

<https://doi.org/10.1038/s41698-024-00760-1>

# Deep mutual learning on hybrid amino acid PET predicts H3K27M mutations in midline gliomas

Check for updates

Yifan Yuan<sup>1,2,3,10</sup>, Guanglei Li<sup>4,10</sup>, Shuhao Mei<sup>1,2,3,10</sup>, Mingtao Hu<sup>5</sup>, Ying-Hua Chu<sup>6</sup>, Yi-Cheng Hsu<sup>6</sup>, Chaolin Li<sup>7</sup>, Jianping Song<sup>1,2,3,8</sup>, Jie Hu<sup>1,2,3</sup>, Danyang Feng<sup>9</sup>, Fang Xie<sup>2,4</sup>, Yihui Guan<sup>2,4</sup>, Qi Yue<sup>1,2,3</sup>✉, Mianxin Liu<sup>5</sup>✉ & Ying Mao<sup>1,2,3</sup>✉

Predicting H3K27M mutation status in midline gliomas non-invasively is of considerable interest, particularly using deep learning with <sup>11</sup>C-methionine (MET) and <sup>18</sup>F-fluoroethyltyrosine (FET) positron emission tomography (PET). To optimise prediction efficiency, we derived an assistance training (AT) scheme to allow mutual benefits between MET and FET learning to boost the predictability but still only require either PET as inputs for predictions. Our method significantly surpassed conventional convolutional neural network (CNN), radiomics-based, and MR-based methods, achieved an area under the curve (AUC) of 0.9343 for MET, and an AUC of 0.8619 for FET during internal cross-validation ( $n = 90$ ). The performance remained high in hold-out testing ( $n = 19$ ) and consecutive testing cohorts ( $n = 21$ ), with AUCs of 0.9205 and 0.7404. The clinical feasibility of the proposed method was confirmed by the agreements to multi-departmental decisions and outcomes in pathology-uncertain cases. The findings position our method as a promising tool for aiding treatment decisions in midline glioma.

Midline glioma, including brainstem glioma and thalamus glioma, is a relatively rare disease, with an incidence of less than 5% in all central nervous system (CNS) tumours<sup>1</sup>. Among them, diffuse midline gliomas (DMGs) with H3K27M mutation are a newly described group of tumours in the World Health Organisation's (WHO) classification of CNS tumours<sup>2</sup>. H3K27M mutation, an independent adverse predictor of overall survival (OS)<sup>3</sup>, was initially reported in diffuse intrinsic pontine gliomas (DIPGs) arising in the brainstems of children<sup>4</sup>. This mutation has further been detected in gliomas of midline structures, such as the thalamus, brainstem, and spinal cord, with a short OS of 11–15 months<sup>4,5</sup>. In addition to conventional surgery, chemotherapy, and radiotherapy, there are a number of clinical trials, including ONC201<sup>6</sup>, validating the efficacy of targeted therapies for H3K27M mutations<sup>7</sup>. Accurate genetic diagnosis is a prerequisite for individualised comprehensive treatment, which often relies on surgical biopsy. However, due to the deep location of the tumour and

proximity to eloquent areas such as midbrain<sup>8,9</sup>, surgery is extremely risky, with high mortality and morbidity<sup>10</sup>. Thus, non-invasive prediction of H3K27M mutations is promising to avoid the enormous risks associated with surgery and to directly guide individualised treatment options, such as targeted therapy.

In previous studies, magnetic resonance imaging (MRI) was regarded as key to establishing a diagnosis for midline glioma, defining the extent of the tumour, and guiding biopsy when indicated. These works integrated artificial intelligence (AI) methods, typically machine learning on radiomic features extracted from MRI, to predict H3K27M mutation status with relatively good performance<sup>11</sup>. Nevertheless, the enhanced signal on post-contrast T1 images, which relies on blood–brain barrier intact, and the high signal area in T2-FLAIR may not be precise enough to delineate tumour infiltration. This is essential because the characteristics derived from MRI were the indirect outcomes instead of hallmarks in molecular biological

<sup>1</sup>Department of Neurosurgery, Huashan Hospital, Fudan University, Neurosurgical Institute of Fudan University, Shanghai, China. <sup>2</sup>National Center for Neurological Disorders, Shanghai, China. <sup>3</sup>Shanghai Clinical Medical Center of Neurosurgery, Shanghai Key Laboratory of Brain Function Restoration and Neural Regeneration, Shanghai, China. <sup>4</sup>Department of Nuclear Medicine & PET Center, Huashan Hospital, Fudan University, Shanghai, China. <sup>5</sup>Shanghai Artificial Intelligence Laboratory, Shanghai, China. <sup>6</sup>MR Collaboration, Siemens Healthineers Ltd., Shanghai, China. <sup>7</sup>School of Education, Guangzhou University, Guangzhou, China. <sup>8</sup>Department of Neurosurgery, Fudan University Huashan Hospital Fujian Campus, Fujian Medical University, The First Affiliated Hospital Binhai Campus, National Regional Medical Center, Fuzhou, Fujian, China. <sup>9</sup>Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, 200433, China. <sup>10</sup>These authors contributed equally: Yifan Yuan, Guanglei Li, Shuhao Mei. ✉ e-mail: [Yueqi1989@126.com](mailto:Yueqi1989@126.com); [liumianxin@pmlab.org.cn](mailto:liumianxin@pmlab.org.cn); [maoying@fudan.edu.cn](mailto:maoying@fudan.edu.cn)

changes of tumours. In addition, due to the anatomic structure of midline glioma, MR signal could be interfered with by skull-base bone and artery, which may further compromise the prediction effect. Therefore, additional non-invasive imaging biomarkers are still needed that could contribute to disease prediction and prognosis and guide the treatment of patients with midline gliomas. Amino acid positron emission tomography (PET), including  $^{11}\text{C}$ -methionine ( $^{11}\text{C}$ -MET, MET in short) and  $^{18}\text{F}$ -fluoroethyltyrosine ( $^{18}\text{F}$ -FET, FET in short), reflects the tumour's subjective metabolic profile and thus serves as an optimal technique for evaluating midline glioma. Previous studies have preliminarily revealed that imaging characteristics such as higher tumour–brain ratios, larger metabolic tumour volume and higher total lesion methionine uptake were significantly associated with shorter progression-free survival and OS of midline gliomas<sup>12,13</sup>. A recent study has further suggested that methionine metabolism represents a vulnerability in H3K27M mutant glioma by modulating the epigenome, supported by evidence from enrichment analysis and cell culture analysis<sup>14</sup>. Thus, 11C-MET, a tracer specifically designed for methionine metabolism, may partly establish an underlying biological relationship between H3K27M evaluation and medical imaging prediction. These emerging evidence indicate amino acid PET's promising value in predicting key genes and guiding therapeutic interventions. However, studies combining amino acid PET and AI remain blank, potentially due to limited cases caused by the low incidence rate of midline gliomas and the intricate technical requisites for amino acid PET tracer synthesis. It can be expected that developing advanced AI methods on amino acid PETs, such as deep learning, would significantly help advance research and enhance the diagnostic and therapeutic options for patients with midline gliomas.

In the present study, we proposed a deep learning framework adapted to both FET and MET to predict the H3K27M mutation status of midline gliomas with a high efficacy. To push the boundary of model predictability, we specifically designed a training scheme using the common diagnostic information from both FET and MET, supported by previous evidence on the high similarity between FET and MET<sup>15–17</sup>, to realise a “deep mutual assistant training” between FET- and MET-based model, but without losing their respective specificities. Using multimodal imaging and clinical data from 133 DMG patients, we validated the effectiveness of the proposed method and demonstrated its advantages over conventional methods.

## Methods

### Human study, ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The Ethics Committees of Huashan Hospital and Shanghai Gamma Hospital approved this retrospective study (KY2015-256). All patients or their legal guardians provided written informed consent.

### Study design

Data from an internal cross-validation cohort (90 patients, aged 6–68 years) were examined to build and validate the H3K27M-mutant prediction model. A hold-out testing cohort (19 patients, aged 6–68 years) was then studied to test the established prediction model in out-of-sample unseen data (Fig. 1A). As we used five-fold cross-validation in the internal cohort, the sample size of the testing cohort (19) is roughly equal to the validation data set in the internal cohort (90/5 = 18). Another 3 pathology-uncertain patients were included to demonstrate the clinical feasibility of the proposed method. All the above patients with midline gliomas were recruited from August 2016 to May 2023. To demonstrate the model's prediction capacity under the real-world setting, we further collected a consecutive testing dataset in the same hospital after the model was trained ( $n = 21$ , from June 2023 to June 2024). Patient inclusion criteria were (1) diagnosis of midline glioma based on MRI and (2) no preoperative treatment.

### Imaging acquisition

MRI was performed within 14 days before surgery. Routine clinical sequences, including contrast-enhanced T1-weighted (T1C) images, T2 and T2-FLAIR, were acquired on a 3 T Ingenia MRI scanner (Koninklijke

Philips N.V., Netherlands) for patients recruited on the west campus of Huashan Hospital. A 3 T Verio MRI scanner (Siemens Healthineers, Erlangen, Germany) was used for patients recruited on the main campus of Huashan Hospital, and a 1.5 T Signa HDxt MRI scanner (GE Healthcare, Milwaukee, WI, USA) was used for patients recruited at Shanghai Gamma Hospital.

