

OPEN

# Prediction of IDH and TERT promoter mutations in low-grade glioma from magnetic resonance images using a convolutional neural network

Ryohei Fukuma<sup>1,2</sup>, Takufumi Yanagisawa<sup>1,2,3\*</sup>, Manabu Kinoshita<sup>1\*</sup>, Takashi Shinozaki<sup>4,22</sup>, Hideyuki Arita<sup>1,5,6</sup>, Atsushi Kawaguchi<sup>7</sup>, Masamichi Takahashi<sup>8</sup>, Yoshitaka Narita<sup>8</sup>, Yuzo Terakawa<sup>5,9</sup>, Naohiro Tsuyuguchi<sup>5,9,10</sup>, Yoshiko Okita<sup>21</sup>, Masahiro Nonaka<sup>5,11,12</sup>, Shusuke Moriuchi<sup>5,11,13</sup>, Masatoshi Takagaki<sup>1,5</sup>, Yasunori Fujimoto<sup>1,5</sup>, Junya Fukai<sup>5,14</sup>, Shuichi Izumoto<sup>5,10</sup>, Kenichi Ishibashi<sup>5,9</sup>, Yoshikazu Nakajima<sup>5,15</sup>, Tomoko Shofuda<sup>5,16</sup>, Daisuke Kanematsu<sup>5,17</sup>, Ema Yoshioka<sup>5,16</sup>, Yoshinori Kodama<sup>18</sup>, Masayuki Mano<sup>5,19</sup>, Kanji Mori<sup>5,20</sup>, Koichi Ichimura<sup>6</sup>, Yonehiro Kanemura<sup>5,17</sup> & Haruhiko Kishima<sup>1</sup>

Identification of genotypes is crucial for treatment of glioma. Here, we developed a method to predict tumor genotypes using a pretrained convolutional neural network (CNN) from magnetic resonance (MR) images and compared the accuracy to that of a diagnosis based on conventional radiomic features and patient age. Multisite preoperative MR images of 164 patients with grade II/III glioma were grouped by IDH and TERT promoter (pTERT) mutations as follows: (1) IDH wild type, (2) IDH and pTERT co-mutations, (3) IDH mutant and pTERT wild type. We applied a CNN (AlexNet) to four types of MR sequence and obtained the CNN texture features to classify the groups with a linear support vector machine. The classification was also performed using conventional radiomic features and/or patient

<sup>1</sup>Department of Neurosurgery, Graduate School of Medicine, Osaka University, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan. <sup>2</sup>Department of Neuroinformatics, ATR Computational Neuroscience Laboratories, 2-2-2 Hikaridai, Seika-cho, Kyoto, 619-0288, Japan. <sup>3</sup>Institute for Advanced Co-creation studies, Osaka University, 2-2 Yamadaoka, Suita, Osaka, 565-0871, Japan. <sup>4</sup>Center for Information and Neural Networks, National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita, Osaka, 565-0871, Japan. <sup>5</sup>Kansai Molecular Diagnosis Network for CNS Tumors, Osaka, 540-0006, Japan. <sup>6</sup>Division of Brain Tumor Translational Research, National Cancer Center Research Institute, Tokyo, 104-0045, Japan. <sup>7</sup>Education and Research Center for Community Medicine, Faculty of Medicine, Saga University, Saga, 849-8501, Japan. <sup>8</sup>Department of Neurosurgery and Neuro-Oncology, National Cancer Center Hospital, Tokyo, 104-0045, Japan. <sup>9</sup>Department of Neurosurgery, Osaka City General Hospital, Osaka, 534-0021, Japan. <sup>10</sup>Department of Neurosurgery, Kindai University Faculty of Medicine, Sayama, 589-8511, Japan. <sup>11</sup>Department of Neurosurgery, National Hospital Organization Osaka National Hospital, Osaka, 540-0006, Japan. <sup>12</sup>Department of Neurosurgery, Kansai Medical University, Hirakata, 573-1191, Japan. <sup>13</sup>Department of Neurosurgery, Rinku General Medical Center, Izumisano, 598-8577, Japan. <sup>14</sup>Department of Neurological Surgery, Wakayama Medical University School of Medicine, Wakayama, 641-0012, Japan. <sup>15</sup>Department of Neurosurgery, Sakai City Medical Center, Sakai, 593-8304, Japan. <sup>16</sup>Division of Stem Cell Research, Department of Biomedical Research and Innovation, Institute for Clinical Research, National Hospital Organization Osaka National Hospital, Osaka, 540-0006, Japan. <sup>17</sup>Division of Regenerative Medicine, Department of Biomedical Research and Innovation, Institute for Clinical Research, National Hospital Organization Osaka National Hospital, Osaka, 540-0006, Japan. <sup>18</sup>Kobe University Graduate School of Medicine, Department of Diagnostic Pathology, 7-5-1 Kusunoki-cho Chuo-ku, Kobe, 650-0017, Japan. <sup>19</sup>Department of Central Laboratory and Surgical Pathology, National Hospital Organization Osaka National Hospital, Osaka, 540-0006, Japan. <sup>20</sup>Department of Neurosurgery, Kansai Rosai Hospital, Amagasaki, 660-8511, Japan. <sup>21</sup>Department of Neurosurgery, Osaka International Cancer Institute, Osaka Prefectural Hospital Organization, Osaka, 541-8567, Japan. <sup>22</sup>Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka, 565-0871, Japan. \*email: [tyanagisawa@nursurg.med.osaka-u.ac.jp](mailto:tyanagisawa@nursurg.med.osaka-u.ac.jp); [mail@manabukinoshita.com](mailto:mail@manabukinoshita.com)

age. Using all features, we succeeded in classifying patients with an accuracy of 63.1%, which was significantly higher than the accuracy obtained from using either the radiomic features or patient age alone. In particular, prediction of the pTERT mutation was significantly improved by the CNN texture features. In conclusion, the pretrained CNN texture features capture the information of IDH and TERT genotypes in grade II/III gliomas better than the conventional radiomic features.

