

# Data Harmonization with StyleTransfer-GANs: Enhancing Non-Invasive IDH Classification in Brain Tumors

Chandan Ganesh B. Y.<sup>a</sup>, Jason Bowerman<sup>a</sup>, Nghi C. D. Truong<sup>a</sup>, Benjamin C. Wagner<sup>a</sup>, Divya D. Reddy<sup>a</sup>, James M. Holcomb<sup>a</sup>, Niloufar Saadat<sup>a</sup>, Kimmo J. Hatanpaa<sup>b</sup>, Toral R. Patel<sup>c</sup>, Baowei Fei<sup>d</sup>, Matthew D. Lee<sup>e</sup>, Rajan Jain<sup>e,f</sup>, Richard J. Bruce<sup>g</sup>, Marco C. Pinho<sup>a</sup>, Ananth J. Madhuranthakam<sup>h</sup>, Joseph A. Maldjian<sup>a</sup>

<sup>a</sup>Department of Radiology, UT Southwestern Medical Center, <sup>b</sup>Department of Pathology, UT Southwestern Medical Center, <sup>c</sup>Department of Neurological Surgery, UT Southwestern Medical Center, <sup>d</sup>Department of Bioengineering, UT Dallas, <sup>e</sup>Department of Radiology, NYU Grossman School of Medicine, <sup>f</sup>Department of Neurosurgery, NYU Grossman School of Medicine, <sup>g</sup>Department of Radiology, University of Wisconsin School of Medicine, <sup>h</sup>Department of Radiology, Mayo Clinic

## ABSTRACT

Isocitrate dehydrogenase (IDH) mutation status has emerged as an important prognostic marker in brain gliomas. Accurate non-invasive determination of IDH mutation status is crucial for effective therapy and prognosis. However, the variability in imaging protocols across institutions hinders the reliability of deep learning (DL) models used for IDH classification. To address data heterogeneity, a StyleTransfer-GAN (*ST-net*) was developed to harmonize multi-site MRI data. *ST-net* was applied to harmonize MR images from diverse databases while preserving critical imaging features. A high performing deep-learning network (*MC-net*) was used to assess the impact of style transfer on IDH classification. The results demonstrated that *ST-net* effectively aligned style features across datasets and minimized discrepancies caused by varied imaging protocols. The IDH classification accuracy improved post-harmonization, with enhancements in both sensitivity and specificity depending on the style transfer reference. Our study highlights the potential of data harmonization using GANs to improve the generalizability and clinical utility of DL models in neuro-oncology. It also provides a scalable solution adaptable to various neuroimaging tasks.

**Keywords:** IDH mutation, deep-learning, data-harmonization, StyleTransfer-GAN, StarGANv2.

## 1. INTRODUCTION

Gliomas, the most common primary brain tumors, pose significant challenges in diagnosis and treatment. The identification of isocitrate dehydrogenase (IDH) mutation status as a marker for therapy and prognosis has highlighted the need for reliable non-invasive approaches to determine IDH status.[1] Recent advances in deep learning (DL) have demonstrated substantial success in various tasks, including molecular profiling of brain tumors.[2-6]. Despite these advances, challenges remain in the accuracy and generalizability of these models, often due to variations in imaging protocols across different centers.[7] This issue, known as data heterogeneity, can significantly impact the performance of DL models, making them less reliable in clinical practice.[8-10]

To address these challenges, Data Harmonization (DH) techniques have emerged as a promising solution. Among these, StyleTransfer, a DL-based approach has shown significant potential for DH of MR Images.[11, 12] StyleTransfer creates uniform datasets by aligning the styles of different images. It can preserve critical imaging features and minimize discrepancies caused by variations in MRI acquisition protocols, and scanner types.[11, 12] In this study, we developed and tested a DL-based StyleTransfer model to address the issue of data heterogeneity and improve the performance of IDH classification. This approach underscores the importance of data harmonization in developing robust and generalizable DL models in medical imaging.

