

Attention-based Vision Transformer Enables Early Detection of Radiotherapy-Induced Toxicity in Magnetic Resonance Images of a Preclinical Model

Manish Kakar, PhD¹ , Bao Ngoc Huynh, PhD^{1,2} , Olga Zlygosteva, PhD³, Inga Solgård Juvkam, PhD^{1,4}, Nina Edin, PhD³, Oliver Tomic, PhD², Cecilia Marie Futsaether, PhD², and Eirik Malinen, PhD^{1,3}

Abstract

Introduction: Early identification of patients at risk for toxicity induced by radiotherapy (RT) is essential for developing personalized treatments and mitigation plans. Preclinical models with relevant endpoints are critical for systematic evaluation of normal tissue responses. This study aims to determine whether attention-based vision transformers can classify MR images of irradiated and control mice, potentially aiding early identification of individuals at risk of developing toxicity.

Method: C57BL/6J mice ($n = 14$) were subjected to 66 Gy of fractionated RT targeting the oral cavity, swallowing muscles, and salivary glands. A control group ($n = 15$) received no irradiation but was otherwise treated identically. T2-weighted MR images were obtained 3–5 days post-irradiation. Late toxicity in terms of saliva production in individual mice was assessed at day 105 after treatment. A pre-trained vision transformer model (ViT Base 16) was employed to classify the images into control and irradiated groups.

Results: The ViT Base 16 model classified the MR images with an accuracy of 69%, with identical overall performance for control and irradiated animals. The ViT's model predictions showed a significant correlation with late toxicity ($r = 0.65$, $p < 0.01$). One of the attention maps from the ViT model highlighted the irradiated regions of the animals.

Conclusions: Attention-based vision transformers using MRI have the potential to predict individuals at risk of developing early toxicity. This approach may enhance personalized treatment and follow-up strategies in head and neck cancer radiotherapy.

Keywords

attention-based vision transformer, radiotherapy, toxicity, mouse model, MRI, deep learning, explainable AI

Received: 10 January 2025; revised: 7 March 2025; accepted: 20 March 2025

Introduction

Patients with head and neck (H&N) cancer undergoing radiotherapy (RT) may experience serious side effects from radiation-induced damage to healthy tissue. Early adverse effects from RT include dermatitis and oral mucositis, which can occur during or shortly after RT and may require unwarranted treatment pause.¹ RT-induced dermatitis affects up to 95% of H&N cancer patients,² while oral mucositis affects about 80% and may require painkillers and in the worst cases also gastrostomy tube to aid in feeding.^{3,4} Late adverse effects include tissue fibrosis, osteoradionecrosis, and hypo-functioning salivary glands, all of which can significantly lower the patient's quality of life.^{1,5} Hypo-functioning salivary glands

leading to hyposalivation and/or xerostomia is estimated to

¹Department of Radiation Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway

²Faculty of Science and Technology, Norwegian University of Life Sciences, Ås, Norway

³Department of Physics, University of Oslo, Oslo, Norway

⁴Institute for Oral Biology, Faculty of Dentistry, University of Oslo, Oslo, Norway

Corresponding Author:

Manish Kakar, Department of Radiation Biology, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway.
Email: kmi@ous-hf.no



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and

distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

occur in more than 80% of H&N cancer patients undergoing RT, and can lead to negative effects on oral health causing loss of taste, difficulties in eating, speaking, and swallowing.^{6,7} Currently, few methods exist to identify patients at risk of early-stage toxicity, limiting opportunities to mitigate side effects.

Magnetic resonance imaging (MRI) is a non-invasive and versatile medical imaging technique with excellent soft tissue contrast, useful for assessing tumors and organs at risk visualizing tissue boundaries and extracting quantitative image biomarkers with prognostic or predictive potential.⁸ MR technology and sequences have improved significantly during the last decades, enabling evaluation of tissue properties such as diffusion and perfusion. T2-weighted MR sequences are standard for cancer diagnosis, as they highlight variations in water content, making tumors often appear hyperintense while also revealing hypointense features like fibrosis. While MRI can identify RT-induced tissue injuries typically occurring months to years after treatment,⁹ its usefulness in detecting toxicity during or early after treatment appears to be very limited. Thus, developing an MRI-based approach for detecting and predicting tissue effects and toxicities could offer significant benefits.