Identification of IDH1/2 (IDH) mutation and/or TERT promoter (pTERT) mutation is crucial for reaching a correct diagnosis and choosing the most appropriate treatment for patients harboring WHO grade II/III gliomas. In fact, the WHO 2016 brain tumor classification requires both IDH mutation status and 1p/19q chromosomal codeletion status to determine the diagnosis of this tumor<sup>1–4</sup>; 1p/19q chromosomal codeletion status can in some ways be considered equivalent to pTERT mutation status<sup>1,3,5–7</sup>. Although determination of the molecular characteristics of these neoplasms prior to surgical intervention is thought to be beneficial for patient care, this information can currently be obtained only after surgically removing the tissue. Several previous studies challenged this issue by pursuing preoperative molecular characterization of WHO grade II/III gliomas using magnetic resonance imaging (MRI) and radiomic analysis<sup>8–12</sup>. This scientific objective was further pursued by attempting to boost the diagnostic accuracy through machine learning algorithms such as a deep neural network (DNN)<sup>13,14</sup> or convolutional neural network (CNN)<sup>15,16</sup>. Indeed, a recent study by Chang *et al.* applied DNN to MR images of WHO grade II to IV gliomas and demonstrated that IDH1/2 mutations can be successfully classified from MR images<sup>17</sup>. However, the question remains as to whether a DNN or CNN can enhance diagnostic accuracy in the radiogenomics of gliomas compared to conventional radiomic analysis. The effect on the diagnostic accuracy of adding clinical information, such as patient age, is also not completely understood.

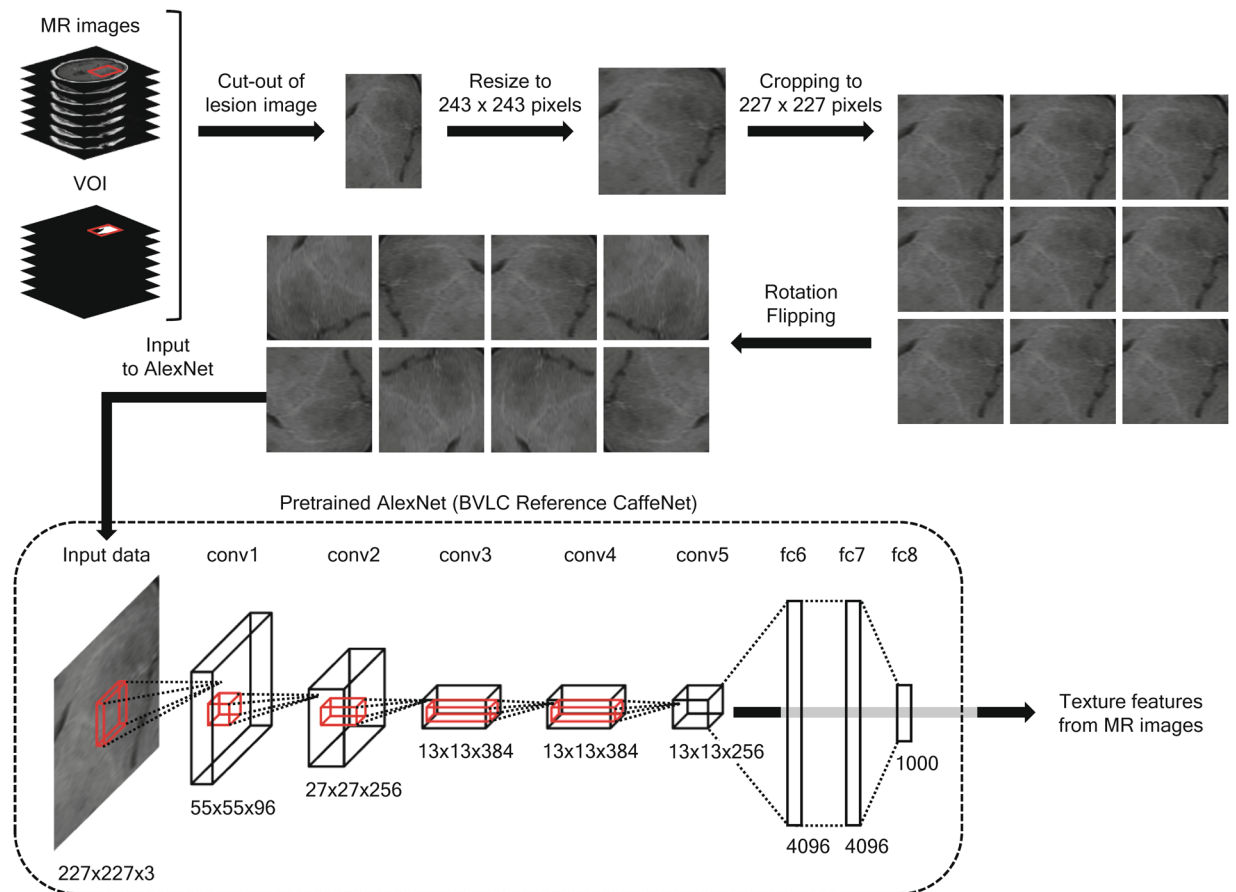
In this study, we sought to reveal a positive impact of sophisticated machine learning algorithms on the radiomic diagnostic accuracy of predicting the molecular characteristics of WHO grade II/III gliomas. More specifically, the diagnostic accuracy concerning the molecular characteristics of WHO grade II/III gliomas was compared between a conventional radiomic image texture extraction technique and an automatic image analysis technique using the CNN trained by natural scenes, the AlexNet<sup>18</sup>. The impact of adding clinical information to model the prediction of molecular characteristics for this neoplasm was also evaluated (Fig. 1).