MET PET and FET PET scans were performed using a Biograph 128 PET/ computed tomography (CT) system (Siemens, Erlangen, Germany). The internal time between an MRI scan and a PET scan is less than 14 days. Before PET, 4 h fasting was required before tracer injection, and CT was performed with a 1.5-mm isotropic spatial resolution for attenuation correction and image fusion. Data acquisition for MET began 10 min after an intravenous bolus injection of 370 MBq (10 mCi) MET. MET PET images were obtained in a 20-min static scan in three-dimensional (3D) acquisition mode. For FET,  $185 \pm 10$  MBq (5 mCi) FET was intravenously injected before the PET/CT scan. A static scan was performed 20 min after the tracer injection and lasted 20 min. Both PET images were reconstructed with filtered back projection (FBP) and full width at a maximum of 3.5 mm using the manufacturer's workstation and post-reconstruction with Gaussian 3D filter smoothing. The standardised uptake value (SUV) was computed as the uptakes divided by the injected dose and the total body weight.

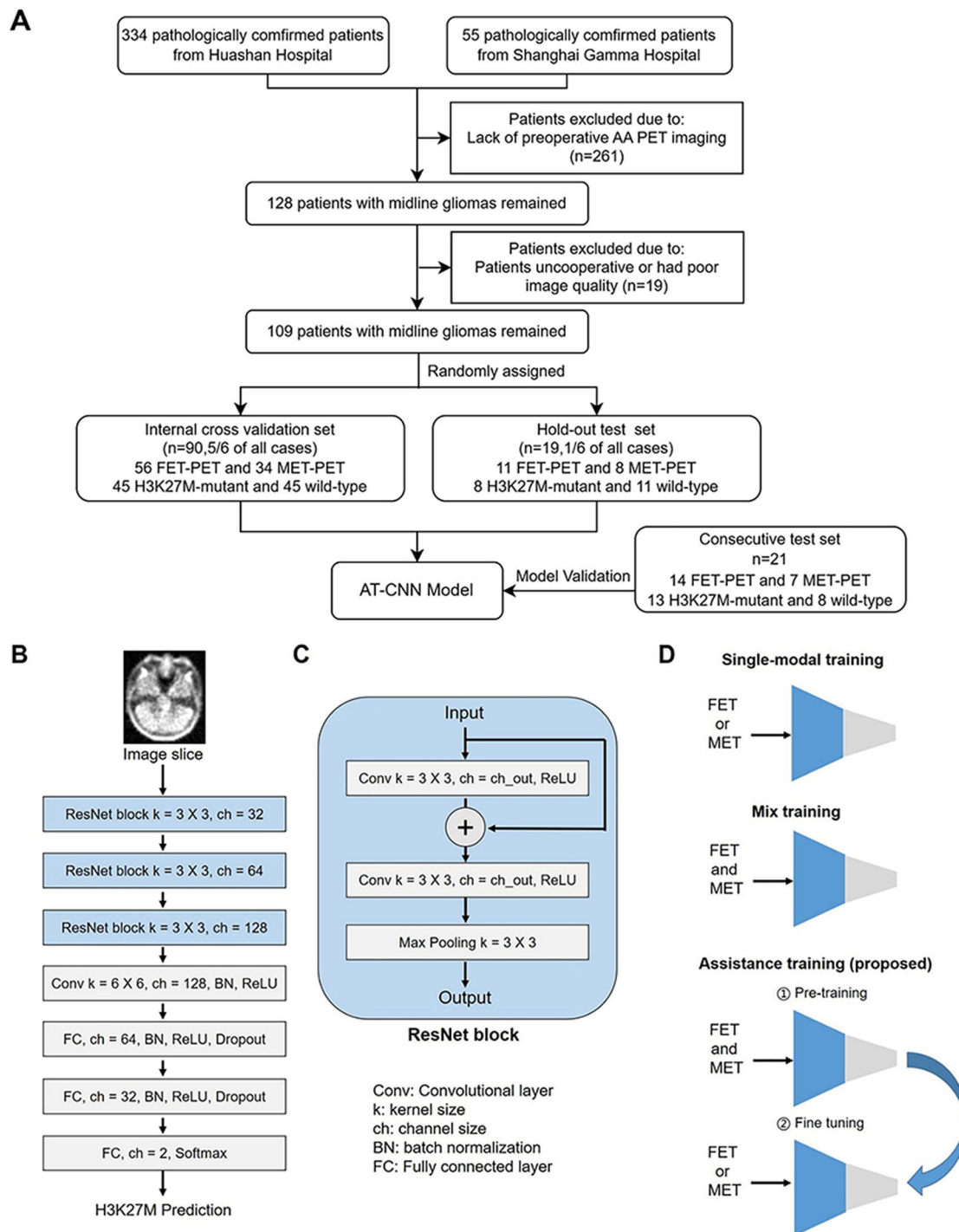
### Pathology and H3K27M status analysis

Tumours were graded according to the 2021 WHO Classification of Tumours of the Central Nervous System. H3K27M and IDH statuses (mutant or wild-type) were determined by immunohistochemical staining using a mutation-specific antibody (Abcam ab190631 for H3K27M and ab172964 for IDH).

### Data preprocessing and feature extraction

For the PET-only experiments, after PET acquisition and reconstruction, we converted the original intensity into the standardised untaken value ratio (SUVr) by dividing the averaged regional value within a large crescent-shaped volume of interest, which was drawn in the normal cerebral hemisphere at the level of the centrum semiovale, including cortical and white matter<sup>12</sup>. The used volume of interest is shown in Supplementary Fig. 1. The principle of implementing SUVr is to achieve an approximated “tumour-to-brain ratio”. Also, it helps the normalisation between two types of PETs (there is a discrepancy in SUV level in two PETs). The T1 template in MNI space was wrapped, using affine registration, to individual space and used as a mask applied to individual PETs to remove extracranial noise. A manually drawn mask for the thalamus and brain stems in the MNI space is co-registered into individual PET spaces along this process. The inter-observer inter-rater reliability of this mask was evaluated according to the Cohen's kappa statistic ( $\kappa = 0.842$ , indicating high reliability. See Supplementary Table 1 for details). The fields of view in different PETs were aligned by image cropping (into  $188 \times 188 \times 148$  pixels). Axial slices containing thalamus and brain stem masks (with region sizes  $>100$  pixels in the whole slice to ensure relevant information) were selected. The slices were further cropped according to the bounding box of the intracranial region and resized to  $188 \times 188$  pixels.

For the MRI-vs-PET comparison, individual PETs (SUVr-converted), T2-FLAIR, and the template T1 brain were registered into individual T1C space to generate alignment in slice numbers and spatial contexts. The PET images then underwent head masking and volume resizing. The MRI images were processed with head masking, intensity normalisation, and volume resizing to perform spatial and intensity normalisation over the multi-centre MR data and align PET and MR. The bias field effect was adjusted using N4 bias correction with ANTS toolbox<sup>18</sup>. In this process, individual T1C, T2-FLAIR, and PET data were resized/resampled to  $188 \times 188 \times 180$  pixels. Intensity normalisation for the MRI images (T1C or T2-FLAIR) used 0-to-100 normalisation over the whole brain. The following slice-based operations were described above.



**Fig. 1 | Illustrations for key methodology components of the study.** **A** The workflow of the patient recruitment and exclusion. **B** Overall data flow in the implemented CNN architecture. **C** Structure of the “ResNet” block. **D** Schematic demonstrations for single-modal, mixed and the proposed assistance training schemes.

**CNN architecture**

A 2D convolutional neural network (CNN) was implemented to predict H3K27M mutants. The CNN was based on classical ResNet<sup>19</sup> blocks using slices of single-modality data as input; such 2D slice-based training potentially increases the training samples and could be more realistic. When using 3D CNN, a 3D image and H3K27M state would constitute a pair of samples, so that the number of samples available for training would significantly decrease (to the same number as the number of patients). Therefore, it is more reasonable to pair the 2D slices with the subject’s gene status to

constitute multiple pairs of samples and train a capable 2D CNN first and then integrate the slice-based decisions into an individual’s diagnosis. Figure 1B, C shows the architecture and parameter settings.

We trained the CNN using single-modality data, with training epochs: 15, learning rate: 0.001, and batch size: 64. We used the Adam algorithm to automatically optimise the trainable parameters, with a weight decay of 0.005. Because the sample size was slightly imbalanced between groups, weighted cross-entropy was used as the loss function, with weights adaptively configured using the inversed class ratio within

the training set. Other parameters of the neural network models were initialised with random weights. The training was accelerated with an Nvidia GTX 1660 GPU. During the training, we implemented random translation, rotation, horizontal flipping and vertical flipping of the inputted image to increase the model's generalisability.

### Assistance training scheme

We further designed an assistance training (AT) scheme to allow mutual assistance between MET and FET data (Fig. 1D) to build a better single-modal PET-based CNN model. AT consists of mixed training using both PET data and a fine-tuning using specific type of PET data. The advantage of this training design is theoretically guaranteed by MET-FET correlation and transfer learning. The existence of the common diagnostic information from both FET and MET supports that we can utilise the knowledge from FET (encoded inside the model parameters) to facilitate the learning of MET and vice versa. Specifically, FET samples could play as pseudo-MET samples and could assist the training for the MET-based model during the joint pre-training by increasing the learning sample size and improving the learning of the robust common diagnostic information. The additional fine-tuning could further transfer the optimal common model into a strong modality-specific model. In a similar rationale, MET could benefit from the FET-based model training.

We used two training stages. In the first stage, we trained the model on a mixed dataset from both FET and MET, expecting the model to first learn common and generalisable features in augmented samples with both MET and FET data. In the second stage, the model should adapt to one targeted PET type by enhancing the learning of specific features, and the pre-learned common features should help the model perform better than single-modal training only (see Fig. 5). Operationally, we first trained the CNN model using both MET and FET data from the training set using a learning rate of 0.001 in the first 8 epochs, then fine-tuned the model using a targeted PET type with a learning rate of 0.001 for another 7 epochs. The weights for cross-entropy loss were estimated only in the targeted PET and kept the same throughout the training for the smoothness of the supervision signal. Using only the first stage would not yield optimal predictions because heterogeneity between MET and FET hinders the best fit in the targeted modality (Figs. 1D and 3. "mixed training").