## 2. MATERIALS & METHODS

### 1.1 Data and Pre-processing

Brain tumor MRI and genomic information were obtained from five publicly available and three in-house/collaborator databases. [13-18] Subjects were screened for the availability of IDH mutation status and multi-contrast MR images. Only pre-operative studies were included in the dataset (*Table 1*). Multi-contrast native space MRI from TCIA, IvyGAP, UTSW, NYU, UWM and UPenn were pre-processed using the FeTS platform.[19] It co-registers the images onto the SRI24 template space[20] and performs multi-class brain tumor segmentation. The EGD and UCSF data (only skull stripped data) were readily available as pre-processed images.[16, 18] Additionally, all datasets were z-score normalized.[21, 22] Multi-label FeTS tumor segmentations were combined to generate whole-tumor masks. The generated Whole-tumor masks for IDH-mutated and IDH wild-type cases were labelled with 1 and 2 respectively.[2] These tumor masks were used as the ground truth in training the IDH classification network.

Table 1: Distribution of IDH Status of all datasets included in the study

	Public databases					In-house / collaborator institutions			TOTAL
	TCIA	IvyGAP	EGD	Upenn	UCSF	UTSW	NYU	UWM	
<b>Mutant</b>	92	2	150	11	103	106	47	16	527
<b>Wildtype</b>	112	21	306	388	392	254	130	188	1791
<b>TOTAL</b>	204	23	456	399	495	360	177	204	<b>2318</b>

### 1.2 Training and Testing data

Two DL networks were used in this study. A pre-trained IDH classification network, *MC-net*[2] and, A new implementation of the StyleTransfer network, *ST-net*. The training & testing cohorts for both the networks were kept constant to avoid the problem of data leakage.[23, 24] A combination of TCIA + UTSW datasets were used for training and model optimization. MRI from NYU, UWM, EGD, UPenn, and UCSF were held-out for testing.

#### 1.2.1 ST-net

The *ST-net*, inspired by the *StarGANv2* model, was developed separately for each MRI contrast to preserve specific characteristics while harmonizing style-related variations.[12] The 2D network architecture treats each image as belonging to a unique "domain," separating it into content (anatomical structures) and style (non-biological variations).[12] The *ST-net* consists of a generator, a discriminator, a style encoder, and a mapping network. The generator produces a harmonized image by retaining the input's anatomical content while adopting the style of a reference image. This is guided by the style encoder and controlled by the discriminator. *ST-net* was trained using adversarial, cycle-consistency, and style

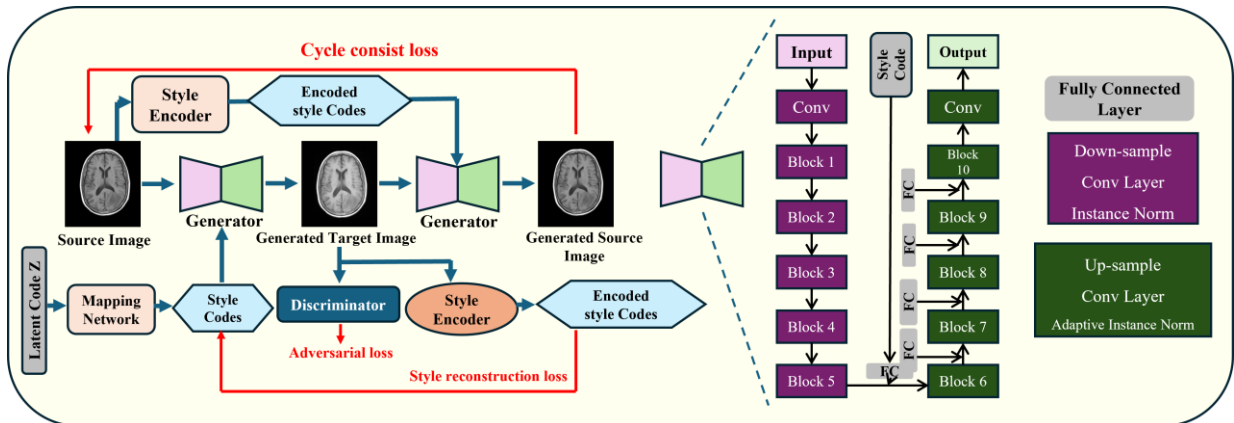


Figure 1: Network architecture of ST-Net

reconstruction losses. Both training and harmonization were performed in 2D, with slices stacked to create harmonized 3D images. A 2D approach was chosen to balance GPU memory and high-resolution image processing.