## Results

**Optimization of a convolutional neural network for glioma feature extraction from a magnetic resonance image.** The CNN (AlexNet) layer was first tuned for glioma MR image feature extraction. In order to extract texture features (Fig. 1), MR images were normalized in their intensity, and from the slice locating center of the lesion, a lesion image was cut out at the boundary of the lesion for each patient and for each sequence (Fig. 2): T1-weighted (T1W) imaging, T2-weighted (T2W) imaging, gadolinium-enhanced T1-weighted (GdT1W) imaging, and fluid-attenuated inversion recovery (FLAIR) imaging. Texture features of the lesion image were acquired from neurons from each layer between data input to probability output. Linear support vector machine (SVM) models were trained to distinguish between brain tumor lesions and normal tissues, using the texture features of images of brain tumor lesions and images of normal tissue. Figure 3 shows the classification accuracies for each CNN layer using 10-fold cross-validation. The classification accuracy between the lesion and normal tissue was highest when the texture features from conv5 were used to train the SVM classifier ( $98.4 \pm 1.9\%$  ( $0.999 \pm 0.003$ ) [mean  $\pm$  95% confidence interval (mean  $\pm$  95% confidence interval of area under the precision-recall curve)]). As a result, the texture features from AlexNet conv5 were further applied in this study.

**Comparison of molecular diagnostic accuracy of WHO grade II/III gliomas between radiomics and the convolutional neural network.** Combinations of the following sets of information were fed to SVMs in an attempt to train an SVM classifier for molecular subtypes of WHO grade II/III gliomas: (1) the age of the patient, (2) 61 conventional radiomic features based on MRI<sup>8</sup> and 3 location parameters (radiomic features), and (3) 4000 texture features selected from AlexNet conv5 analysis of T1W, T2W, GdT1W, and FLAIR images (CNN features). The three molecular subtypes consisted of (1) IDH wild type ( $n = 56$ ), (2) IDH and pTERT mutations ( $n = 54$ ), and (3) IDH mutant and pTERT wild type ( $n = 54$ ). The molecular subtypes were successfully classified using the following information: patient age, radiomic features and CNN feature; classification accuracies, balanced among classes, were significantly higher than chance ( $p < 0.01$ , one-tailed Welch's *t*-test; Fig. 4a). Classification accuracies using the information differed significantly according to the combination of information (age:  $48.7 \pm 7.1\%$  ( $0.486 \pm 0.063$ ); conventional radiomic features with 3 location parameters:  $54.8 \pm 9.2\%$  ( $0.604 \pm 0.081$ ); CNN features:  $62.1 \pm 6.0\%$  ( $0.636 \pm 0.071$ ); all features:  $63.1 \pm 8.0\%$  ( $0.649 \pm 0.090$ );  $p = 0.0446$ ,  $F(3,36) = 3.9963$ , one-way ANOVA, Bonferroni corrected). Among the three types of features, the accuracy using the CNN features was the highest and significantly higher than that using patient age ( $p = 0.0386$ , post hoc Tukey-Kramer test, Fig. 4a). Notably, the accuracies were not significantly different between the analysis using the conventional radiomic features and patient age; combination of these features did not improve the classification accuracy ( $54.3 \pm 6.8\%$  ( $0.613 \pm 0.055$ )). Moreover, the classification accuracy was highest when all of the features were used together for the classification, although the accuracy using all features was not significantly different from that using CNN features alone (Fig. 4a). It therefore appears that among the three types of features, the CNN-based features were the most informative in classifying the molecular subtypes of WHO grade II/III gliomas.

**Comparison of classification accuracy for IDH mutation of WHO grade II/III gliomas.** To clarify how the IDH mutation and pTERT mutation affect classification accuracy, we classified the IDH mutation of WHO grade II/III gliomas. The balanced classification accuracies for IDH mutation were significantly higher



**Figure 1.** Feature extraction by AlexNet. The lesion image was cut out using a VOI and resized to  $243 \times 243$  pixels. The image was cropped to  $227 \times 227$  pixels with a shift of  $\pm 8$  pixels and rotated/flipped for data augmentation. The augmented image was input to the pretrained AlexNet to acquire the texture features for classification.

