








Article

Convolutional Neural Networks to Detect Vestibular Schwannomas on Single MRI Slices: A Feasibility Study

Carole Koechli ¹, Erwin Vu ², Philipp Sager ¹, Lukas Näf ³, Tim Fischer ³, Paul M. Putora ^{2,4}, Felix Ehret ^{5,6,7}, Christoph Fürweger ^{7,8}, Christina Schröder ¹, Robert Förster ¹, Daniel R. Zwahlen ¹, Alexander Muacevic ⁷ and Paul Windisch ^{1,7,*}

- ¹ Department of Radiation Oncology, Kantonsspital Winterthur, 8401 Winterthur, Switzerland; carole.koechli@uzh.ch (C.K.); philipp.sager@student.unisg.ch (P.S.); christina.schroeder@ksw.ch (C.S.); robert.foerster@ksw.ch (R.F.); daniel.zwahlen@ksw.ch (D.R.Z.)
- ² Department of Radiation Oncology, Kantonsspital St. Gallen, 9007 St. Gallen, Switzerland; erwin.vu@kssg.ch (E.V.); paulmartin.putora@kssg.ch (P.M.P.)
- ³ Department of Radiology, Kantonsspital St. Gallen, 9007 St. Gallen, Switzerland; lukas.naef@kssg.ch (L.N.); tim.fischer@kssg.ch (T.F.)
- ⁴ Department of Radiation Oncology, University of Bern, 3010 Bern, Switzerland
- ⁵ Berlin Institute of Health at Charité—Universitätsmedizin Berlin, 10117 Berlin, Germany; felix.ehret@charite.de
- ⁶ Charité—Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Department of Radiation Oncology, 13353 Berlin, Germany
- ⁷ European Radiosurgery Center, 81377 Munich, Germany; christoph.fuerweger@erc-munich.com (C.F.); alexander.muacevic@erc-munich.com (A.M.)
- ⁸ Department of Stereotaxy and Functional Neurosurgery, University of Cologne, Faculty of Medicine and University Hospital Cologne, 50937 Cologne, Germany
- * Correspondence: paul.windisch@ksw.ch; Tel.: +41-52-266-2653



Citation: Koechli, C.; Vu, E.; Sager, P.; Näf, L.; Fischer, T.; Putora, P.M.; Ehret, F.; Fürweger, C.; Schröder, C.; Förster, R.; et al. Convolutional Neural Networks to Detect Vestibular Schwannomas on Single MRI Slices: A Feasibility Study. *Cancers* **2022**, *14*, 2069. <https://doi.org/10.3390/cancers14092069>

Academic Editors: Brigitta G. Baumert and Salvatore Cappabianca

Received: 10 March 2022

Accepted: 19 April 2022

Published: 20 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Simple Summary: Due to the fact that they take inter-slice information into account, 3D- and 2.5D-convolutional neural networks (CNNs) potentially perform better in tumor detection tasks than 2D-CNNs. However, this potential benefit is at the expense of increased computational power and the need for segmentations as an input. Therefore, in this study we aimed to detect vestibular schwannomas (VSs) in individual magnetic resonance imaging (MRI) slices by using a 2D-CNN. We retrained (539 patients) and internally validated (94 patients) a pretrained CNN using contrast-enhanced MRI slices from one institution. Furthermore, we externally validated the CNN using contrast-enhanced MRI slices from another institution. This resulted in an accuracy of 0.949 (95% CI 0.935–0.963) and 0.912 (95% CI 0.866–0.958) for the internal and external validation, respectively. Our findings indicate that 2D-CNNs might be a promising alternative to 2.5-/3D-CNNs for certain tasks thanks to the decreased requirement for computational power and the fact that there is no need for segmentations.

Abstract: In this study, we aimed to detect vestibular schwannomas (VSs) in individual magnetic resonance imaging (MRI) slices by using a 2D-CNN. A pretrained CNN (ResNet-34) was retrained and internally validated using contrast-enhanced T1-weighted (T1c) MRI slices from one institution. In a second step, the model was externally validated using T1c- and T1-weighted (T1) slices from a different institution. As a substitute, bisected slices were used with and without tumors originating from whole transversal slices that contained part of the unilateral VS. The model predictions were assessed based on the categorical accuracy and confusion matrices. A total of 539, 94, and 74 patients were included for training, internal validation, and external T1c validation, respectively. This resulted in an accuracy of 0.949 (95% CI 0.935–0.963) for the internal validation and 0.912 (95% CI 0.866–0.958) for the external T1c validation. We suggest that 2D-CNNs might be a promising alternative to 2.5-/3D-CNNs for certain tasks thanks to the decreased demand for computational power and the fact that there is no need for segmentations. However, further research is needed on the difference between 2D-CNNs and more complex architectures.

Keywords: artificial intelligence; deep learning; machine learning; vestibular; schwannoma; neuro-oncology

1. Introduction

Recent advances have allowed for a more widespread application of machine learning (ML) techniques in various fields thanks to improved computational power and user-optimized software libraries [1,2]. The possible applications of ML techniques include automatic tumor detection, segmentation, and classification to diagnose brain tumors based on magnetic resonance imaging (MRI) [3,4]. Accordingly, automatic tumor detection and classification could potentially support radiologists by increasing the efficiency of the clinical workflow and providing a second opinion. Currently, to the best of our knowledge, no commercially available or established implementation can detect or classify different brain tumor entities.

A shift in the investigated ML techniques was reported from traditional and classical ML to deep learning techniques to detect brain tumors based on MRI [5]. Accordingly, 2.5D- and 3D-convolutional neural networks (CNNs) have been implemented in several publications to automatically perform substeps of a tumor diagnosis based on MRI [6–10]. Various 3D-CNNs reportedly performed well in subtasks of a tumor diagnosis using the publicly available Multimodal Brain Tumor Segmentation Challenge (BraTS) data set [8–10].

There are distinct benefits and downsides with 3D- and 2D-CNNs. On the one hand, 3D-CNNs take inter-slice information into account, potentially resulting in a superior performance in comparison with 2D-CNNs. On the other hand, 2D-CNNs achieve a higher efficiency in memory and computational power compared with 3D-CNNs [11]. Moreover, many CNNs require segmentations that indicate the region of interest to the algorithm. However, it tends to be less time-consuming to select tumor containing slices as an input for 2D-CNNs than to contour the region of interest as an input for 3D-CNNs [7,12,13]. In addition, data augmentation techniques are better established for 2D-CNNs compared with 3D-CNNs [14,15].

Several techniques have been investigated that automatically segment the regions of interest in MRI slices [16,17]. For example, Kadry et al. used the U-Net scheme to extract ischemic stroke lesions from MRI slices [17].

Various architectures of convolutional neural networks have been investigated for tumor classification and detection including AlexNet, GoogLeNet, and VGGNet [18]. Rehman et al. observed that the highest accuracy was achieved in terms of classification and detection of benign brain tumors using the VGG16 architecture [18].

Vestibular schwannomas (VSs) represent an optimal tumor entity to examine early automatic tumor detection feasibility studies thanks to their relatively constant location compared with most other entities. VSs are located in the cerebellopontine angle and only vary in size, shape, and intratumoral homogeneity [19]. Thus, it is likely that less data are required for training.