### 3. RESULTS

#### 3.1 ST-net

Two subjects (1 mutated & 1 wildtype) from the UTSW database were carefully selected as reference images for data harmonization. The selected images were acquired on a 3T Philips and Siemens scanners respectively and had high resolution multi-contrast MRI. The testing cohort were style-transferred to each of the reference images separately. The impact of style transfer on the MRI images was evaluated using several image comparison metrics (Table 2). The JS Divergence values across MRI contrasts (T1-pre, T1CE, T2, T2-FLAIR) were relatively low, ranging from 0.39 to 0.44, indicating minimal differences in intensity distributions between the original and style-transferred images. High SSIM values (0.90 to 0.94) suggest that *ST-net* effectively preserved the structural integrity of the MRI images, maintaining essential anatomical features without significant deviations. PSNR values, which varied between 22.23 dB and 25.46 dB, indicate that the overall quality of the style-transferred images was good, despite some noise or variation introduced during the process.

The intensity correlation between the original and style-transferred images remained consistently high (0.94 to 0.96), suggesting that the style transfer process successfully preserved the overall intensity patterns across the images. Figure 2 shows examples of the style-transferred images and the uniform histograms, illustrating how *ST-net* successfully aligned the intensity profiles. This consistency in intensity distribution ensures that critical imaging features were preserved, enabling reliable comparisons and analyses across datasets. The visual confirmation supports the quantitative results, demonstrating that *ST-net* effectively harmonizes images while maintaining essential imaging features. Figure 3 depicts the t-SNE plots of style codes extracted from different databases before & after harmonization, revealing clear clustering post-harmonization. This clustering indicates that *ST-net* successfully reduced inter-dataset variability, facilitating better integration of multi-site data.

Table 2: Example image comparison metrics → Before v/s After style transfer to an **IDH mutated** subject

	JS_Divergence	SSIM	PSNR (dB)	Intensity Correlation
<b>T1-pre</b>	0.44 +/- 0.15	0.93 +/- 0.03	23.75 +/- 2.5	0.95 +/- 0.02
<b>T1CE</b>	0.39 +/- 0.17	0.94 +/- 0.02	25.46 +/- 2.8	0.94 +/- 0.02
<b>T2</b>	0.41 +/- 0.17	0.94 +/- 0.02	24.72 +/- 3.5	0.96 +/- 0.02
<b>T2-Flair</b>	0.43 +/- 0.15	0.92 +/- 0.02	23.88 +/- 2.8	0.95/- 0.01

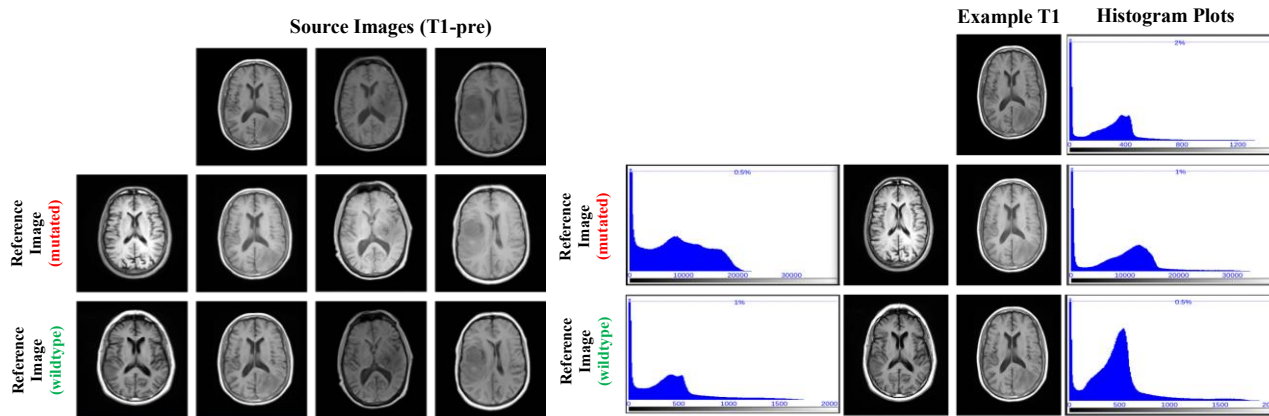


Figure 2: Examples of the style-transferred images and Comparison of intensity histograms