### Radiomics

The open-source toolbox "Pyradiomics" (<https://pyradiomics.readthedocs.io/en/latest/>)<sup>20</sup> was used to compute radiomics features from the FET and MET data. Note that the shape features were not extracted, as a tumour annotation was not needed for our proposed methodology and thus, we will compare all the radiomics methods under the same condition for reasonable comparisons. The other features were computed within the thalamus and brain stem masks in each slice. All other features automatically provided in the toolbox were implemented, resulting in 93 features per slide. The extracted features were fed into a conventional machine-learning classifier. Results from three methods are presented in the main text: (1) a logistic regression (LR) model with ElasticNet penalty for classification, which combines L1 and L2 penalties as the loss function (ratio: 0.5); (2) a non-linear support vector machine (SVM) with radial basis function or polynomial kernel; and (3) the random forest (RF) method, which represents ensemble learning. We set the number of trees in the forest to 100. In this setting, LR and RF performed automatic feature selections during the model optimisation (using sparseness constraint and Gini impurity, respectively). The number of selected features is thus adaptively changing. The best-performing results are reported. All three machine-learning methods were implemented with the open-source "scikit-learn" toolbox<sup>21</sup>. To address the imbalanced sample size between classes, weights were automatically set on the classes according to the inverse proportion of class frequencies in the training data.

### Internal cross-validation and hold-on testing using PET

For the internal cross-validation experiment, we used individual-level fivefold cross-validation to build and validate the prediction model. Thirty-

four participants with MET (1517 slices, 17 mutants) and 58 with FET (2638 slices, 28 mutants) were evaluated. The available participants were randomly separated into five equal folds, and the slices from their imaging data were correspondingly assigned to the same fold (for MET and FET). Slices from four folds were used as the training set; slices from the remaining fold were used as the validation set. The roles of the folds were switched until all folds were used once as the validation set. This cross-validation design prevents intra-individual correlation in slices from the same participant from artificially boosting the performance. In this experiment, we were using FET and MET from roughly 72 patients (80% of the 90 patients) for the pre-training stage of AT.

The hold-on testing cohort only involved the PET data and partially assessed the out-of-sample generalisability of the methods in the internal cross-validation set. We included eight participants (363 slices, 3 mutants) with MET and 11 (505 slices, 5 mutants) with FET. Models of different methods were retrained based on all the data inside the internal cross-validation dataset (90 participants). The pre-training of AT therefore involved all 90 patients.

### MRI-vs-PET comparison

We compared the single-modal CNN methods working on enhanced T1 contrast (T1C), T2-FLAIR, or PET images, from participants with all three data modalities. Among all available data, by matching PET with MRI data, we yielded 3805 slices from 47 participants with FET (22 mutants) and 2076 slices from 26 participants with MET (17 mutants). Because the MET dataset contained fewer participants, we performed only individual-level fivefold cross-validation on the FET dataset. The data in the MET dataset were used as assistance for the FET-based training process. Here, we included FET from around 38 patients (80% of 47) MET from 26 participants at the pre-training of AT. To obtain comparable results, we also used the T1C and T2-FLAIR data in the MET dataset as augmented data (from 26 participants) for the T1C-based and T2-FLAIR-based training.

### Consecutive testing

In the consecutive test dataset, we directly applied the final model trained using all data in the internal cross-validation dataset without any further tuning, which proposes a higher requirement on the model's reliability. In the consecutive testing dataset, 7 MET (336 slices, 6 mutants) and 14 FET (596 slices, 7 mutants) data were included. We only investigated AT-CNN in the consecutive testing dataset, as the method comparison has been comprehensively conducted with the above-mentioned experiments. In addition, we only focused on the individual-level prediction performances, which match the realistic requirement.

The PET scanning variation could be often observed across different scanners and collection sites. During the FET image collection, we found that different reconstruction protocols were applied at the hospital, which offers us a chance to investigate the reliability of our model and the effect of the variations in sampling parameters. Five data (246 slices, 3 mutants) were reconstructed using FBP as mentioned above, and 9 data (350 slices, 4 mutants) used ordered subset expectation maximisation (OSEM).

### Validation metrics

For all experiments, we assessed the prediction performance at both the slice and individual levels, expect the consecutive testing cohort. At the slice level, for slice-based predictions from each round, we computed four metrics from different aspects to evaluate the performance: accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). The accuracy, sensitivity, and specificity are computed under the probability threshold of 0.5.

To obtain the individual-level metrics, we integrated the predictions for slices from the same participant. The individual-level prediction probability was given by the averaged prediction probability among all slices for each class and the final prediction corresponding to the class with the probability being larger than 50%. Because of the limited sample size, we integrated the individual-level predictions from each fold to compute the metrics on

**Table 1 | Participants' demographic and imaging features**

	Whole dataset	H3K27Mmut	H3K27Mwt	MET-PET	FET-PET
No. of patients	109	53	56	42	67
Age					
Range	6–68	6–67	16–68	7–67	6–68
Mean	37.8	34.5	40.9	39.7	36.6
Gender					
Female	52(47.71%)	27(50.94%)	25(44.64%)	20(47.62%)	32(47.76%)
Male	57(52.29%)	26(49.06%)	31(55.36%)	22(52.38%)	35(52.24%)
Tumour location					
Thalamus	41(37.61%)	19(35.85%)	22(39.29%)	15(35.71%)	26(38.81%)
Midbrain	1(0.92%)	0(0.00%)	1(1.79%)	0(0.00%)	1(1.49%)
Pons	37(33.94%)	16(30.19%)	21(37.50%)	12(28.57%)	25(37.31%)
Medulla	8(7.34%)	5(9.43%)	3(5.36%)	7(16.67%)	1(1.49%)
Multiple Regions	22(20.18%)	13(24.53%)	9(16.07%)	8(19.05%)	14(20.90%)
H3K27M status					
H3 K27M <sub>mut</sub>	53(48.62%)	N/A	N/A	20(47.62%)	33(49.25%)
H3 K27M <sub>wt</sub>	56(51.38%)	N/A	N/A	22(52.38%)	34(50.75%)
IDH Status					
IDH mutant	12(11.01%)	0(0.00%)	12(21.43%)	6(14.29%)	6(8.96%)
IDH wildtype	97(88.99%)	53(100.00%)	44(78.57)	36(85.71%)	61(91.04%)
MR imaging features					
Enhanced	43(39.45%)	22(41.51%)	21(37.50%)	17(40.48%)	26(38.81%)
Nonenhanced	66(60.55%)	31(58.49%)	35(62.50%)	25(59.52%)	41(61.19%)

predictions for all participants during internal cross-validation. Otherwise, the small sample size would artificially induce large variations in estimated metrics across the folds.

We did not respectively compute individual-level AUCs for hold-out and consecutive testing cohorts for MET and FET because the limited sample size could bias the AUC estimations<sup>22</sup>. Alternatively, we computed the overall individual-level metrics by combining the predictions from subjects as many as possible in the testing cohorts (using both FET and MET data, or both FBP and OSEM data) for a reasonable and comprehensive quantification. In the consecutive testing cohort, we did not compute the slice-level prediction performance.

### Feature analysis for AT-CNN

We implemented the Grad-CAM (class activation map) method to explore the explainable feature hidden in the CNN models. The toolbox is publicly available at <https://github.com/jacobgil/pytorch-grad-cam>. Briefly, the Grad-CAM tracks the gradient of the prediction probability for the mutant class with respect to the feature maps in intermediate layers of CNN. The gradient can be interpreted as the amount of influence of the intermediate feature on the outputted decision-making and thus can be regarded as the importance of each feature element. The Grad-CAM map was computed by the product between the gradient map and the feature map at individual data. We derived the Grad-CAM map from accurately predicted mutant slices. The Grad-CAM map was further smoothed using “aug-smooth” and “eigen-smooth” configurations. Due to the variations in tumour locations, we presented the typical cases rather than a group-level result.

### Statistical analysis

For slice-level metrics in the internal cross-validation and MR-vs-PET comparison, the mean and standard deviation of the performance metrics from the cross-validation were calculated. These metrics were compared among the methods using one-sided two-sample *t*-tests. For individual-level metrics and testing cohort results, the metrics to be compared were values without variance, as all folds were integrated to guarantee the sample size or only one-time

testing was conducted. To further compare the individual-level metrics, we used a one-sided Delong’s test<sup>23</sup> to test the improvement of our proposed method over other methods in terms of the AUCs. Since we aimed only to assess the significance of the improvements based on *t*-tests and Delong’s test, significant results ( $P < 0.05$ ) may not appear due to a large variance in the metrics from the compared method, although the proposed method showed great improvements. To test whether the AUC is higher than the chance level, a permutation test was implemented. The distribution of chance-level is estimated with 1000 times permutation on the label. And *P*-value is the probability that the real AUC is higher than the chance-level AUCs.