Sager et al. investigated the feasibility of VS detection using single MRI slices [20]. MRI was selected as the imaging modality because it is the method of choice to detect VSs in clinical practice [19]. A 2D-CNN was trained and validated without the use of segmentations. However, the study used the laterality of the VS as a surrogate for the tumor detection. A high accuracy was obtained for the prediction of laterality in the internal and external validation cohort, with only a minor reduction in accuracy in the latter. The current study constituted a continuation of the work by Sager et al. using the same data but without the use of laterality as a surrogate [20].

The aim of this work was, therefore, to study the performance of a computationally inexpensive 2D-CNN to detect VSs based on single transversal MRI slices. For this purpose, MRI slices were used to train, internally validate, and externally validate the 2D-CNN for VS detection. Advancements in artificial intelligence were used to investi-

gate the region of interest for the model to detect the VS. Additionally, the 2D-CNN was externally validated using MRI slices without contrast enhancement. This was to test the hypothesis as to whether the model could generalize when applied to the MRI slices without contrast enhancement.

2. Materials and Methods

A retrospective feasibility study was conducted using the same MRI slices as Sager et al. [20]. The cohorts included an internal training, an internal validation, and an external validation data set. The internal data sets were supplied by the European Radiosurgery Center in Munich (Munich, Germany) whereas the external validation data set was from the Kantonsspital St. Gallen (St. Gallen, Switzerland). The workflow is illustrated in Figure 1.

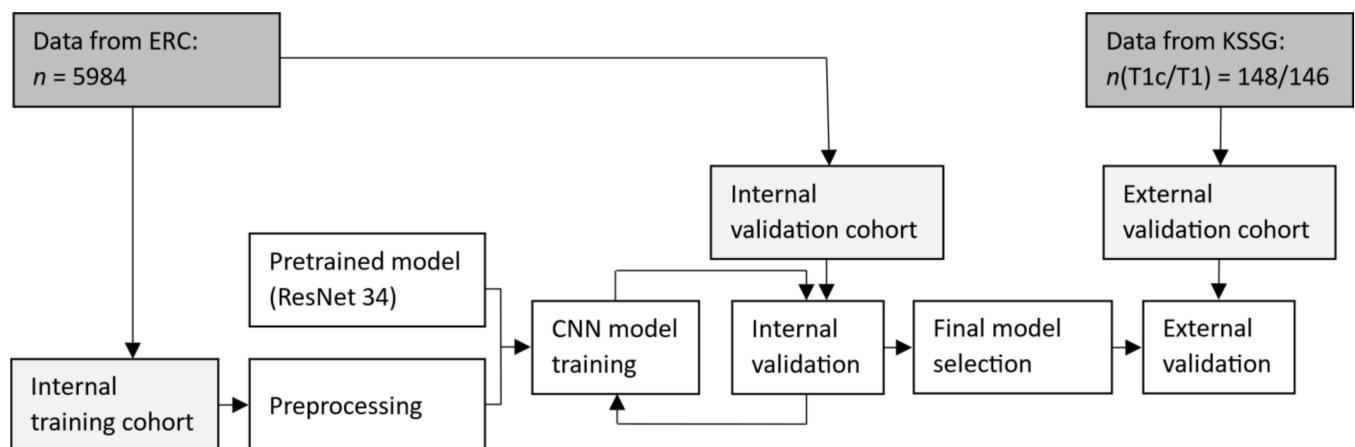


Figure 1. Workflow of the training, internal validation, and external validation. ERC: European Radiosurgery Center in Munich; KSSG: Kantonsspital St. Gallen; n : number of slices; T1c: contrast-enhanced T1-weighted; T1: T1-weighted.

Overall, the MRI slices were included and excluded as follows: Only patients with a T1-weighted (T1) or contrast-enhanced T1-weighted (T1c) MRI scan of a VS diagnosed by a board-certified radiologist were considered. Hereof, only transversal MRI slices were collected that contained part of the tumor. Accordingly, one or several slices were selected per patient, depending on the tumor size and slice thickness. The internal data were split into training and validation sets before and after a set reference date without a cross-validation. This was to ensure that no slices from the same patient could be present in both the training and validation sets. The MRI slices of patients with bilateral VSs were omitted. Moreover, slices were excluded with a significant tumor spread to the contralateral hemisphere. This elimination step was performed by a radiation oncology resident and a medical student by visually assessing all MRI slices until a consensus was reached. As the internal data set consisted of patients assigned from different clinics in various countries, heterogeneous imaging protocols were used.

For the training and internal validation, T1c MRI slices were included independent of whether or not the patient had received any prior treatment to the tumor before the images had been acquired. Magnetic resonance (MR) images were omitted with artifacts that strongly reduced the overall image quality.

For the external validation, T1c and T1 MRI slices were consecutively sampled from the radiology database. The MR images were acquired on Siemens devices with field strengths between 1.5 and 3 T and a variable slice thickness ≥ 1 mm. Specifically, T1 MRI slices were included to assess whether the CNN could generalize the slices with an acquisition sequence on which the model had not been trained. Patients with pretreated VSs were excluded from the external validation cohort. This was to prevent the CNN from considering treatment-related modifications when predicting the presence of the VS.

The MRI slices were preprocessed by labeling and bisecting the images. First, all MR images were labeled from the internal and external cohort to specify whether the VS was located on the left or right side. For the internal data set, the labeling was performed by a radiation oncology resident and a medical student. For the external data set, two board-certified radiologists classified and revised the MRI slices containing the VS, respectively. A Python (version 3.8.6) algorithm was then programmed to automatically split the MRI slices into a left and right image side. By splitting the images, two sliced images were obtained; one hemisphere with and one without a VS, respectively, as illustrated in Figure 2.

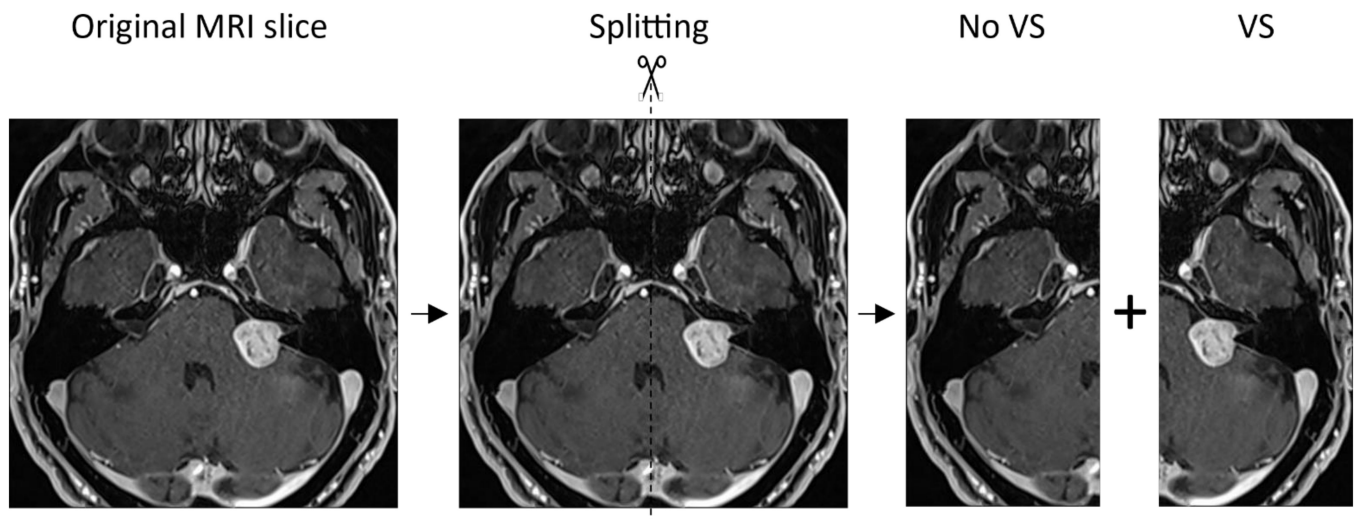


Figure 2. Workflow of splitting the magnetic resonance imaging (MRI) slices into a left and right image, resulting in a hemisphere with and without a tumor.