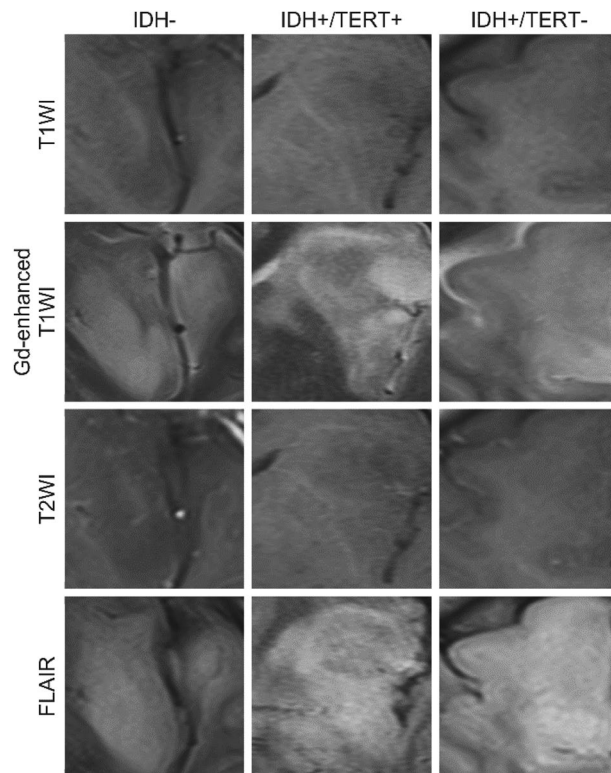
than that expected by chance for each feature ( $p < 0.01$ , one-tailed Welch's  $t$ -test, Fig. 4b). Although the balanced classification accuracies were highest with all the features combined (age:  $65.4 \pm 7.5\%$  ( $0.558 \pm 0.090$ ); conventional radiomic features with 3 location parameters:  $71.7 \pm 8.3\%$  ( $0.718 \pm 0.139$ ); CNN features:  $69.6 \pm 5.6\%$  ( $0.619 \pm 0.132$ ); all features:  $73.1 \pm 9.4\%$  ( $0.699 \pm 0.145$ ); Fig. 4b), a statistically significant difference was not observed among these accuracies ( $p = 1.2940$ ,  $F(3,36) = 0.9402$ , one-way ANOVA, Bonferroni corrected). These results suggest that conventional radiomic features and CNN-based texture features extract similar information from MR images concerning IDH mutations in WHO grade II/III gliomas.

### Comparison of classification accuracy for TERT promoter mutation in IDH-mutated WHO grade II/III gliomas.

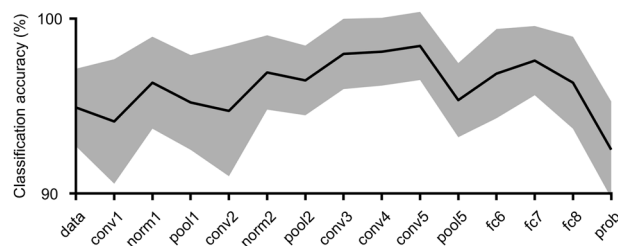
Finally, classification of the pTERT mutation in IDH-mutated WHO grade II/III gliomas was pursued. The pTERT mutation was successfully classified using each of the features with an accuracy greater than chance ( $p < 0.01$ , one-tailed Welch's  $t$ -test, Fig. 4c) except using conventional radiomic features with 3 location parameters ( $p = 0.257$ ,  $t(17.4) = 0.6674$ ). Classification accuracies, balanced among classes, differed significantly based on the features used (age:  $61.8 \pm 7.7\%$  ( $0.621 \pm 0.122$ ); conventional radiomic features with 3 location parameters:  $59.0 \pm 9.0\%$  ( $0.656 \pm 0.113$ ); CNN features:  $84.0 \pm 9.3\%$  ( $0.868 \pm 0.099$ ); all features:  $79.8 \pm 11.0\%$  ( $0.861 \pm 0.116$ );  $p = 0.0003$ ,  $F(3,36) = 9.3406$ , one-way ANOVA, Bonferroni corrected). The classification accuracies using CNN base texture features and all features together were both significantly higher than those using patient age alone or conventional radiomic features with 3 location parameters (patient age vs. CNN base texture features,  $p = 0.0028$ ; patient age vs. all features,  $p = 0.0192$ ; conventional radiomic features with 3 location parameters vs. CNN base texture features,  $p = 0.0007$ ; conventional radiomic features with 3 location parameters vs. all features  $p = 0.0053$ , Tukey-Kramer post hoc test). These results suggest that the CNN features succeeded in capturing some of the characteristics of the pTERT mutation that were not captured by conventional radiomic features and patient age.

### Discussion

Noninvasive molecular characterization of neoplasms via radiomics or convolutional neural network (CNN) analysis of radiographic images is a rapidly expanding research field. It is believed that the development of this type of technology could be a substitute for a large portion of – if not all – invasive procedures involved in acquiring information concerning the fundamental nature(s) of neoplastic disease. This diagnostic concept has been



**Figure 2.** Example of input images. Representative lesion images of each sequence for each molecular subtype of WHO grade II/III gliomas were shown.

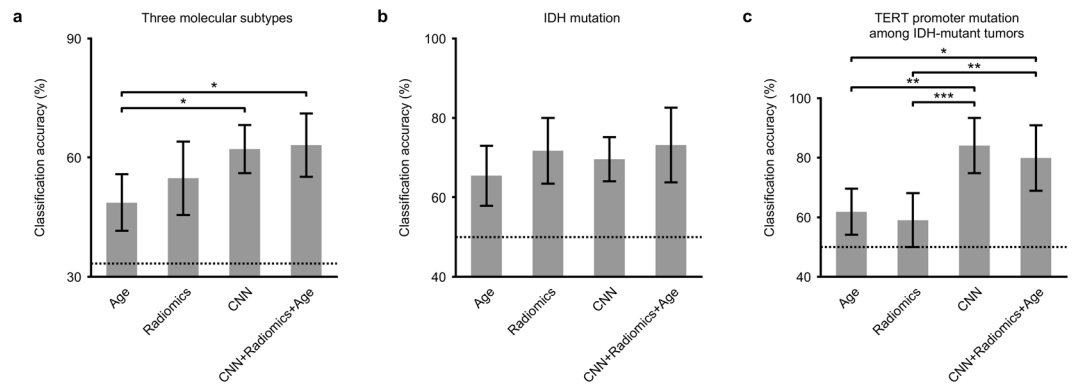


**Figure 3.** Classification accuracy of lesions from normal tissue. The average classification accuracies in all folds of cross-validation are shown with 95% confidence intervals.

widely explored for lung cancers, breast cancers and, recently, gliomas, for all of which genetic characterization of the tumor is crucial for determining an accurate diagnosis and selecting the most appropriate treatment strategy<sup>19–29</sup>. Historically, this proof-of-concept strategy was initially tested by introducing the idea of ‘radiomics’ – a coinage based on ‘genomics’ and ‘proteomics’ – that consists of handling numerous texture features of radiographic images to construct an objective-driven prediction model<sup>30</sup>. Combined with automatic segmentation of tumor<sup>31</sup>, these radiomic approaches may be a powerful tool in determining diagnosis and treatment in the future.