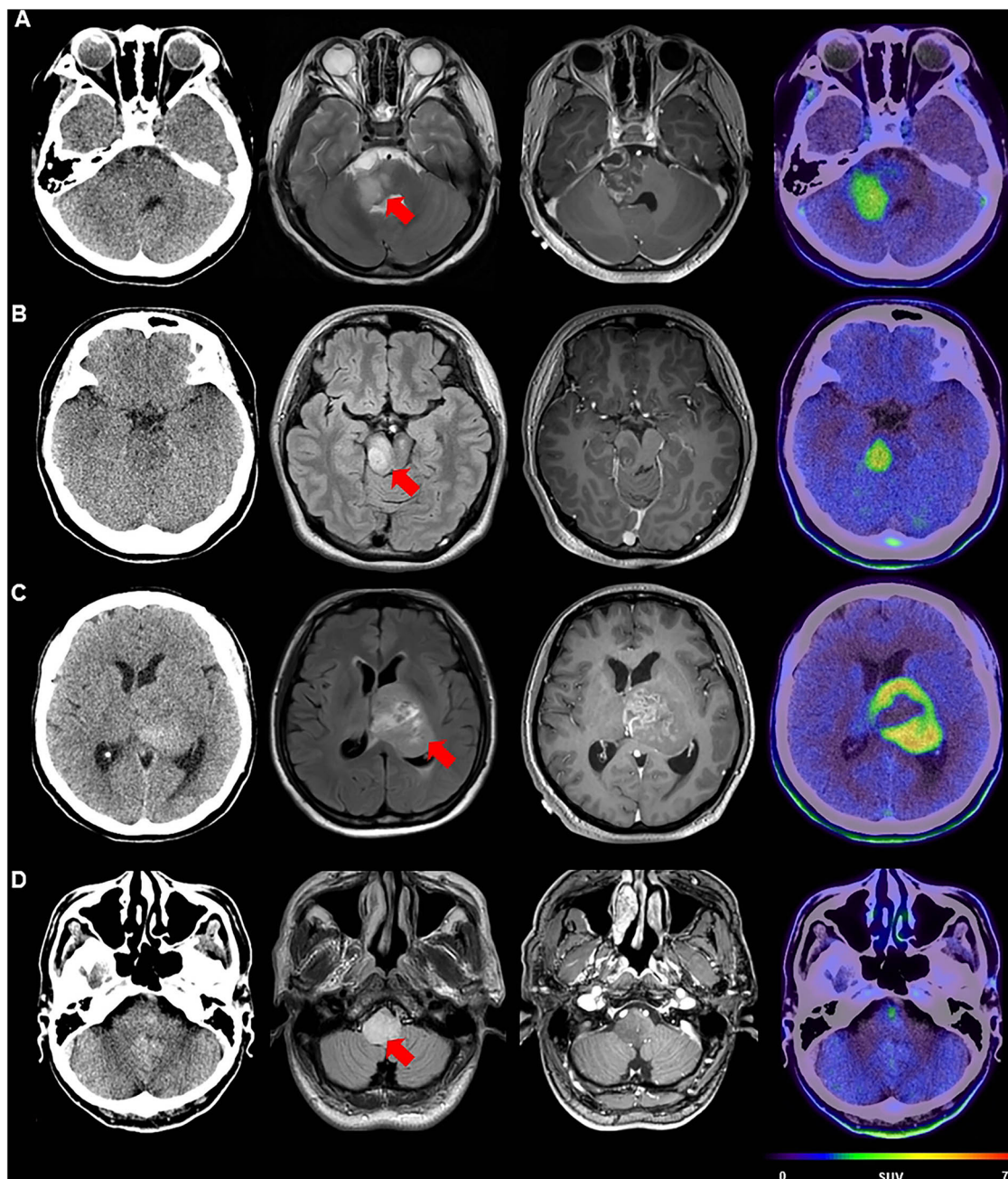
## Results

### Patient demographics and imaging characteristics

We enrolled 389 patients with midline gliomas (Fig. 1A); 280 were excluded for a lack of amino acid PET ( $n = 261$ ) or poor imaging quality ( $n = 19$ ). Finally, 53 patients with H3K27M-mutant DMGs ( $34.5 \pm 16.2$  years old; 27 women, 26 men) and 56 with wild-type gliomas ( $40.9 \pm 16.2$  years old; 31 men, 25 women) were consecutively enrolled. Patients were randomly separated into the internal cross-validation ( $n = 90$ ) and test ( $n = 19$ ) cohorts via computer-generated random numbers.

More than one-third (19/53, 35.85%) of the H3K27M-mutant (denoted as “H3K27M<sub>mut</sub>”) DMGs were in the thalamus, followed by the pons (16/53, 30.19%) and multiple regions (13/53, 24.53%), with similar distribution frequencies in wild-type (denoted as “H3K27M<sub>wt</sub>”) midline gliomas. According to the 2021 WHO central nervous system oncology classification, H3K27M<sub>mut</sub> midline gliomas were totally diagnosed as DMGs, WHO grade 4, whereas only half (28/56) of the wild-type midline gliomas were WHO grade 4, with two WHO grade 3 cases and 26 low-grade gliomas. IDH mutation was detected in 21.43% (12/56) of the H3K27M<sub>wt</sub> gliomas, but not in the DMGs (Table 1).

No significant statistical difference was found in the relationship between H3K27M status and conventional PET statistics, including the max SUV within the tumour region, SUV<sub>max</sub> (MET:  $3.81 \pm 0.99$  in H3K27M<sub>mut</sub> and  $3.70 \pm 2.09$  in H3K27M<sub>wt</sub>, two-sample *t*-test,  $P = 0.84$ ; FET:  $3.36 \pm 1.41$



**Fig. 2 | Representative images of H3K27M-mutant and wild-type midline gliomas (from left to right; images are CT, T2-Flair, gadolinium-enhanced T1 and amino acid PET: 18F-FET in A and C; 11C-MET in B and D). A** Nine-year-old boy with diffuse midline glioma, H3K27M<sub>mut</sub>; **B** 21-year-old man with diffuse

midline glioma, H3K27M<sub>mut</sub>; **C** 39-year-old woman with left thalamus glioblastoma, WHO Grade 4, H3K27M<sub>wt</sub>; **D** 55-year-old man with medulla astrocytoma, WHO Grade 2, IDH<sub>mut</sub>, H3K27M<sub>wt</sub>. The red arrow indicates the location of the tumour.

in H3K27M<sub>mut</sub> and  $2.98 \pm 1.53$  in H3K27M<sub>wt</sub>, two-sample *t*-test,  $P = 0.31$ ) and the maximum tumour to brain ratio (MET:  $3.82 \pm 1.25$  in H3K27M<sub>mut</sub> and  $2.86 \pm 1.05$  in H3K27M<sub>wt</sub>, two-sample *t*-test,  $P = 0.09$ ; FET:  $3.39 \pm 1.45$  in H3K27M<sub>mut</sub> and  $2.98 \pm 1.73$  in H3K27M<sub>wt</sub>, two-sample *t*-test,  $P = 0.31$ ). Representative images of H3K27M mutant and wild-type cases were demonstrated in Fig. 2.

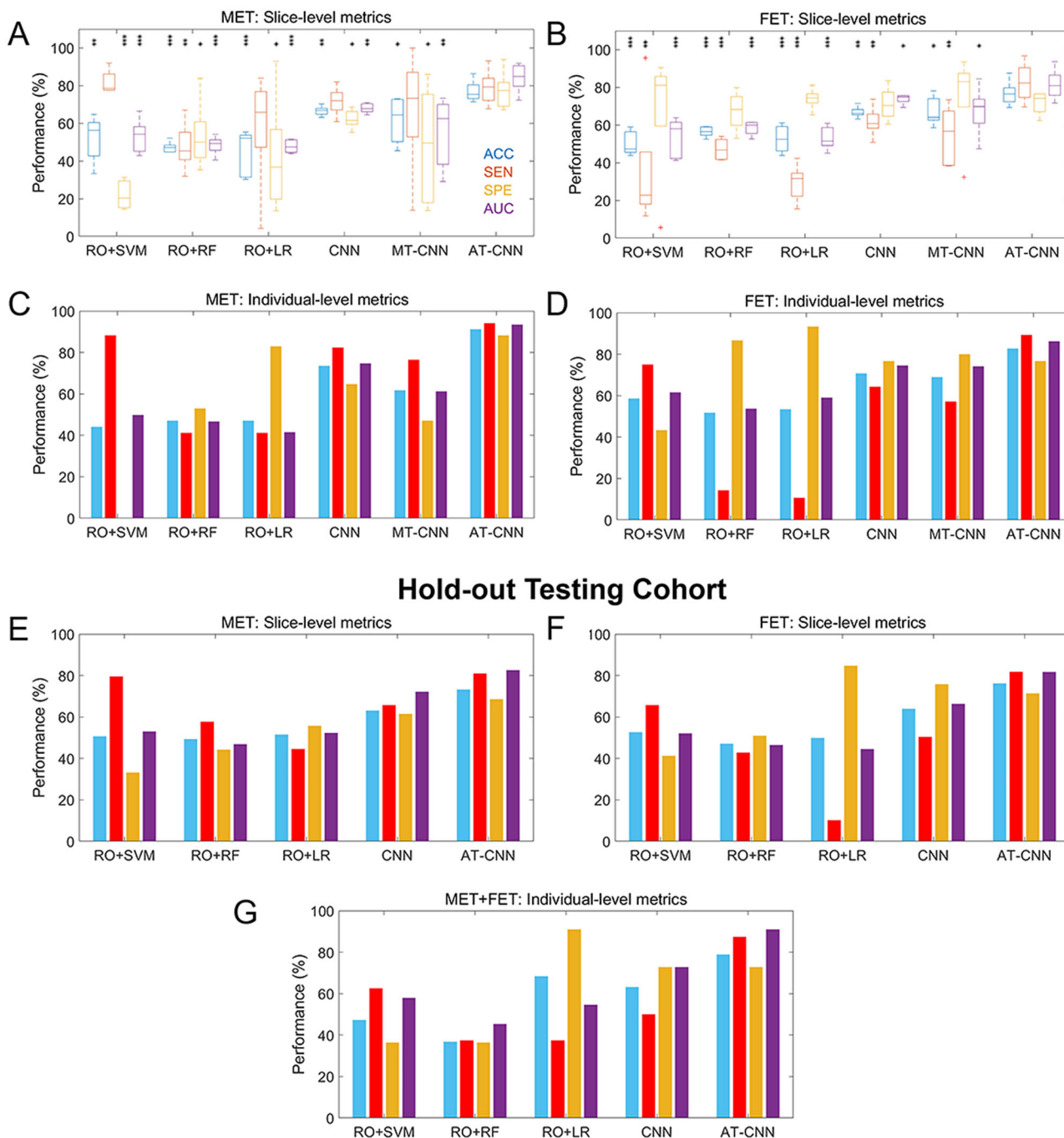
**Prediction results for internal cross-validation cohort**

In the internal cross-validation cohort, we tested on the radiomics-based (with machine learning classifiers, support vector machine [SVM], random forest [RF], and logistic regression [LR]) and CNN-based methods using MET and FET separately as inputs (Fig. 3A–D and Supplementary Table 2). First, for both MET and FET, all radiomics-

based methods exhibited worse predictability than the CNN-based methods. The AUCs from the radiomics-based methods ranged from 0.4781–0.5836 at the slice level and 0.4152–0.6155 at the individual level. The AUCs lower than 0.5 could be attributed to overfitting issues or insufficiency of extracting useful information when using these methods (we have tried other configurations or parameters while the results remain similar). In comparison, the CNN yielded an AUC of  $0.6803 \pm 0.0236$  using MET and  $0.7384 \pm 0.0238$  using FET at the slice level, while an AUC of 0.7474 for MET and 0.7452 for FET at the individual level. The restricted predictability of CNN can be attributed to the limited training samples.