The CNN was trained using Python, the fastai library (version 2.4.1), and the PyTorch library (version 1.7.0) [21,22]. The CNN training included the data augmentation and fine-tuning of a pretrained model. All settings were selected according to the precursor publication by Sager et al. to allow the assessment of laterality as a surrogate for tumor detection [20]. First, the data were augmented using RandomResizedCrop and aug_transforms [23]. RandomResizedCrop amplifies data by varying the size and image section [23]. A minimum scale and resample were selected equal to 0.9 and 224×224 pixels, respectively. The resampling was chosen as a tradeoff between a fast computational time and maintaining sufficient information. aug_transforms contains a standard set of augmentations including random flipping, zooming, warping, maximum rotation, and maximum change of brightness and contrast (i.e., lighting) [23]. For this purpose, the following settings were defined: no flipping, minimum zooming = 1, maximum zooming = 1.1, no warping, maximum rotation = 15° , and maximum lighting = 0.1. The existing ResNet-34 was selected as a pretrained model, previously trained using the ImageNet data set [24]. It consisted of 34 layers [25]. This specific pretrained model was selected due to its demonstrated ability to classify images [24] as well as the possibility of importing ResNet architectures of varying complexities into different popular deep learning libraries [24]. Finally, the CNN was fine-tuned using a variable learning rate, flattened cross-entropy loss as the loss function, and Adam as the optimizer [26]. For the first five epochs, the body weights of the CNN were frozen and only the head was trained. Subsequently, the body weights were unfrozen and the CNN was trained for another 15 epochs. The training and validation loss were observed to investigate truncation and overfitting.

The performance of the CNN was assessed based on the internal and external data set. For the external validation set, MR images with and without contrast enhancement were separately evaluated. The flattened cross-entropy loss was determined for the 15 unfrozen epochs regarding the training and internal validation cohort. Furthermore, the following performance metrics were calculated for the internal, external T1c, and ex-

ternal T1 validation cohort: the accuracy and corresponding 95% confidence interval (CI); the sensitivity; the specificity; and the F1 score. The 95% CI was computed as follows: $CI = 1.96 * \sqrt{((accuracy * (1 - accuracy)) / n)}$ with n = size of the corresponding validation cohort. Additionally, the confusion matrices were plotted for the external validation cohort with and without contrast enhancement, respectively. Moreover, all external T1c and T1 slices were sorted and displayed, respectively, based on whether they were classified correctly or incorrectly. Finally, the gradient-weighted class activation mapping (Grad-CAM) images were evaluated for five correctly and five incorrectly classified sample images [27]. In doing so, the focus of the CNN was determined during the prediction-making process. Supplementary Material S1 includes the programming code used to split the MRI slices and train, validate, and assess the network.

A literature search was further conducted to compare the findings of this study with previous publications. For this purpose, the Medical Literature Analysis and Retrieval System Online (MEDLINE) database was searched on 31 January 2022 using the PubMed interface. The query used was taken from the precursor publication by Sager et al.: “((vestibular schwannoma[Title]) OR (vestibular schwannomas[Title])) AND ((deep[Title]) OR (network[Title]) OR (networks[Title]) OR (artificial intelligence[Title]) OR (machine learning[Title]))” [20].

Written informed consent was received from all patients to analyze the anonymized clinical and imaging data. Institutional policies were followed when working with the imaging data. For the internal cohort, approval was granted by the institutional review board of the Ludwig Maximilian University of Munich (project 20–437, 22 June 2020) as part of a larger project. This publication is a side project of a larger project modelling the outcomes after radiosurgery. For the external validation cohort, the Ethics Committee of Eastern Switzerland (EKOS 21/041, 9 March 2021) waived the ethical approval, stating that single MRI slices preserve adequate anonymization.

3. Results

The characteristics of the data sets are summarized in Table 1, including the number of patients and corresponding MRI slices as well as the VS location. In the training and internal validation cohort, 18 patients (2.8%) previously received radiotherapy and 112 patients (17.7%) formerly underwent a (partial) resection of the VS or suffered a relapse. In the external validation cohort, a few patients did not have both a T1 and a T1c image available. Therefore, the external validation cohort included 82 patients in total (65 patients with a T1 and T1c MR image, eight patients with a T1 MR image only, and nine patients with a T1c MR image only). Overall, no MRI slices were identified with a significant crossing to the contralateral image side. Only two MRI slices were identified in the training, external T1c, and external T1 data set, respectively, with an insignificant intersection.

Table 1. Characteristics of the included patients and the corresponding magnetic resonance slices (MRI) slices. T1c: contrast-enhanced T1-weighted; T1: T1-weighted.

Characteristics	Training (T1c)	Internal Validation (T1c)	External Validation (T1c)	External Validation (T1)
Number of patients	539	94	74	73
Number of MRI slices/bisected MRI slices	2538/5076	454/908	74/148	73/146
Tumor location				
Left (number of patients/MRI slices/bisected MRI slices)	278/1307/2614	54/270/540	31/31/62	34/39/78
Right (number of patients/MRI slices/bisected MRI slices)	261/1231/2462	40/184/368	43/43/86	39/39/78

The flattened cross-entropy loss is depicted in Figure 3 for the training and internal validation. Table 2 summarizes the performance metrics of the internal, external T1c, and external T1 validation. The confusion matrix is illustrated in Figure 4 for the external validation using the T1c images and T1 images, respectively. The slices are shown in Figures 5 and 6 for the external T1c and T1 validations. The slices were grouped based on whether they were correctly or incorrectly classified by the CNN. They included VSs of varying sizes and shapes as well as blood vessels in the cerebellopontine angle.