Regarding WHO grade II/III gliomas, several attempts have shown that this approach holds promise for identifying genetic alterations in gliomas. In fact, IDH mutations can be readily predicted with an accuracy greater than 80%. These studies have also highlighted the limitation of this approach. Notably, pTERT mutation or 1p/19q codeletion, either of which are crucial information for discriminating oligodendroglial tumors from astrocytic tumors, cannot be predicted within a reasonable accuracy. The introduction of the CNN was thought to enable feature extraction of images that were overlooked by handcrafted feature extraction used in radiomics. The current investigation indeed reveals that a CNN can provide valuable texture information, especially for discriminating the pTERT mutation, a task that radiomics alone cannot achieve successfully. Indeed, Fig. 4c shows that the CNN enabled pTERT mutation classification accuracy to jump from 59% to 84% when the analyzed cohort was confined to IDH mutant tumors. The three-class classification accuracy also improved from 55% to 63% with the aid of the CNN. These observations strongly suggest that CNN analysis is superior to radiomics in detecting some types of genetic alterations. It is also noteworthy that there are genetic alterations that CNN analysis has no advantage in detecting compared to radiomics. As in this investigation, CNNs usually analyze two-dimensional images, while radiomics performs three-dimensional data acquisition, and this is a theoretical difference in which





**Figure 4.** Classification accuracy of genotype status for different features. (a–c) Each bar shows the averaged classification accuracy of three molecular subtypes consisting of 1) IDH wild type, 2) IDH and pTERT comutated, and 3) IDH mutant and pTERT wild type (a), the IDH mutation alone (b), or the pTERT mutation alone (c). The label of each bar denotes the features used for the classification: Age, patient age; Radiomics, conventional radiomic features from MR images with three location parameters; CNN, features extracted by AlexNet; CNN + Radiomics + Age, all of these features. The average was calculated from the accuracies, which were balanced among classes, for each test dataset in 10-fold nested cross-validation. Error bars show 95% confidence intervals of the classification accuracy. Dotted lines denote chance level. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$  significant difference among different features (one-way ANOVA with a Tukey-Kramer post hoc test).

radiomics may be superior to CNN analysis in some cases. The conclusion that can be drawn from this phenomenon is that the use of, and results from, a CNN should always be critically and quantitatively evaluated to ensure its superiority over conventional radiomics.

The CNN used in this investigation was a natural scene pretrained CNN. It is possible that training the CNN with glioma images could improve genetic alteration classification accuracy, and this scientific question should be pursued in future investigations. Indeed, a similar approach has been reported, focusing on the IDH mutation status of WHO grade II–IV gliomas<sup>17</sup>. The scientific and clinical significance of the current research is that the investigated cohort is tightly restricted to WHO grade II/III gliomas. Because age alone has a great impact on classifying IDH mutation status, clinicians are already aware that IDH wild-type and mutant tumors exhibit completely different clinical presentations. More specifically, grade IV gliomas (glioblastomas) are mostly IDH wild-type tumors and are prone to showing contrast enhancement and elderly age distribution. IDH mutant tumors are mainly grade II or III gliomas and are less likely to show massive contrast enhancement and tend to occur in younger patients. If the cohort of the analysis, regardless of the technique used for image analysis (i.e., radiomics-based or CNN-based analysis), was performed without carefully discriminating the two, IDH mutation can be easily classified by age or contrast enhancement of the tumor alone, which adds no value in clinical practice, as tumor contrast enhancement has long been considered a hallmark of glioblastoma (i.e., IDH wild-type tumor) in many cases. The real clinical value of this technique appears when it achieves accurate diagnosis in situations where conventional knowledge is unable to do so. The current investigation took this clinical question seriously and raised the bar for predicting molecular alterations in gliomas by restricting the analyzed cohort to WHO grade II/III gliomas and was able to prove not only that IDH mutation can be classified at a reasonable accuracy but also that the classification accuracy of the pTERT mutation can be improved by a CNN, which warrants future investigations for training the CNN itself with a WHO grade II/III glioma-restricted cohort.

## Materials and Methods

**Patient cohorts.** This study adhered to the Declaration of Helsinki and was performed in accordance with protocols approved by the internal ethical review boards of the Osaka International Cancer Institute (No. 1306055036), Osaka University Graduate School of Medicine (No. 13244), and all collaborating institutes. All patients were enrolled after obtaining written informed consent.

The inclusion criteria was as follows: (1) 20 years of age or older; (2) frozen or fresh tissue available for genomic analysis; (3) pre-operative MRI available for radiomic analysis; and (4) local diagnosis of lower grade (WHO grade II/III) glioma based on the fourth edition of the WHO classification<sup>32</sup>. A total of 199 cases from 11 institutions were collected and included for analysis. MR images were acquired using either a 1.5 T or 3.0 T MRI scanner according to the protocols at each institution with a wide variety of MR vendors. A total of 164 datasets comprising T1W, T2W, FLAIR, and GdT1W images, were sent for further analysis. All of the images were acquired in the axial plane.