Former studies evidenced that MET and FET imaging potentially share features in imaging manifestations<sup>15–17</sup>. This potentially provides a chance to

### Internal Cross-Validation Cohort



**Fig. 3 | Results in internal cross-validation and hold-out testing cohort from the different methods. A–D** Internal cross-validation and **E–G** Hold-out testing cohort. **A, B** Boxplots for the slice-level results from MET and FET, respectively. **C, D** Bar plots for the individual-level results from MET and FET, respectively. **E, F** Corresponding slice-level results for **A, B** in the hold-out testing cohort, as bar plots. **G** Bar plots for the individual-level results from combined MET and FET data

in the hold-out testing cohort. \* Results from CNN with assistance training (AT-CNN) were significantly higher than those of the indicated method at  $P < 0.05$ . \*\*  $P < 0.01$ . \*\*\*  $P < 0.001$ . RO radiomics, SVM support vector machine, RF random forest, LR logistic regression, MT mixed training, AT assistance training. Detailed performance values are provided in Supplementary Table 2 for the internal cross-validation cohort and in Supplementary Table 3 for the hold-out testing cohort.

increase the training samples by using both MET and FET. We thus tested this by building a model trained on mixed data from both MET and FET and letting the model make predictions using either MET or FET (corresponding results denoted as “MT-CNN”, Fig. 3A–D and Supplementary Table 2). We combined the training and used validating splits implemented in the respective FET and MET experiments so that the comparison could be

fair. This model, simultaneously using information from MET and FET, generated satisfactory predictions, which were only supported by nearly 60–70% accuracy and AUCs at the individual level, which is higher than the theoretical chance level (i.e.  $AUC = 0.5$ ). The results of this mixed model verified the existence of common information between MET and FET. However, compared with the results from CNNs trained on single PET

modalities, the mixed model performed poorly. This may be because a model trained on both MET and FET could better learn the common and generalisable features (combining MET and FET led to learning more samples) for predicting H3K27M mutations, but some modality-specific features were ignored. In addition, it can be observed that the results tend to be biased to FET, where the samples could be more sufficient. Therefore, we further designed an assistance training scheme to allow the model to first learn the common features using both modalities and then adapt to specific PET types by further learning the targeted modality (“AT-CNN” in Fig. 3A–D and Supplementary Table 2). The AT-CNN provides improved predictability than that of the originals without AT (“CNN” and “MT-CNN”). At the slice level, the AUCs were significantly improved ( $P = 0.0065$  for MET,  $P = 0.04$  for FET). At the individual level, the AT-CNN reached 91.18% accuracy for MET and 82.76% accuracy for FET, yielding increases of 17.7% and 12.1%, respectively, compared with that of CNN. The AUCs of AT-CNN were 0.9343 for MET and 0.8619 for FET, which shows different levels of significant improvements when compared to other methods (Fig. 4A, B). These results demonstrate the effectiveness of our proposed AT scheme. In general, the MET-based models could be more benefit from AT than FET-based models, as the usable training sample is remarkably expanded.

### Prediction results for hold-out testing cohort

We used a hold-out testing dataset to verify the internal cross-validation results. The MT-CNN was excluded in this validation because the mixed training results only provide supporting evidence for our proposed AT-CNN method. The slice-level metrics for MET and FET yielded consistent evidence supporting the advantage of our proposed method (Fig. 3E, F and Supplementary Table 3). The radiomics-based methods exhibited poor predictability in most cases, obtaining accuracies and AUCs of around 50%. Conversely, the CNN showed superior performance over all radiomics-based methods, and the AT-CNN further boosted the prediction precision. At the individual level, for a reliable estimation, the overall prediction performance was computed by combining both MET- and FET-based predictions (all 19 patients) from the same method in the external validation (Fig. 3G and Supplementary Table 3). The respective results in MET and FET can be found in Supplementary Table 3. The AT-CNN retained the top position over other methods, achieving 78.95% accuracy, 87.50% sensitivity, 72.73% specificity, and an AUC of 0.9205. This AUC is significantly higher than the chance level with  $P < 0.001$ . Figure 4C shows the corresponding ROC curves, which demonstrate the significant (or nearly significant) advantage of the AT-CNN over all competing methods.

### Prediction results for the MRI-vs-PET comparison

The following analysis compared the PET-based and MR-based H3K27M prediction methods. We selected participants with T1C, T2-FLAIR, and PET data for this experiment. Owing to data availability, the FET dataset was used for the major validation; the MET dataset was used for the AT (see Methods). We restricted the comparison within the single-modal model to test the advantages of PET-based analysis.

The FET-based model demonstrated higher predictability than the MR-based models, supported by slice- and individual-level evidence (Fig. 5 and Supplementary Table 4). At the slice level, using only the training samples from the FET dataset (without AT or augmented data), the T1C-based and T2-FLAIR-based models yielded AUCs of  $0.6276 \pm 0.0494$  and  $0.6253 \pm 0.0948$ , respectively, while the FET-based model yielded an AUC of  $0.7316 \pm 0.1046$ . Implementing data augmentation boosted the performances of the T1C-based and T2-FLAIR-based models by different degrees, but they still failed to outperform the FET-based model with AT. Similar trends occurred at the individual level, where FET-based CNN with and without AT achieved the best

performances when compared to their corresponding versions based on T1C and T2-FLAIR. The final model, “FET AT”, yielded 82.98% accuracy and an AUC of 0.9182, which were significantly higher than the results from MR-based models (with data augments) with increases of approximately 0.20–0.25. The ROC analysis supported the improved AUC exhibited by our method (Fig. 4D).

### Prediction results for consecutive testing cohorts and reliability of sampling protocol

We further applied the established model to a consecutive testing cohort without any tuning, which strictly examines the reliability of the AT-CNN models under a real-world setting (see Methods). The consecutive testing is the most difficult setting among all the presented validations, and it could provide the strongest evidence about our method’s capacity. In this dataset, the AT-CNN achieves an accuracy of 71.43%, a sensitivity of 66.67%, and a specificity of 100.00% on 7 participants with MET scans and an accuracy of 71.43%, a sensitivity of 71.43%, and a specificity of 71.43% on 14 participants with FET scans. Note that the FETs were collected using different reconstruction protocols. Our model yields 80% accuracy, 66.67% sensitivity, and 100.00% specificity on 5 FBP-based FET data and 66.67% accuracy, 75.00% sensitivity, and 60.00% specificity on 9 OSEM-based FET data. Combining all data, we estimated that our model obtains an accuracy of 71.43%, a sensitivity of 69.23%, a specificity of 75.00% and an AUC of 0.7404 ( $P = 0.028$ , indicating predictability being significantly higher than chance level) in the consecutive testing (Fig. 4E).

Note that the predictions for hold-out testing and consecutive testing were essentially performed by the same model, we also did metrics statistics using integrated data of both cohorts ( $n = 40$ ). AT-CNN yields an accuracy of 75.00%, a sensitivity of 76.19%, a specificity of 73.68% and an AUC of 0.8170 ( $P < 0.001$ ) under this setting (Fig. 4F).