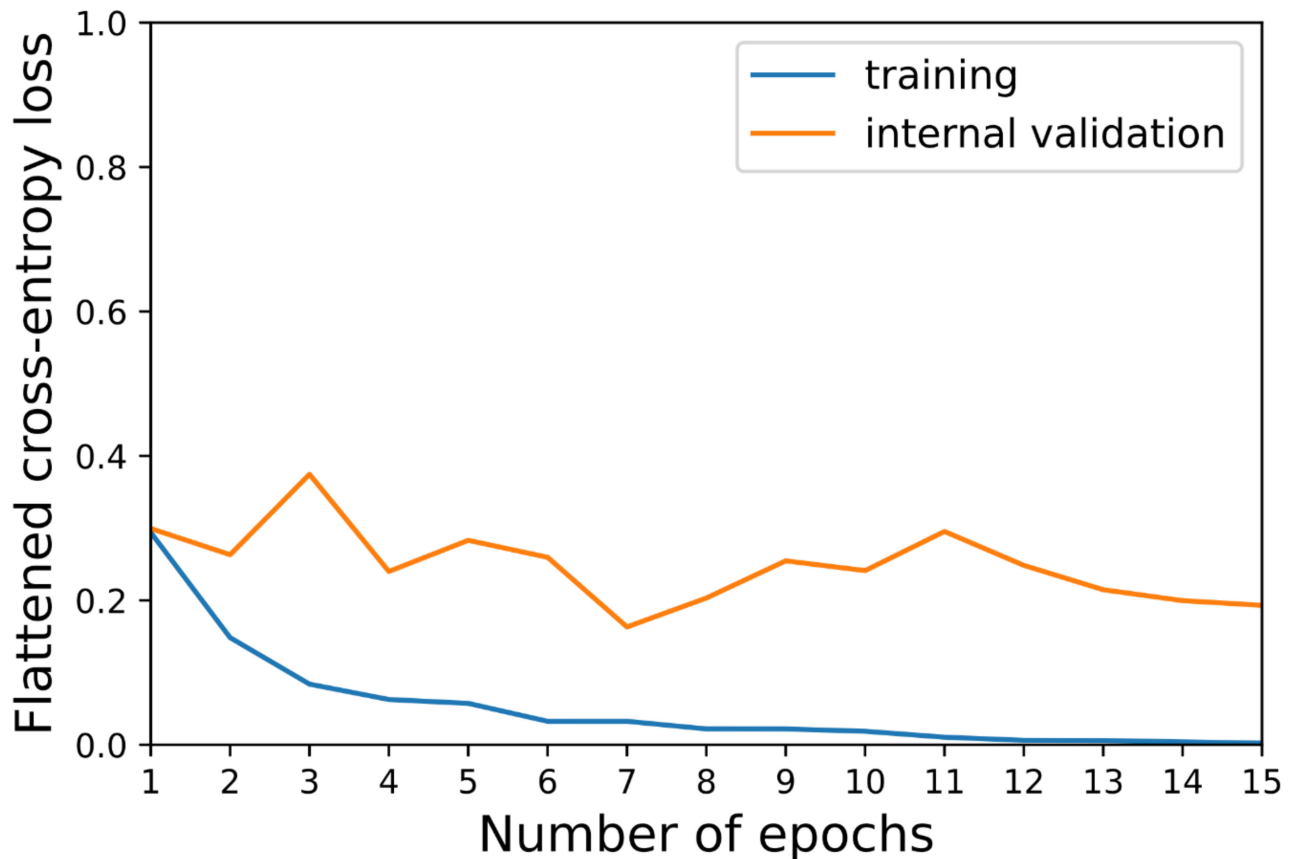


Figure 3. Flattened cross-entropy loss of the training (blue) and internal validation (orange) data set across the 15 unfrozen epochs.

Table 2. Performance metrics of the internal, external T1c, and external T1 validation cohort. 95% CI: 95% confidence interval.

Data Set	Accuracy (95% CI)	Sensitivity	Specificity	F1 Score
Internal validation	0.949 (95% CI 0.935–0.963)	0.916	0.982	0.948
External T1c validation	0.912 (95% CI 0.866–0.958)	0.851	0.973	0.906
External T1 validation	0.514 (95% CI 0.433–0.595)	0.055	0.973	0.101

Figure 7 shows five correctly and incorrectly classified examples of Grad-CAM images and the corresponding MRI slices for the external T1c validation, respectively. The CNN seemed to focus on a larger area surrounding the tumor including the cerebellopontine angle for the correctly classified slices (slices #1–#5). In contrast, the model mostly concentrated on a larger image section (slices #1, #2, #4, and #5) in the case of the incorrectly classified images. However, in the case of one incorrectly classified image (slice #3), the model focused on the cerebellopontine angle without detecting the VS.

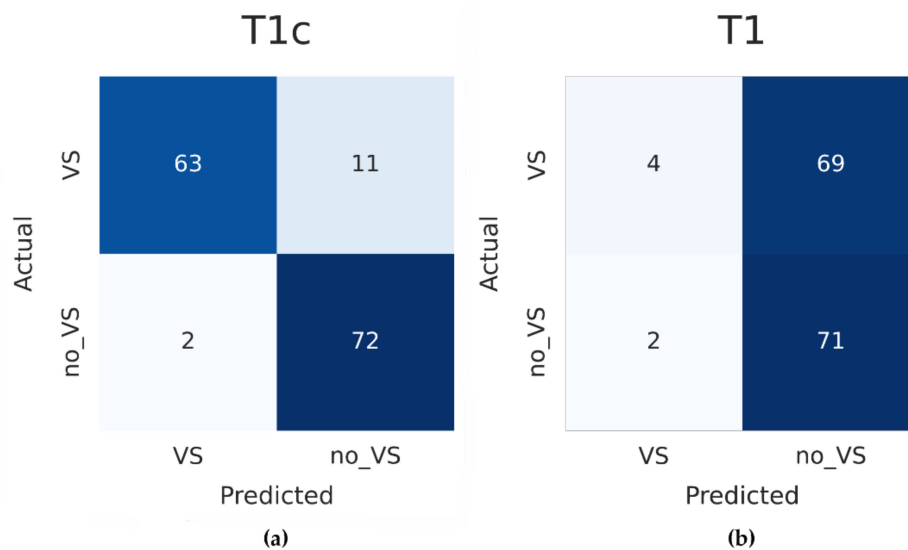


Figure 4. (a) Confusion matrix of the external validation with T1c MRI slices; (b) confusion matrix of the external validation with T1 MRI slices. VS: vestibular schwannoma.