During the last decade, we have witnessed remarkable progress in artificial intelligence (AI) from machine learning to deep learning, with valuable applications in medical imaging.¹⁰ Recent developments in deep learning have shown promise for image-based outcome prediction in terms of radiotherapy-induced toxicity.^{11,12} Moreover, transfer learning could be a useful approach, particularly due to the challenges involved in obtaining large-scale datasets.¹³ In addition, explainable AI (XAI) has been introduced to provide models that are interpretable and more trustworthy.¹⁴ The attention-based vision transformer (ABVT)¹⁵ is a new type of neural network architecture, inspired by the transformer-based large language models like ChatGPT (Chat Generative Pre-trained Transformer).¹⁶ ABVTs have the ability to model long-range pixel dependencies, adapt to different input sizes, and leverage parallel processing, making them highly suitable for image classification tasks.¹⁷ Here, a weighted representation of the complete image is created by summing attention scores assigned to image patches via an attention mechanism. By using local attention, model efficiency can be improved and the attention maps determine the significance of individual patches. As the attention score indicates the importance of image patches, patches with high scores may align with classifier predictions. Therefore, in this study, we investigated the use of ABVT together with MRI as a tool to predict toxicity and to enhance explainability.

In this work, we hypothesize that subtle radiographic patterns in early post-RT MR images can indicate the risk of late radiation-induced toxicity, and that ABVT can predict its likelihood. T2 weighted MR images were acquired from a preclinical mouse study that employed a clinically relevant radiation field and fractionated radiation delivery to the head and neck of mice.¹⁸ A pre-trained ABVT model, the vision transformer (ViT) base 16 model,^{15,19} was used for binary classification of MR images into two classes, identifying the images as stemming from irradiated mice or controls. The predictive capability

of the image-based ABVT classifier was then tested against late toxicity, with respect to reduced saliva production, in the same mice. Furthermore, attention maps were generated and compared between control and irradiated mice, to reveal image regions of importance for the model.²⁰ Our assumption was that the model's attention, i.e. the relative weighing of the image information used for the classification task, is related to where in the images the model looks for radiation-induced alterations.

Materials and Methods

Mice and MRI

Nine-week-old C57BL/6J female mice purchased from Janvier labs (France), were kept in a 12-h light/12-h dark cycle under pathogen-free conditions and fed a standard commercial fodder with water given ad libitum (see¹⁸ for animal handling). Standard housing with nesting material and refuge was provided. All experiments were approved by Norwegian Food Safety Authority (ID 27931) and performed in accordance with directive 2010/63/EU on the protection of animals used for scientific purposes. Euthanasia was performed through overdose of anesthetic (Pentobarbital, Exagon[®] Vet) by intraperitoneal injection under terminal gas anesthesia (Sevoflurane 4% with O₂). Efforts to minimize the number of animals utilized and decrease their suffering was made. The animals were closely monitored throughout the experimental timeline and were adequately cared for. The reporting of this study conforms to ARRIVE 2.0 guidelines.^{21,22}

At the onset of experiments, animals were 12-weeks-old. A total of 29 mice were randomly assigned to a control group (n = 15) and an irradiated group (n = 14). X-irradiation was performed with a Faxitron Multirad 225 system (Faxitron Bioptics, Tucson, AZ, USA). A total dose of 66 Gy was given in 10 fractions with the following settings: 100 kV X-ray potential, 15 mA current, 2.0 mm Al filter, and 0.68Gy/min dose rate. The radiation field ($1.5 \times 0.75 \text{ cm}^2$), covering the oral cavity, swallowing muscles and salivary glands, was delivered over five days (two fractions per day) or twelve days (one fraction per day) (see¹⁸). The control group was not irradiated but was otherwise treated identically. Saliva production, used as a late toxicity endpoint, was assessed 105 days after RT. Saliva production was stimulated by intraperitoneal administration of 0.375 mg/kg pilocarpine²³ (Pilocarpine hydrochloride, Sigma) to 8 control and 6 irradiated mice under injection anesthesia (Zoletil-mix: Narcoxyl or Rompun[®] + Torbugesic[®] + Zoletil[®]), as described in¹⁸ and.²⁴ Saliva was collected into a cotton swab for 15 min. The swab was then centrifuged and the obtained volume was measured. The two fractionation schedules did not result in significant differences in saliva production.

MRI of the H&N region was performed using a 7.05 T Biospec scanner (Bruker Medical systems, Germany) with a fast T2 weighted spin-echo sequence, TurboRARE, with TE = 31 ms, TR = 3100 ms, and voxel size of $0.12 \times 0.12 \times 0.70 \text{ mm}^3$ in the sagittal, coronal and axial plane. Body