### Clinical feasibility test of AT-CNN

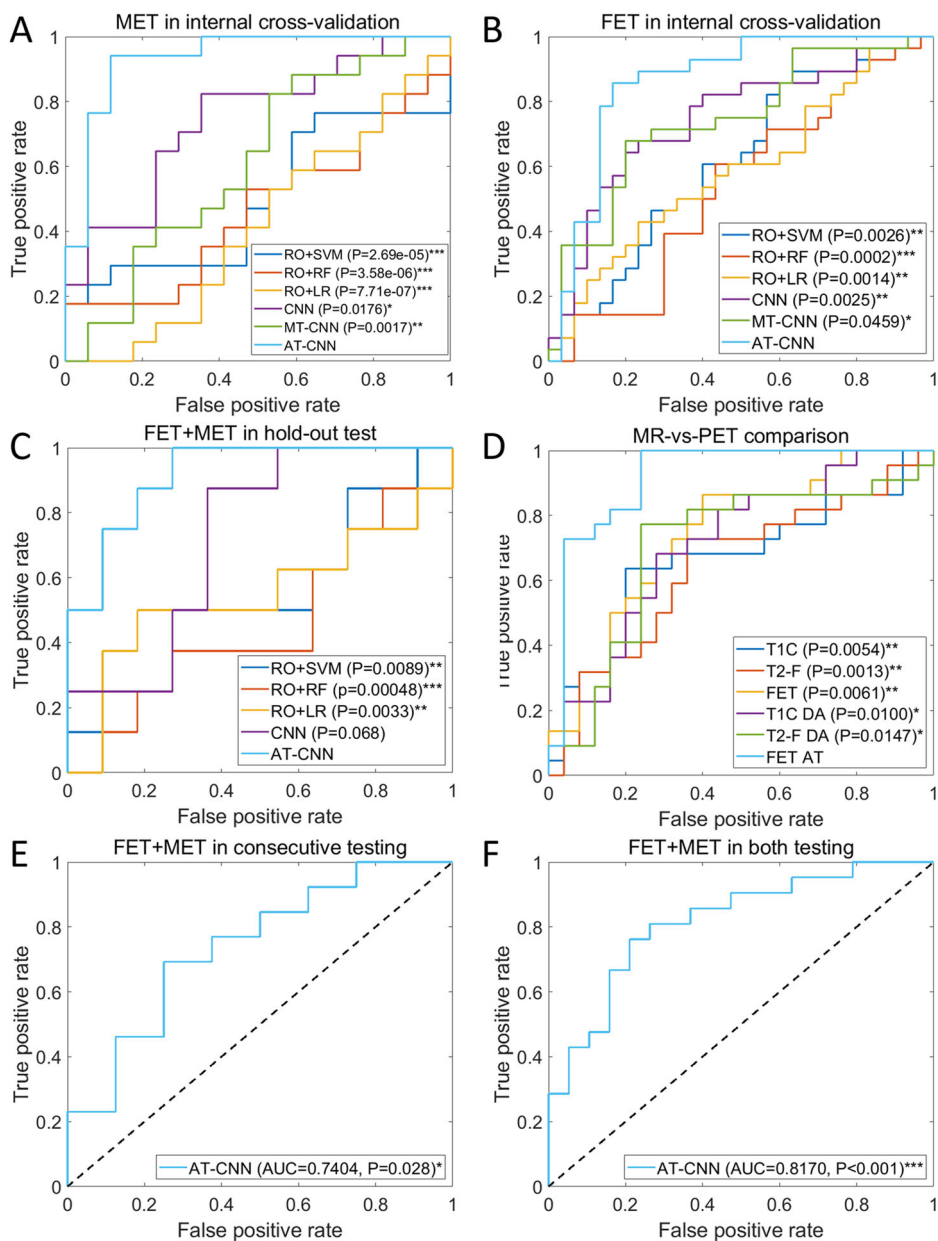
Here, we presented three cases (all FET data) to test the clinical feasibility and demonstrate the application of AT-CNN. These patients were pathology-uncertain cases, and comprehensive treatment decisions were derived by multiple-department treatment (MDT); we expected the AT-CNN can offer aligned guidance (Fig. 6). In this analysis, we kept using the same fixed model as above without tuning. In the following analysis, we used a cut-off probability of 60.92%, informed by the ROC curve in Fig. 4C.

The first case is a 59-year-old female with a lesion located in the right thalamus. Due to tumour heterogeneity, postoperative pathology revealed glial proliferation with nuclear atypia. After the MDT discussion, midline glioma (H3K27M<sub>wt</sub>) was first considered; the patient opted for follow-up observation. Independently, AT-CNN gave a negative prediction (45.76% probability of supporting H3K27M<sub>mut</sub>). Currently, overall survival (OS) exceeds 40 months, and the latest MRI examination shows the lesion has progressed to both thalami, confirming the prediction of H3K27M<sub>wt</sub> glioma.

The second case is a 35-year-old male with a mass in the left midbrain, pons, and cerebellum. Considering the diffuse high signal on T2-FLAIR, the surgeon chose the cerebellum as a relatively safe target for biopsy. Postoperative pathology indicated cortical tissue with cellular atypia, tending to Grade 2 astrocytoma, however DIPG cannot be excluded. The AT-CNN predicted the probability of the tumour being an H3K27M mutant type as 79.65%, aligned with the MDT discussion. Currently, the patient is receiving postoperative radiotherapy and chemotherapy. The OS was less than 13 months, being consistent with the epidemiology of DMG.

The third case is a 7-year-old child presented with hydrocephalus, and the lesion was located in the pons and left middle cerebellar peduncle. After MDT discussion and extensive communication with the child’s parents, non-invasive treatments, including radiotherapy and chemotherapy, were chosen. AT-CNN indicated the patient has an

**Fig. 4 | Receiver operating characteristic (ROC) curves of individual-level predictions from the different methods and experiments.** A, B ROC curves and Delong's test results for the internal cross-validation cohort using MET and FET. C Overall (MET and FET combined) results for the hold-out testing cohort. D MRI-vs-PET comparison results. P-values from one-sided Delong's test were stated in the figure legends. \* $P < 0.05$ . \*\* $P < 0.01$ . \*\*\* $P < 0.001$ . E Overall results for consecutive testing cohorts. F Overall results for combined data from hold-out and consecutive testing cohorts. P-values from the permutation test. \* $P < 0.05$ . \*\*\* $P < 0.001$ .



87.83% probability of being H3K27M<sub>mut</sub> DMG, also suggesting a poor prognosis. In the third month after radiotherapy, the symptom was alleviated, and MRI indicated tumour regression. Currently, progression-free survival (PFS) is over 6 months, and the patient is undergoing follow-up.

### Explorations on the potential biological basis

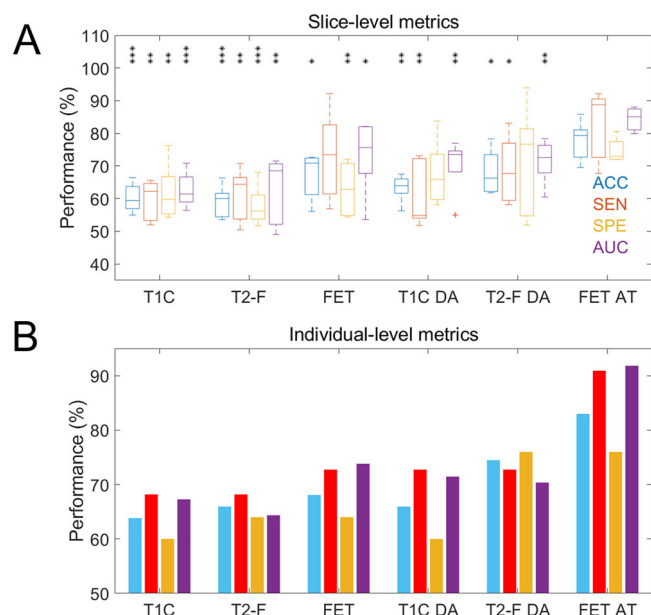
The biological basis underlying the success of prediction from MET and FET to H3K27M status is of broad interest. Using Grad-CAM (see Methods), we preliminarily investigated the predictive features encoded in the deep layers. In Fig. 7, in the exemplified cases, for both FET and MET, the predictive features are located on the high-uptake regions near the brain middle line (the tumour regions). Note that the predictive area covers broader than the original high-uptake region, which may imply that the AT-CNN is also considering the surrounding tissues of tumours.

Step further, based on the transport selectivity of amino acids and the findings in literature<sup>24</sup>, LAT1 (encoded by the SLC7A5 gene) is recognised as the primary transporter for both MET and FET. To

investigate this, we utilised the CPTAC and OpenPBTa databases in Paediatric cBioPortal to examine the mRNA and protein expression levels of SLC7A5 in H3K27M wild-type and mutant samples. In Supplementary Fig. 2, our analysis did not reveal significant differences in the expression levels of SLC7A5 between wild-type and mutant samples. Specifically, the CPTAC database results are WT vs. MT:  $5.05 \pm 1.09$  vs.  $5.30 \pm 0.661$ ,  $P = 0.58$ ,  $n = 15$  vs.  $n = 7$ , for mRNA Expression, and WT vs. MT:  $-0.164 \pm 0.674$  vs.  $-0.058 \pm 0.574$ ,  $P = 0.72$ ,  $n = 18$  vs.  $n = 7$  for protein expression. Similarly, the OpenPBTa database indicates WT vs. MT:  $5.25 \pm 1.03$  vs.  $5.45 \pm 0.883$ ,  $P = 0.22$ ,  $n = 62$  vs.  $n = 66$  in mRNA expression. This negative result could be attributed to the limited sample size and the coarse bulk mRNA and protein expression levels reflecting the average outcomes of the entire tumour tissue sample.

### Discussion

In the presented work, we proposed a non-invasive AT-CNN model to predict the H3K27M mutation status with preoperative amino acid PET, achieving 91.18% accuracy with an AUC of 0.9343 for MET and



**Fig. 5 | MRI-vs-PET experiment results from different configurations.** **A** Boxplots for the slice-level results; **B** bar plots for the individual-level results. \* results with PET-based AT were significantly higher than those of the indicated method at  $P < 0.05$ . \*\* $P < 0.01$ . \*\*\* $P < 0.001$ . T2-F: T2-FLAIR. DA data augmentation, AT assistance training. Detailed performance values are provided in Supplementary Table 4.