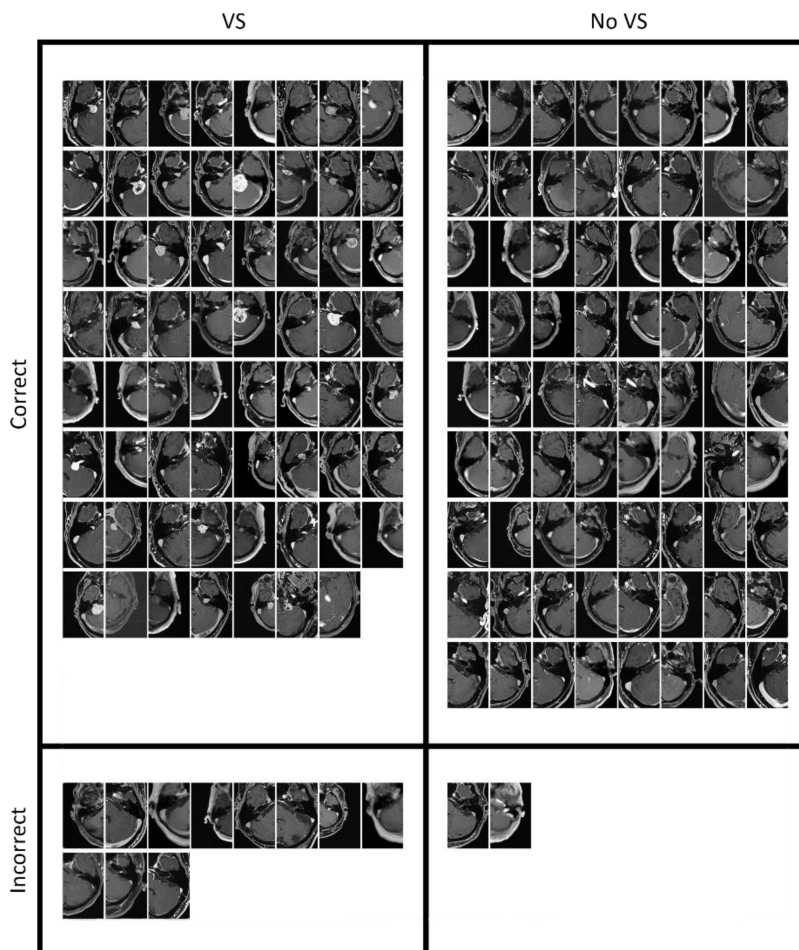


Figure 5. All MRI slices of the external validation data set with the T1c sequence. The images are sorted based on whether they contained a VS and whether they were correctly classified.

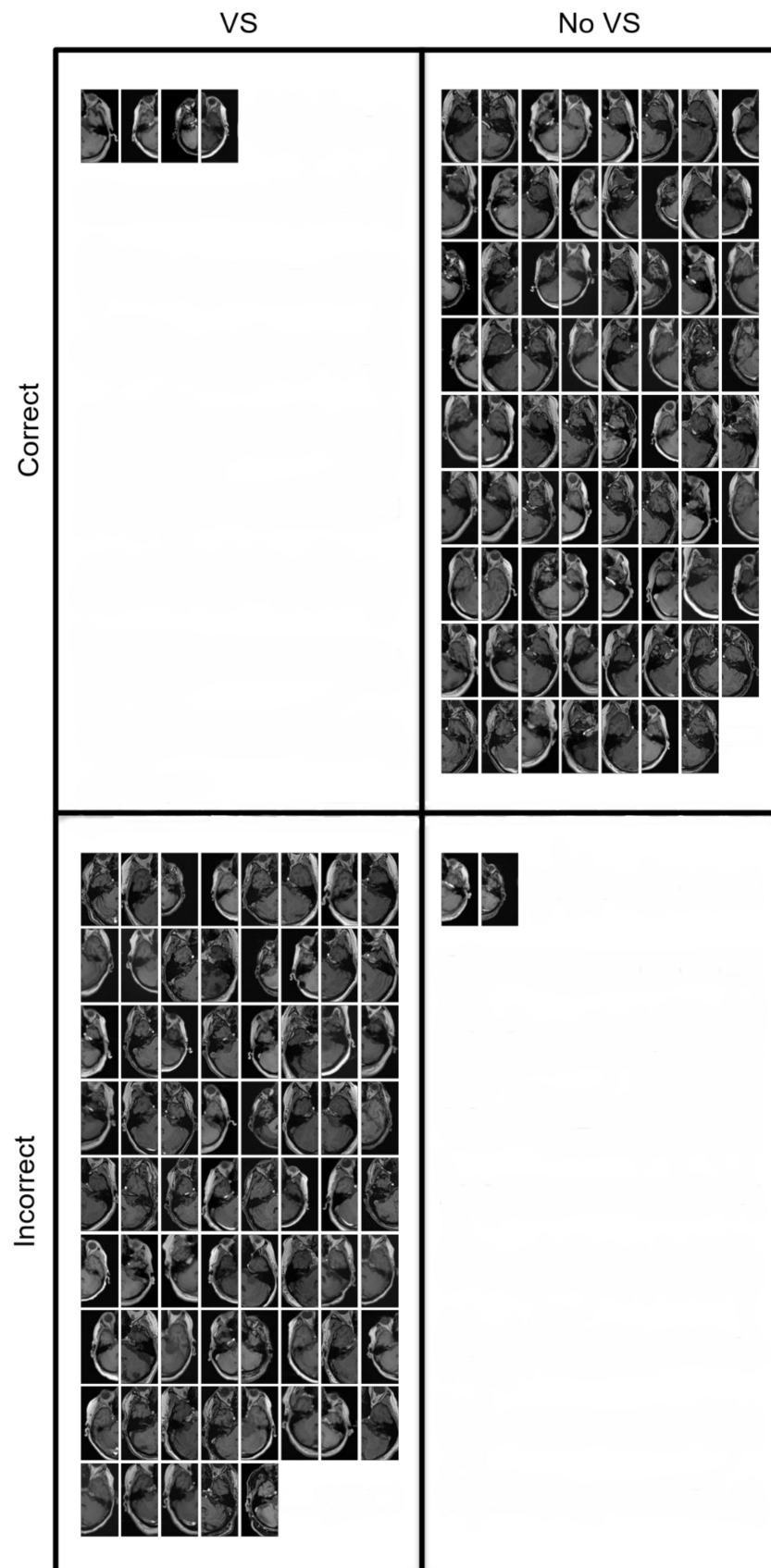


Figure 6. All MRI slices of the external validation data set with the T1 sequence. The images are sorted based on whether they contained a VS and whether they were correctly classified.



Figure 7. Sample MRI slices of the T1c data set with and without gradient-weighted class activation mapping (Grad-CAM). (a) The correctly classified images are shown above; (b) the incorrectly classified slices are shown below. Bright yellow and purple correspond with high and low activation, respectively.

The literature search resulted in 18 publications; no paper reported on the detection of VSs apart from the previously mentioned study by Sager et al. [20,28–44]. Out of all the search results, five publications investigated the automatic segmentation and volumetry of VSs [28–32]. This was, amongst others, to detect tumor growth for clinical management and to evaluate the treatment response after radiosurgery [28–32]. Seven papers examined the clinical factors of VSs including the prediction of postoperative complications, prolonged length of stay after resection, hearing preservation, recurrence, decision-making on treatment, and blood supply to the VS [33–39]. One paper focused on the segmentation and cystic components of a VS as well as the tumor response after radiosurgery [40]. Three reports explored the genetic aspects including gene expression, disease-related genes, pathways, and potential therapeutic drugs [41–43]. One publication reviewed the literature regarding the prophylaxis of deep venous thromboses [44].

4. Discussion

Overall, a high accuracy was achieved for both the internal and external validation using contrast enhancement, with only a minor reduction in accuracy in the latter. When comparing these results with those of Sager et al., the performance appeared to be fairly similar [20]. An accuracy of 0.949 (95% CI 0.935–0.963) was obtained for the internal validation in our work compared with 0.974 (95% CI: 0.960–0.988) in the previous study. Likewise, an accuracy of 0.912 (95% CI 0.866–0.958) was obtained for the external T1c validation in the present study in comparison with 0.928 (95% CI: 0.869–0.987) in Sager et al. [20].