**Diagnosis, central pathology and genetic analysis.** All cases analyzed in this study were the same as the cases that we previously reported<sup>8</sup>. A senior neuropathologist performed central pathological reviews of all cases. Integrated diagnosis was made based on the microscopic histological diagnosis and the status of *IDH1/2* and 1p/19q copy number in compliance with the CNS WHO2016<sup>33</sup>. The following two laboratories conducted the genetic analysis: the Osaka National Hospital in Osaka, Japan, and the National Cancer Center Research Institute in Tokyo, Japan. Hotspot mutations of *IDH1/2* (codon 132 of *IDH1* and codon 172 of *IDH2*) and the *TERT* promoter (pTERT, termed C228 and C250) were assessed by Sanger sequencing and/or pyrosequencing at

either laboratory (detailed information can be found in previous publications)<sup>6,8</sup>. This study included 56 samples with IDH wild type, 54 samples with IDH- and pTERT-mutated, and 54 samples with IDH-mutated but pTERT wild type, of which 31, 34, and 25 samples were acquired by 1.5 T scanner, respectively. There was no significant relationship between the molecular subtypes and the MRI scanners (1.5 T ( $n=90$ ), or 3.0 T ( $n=74$ )) ( $p=0.219$ ,  $X^2(2, n=164)=3.0367$ , chi-squared test).

**Tumor segmentation and conventional radiomic analysis.** The radiomic analyses (radiomics) were performed using the method reported in our previous study<sup>8</sup>. An in-house-developed MATLAB-based (Mathworks, Natick, MA) image analysis software was used in combination with the Oxford Centre for Functional MRI of the Brain (FMRIB) Linear Image Registration Tool (FLIRT) provided by the FMRIB Software Library (FSL)<sup>34–36</sup>. First, MRIConvert (University of Oregon Lewis Center for Neuroimaging: <http://lcnf.uoregon.edu/~jolinda/MRIConvert/>) converted all Digital Imaging and Communications in Medicine (DICOM) format images to Neuroimaging Informatics Technology Initiative (NIfTI) format. All NIfTI files were sent to intensity normalization processing. Voxels that were in the top 0.1% in intensity histogram were considered high signal noise and therefore deleted, and the remaining 99.9% were reallocated to 256 grayscale among noncontrast T1W, contrast-enhanced GdT1W, and FLAIR, but not for T2W images. For T2W images, 100% of the data range was reallocated to 256 grayscale. T2Edge images were constructed by applying a Prewitt filter to T2W images. Gdzscore images were also constructed by performing a voxelwise contrast enhancement calculation using non-contrast and contrast-enhanced GdT1W images. Detailed methods can be found in Table S1. Tumors were segmented by an experienced surgical neuro-oncologist manually tracing high-intensity lesions on T2W images in three dimensions. Next, T2W images were registered to the standard MNI152 space, and the segmented tumors were mapped on the MNI152 coordinate system using the conversion matrix of T2W to MNI152. FSL-FLIRT was used for image registration with 12 degrees of freedom via mutual information algorithm.

Furthermore, all image sequences obtained from a single subject were coregistered with each other, thus enabling the created T2W-based lesions to be applied to different sequences. Three different aspects of the tumor were measured: (1) histogram-based first-order texture; (2) shape; and (3) location (Table S1). Lesion location information was also calculated based on the weight center of the VOI in MNI152 space, which consisted of 3 values corresponding to the  $x$ ,  $y$ ,  $z$  coordinates of the point located within MNI152.

**Image preprocessing.** The intensity of each MR image was normalized among the entire slices so that the 2.5th and 97.5th percentiles of the voxels became the minimum and maximum values, respectively. The VOI of the lesion on T2W images was coregistered to the T1W and FLAIR images to identify the tumor area within each of these images. GdT1W images were resliced to the T1WI space to use the VOI created for the T1W images.

**CNN feature extraction.** For each MRI sequence, an axial slice that contained the lesion was selected at first to extract CNN texture features using the VOI of the lesion coregistered to the sequence. On the center slice of the VOI in the axial direction, an image was cut out by the boundary box of the VOI (lesion image) and resized to  $243 \times 243$  pixels (because each lesion was a different size). To artificially enlarge the dataset<sup>18</sup>, data augmentation was performed by cropping, flipping, and rotating the lesion image (Fig. 1). For the cropping, nine images of  $227 \times 227$  pixels were cropped from a single resized lesion image of  $243 \times 243$  pixels with shift of 0 or  $\pm 8$  pixels in the  $x/y$  direction from its center. The cropped images were then rotated by 0, 90, 180, or 270 degrees and either horizontally flipped or not (resulting in 72 images) prior to inputting them to the pretrained BVLC CaffeNet running on Chainer<sup>37</sup>. The BVLC CaffeNet was AlexNet<sup>18</sup> with a minor variation developed by Berkeley Vision and Learning Center, who also trained the weights of the network using the ILSVRC-2012 ImageNet dataset on Caffe<sup>38</sup>. Outputs of neurons in each layer other than the ReLU and dropout layers were extracted as CNN texture features.

To extract texture features of normal tissue, the VOIs modeled on the T2W images were flipped in the sagittal plane and allocated manually in another hemisphere to define a normal tissue region. The allocated normal tissue region on the T2W images was coregistered to the MR images of different sequences; in exactly the same way, to extract the texture features of the lesion, the normal image was cut out using the coregistered normal tissue region as the VOI and augmented before extracting the texture features. During this process, the normal tissue region could not be allocated without overlap between the lesion image and the normal image in 37 patients, so these patients were excluded from the analysis decoding the lesion from normal tissue.