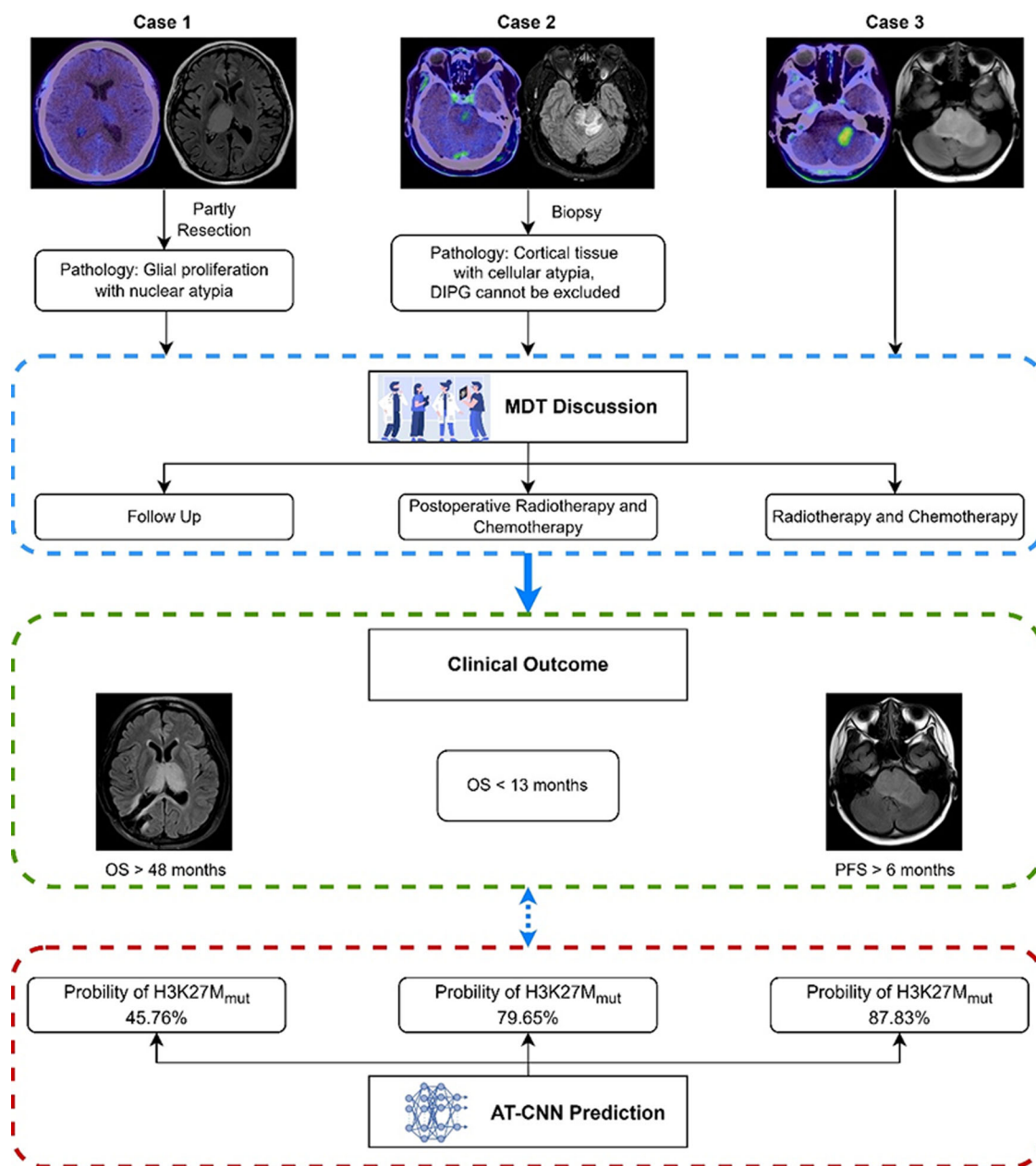
82.76% accuracy with an AUC of 0.8619 for FET in the internal cross-validation cohort. The predictability is robustly high when testing the method on the hold-out test cohort using both PET types (accuracy = 78.95%, sensitivity = 87.50%, specificity = 72.73% and AUC = 0.9205), which shows significant advantages over other competing methods. In addition, the relatively high reliability of our method was directly supported by the results using a consecutive test dataset, containing both types of PET and even data collected with different protocols (accuracy = 71.43%, sensitivity = 69.23%, specificity = 75.00%, and AUC = 0.7404). Finally, using MDT-based decisions and clinical outcomes from three exemplified patients, we demonstrated our model can be clinically feasible. The biological basis underlying the success of our method was preliminarily explored. To the best of our knowledge, our study is the first to establish effective deep-learning models on amino acid PET to accurately predict H3K27M mutation. The method is promising to benefit from accurate pathological diagnosis and identification of key genes as the foundation for the whole treatment for midline gliomas.

Previous studies of non-invasively predicting H3K27M status in patients with midline gliomas were primarily based on radiomics methods and other imaging modalities. Pan et al. recruited 151 patients and implemented RF on radiomics features from pre- and post-contrast T1 and T2 images<sup>25</sup>. They achieved 84.44% accuracy and an AUC of 0.8298 in the test cohort. Su et al. applied a tree-based pipeline optimisation tool on radiomics from 100 T2-FLAIR data points and tested it on 22 independent patients, obtaining an AUC of 0.85<sup>26</sup>. Compared with MRI, MET, and FET, serving as exogenous radiotracers, exhibit the remarkable ability to be actively uptaken by tumour cells. Studies have shown that amino acid PET scans surpassed conventional MRI for tumour delineation, especially in areas outside enhanced regions<sup>27</sup>. To support the advantage of amino acid PET, we performed direct comparisons between radiomics and CNN and between MRI and PET on our data and observed the superiority of CNN- and PET-based methods over radiomics- and MRI-based methods, respectively.

In terms of methodology, our proposed method has several novelties and merits. First, being different from previous radiomics-based methods, which heavily depend on tumour segmentation, our proposed method does not require accurate tumour segmentations or fine tissue registrations. A rough localisation of the brainstem and thalamus for slice selections is sufficient for our method to work with high accuracy. In our testing conditions, our fully automatic procedure, from preprocessing to the prediction generation, can be completed within two minutes for one subject, which could significantly improve the diagnosis efficiency. Second, the AT is proposed based on previous observations of common diagnostic information for FET and MET<sup>15–17</sup>. AT allows mutual assistance between FET and MET during the training process to significantly boost prediction performance. The success of our AT scheme also supports the existence of commonness between FET and MET. Additionally, it uses information from multimodal PET data but does not require input from multimodal PET during the inferences, which fits clinical practices where patients will rarely undergo multiple PET scans. Our AT thus offers practical benefits for applications of MET and FET. Note that the sampling of MET is restricted by its imaging mechanism, as the half-life period of the C marker is quite short compared to the F marker used in FET, so often MET could be applied to fewer patients than FET. Using the AT scheme, the shortage of training samples would be eased as two modality data can be flexibly collected and both can promote the model. A reliable model would be established with AT at a faster pace than conventional methods where only one type of data can be used. To conclude, our method provides a novel and flexible multimodal prediction framework. In principle, more PET data modalities can be included to form a larger dataset for the AT, the model can then be transferred to improve the prediction based on any of the modalities once the common information among the PET data is determined.

Recently, several novel treatment strategies and clinical trials have been conducted in diffuse midline glioma. Among them, dordaviprone (also referred to as ONC201 and TIC10) is a first-in-class, brain penetrant, small molecule of the imipridone family, with reported case study benefits for the treatment of H3K27M<sub>mut</sub> DMG<sup>28</sup>. Other treatments include molecular-based target therapy, immunotherapy<sup>29</sup>, tumour treating field<sup>30</sup> and boron neutron capture therapy<sup>31</sup>. Drugs targeting deregulated H3 mutations, such as GSK-J4, an H3K27M demethylase inhibitor that increases H3K27me3 in vivo, have been tested in humans<sup>32</sup>. These therapies aim to induce tumour cell death, arrest the cell cycle, modulate chromatin expression, and limit differentiation. However, molecular classification of tumours, especially for the key genes like H3K27M status, is inevitable for all these novel treatment strategies. Thus, our proposed non-invasive model to predict H3K27M status with high efficiency and user-friendly is urgent in clinical practice, and may benefit those patients not available for surgery.

Despite the success in the prediction of H3K27M mutant status, the biological basis underlying this success remains a puzzle. Potentially due to the limited sample size and the coarse bulk mRNA and protein expression levels reflecting the average outcomes of the entire tumour tissue sample, we did not detect significant differences in the transporter for MET and FET between wide type and mutant groups. In general, these findings align with our observations that there is no significant difference in max SUV uptake levels between the two patient groups, which is coincident with previous investigations using MET and FET based on conventional uptake metrics<sup>12,24,33</sup>. However, our results suggest that there exist certain changes in the imaging but the changes are not captured by simple uptake level differences. The advantage of using the deep learning method is that they can model more complex features, such as high-order co-variance between uptake changes in different places. As indicated by the Grad-CAM analysis, we detected that the CNN would utilise the abstract feature from the spatial characteristics from the image manifestations of the tumour and its surrounding

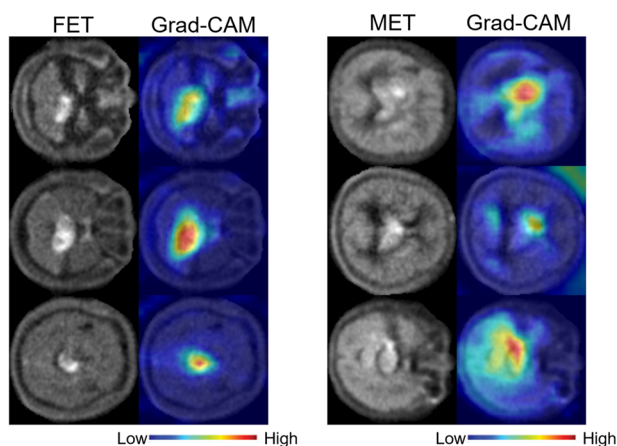


**Fig. 6 | The consistency between AT-CNN prediction and MDT-based decisions.** Three pathology-uncertain exemplary cases of midline glioma are depicted, with the probability of H3K27M mutant from the model prediction, the corresponding MDT-based decision, the received treatment, and current clinical outcomes.

tissues in MET and PET. However, this limited interpretability of CNN could not provide more intuitions on the exact feature and prevents further biologically meaningful explorations. We expect the development of analysing tools for the interpretability of deep learning models, hypothesis-driven analysis, and cell-level experiments would result in a solution to this puzzle.