In contrast, the CNN could not generalize the slices without contrast enhancement, which was contrary to the previous study. The accuracy was 0.514 (95% CI 0.433–0.595) in this study versus 0.795 (95% CI: 0.702–0.888) in Sager et al. [20]. We hypothesized that the network compared the anatomy of the two hemispheres, if available, to locate the tumor when no contrast enhancement was used. However, no comparison could be made in the case of split slices and the performance of the network dropped to being almost random. The performance of the model could be further analyzed when trained with non-contrast-enhanced slices.

These findings illustrated that, on one hand, laterality was an appropriate surrogate for tumor detection based on T1c slices. On the other hand, there were pitfalls to this approach; e.g., when using slices without contrast enhancement, which led to a significant increase in the prediction performance of the surrogate compared with the actual task.

There are many possible reasons for the high accuracy of our results regarding the contrast-enhanced slices of the external validation data. Extensive data augmentation was implemented for the training of the CNN. Moreover, the training data set contained highly heterogeneous MRI studies.

Only a few MRI slices were classified incorrectly regarding the contrast-enhanced external validation. The misclassified images mostly included slices with fairly small tumor portions or blood vessels in the cerebellopontine angle. Notably, the CNN tended to show more false-negative results (eleven slices) than false-positive results (two slices), as seen in the corresponding confusion matrix. Today, many MRIs use relatively thin slices. Consequently, most MRI scans include at least one slice with a larger tumor portion. Therefore, the CNN could be expected to detect the VS in the vast majority of cases.

Grad-CAM can be used to explain the decision-making process of ML models [27]. Thereby, a more widespread application of ML is supported in general but also in particular in medicine by removing its reputation as a black box [45]. Additionally, Grad-CAM can support the development of ML models [46]. Grad-CAM was implemented in this work to identify the region of interest to detect the VS. Additionally, Grad-CAM observations led to the hypothesis that a CNN could be applied to MRI slices without contrast enhancement by Sager et al. [20].

This study had several limitations. Firstly, split MRI slices were used to assess the feasibility of automatic tumor detection rather than entire MRI slices. However, this allowed a direct comparison with the results of Sager et al. [20]. Additionally, fewer MRI slices were needed to train and validate the model to assess the feasibility of the VS detection. Secondly, MR images were included of pretreated patients for training and internal validation. Therefore, a smaller amount of training data were required at the expense of consistency between the cohorts used for training and internal validation versus external validation. Thirdly, the study comprised a qualitative assessment of Grad-CAM to investigate the region of interest of the CNN whilst lacking a quantitative assessment. Further investigations are needed to determine the optimal hyperparameter settings for training.

This study had several strengths. The CNN was trained using highly inhomogeneous data including MRI scans from internationally assigned patients and the validation cohorts contained studies from different scanners with different field strengths and slice thicknesses. Furthermore, the training data were augmented using various techniques and the CNN was internally and externally validated with a sizable number of patients. Thereby, the

ability of the model to generalize the data from other institutions was assessed. This is, unfortunately, not frequently demonstrated in medical ML research [47].

Automatic tumor detection can potentially increase the efficiency of the clinical workflow by supporting radiologists. It can also provide a second opinion on the existence of a tumor. VS detection is of major significance to offer adequate treatment including surgery, fractionated radiotherapy, radiosurgery, and observations [19]. For example, single fraction stereotactic radiosurgery represents a highly effective treatment with little treatment-related toxicity for small to middle-sized VSs [48].

To build on the findings of this study, the performance of the network on whole MRI slices could be tested in a cohort that also contains bilateral VSs. Moreover, it could be investigated whether training on whole slices with and without VSs results in a different performance compared with using split images. Additionally, a model could be established that combines several VS predictions based on single slices from one patient into one final VS prediction. This combination would constitute a computationally inexpensive way to benefit from an inter-slice context without needing a more powerful graphics processing unit (GPU). Finally, a comparison between this model and a 3D-CNN could analyze the performance of the two models. The research results indicate that 3D-CNNs may perform better for certain tasks compared with 2D-CNNs [49].

5. Conclusions

This work demonstrated that it was feasible to detect VSs in split contrast-enhanced MRI slices without the need for segmentations. An accuracy of 0.912 was obtained for the external contrast-enhanced validation data set by using the mentioned training procedure and data; therefore, 2D-CNNs may constitute a promising option for tumor detection thanks to a potentially increased efficiency in data labeling and model training and no need for segmentations. However, further investigations are required on the comparison of 2D-CNNs with more complex architectures, in particular for more challenging research questions. Moreover, it needs to be examined whether split slices constitute an adequate replacement for tumor detection based on unmodified MRI slices.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers14092069/s1>, Supplementary Material S1: The programming code used to split the MRI slices and train, validate, and assess the network.

Author Contributions: Conceptualization, P.W.; data curation, P.S., C.S. and R.F.; formal analysis, C.K. and P.W.; funding acquisition, P.W.; investigation, F.E., C.F. and A.M.; methodology, C.K. and P.W.; project administration, P.W.; resources, D.R.Z. and A.M.; software, C.K. and P.W.; supervision, R.F., D.R.Z. and P.W.; validation, E.V., L.N., T.F. and P.M.P.; visualization, C.K.; writing—original draft, C.K.; writing—review and editing, P.S., P.M.P., F.E., C.F., A.M. and P.W. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by a BRIDGE—Proof of Concept research grant from the Swiss National Science Foundation (SNSF), grant number 195054, and Innosuisse.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki. All imaging data were handled in accordance with institutional policies, and approval by an institutional review board was obtained for the training as well as the internal validation cohort from the Ludwig Maximilian University of Munich (project 20-437, 22 June 2020) for a project on outcome modeling after radiosurgery, of which this study is a subproject. Ethical approval for the images in the validation cohort was waived by the Ethics Committee of Eastern Switzerland (EKOS 21/041, 9 March 2021) due to the fact that using single MRI slices constitutes sufficient anonymization.

Informed Consent Statement: Written informed consent for the analysis of the anonymized clinical and imaging data was obtained from all patients.

Data Availability Statement: The corresponding author will provide the data presented in this study upon reasonable request. In Figures 5 and 6, we provided the data used to externally validate the CNN.

Acknowledgments: F.E. is a participant in the BIH Charité Junior Clinician Scientist Program funded by the Charité—Universitätsmedizin Berlin, and the Berlin Institute of Health at Charité (BIH).

Conflicts of Interest: P.W. has a patent application titled ‘Method for detection of neurological abnormalities’. C.F. and F.E. received speaker honoraria from Accuray outside of the submitted work. The remaining authors declare no conflict of interest.