**Classification.** To select the appropriate layer for decoding the genotype status for each layer of the CNN, classification analysis between the lesion and normal tissue was performed in 127 patients using linear SVMs. For each layer, 1000 neurons were first randomly selected to apply SVM directly, and the texture features from these neurons were concatenated among the MR images (resulting in 4000 features) to train the SVMs. A 10-fold nested cross validation was adopted. Each test dataset of the outer cross-validation was classified by a linear SVM that was trained using the training dataset of the outer cross-validation, with a cost parameter optimized by another cross-validation within the training dataset. The division of the dataset into the test dataset and training dataset was controlled so that the two datasets did not contain training samples from the same patient. The classification accuracy accounted for only the decoding results using the texture feature from the center of the lesion image or the normal image, with neither rotation nor horizontal flip applied. For each test dataset of the outer cross-validation, area under the precision-recall curve (AUPRC) was calculated using distances of the samples to the separating hyperplane.

The genotype status of each patient was classified using the SVM for the following three conditions: three molecular subtypes (IDH wild type, IDH and pTERT mutations, or IDH mutation with pTERT wild type; totally,  $n=164$ ); IDH (mutation or wild type;  $n=164$ ); and pTERT (mutation or wild type;  $n=108$ ). Again, 10-fold

nested cross validation was applied in the same manner as in the classification analysis between the lesion and normal tissue. For further analysis, the division of the dataset was kept the same during the classification analyses for the same status of genotypes but with different features. In the analysis using the CNN texture features, the number of features for building the SVM was reduced to 4000 by selecting the features with the highest *F*-statistics, which were calculated only among the training dataset. Moreover, classification accuracy using the CNN texture features again accounted for only the classification using features extracted from the center of the lesion image, with neither rotation nor horizontal flip. Each classification accuracy was balanced among classes, so that it was not affected by an imbalanced number of samples among classes. Finally, AUPRCs were again calculated. In the classification analysis of the three molecular subtypes, distance to the hyperplane that separated the target subtype from the other two subtypes was used to calculate the AUPRC; the AUPRCs of the three subtypes were then averaged within the same test dataset of the outer cross-validation.

**Statistics.** For each decoding condition for the three molecular subtypes, classification analysis was performed four times using the following features: patient age, 64 radiomic features (including lesion locations), CNN features, and the combination of these three features. The classification accuracies of the 10-fold validations were compared using one-way analysis of variance (ANOVA) with a Tukey-Kramer post hoc test. Bonferroni correction was applied to the *p*-value of the ANOVA for the three classifications of genotype status. For each classification analysis, the significance of the classification accuracy was tested by a one-tailed Welch's *t*-test compared to the classification accuracy expected by chance. The chance classification accuracy was estimated by another classification analysis using the age feature with a shuffled label for the genotype status.

### Data availability

The data are not publicly available because they contain information that could compromise research participants' privacy and/or consent.

Received: 12 September 2019; Accepted: 3 December 2019;

Published online: 30 December 2019

### References

1. Suzuki, H. *et al.* Mutational landscape and clonal architecture in grade II and III gliomas. *Nat. Genet.* **47**, 458–468, <https://doi.org/10.1038/ng.3273> (2015).
2. Cairncross, G. *et al.* Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. *J. Clin. Oncol.* **31**, 337–343, <https://doi.org/10.1200/JCO.2012.43.2674> (2013).
3. Eckel-Passow, J. E. *et al.* Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N. Engl. J. Med.* **372**, 2499–2508, <https://doi.org/10.1056/NEJMoa1407279> (2015).
4. Buckner, J. C. *et al.* Radiation plus procarbazine, CCNU, and vincristine in low-grade glioma. *N. Engl. J. Med.* **374**, 1344–1355, <https://doi.org/10.1056/NEJMoa1500925> (2016).
5. Arita, H. *et al.* Upregulating mutations in the TERT promoter commonly occur in adult malignant gliomas and are strongly associated with total 1p19q loss. *Acta Neuropathol.* **126**, 267–276, <https://doi.org/10.1007/s00401-013-1141-6> (2013).
6. Arita, H. *et al.* A combination of TERT promoter mutation and MGMT methylation status predicts clinically relevant subgroups of newly diagnosed glioblastomas. *Acta Neuropathol. Commun.* **4**, 79, <https://doi.org/10.1186/s40478-016-0351-2> (2016).
7. Cancer Genome Atlas Research, N. *et al.* Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498, <https://doi.org/10.1056/NEJMoa1402121> (2015).
8. Arita, H. *et al.* Lesion location implemented magnetic resonance imaging radiomics for predicting IDH and TERT promoter mutations in grade II/III gliomas. *Sci. Rep.* **8**, 11773, <https://doi.org/10.1038/s41598-018-30273-4> (2018).
9. Kickingeder, P. *et al.* IDH mutation status is associated with a distinct hypoxia/angiogenesis transcriptome signature which is non-invasively predictable with rCBV imaging in human glioma. *Sci. Rep.* **5**, 16238, <https://doi.org/10.1038/srep16238> (2015).
10. Leu, K. *et al.* Perfusion and diffusion MRI signatures in histologic and genetic subtypes of WHO grade II-III diffuse gliomas. *J. Neurooncol.* **134**, 177–188, <https://doi.org/10.1007/s11060-017-2506-9> (2017).
11. Choi, C. *et al.* 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat. Med.* **18**, 624–629, <https://doi.org/10.1038/nm.2682> (2012).
12. Andronesi, O. C. *et al.* Detection of 2-hydroxyglutarate in IDH-mutated glioma patients by *in vivo* spectral-editing and 2D correlation magnetic resonance spectroscopy. *Sci. Transl. Med.* **4**, 116ra114, <https://doi.org/10.1126/scitranslmed.3002693> (2012).
13. Titano, J. J. *et al.* Automated deep-neural-network surveillance of cranial images for acute neurologic events. *Nat. Med.* **24**, 1337–1341, <https://doi.org/10.1038/s41591-018-0147-y> (2018).
14. Chilamkurthy, S. *et al.* Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* **392**, 2388–2396, [https://doi.org/10.1016/S0140-6736\(18\)31645-3](https://doi.org/10.1016/S0140-6736(18)31645-3) (2018).
15. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *P IEEE* **86**, 2278–2324, <https://doi.org/10.1109/5.726791> (1998).
16. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444, <https://doi.org/10.1038/nature14539> (2015).
17. Chang, K. *et al.* Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR Imaging. *Clin. Cancer Res.* **24**, 1073–1081, <https://doi.org/10.1158/1078-0432.CCR-17-2236> (2018).
18. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105 (2012).
19. Aerts, H. J. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006, <https://doi.org/10.1038/ncomms5006> (2014).
20. Li, H. *et al.* MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *Radiology* **281**, 382–391, <https://doi.org/10.1148/radiol.2016152110> (2016).
21. Toney, L. K. & Vesselle, H. J. Neural networks for nodal staging of non-small cell lung cancer with FDG PET and CT: importance of combining uptake values and sizes of nodes and primary tumor. *Radiology* **270**, 91–98, <https://doi.org/10.1148/radiol.13122427> (2014).
22. Gevaert, O. *et al.* Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. *Radiology* **264**, 387–396, <https://doi.org/10.1148/radiol.12111607> (2012).
23. Liu, Y. *et al.* Relationship between glioblastoma heterogeneity and survival time: an MR Imaging texture analysis. *AJNR Am. J. Neuroradiol.* **38**, 1695–1701, <https://doi.org/10.3174/ajnr.A5279> (2017).