Figure 3: T-distributed stochastic neighbor embedding (t-SNE) representation of the style codes extracted from **T1-images** of different databases before and after the harmonization

### 3.2 MC-net

The trained *MC-net* was tested on held-out datasets before and after style transfer (Table 3). *MC-net* results indicate that data harmonization helped in enhancing the IDH classification accuracy. Style-transfer to an IDH-mutated subject led to improved sensitivity, allowing for better identification of IDH-mutated tumors. Conversely, style-transfer to an IDH-wildtype subject improved the specificity, enabling accurate identification of IDH-wildtype cases. *MC-net* achieved AUC values of 0.943, 0.952 and 0.948 for Exp 1, 2 & 3 respectively. Data harmonization showed improved performance. These improvements highlight the effectiveness of style transfer in refining the model's performance across diverse datasets.

Table 3: IDH Classification accuracies for before and after data harmonization

Exp	Train group	Testing type	Metrics	NYU	UWM	EGD	UPenn	UCSF	Overall
1.	TCIA + UTSW	Real MR Images	Accuracy	<b>89.3</b>	<b>91.7</b>	<b>92.5</b>	<b>95.0</b>	<b>94.1</b>	<b>93.0</b>
			Sensitivity	89.4	68.8	83.3	81.8	89.3	85.6
			Specificity	89.2	93.6	97.1	95.4	95.4	94.9
2.	TCIA + UTSW	After StyleTransfer to <b>mutated case</b>	Accuracy	<b>89.8</b>	<b>90.7</b>	<b>93.0</b>	<b>95.2</b>	<b>94.3</b>	<b>93.3</b>
			Sensitivity	91.5	68.8	89.3	81.8	88.3	88.1
			Specificity	89.2	92.6	94.8	95.6	95.9	94.5
3.	TCIA + UTSW	After StyleTransfer to <b>wildtype case</b>	Accuracy	<b>90.4</b>	<b>92.2</b>	<b>92.1</b>	<b>95.7</b>	<b>94.3</b>	<b>93.5</b>
			Sensitivity	87.2	68.8	80.7	81.8	88.3	83.5
			Specificity	91.5	94.1	97.7	96.1	95.9	95.7

## 4. DISCUSSION

We developed and evaluated a style-transfer deep-learning network (*ST-net*) to enhance IDH classification accuracy. This study demonstrates the potential of DH using GANs to address variability in MR images across different institutions. *ST-net* showed promising results that underscore the importance of DH. One of the key findings of our study is the preservation of essential imaging features after style transfer. *ST-net* successfully reduced inter-dataset variability (Fig 2 & Table 2). The t-SNE plots of the style codes revealed clear clustering after harmonization, indicating that *ST-net* effectively aligned style features across datasets and minimized data heterogeneity. Histogram analysis further confirmed that *ST-net* reduced variability in intensity distribution, ensuring consistent and comparable imaging data. The low JS-Divergence scores

indicated minimal changes between the original and style-transferred images, while high SSIM and intensity correlation values showed that the anatomical integrity of the images was preserved.

Another important finding of our study is the improvement in IDH classification after applying style transfer. By reducing image variability, *ST-net* allowed *MC-net* to predict on more uniform input features, leading to better performance. It enhanced the consistency and reliability of IDH predictions, increasing overall classification accuracy and AUC. The results indicate that DH improves the model's applicability in multi-site studies, ensuring consistent outcomes despite data heterogeneity. Notably, style transfer to an IDH-mutated reference image enhanced sensitivity, while style transfer to an IDH-wildtype reference improved specificity. These findings suggest that the choice of reference image in style transfer makes a difference in the classification performance. It also suggests that DH can be tailored to clinical needs, enhancing either sensitivity or specificity as required. Our study seamlessly applies style-transfer to enhance IDH classification. This method is particularly unique as it allows for a nuanced interpretation of MRI data. *ST-net* helps maintain focus on primary imaging features while mitigating the impact of inter-site variability. DH refines and augments the primary imaging data, ensuring consistency without compromising the integrity of critical features. This balance between DH & feature preservation is critical for clinical applications of DL models. It keeps the focus on the imaging characteristics most indicative of IDH mutations while improving data consistency. The success of *ST-net* in mitigating data heterogeneity across multi-site studies emphasizes its potential role in advancing the field of non-invasive IDH classification.

