGG	HGG
Single	Shape Trans.	85.3	72.0	67.8	47.8	96.2
Single	MLP	85.3	76.6	75.0	60.9	92.4
Double	Shape Trans.	85.3	72.0	67.8	47.8	96.2
Double	MLP	85.3	76.6	75.0	60.9	92.4

Table 8. Noise-free hybrid simulation comparison of the direct coefficient-space (S) and spatial-roundtrip (D) implementations, balanced accuracy (%), mean [95% CI] over 20 independent training runs.

N	MLP-S	Shape-S	MLP-D	Shape-D
50	56.3 [50.9, 61.7]	65.8 [63.4, 68.2]	58.4 [52.5, 64.3]	66.3 [65.0, 67.7]
100	63.4 [58.2, 68.6]	66.4 [64.0, 68.7]	66.3 [61.9, 70.8]	67.8 [66.1, 69.6]
250	69.8 [65.2, 74.5]	67.5 [65.1, 69.9]	71.5 [67.2, 75.8]	67.8 [65.5, 70.1]
1,000	69.6 [64.9, 74.4]	68.5 [63.6, 73.4]	68.4 [64.2, 72.6]	66.8 [62.8, 70.8]

D. Simulation with Noise

Table 9. Hybrid simulation with contour vertex noise ($\sigma=0.10$), balanced accuracy (%), mean [95% CI] over 20 independent training runs. Shape-based and pixel-based classifiers are compared on the same generated examples.

N	MLP-S	Shape-S	MLP-D	Shape-D	CNN	ViT-Tiny
50	58.7 [51.2, 66.3]	66.9 [63.5, 70.3]	61.0 [53.4, 68.5]	69.4 [67.9, 70.9]	59.7 [57.1, 62.2]	50.0 [50.0, 50.0]
100	57.8 [50.5, 65.2]	66.5 [64.2, 68.8]	56.1 [49.1, 63.1]	68.8 [66.7, 71.0]	66.4 [64.5, 68.2]	50.0 [49.9, 50.0]
250	54.8 [47.5, 62.2]	67.6 [65.0, 70.2]	59.0 [52.3, 65.7]	68.0 [66.1, 70.0]	66.5 [64.8, 68.3]	50.5 [49.5, 51.5]
1,000	59.9 [54.4, 65.5]	65.6 [62.3, 68.9]	60.8 [54.6, 67.1]	68.9 [66.4, 71.4]	64.6 [62.8, 66.5]	52.8 [50.3, 55.3]

E. Simulation Attention by Noise and Tokenization Path

Figure 4 shows the paired change in first-layer Shape Transformer attention from the noise-free condition to $\sigma=0.10$ for the single and double tokenization paths at each reported sample size. Each curve is noisy-minus-clean attention from the pose-token query to each harmonic-token key, expressed in attention percentage points, paired by training seed, averaged first over heads and the fixed 2,000-example test set, and then summarized over the same 20 seeds used in Tables 2 and 9. Positive values indicate that noise shifts more pose-query attention toward a harmonic; negative values indicate a reduction. Pose-token self-attention decreases significantly under noise for the double path at every sample size. For the single path, the decrease is significant at $N=50$ and $N=1,000$; the intervals at $N=100$ and $N=250$ include zero. These shifts describe changes in learned attention allocation and should not by themselves be interpreted as causal feature importance.

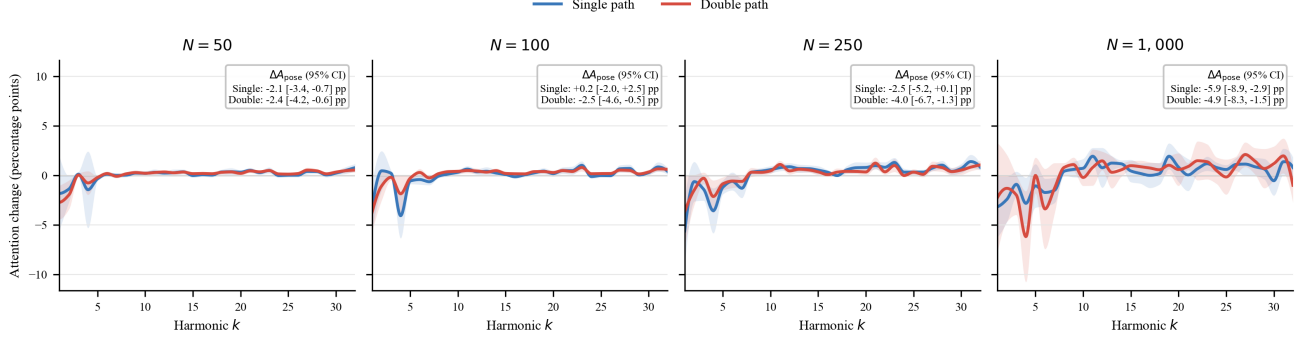


Figure 4. Paired noisy-minus-clean change in pose-query attention to harmonics, in attention percentage points, for the single and double tokenization paths across sample sizes. Both paths are overlaid in each sample-size panel. Solid curves are means over 20 matched seeds and shaded regions are 95% Student- t confidence intervals. Curves use shape-preserving interpolation for display; summaries are computed at the original integer harmonics. Each annotation gives the paired percentage-point change in pose-token self-attention and its 95% confidence interval for both paths. Models were trained separately at each noise level.

F. Simulated Shape Examples

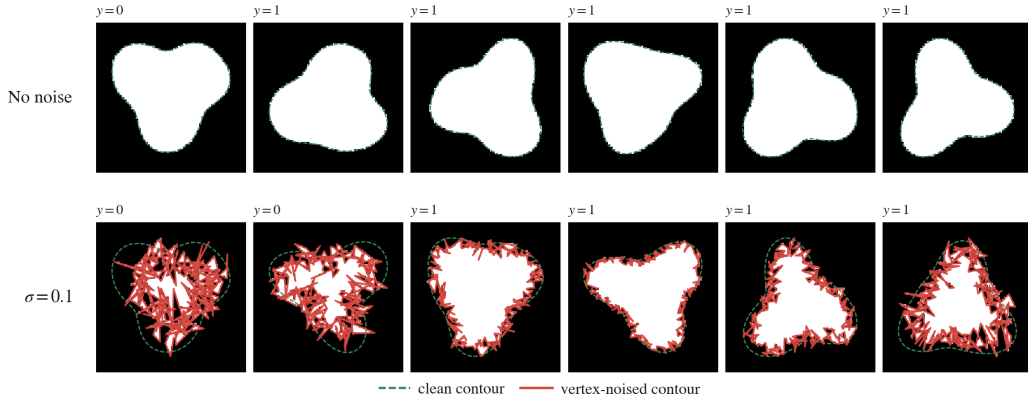


Figure 5. Example simulated shapes. Each cell shows a 64×64 raster with the noise-free contour (green dashed) and the vertex-noised contour (red). Labels combine deformation evidence from scale and translation with class-dependent mid-frequency bumps; pixel-based models read the raster, while shape-based models read the deformation and residual shape tokens.

G. Full Hyperparameter-Sweep Tables

The MLP and Shape Transformer sweeps on BraTS cover the reported token-classifier configurations. The full per-configuration logs are in the code repository. The completed hybrid simulation aggregates are stored under `results/hybrid.sim/`, including `sim.single.ci.results.json`, `sim.double.ci.results.json`, and `pixel.ci.results.json`.