24. Zhou, H. *et al.* MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol.* **19**, 862–870, <https://doi.org/10.1093/neuonc/now256> (2017).
25. Hu, L. S. *et al.* Radiogenomics to characterize regional genetic heterogeneity in glioblastoma. *Neuro Oncol.* **19**, 128–137, <https://doi.org/10.1093/neuonc/now135> (2017).
26. Zhang, B. *et al.* Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol.* **19**, 109–117, <https://doi.org/10.1093/neuonc/now121> (2017).
27. Smits, M. & van den Bent, M. J. Imaging correlates of adult glioma genotypes. *Radiology* **284**, 316–331, <https://doi.org/10.1148/radiol.2017151930> (2017).
28. Kickingereder, P. *et al.* Radiomic profiling of glioblastoma: identifying an imaging predictor of patient survival with improved performance over established clinical and radiologic risk models. *Radiology* **280**, 880–889, <https://doi.org/10.1148/radiol.2016160845> (2016).
29. Van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107, <https://doi.org/10.1158/0008-5472.CAN-17-0339> (2017).
30. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446, <https://doi.org/10.1016/j.ejca.2011.11.036> (2012).
31. Havaei, M. *et al.* Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31, <https://doi.org/10.1016/j.media.2016.05.004> (2017).
32. Louis, D. N. & International Agency for Research on, C. *WHO classification of tumours of the central nervous system. 4th ed edn*, (International Agency for Research on Cancer, 2007).
33. Louis, D. N., Ohgaki, H., Wiestler, O. D. & Cavenee, W. K. *WHO classification of tumours of the central nervous system*. (International Agency for Research on Cancer, 2016).
34. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
35. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).
36. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72, <https://doi.org/10.1016/j.neuroimage.2009.06.060> (2009).
37. Tokui, S., Oono, K., Hido, S. & Clayton, J. Chainer: a next-generation open source framework for deep learning. In *Proceedings of workshop on machine learning systems (LearningSys) in the twenty-ninth annual conference on neural information processing systems (NIPS)*. 1–6 (2015).
38. Jia, Y. *et al.* Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*. 675–678 (ACM, 2014).

## Acknowledgements

This research was conducted with support from the Brain/MINDS from AMED; SIP (Innovative AI Hospital System) of NIBIOHN; JST PRESTO (15654740); CREST (JPMJCR18A5); ERATO (JPMJER1801); the TERUMO Foundation for Life Sciences and Arts; the Ministry of Health, Labor, and Welfare (29200801); the Japan Foundation of Aging and Health; Grants-in-Aid for Scientific Research KAKENHI (JP17H06032 and JP15H05710, JP18H04085, 16K10778, 17H05308); MSD Life Science Foundation; a Research Grant from the Takeda Science Foundation; the Uehara Memorial Foundation; and The Canon Foundation

## Author contributions

R.F. implemented the analysis, analyzed the data, interpreted the data, wrote and reviewed the manuscript. T.Y. and M.K. designed experiment, interpreted the data, wrote and reviewed the manuscript. T.S. implemented the analysis and reviewed the manuscript. H.A., A.K., M.T., Y.N., Y.T., N.T., Y.O., M.N., S.M., M.T., Y.F., J.F., S.I., K.I., Y.N., T.S., D.K., E.Y., Y.K., M.M., K.M. and K.I. reviewed the manuscript. YK and HK interpreted the data and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-56767-3>.

**Correspondence** and requests for materials should be addressed to T.Y. or M.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019