Our study remains limited in certain aspects. Although we implicitly augmented the data using AT and slice-based learning, the sample size for building the model was still small. With more data, the model and validation results will be more reliable, and a more advanced deep learning module for mining complex and long-range spatial features, such as Transformer<sup>34</sup> may be applied. As the tumour information could be more essential in 3D and the “Pyradiomics” toolbox is designed to be more suitable for 3D inputs, the potential of 3D-based CNN and radiomics methods should also be further explored when the sample size reaches a certain level. Further, our study included both adults and

children, where the PET image characteristics could be different. Though our model satisfactorily adapts to both situations, a detailed exploration of the two subgroups could be needed when the sample size in each age group is large enough. Additionally, the MRI data used in this study were collected from multiple sites, which potentially influences the predictability of MRI-based models, though we have implemented several normalisations. A more systematic comparison could be needed when data is available. Correspondingly, our method’s performances remarkably dropped in the OSEM FET data in the consecutive test. Therefore, one should admit that these exists an effect on the model’s performance from the sampling protocol’s variation. If applying our method to multi-centre data, we anticipate that proper data normalisation or further fine-tuning would be necessary. What is more, we did not explore the model using combined MRI and PET data, which should allow for the construction of a more comprehensive and precise prediction model. Developing methods to optimally combine different data



**Fig. 7 | The predictive features from the deep layer of AT-CNNs for FET and MET, respectively.** The exemplified PET slices (in greyscale) are visualised with the associated Grad-CAM maps (as colour maps).

modalities remains an open and important topic in methodology, and more focused studies are warranted.

In conclusion, we constructed the first deep learning algorithm on amino acid PET to non-invasively predict H3K27M mutations of midline gliomas. The model intelligently combined the strengths of both FET and MET modalities, resulting in sufficiently high diagnostic accuracy that significantly outperformed those models based on MRI or radiomics and other conventional methodological schemes. The model is applicable to clinical scenarios and will promisingly aid in genetic diagnosis, surgical strategy, and adjuvant treatment decision-making for midline gliomas.

### Material availability

Due to the restriction of the hospital, the data is available upon reasonable request from the corresponding authors. The code and model is publicly available via [GitHub](#).

### Data availability

Due to the restriction of the hospital, the data is available upon reasonable request from the corresponding authors. The code and model is publicly available via [GitHub](#).

Received: 27 February 2024; Accepted: 11 November 2024;

Published online: 25 November 2024

### References

- Patil, N. et al. Epidemiology of brainstem high-grade gliomas in children and adolescents in the United States, 2000–2017. *Neurooncol.* **23**, 990–998 (2021).
- Louis, D. N. et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neurooncol.* **23**, 1231–1251 (2021).
- Higginbottom, S. L., Tomaskovic-Crook, E. & Crook, J. M. Considerations for modelling diffuse high-grade gliomas and developing clinically relevant therapies. *Cancer Metast. Rev.* **42**, 507–541 (2023).
- Mackay, A. et al. Integrated molecular meta-analysis of 1,000 pediatric high-grade and diffuse intrinsic pontine glioma. *Cancer Cell* **32**, 520–537 (2017).
- Schulte, J. D. et al. Clinical, radiologic, and genetic characteristics of histone H3 K27M-mutant diffuse midline gliomas in adults. *Neurooncol. Adv.* **2**, vdaa142 (2020).
- Venneti, S. et al. Clinical efficacy of ONC201 in H3K27M-mutant diffuse midline gliomas is driven by disruption of integrated metabolic and epigenetic pathways. *Cancer Discov.* **13**, 2370–2393 (2023).
- Lopez-Perez, C. A. et al. Adult diffuse midline gliomas H3 K27-altered: review of a redefined entity. *J. Neurooncol.* **158**, 369–378 (2022).
- Cohen, K. J., Jabado, N. & Grill, J. Diffuse intrinsic pontine gliomas—current management and new biologic insights. Is there a glimmer of hope? *Neurooncol.* **19**, 1025–1034 (2017).
- Meyronet, D. et al. Characteristics of H3 K27M-mutant gliomas in adults. *Neurooncol.* **19**, 1127–1134 (2017).
- Srikanthan, D. et al. Diffuse intrinsic pontine glioma: current insights and future directions. *Chin. Neurosurg. J.* **7**, 6 (2021).
- Kandemirli, S. G. et al. Machine learning-based multiparametric magnetic resonance imaging radiomics for prediction of H3K27M mutation in midline gliomas. *World Neurosurg.* **151**, e78–e85 (2021).
- Zhao, X. et al. (11)C-methionine PET imaging characteristics in children with diffuse intrinsic pontine gliomas and relationship to survival and H3 K27M mutation status. *Eur. J. Nucl. Med. Mol. Imaging* **50**, 1709–1719 (2023).
- Vettermann, F. J. et al. Dual PET imaging of an H3K27M-mutant glioma with 18F-GE-180 and 18F-FET PET. *Clin. Nucl. Med.* **45**, 992–993 (2020).
- Golbourn, B. J. et al. Loss of MAT2A compromises methionine metabolism and represents a vulnerability in H3K27M mutant glioma by modulating the epigenome. *Nat. Cancer* **3**, 629–648 (2022).
- Grosu, A. L. et al. An interindividual comparison of O-(2-[18F]fluoroethyl)-L-tyrosine (FET)- and L-[methyl-11C]methionine (MET)-PET in patients with brain gliomas and metastases. *Int. J. Radiat. Oncol.* **81**, 1049–1058 (2011).
- Geisler, S. et al. Treatment-related uptake of O-(2-(18F)-fluoroethyl)-L-tyrosine and L-[methyl-(3)H]-methionine after tumor resection in rat glioma models. *J. Nucl. Med.* **60**, 1373–1379 (2019).
- Muthukumar, S., Darden, J., Crowley, J., Witcher, M. & Kiser, J. A comparison of PET tracers in recurrent high-grade gliomas: a systematic review. *Int. J. Mol. Sci.* **24**, 408 (2022).
- Tustison, N. J. et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 770–778 (2016).
- van Griethuysen, J. et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012).
- Hanczar, B. et al. Small-sample precision of ROC-related estimates. *Bioinformatics* **26**, 822–830 (2010).
- DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).
- Tinkle, C. L. et al. Evaluation of (11)C-methionine PET and anatomic MRI associations in diffuse intrinsic pontine glioma. *J. Nucl. Med.* **60**, 312–319 (2019).
- Pan, C. C. et al. A machine learning-based prediction model of H3K27M mutations in brainstem gliomas using conventional MRI and clinical features. *Radiother. Oncol.* **130**, 172–179 (2019).
- Su, X. et al. Automated machine learning based on radiomics features predicts H3 K27M mutation in midline gliomas of the brain. *Neurooncol.* **22**, 393–401 (2020).
- Yuan, Y. et al. Noninvasive delineation of glioma infiltration with combined 7T chemical exchange saturation transfer imaging and MR spectroscopy: a diagnostic accuracy study. *Metabolites* **12**, 901 (2022).
- Jackson, E. R. et al. A review of the anti-tumor potential of current therapeutics targeting the mitochondrial protease ClpP in H3K27-altered, diffuse midline glioma. *Neurooncol.* **26**, S136–S154 (2024).
- Mueller, S. et al. Mass cytometry detects H3.3K27M-specific vaccine responses in diffuse midline glioma. *J. Clin. Invest.* **130**, 6325–6337 (2020).

30. Ibn, E. W. et al. Positioning transclival tumor-treating fields for the treatment of diffuse intrinsic pontine gliomas. *Life—Basel* **13**, 601 (2023).
31. Huang, W. H. et al. Salvage boron neutron capture therapy for pediatric patients with recurrent diffuse midline glioma. *Child Nerv. Syst.* **39**, 1529–1536 (2023).
32. Katagi, H. et al. Radiosensitization by histone H3 demethylase inhibition in diffuse intrinsic pontine glioma. *Clin. Cancer Res.* **25**, 5572–5583 (2019).
33. Lovibond, S., Gewirtz, A. N., Pasquini, L., Krebs, S. & Graham, M. S. The promise of metabolic imaging in diffuse midline glioma. *Neoplasia* **39**, 100896 (2023).
34. Vaswani, A. et al. Attention Is All You Need. In *2017 Advances in Neural Information Processing Systems (NIPS)*. 5998–6008 (2017).

## Acknowledgements

This study is supported in part by National Key Research and Development Program of China, MOST (No. 2023YFC2510000), Noncommunicable Chronic Diseases - National Science and Technology Major Project (No. 2023ZD0511800), Shanghai Science and Technology Commission (No. 20DZ1100800 and 21Y21900600), Shanghai Artificial intelligence laboratory, National Natural Science Foundation of China (No. 82272063, 82127801 and 82227806), Guangzhou Science and Technology Plan Project (No. 202102010495) and Shanghai Hospital Development Centre (Clinical Research Plan of SHDC 2022CRW004).

## Author contributions

Conceptualisation: Y.F.Y., M.X.L., and Y.M.; Methodology: M.T.H. and M.X.L.; Software: Y.H.C., L.C.L., and Y.C.H.; Validation: S.H.M., G.L.L., and J.P.S.; Investigation: Y.F.Y. and G.L.L.; Resources: Q.Y. and F.X.; Data curation: Y.F.Y., S.H.M., and D.Y.F.; Writing—original draft preparation: Y.F.Y., M.X.L., and Q.Y.; Writing—review and editing: M.X.L. and Y.M.; Visualisation: Y.F.Y., F.X., and L.G.L.; Supervision: Y.M.; Funding acquisition: Q.Y., L.X.L., and Y.M.; All authors have read and agreed to the published version of the paper. All authors contributed to the article and approved the submitted version.

## Competing interests

Authors Y.H.C. and Y.C.H. were employed by the company Siemens Healthineers Ltd., China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41698-024-00760-1>.

**Correspondence** and requests for materials should be addressed to Qi Yue, Mianxin Liu or Ying Mao.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024