*MC-net* & *ST-net* were trained on MR images with the skull, leveraging the additional anatomical context it provides. This approach helps maintain the anatomical integrity of the entire cranial region during harmonization, resulting in realistic and clinically relevant images. Retaining the skull allows *ST-net* to better handle variations in head positioning, scanner types, and other acquisition-related differences. A noteworthy finding is that both models performed well on the UCSF dataset, which consisted of only skull-stripped images, demonstrating their robustness and generalizability. *MC-net* showed modest improvement in IDH classification post-style transfer, highlighting its generalizability. The impact of style-transfer on DH is justified in several studies.[11, 12] Our findings also show that DH can play a key role in enhancing IDH classification, offering a more standardized and robust approach to multi-site studies. The *ST-net* + *MC-net* combination is efficient, easy to implement, and integrates seamlessly into clinical workflows without extensive modifications. This approach can significantly enhance diagnostic accuracy without increasing the operational burden.

## 5. CONCLUSION

We developed and evaluated a style-transfer deep-learning network (ST-net) to enhance IDH classification. The combination of ST-net and MC-net was tested on a large and diverse database of 1,731 cases, making it the largest study on non-invasive IDH prediction to date. By leveraging ST-net for style transfer, we mitigated variability in multi-site MRI datasets, improving the accuracy of IDH classification by MC-net. This approach preserved crucial diagnostic features during harmonization and enhanced data consistency across institutions. The marked improvement in IDH classification following style transfer highlights the potential of integrating data harmonization techniques into clinical workflows, offering a scalable solution adaptable to various neuroimaging tasks. While our findings are promising, they also reveal the challenges of applying advanced deep learning techniques in clinical settings. Future work should focus on expanding the harmonization process, using multiple reference images for robust harmonization, and developing 3D models to better capture brain tumor spatial context. Further studies are needed to validate these findings in larger, more diverse cohorts to fully realize the clinical impact. Our results support the potential integration of ST-net and MC-net in clinical practice for non-invasive, accurate, and consistent IDH mutation status across imaging centers.

The successful implementation of ST-net and MC-net represents a significant advancement in neuro-oncology, underscoring the importance of data harmonization in multi-site neuroimaging studies. As neuroimaging increasingly relies on data from multiple institutions, addressing data heterogeneity is crucial for the generalizability and clinical utility of deep learning models. This approach provides a scalable solution adaptable to various imaging modalities and diagnostic tasks, contributing to more reliable and precise diagnostics and ultimately leading to better patient outcomes.

## ACKNOWLEDGEMENTS

This study was funded by R01CA260705 (JM) and NIH/NCI U01CA207091 (AJM, JAM).