References

1. Angra, S.; Ahuja, S. Machine Learning and Its Applications: A Review. In Proceedings of the 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, India, 23–25 March 2017; pp. 57–60.
2. Valliani, A.A.-A.; Ranti, D.; Oermann, E.K. Deep Learning and Neurology: A Systematic Review. *Neurol. Ther.* **2019**, *8*, 351–365. [[CrossRef](#)] [[PubMed](#)]
3. Tiwari, A.; Srivastava, S.; Pant, M. Brain Tumor Segmentation and Classification from Magnetic Resonance Images: Review of Selected Methods from 2014 to 2019. *Pattern Recognit. Lett.* **2020**, *131*, 244–260. [[CrossRef](#)]
4. Ehret, F.; Kaul, D.; Clusmann, H.; Delev, D.; Kernbach, J.M. Machine Learning-Based Radiomics in Neuro-Oncology. *Acta Neurochir. Suppl.* **2022**, *134*, 139–151. [[PubMed](#)]
5. Abd-Ellah, M.K.; Awad, A.I.; Khalaf, A.A.M.; Hamed, H.F.A. A Review on Brain Tumor Diagnosis from MRI Images: Practical Implications, Key Achievements, and Lessons Learned. *Magn. Reson. Imaging* **2019**, *61*, 300–318. [[CrossRef](#)]
6. Wang, G.; Li, W.; Ourselin, S.; Vercauteren, T. Automatic Brain Tumor Segmentation Based on Cascaded Convolutional Neural Networks With Uncertainty Estimation. *Front. Comput. Neurosci.* **2019**, *13*, 56. [[CrossRef](#)]
7. Grøvik, E.; Yi, D.; Iv, M.; Tong, E.; Rubin, D.; Zaharchuk, G. Deep Learning Enables Automatic Detection and Segmentation of Brain Metastases on Multisequence MRI. *J. Magn. Reson. Imaging* **2020**, *51*, 175–182. [[CrossRef](#)]
8. Rehman, A.; Khan, M.A.; Saba, T.; Mehmood, Z.; Tariq, U.; Ayesha, N. Microscopic Brain Tumor Detection and Classification Using 3D CNN and Feature Selection Architecture. *Microsc. Res. Tech.* **2021**, *84*, 133–149. [[CrossRef](#)]
9. Sharif, M.I.; Li, J.P.; Khan, M.A.; Saleem, M.A. Active Deep Neural Network Features Selection for Segmentation and Recognition of Brain Tumors Using MRI Images. *Pattern Recognit. Lett.* **2020**, *129*, 181–189. [[CrossRef](#)]
10. Khan, M.A.; Ashraf, I.; Alhaisoni, M.; Damaševičius, R.; Scherer, R.; Rehman, A.; Bukhari, S.A.C. Multimodal Brain Tumor Classification Using Deep Learning and Robust Feature Selection: A Machine Learning Application for Radiologists. *Diagnostics* **2020**, *10*, 565. [[CrossRef](#)]
11. Liu, Z.; Tang, H.; Lin, Y.; Han, S. Point-Voxel CNN for Efficient 3D Deep Learning. *arXiv* **2019**, arXiv:1907.03739.
12. Zhou, Z.; Sanders, J.W.; Johnson, J.M.; Gule-Monroe, M.K.; Chen, M.M.; Briere, T.M.; Wang, Y.; Son, J.B.; Pagel, M.D.; Li, J.; et al. Computer-Aided Detection of Brain Metastases in T1-Weighted MRI for Stereotactic Radiosurgery Using Deep Learning Single-Shot Detectors. *Radiology* **2020**, *295*, 407–415. [[CrossRef](#)] [[PubMed](#)]
13. Zhang, M.; Young, G.S.; Chen, H.; Li, J.; Qin, L.; McFaline-Figueroa, J.R.; Reardon, D.A.; Cao, X.; Wu, X.; Xu, X. Deep-Learning Detection of Cancer Metastases to the Brain on MRI. *J. Magn. Reson. Imaging* **2020**, *52*, 1227–1236. [[CrossRef](#)] [[PubMed](#)]
14. Xu, J.; Li, M.; Zhu, Z. Automatic Data Augmentation for 3D Medical Image Segmentation. In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2020, Lima, Peru, 4–8 October 2020; Springer International Publishing: New York, NY, USA, 2020; pp. 378–387.
15. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. Mixup: Beyond Empirical Risk Minimization. *arXiv* **2017**, arXiv:1710.09412.
16. Kadry, S.; Damaševičius, R.; Taniar, D.; Rajinikanth, V.; Lawal, I.A. Extraction of Tumour in Breast MRI Using Joint Thresholding and Segmentation—A Study. In Proceedings of the 7th International conference on Bio Signals, Images and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; pp. 1–5.
17. Kadry, S.; Damaševičius, R.; Taniar, D.; Rajinikanth, V.; Lawal, I.A. U-Net Supported Segmentation of Ischemic-Stroke-Lesion from Brain MRI Slices. In Proceedings of the 7th International conference on Bio Signals, Images, and Instrumentation (ICBSII), Chennai, India, 25–27 March 2021; pp. 1–5.
18. Rehman, A.; Naz, S.; Razzak, M.I.; Akram, F.; Imran, M. A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning. *Circuits Syst. Signal Process* **2020**, *39*, 757–775. [[CrossRef](#)]
19. Goldbrunner, R.; Weller, M.; Regis, J.; Lund-Johansen, M.; Stavrinou, P.; Reuss, D.; Evans, D.G.; Lefranc, F.; Sallabanda, K.; Falini, A.; et al. EANO Guideline on the Diagnosis and Treatment of Vestibular Schwannoma. *Neuro Oncol.* **2019**, *22*, 31–45. [[CrossRef](#)]
20. Sager, P.; Näf, L.; Vu, E.; Fischer, T.; Putora, P.M.; Ehret, F.; Fürweger, C.; Schröder, C.; Förster, R.; Zwahlen, D.R.; et al. Convolutional Neural Networks for Classifying Laterality of Vestibular Schwannomas on Single MRI Slices—A Feasibility Study. *Diagnostics* **2021**, *11*, 1676. [[CrossRef](#)]
21. Howard, J.; Gugger, S. Fastai: A Layered API for Deep Learning. *Information* **2020**, *11*, 108. [[CrossRef](#)]
22. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2019; Volume 32.
23. Data Augmentation in Computer Vision. Available online: <https://docs.fast.ai/vision.augment.html> (accessed on 14 November 2021).

24. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A Survey of the Recent Architectures of Deep Convolutional Neural Networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [[CrossRef](#)]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
27. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
28. Shapey, J.; Kujawa, A.; Dorent, R.; Saeed, S.R.; Kitchen, N.; Obholzer, R.; Ourselin, S.; Vercauteren, T.; Thomas, N.W.M. Artificial Intelligence Opportunities for Vestibular Schwannoma Management Using Image Segmentation and Clinical Decision Tools. *World Neurosurg.* **2021**, *149*, 269–270. [[CrossRef](#)]
29. Lee, C.-C.; Lee, W.-K.; Wu, C.-C.; Lu, C.-F.; Yang, H.-C.; Chen, Y.-W.; Chung, W.-Y.; Hu, Y.-S.; Wu, H.-M.; Wu, Y.-T.; et al. Applying Artificial Intelligence to Longitudinal Imaging Analysis of Vestibular Schwannoma Following Radiosurgery. *Sci. Rep.* **2021**, *11*, 3106. [[CrossRef](#)] [[PubMed](#)]
30. Lee, W.-K.; Wu, C.-C.; Lee, C.-C.; Lu, C.-F.; Yang, H.-C.; Huang, T.-H.; Lin, C.-Y.; Chung, W.-Y.; Wang, P.-S.; Wu, H.-M.; et al. Combining Analysis of Multi-Parametric MR Images into a Convolutional Neural Network: Precise Target Delineation for Vestibular Schwannoma Treatment Planning. *Artif. Intell. Med.* **2020**, *107*, 101911. [[CrossRef](#)] [[PubMed](#)]
31. George-Jones, N.A.; Wang, K.; Wang, J.; Hunter, J.B. Automated Detection of Vestibular Schwannoma Growth Using a Two-Dimensional U-Net Convolutional Neural Network. *Laryngoscope* **2021**, *131*, E619–E624. [[CrossRef](#)] [[PubMed](#)]
32. Shapey, J.; Wang, G.; Dorent, R.; Dimitriadis, A.; Li, W.; Paddick, I.; Kitchen, N.; Bisdas, S.; Saeed, S.R.; Ourselin, S.; et al. An Artificial Intelligence Framework for Automatic Segmentation and Volumetry of Vestibular Schwannomas from Contrast-Enhanced T1-Weighted and High-Resolution T2-Weighted MRI. *J. Neurosurg.* **2019**, *134*, 171–179. [[CrossRef](#)]
33. Beyea, J.A.; Newsted, D.; Campbell, R.J.; Nguyen, P.; Alkins, R.D. Response To Letter To The Editor: “Artificial Intelligence and Decision-Making For Vestibular Schwannoma Surgery”. *Otol. Neurotol.* **2022**, *43*, e132–e133. [[CrossRef](#)]
34. Risbud, A.; Tsutsumi, K.; Abouzari, M. Artificial Intelligence and Decision-Making for Vestibular Schwannoma Surgery. *Otol. Neurotol.* **2022**, *43*, e131–e132. [[CrossRef](#)]
35. Dang, S.; Manzoor, N.F.; Chowdhury, N.; Tittman, S.M.; Yancey, K.L.; Monsour, M.A.; O’Malley, M.R.; Rivas, A.; Haynes, D.S.; Bennett, M.L. Investigating Predictors of Increased Length of Stay After Resection of Vestibular Schwannoma Using Machine Learning. *Otol. Neurotol.* **2021**, *42*, e584–e592. [[CrossRef](#)]
36. Cha, D.; Shin, S.H.; Kim, S.H.; Choi, J.Y.; Moon, I.S. Machine Learning Approach for Prediction of Hearing Preservation in Vestibular Schwannoma Surgery. *Sci. Rep.* **2020**, *10*, 7136. [[CrossRef](#)]
37. Abouzari, M.; Goshtasbi, K.; Sarna, B.; Khosravi, P.; Reutershan, T.; Mostaghni, N.; Lin, H.W.; Djalilian, H.R. Prediction of Vestibular Schwannoma Recurrence Using Artificial Neural Network. *Laryngoscope Investig. Otolaryngol.* **2020**, *5*, 278–285. [[CrossRef](#)]
38. Profant, O.; Bureš, Z.; Balogová, Z.; Betka, J.; Fík, Z.; Chovanec, M.; Voráček, J. Decision Making on Vestibular Schwannoma Treatment: Predictions Based on Machine-Learning Analysis. *Sci. Rep.* **2021**, *11*, 18376. [[CrossRef](#)]
39. Song, D.; Zhai, Y.; Tao, X.; Zhao, C.; Wang, M.; Wei, X. Prediction of Blood Supply in Vestibular Schwannomas Using Radiomics Machine Learning Classifiers. *Sci. Rep.* **2021**, *11*, 18872. [[CrossRef](#)] [[PubMed](#)]
40. Huang, C.-Y.; Peng, S.-J.; Wu, H.-M.; Yang, H.-C.; Chen, C.-J.; Wang, M.-C.; Hu, Y.-S.; Chen, Y.-W.; Lin, C.-J.; Guo, W.-Y.; et al. Quantification of Tumor Response of Cystic Vestibular Schwannoma to Gamma Knife Radiosurgery by Using Artificial Intelligence. *J. Neurosurg.* **2021**, 1–9. [[CrossRef](#)] [[PubMed](#)]
41. Huang, Q.; Zhai, S.-J.; Liao, X.-W.; Liu, Y.-C.; Yin, S.-H. Gene Expression, Network Analysis, and Drug Discovery of Neurofibromatosis Type 2-Associated Vestibular Schwannomas Based on Bioinformatics Analysis. *J. Oncol.* **2020**, *2020*, 5976465. [[CrossRef](#)] [[PubMed](#)]
42. Sass, H.C.R.; Borup, R.; Alanin, M.; Nielsen, F.C.; Cayé-Thomasen, P. Gene Expression, Signal Transduction Pathways and Functional Networks Associated with Growth of Sporadic Vestibular Schwannomas. *J. Neurooncol.* **2017**, *131*, 283–292. [[CrossRef](#)] [[PubMed](#)]
43. Agnihotri, S.; Gugel, I.; Remke, M.; Bornemann, A.; Pantazis, G.; Mack, S.C.; Shih, D.; Singh, S.K.; Sabha, N.; Taylor, M.D.; et al. Gene-Expression Profiling Elucidates Molecular Signaling Networks That Can Be Therapeutically Targeted in Vestibular Schwannoma. *J. Neurosurg.* **2014**, *121*, 1434–1445. [[CrossRef](#)] [[PubMed](#)]
44. Killeen, D.E.; Isaacson, B. Deep Venous Thrombosis Chemoprophylaxis in Lateral Skull Base Surgery for Vestibular Schwannoma. *Laryngoscope* **2020**, *130*, 1851–1853. [[CrossRef](#)] [[PubMed](#)]
45. Kundu, S. AI in Medicine Must Be Explainable. *Nat. Med.* **2021**, *27*, 1328. [[CrossRef](#)]
46. Windisch, P.; Weber, P.; Fürweger, C.; Ehret, F.; Kufeld, M.; Zwahlen, D.; Muacevic, A. Implementation of Model Explainability for a Basic Brain Tumor Detection Using Convolutional Neural Networks on MRI Slices. *Neuroradiology* **2020**, *62*, 1515–1518. [[CrossRef](#)]
47. Kim, D.W.; Jang, H.Y.; Kim, K.W.; Shin, Y.; Park, S.H. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean J. Radiol.* **2019**, *20*, 405–410. [[CrossRef](#)]

-
48. Windisch, P.Y.; Tonn, J.-C.; Fürweger, C.; Wowra, B.; Kufeld, M.; Schichor, C.; Muacevic, A. Clinical Results After Single-Fraction Radiosurgery for 1,002 Vestibular Schwannomas. *Cureus* **2019**, *11*, e6390. [[CrossRef](#)]
 49. Starke, S.; Leger, S.; Zwanenburg, A.; Leger, K.; Lohaus, F.; Linge, A.; Schreiber, A.; Kalinauskaite, G.; Tinhofer, I.; Guberina, N.; et al. 2D and 3D Convolutional Neural Networks for Outcome Modelling of Locally Advanced Head and Neck Squamous Cell Carcinoma. *Sci. Rep.* **2020**, *10*, 15625. [[CrossRef](#)] [[PubMed](#)]