## REFERENCES

- [1] D. N. Louis, A. Perry, G. Reifenberger *et al.*, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathol*, 131(6), 803-20 (2016).
- [2] C. G. Bangalore Yogananda, B. C. Wagner, N. C. Truong *et al.*, "MRI-Based Deep Learning Method for Classification of IDH Mutation Status," *Bioengineering*, 10(9), 1045 (2023).
- [3] M. L. Giger, "Machine Learning in Medical Imaging," *J Am Coll Radiol*, 15(3 Pt B), 512-520 (2018).
- [4] X. Zhang, Q. Tian, L. Wang *et al.*, "Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI," *Journal of Magnetic Resonance Imaging*, 48(4), 916-926 (2018).
- [5] B. Zhang, K. Chang, S. Ramkissoon *et al.*, "Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas," *Neuro Oncol*, 19(1), 109-117 (2017).
- [6] P. Chang, J. Grinband, B. D. Weinberg *et al.*, "Deep-Learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas," *AJNR Am J Neuroradiol*, 39(7), 1201-1207 (2018).
- [7] K. E. Keenan, J. G. Delfino, K. V. Jordanova *et al.*, "Challenges in ensuring the generalizability of image quantitation methods for MRI," *Medical physics*, 49(4), 2820-2835 (2022).
- [8] B. S. Guendouzi, S. Ouchani, H. El Assaad *et al.*, "A systematic review of federated learning: Challenges, aggregation methods, and development tools," *Journal of Network and Computer Applications*, 220, 103714 (2023).
- [9] M. Babar, B. Qureshi, and A. Koubaa, "Investigating the impact of data heterogeneity on the performance of federated learning algorithm using medical imaging," *Plos one*, 19(5), e0302539 (2024).
- [10] N. K. Dinsdale, E. Bluemke, V. Sundaresan *et al.*, "Challenges for machine learning in clinical translation of big data imaging studies," *Neuron*, 110(23), 3866-3881 (2022).
- [11] S. Liu, and P.-T. Yap, "Learning multi-site harmonization of magnetic resonance images without traveling human phantoms," *Communications Engineering*, 3(1), 6 (2024).
- [12] M. Liu, A. H. Zhu, P. Maiti *et al.*, "Style transfer generative adversarial networks to harmonize multisite MRI to a single reference image to avoid overcorrection," *Human Brain Mapping*, 44(14), 4875-4892 (2023).
- [13] L. Scarpance, T. Mikkelsen, S. Cha *et al.*, "The Cancer Genome Atlas Glioblastoma Multiforme Collection (TCGA-GBM)(Version 4)[Data set]," *Cancer Imaging Arch*. Published online, (2016).
- [14] N. Pedano, A. Flanders, L. Scarpance *et al.*, "The cancer genome atlas low grade glioma collection (TCGA-LGG)(version 3)[Data set]," *Cancer Imaging Archive*, (2016).
- [15] N. Shah, X. Feng, M. Lankerovich *et al.*, [Data from Ivy GAP [Data set]. The Cancer Imaging Archive], (2016).
- [16] S. R. van der Voort, F. Incekara, M. M. Wijnenga *et al.*, "The Erasmus Glioma Database (EGD): Structural MRI scans, WHO 2016 subtypes, and segmentations of 774 patients with glioma," *Data in brief*, 37, 107191 (2021).
- [17] S. Bakas, C. Sako, H. Akbari *et al.*, "The University of Pennsylvania glioblastoma (UPenn-GBM) cohort: Advanced MRI, clinical, genomics, & radiomics," *Scientific data*, 9(1), 453 (2022).
- [18] E. Calabrese, J. E. Villanueva-Meyer, J. D. Rudie *et al.*, "The university of california san francisco preoperative diffuse glioma mri dataset," *Radiology: Artificial Intelligence*, 4(6), e220058 (2022).
- [19] M. J. Sheller, B. Edwards, G. A. Reina *et al.*, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Sci Rep*, 10(1), 12598 (2020).
- [20] T. Rohlfing, N. M. Zahr, E. V. Sullivan *et al.*, "The SRI24 multichannel atlas of normal adult human brain structure," *Hum Brain Mapp*, 31(5), 798-819 (2010).
- [21] N. J. Tustison, P. A. Cook, A. Klein *et al.*, "Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements," *Neuroimage*, 99, 166-79 (2014).
- [22] N. J. Tustison, B. B. Avants, P. A. Cook *et al.*, "N4ITK: improved N3 bias correction," *IEEE Trans Med Imaging*, 29(6), 1310-20 (2010).

- [23] A. S. Wegmayr V, Buhmann J, Nicholas Petrick; Kensaku Mori, Editor(s), "Classification of brain MRI with big data and deep 3D convolutional neural networks," Published in SPIE Proceedings, Medical Imaging 2018: Computer-Aided Diagnosis, 1057501, (2018).
- [24] J. Y. Xinyang Feng, Zachary C Lipton, Scott A Small, Frank A Provenzano, "Deep Learning on MRI Affirms the Prominence of the Hippocampal Formation in Alzheimer's Disease Classification," bioRxiv, 2018:456277., (2018).