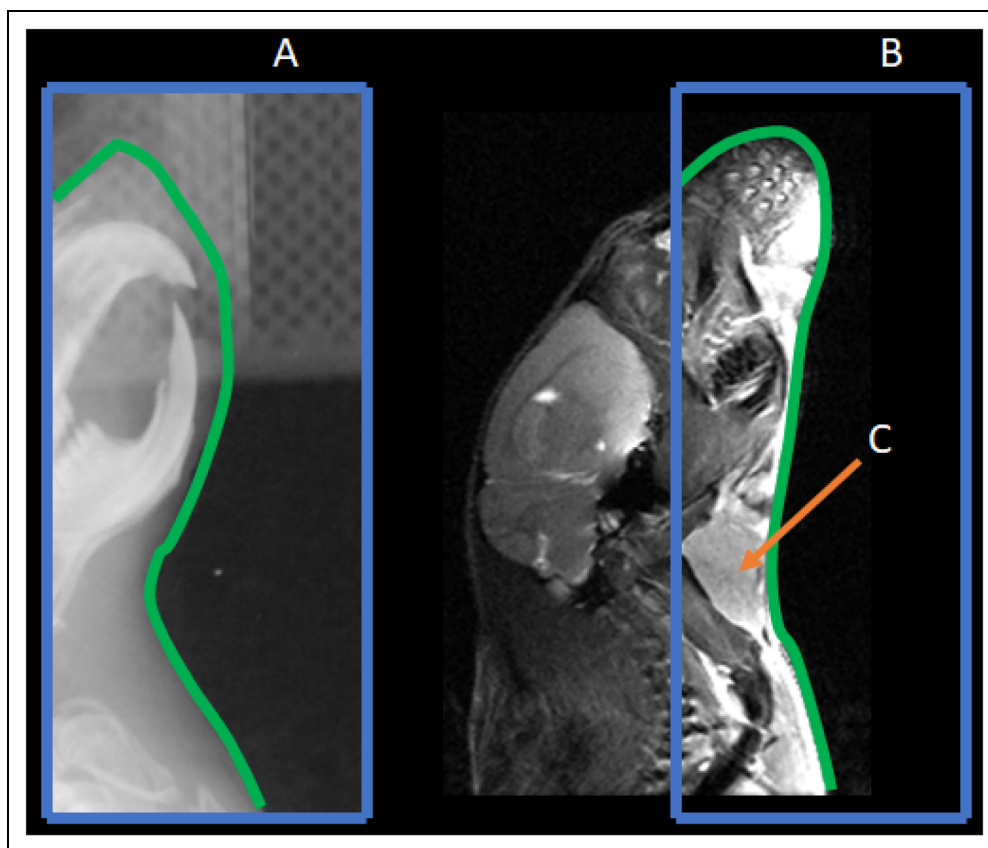


Figure 1. Transmission X-Ray Image Of The Irradiation Field Projected Through A Mouse (A) And A Corresponding T2-Weighted Mr Image (B). The Blue And Green Lines Indicate The Irradiation Field And The Midline Contour, Respectively. In (A), Large Parts Of The Mouse Are Not Seen Due To Lead Shielding. A Submandibular Gland (C), Responsible For Saliva Production, Is Indicated.

temperature was monitored and maintained at 37 degrees Celsius by a feedback-regulated heating fan. Respiration rate was monitored by a respiration probe. Animals were anesthetized during imaging using gas anesthesia with Sevoflurane 4% in O₂. For image analysis, 420 (14 mice×30 slices per mouse) and 450 (15 mice×30 slices) sagittal MR image slices from irradiated and control mice, respectively, were acquired. Figure 1 shows the treated area in a transmission X-ray image from the X-ray source used for treatment together with a T2-weighted MR image.

Data Preparation and Vision Transformer Model

The DICOM MR images were converted to uncompressed TIFF format using a batch conversion tool in IrfanView™ software. Some of the most lateral MR images containing virtually no anatomical information of interest (i.e. images not showing the mouse head or only minor parts of it) were manually removed from the dataset, reducing the image count to 300 for each class. Thereafter, the images were resized from 256 × 256 to 224 × 224 using bilinear interpolation, converted into 3-channel tensors, and then transformed using mean = [0.485, 0.456, 0.406] and standard deviation = [0.229, 0.224, 0.225], as required by PyTorch for the subsequent analysis.²⁵ Using random splitting

with repeated training-test splits, images were randomly assigned into training (approximately 70% of images) and test (approximately 30%) sets, stratified by class to ensure balance between control and irradiated mouse images in each set. To avoid information leakage, all images from a single mouse were assigned to only one set (either training or test set). Given the small sample size (29 mice), this procedure was repeated five times with different splits, creating five distinct training and test sets, and thereby generating five different models.

The vision transformer used herein was proposed by Dosovitskiy et al¹⁵ as an alternative neural network architecture to convolution neural networks (CNN). Given that ViT is a fairly large network (appx. 8.5 million parameters), a transfer learning strategy can be used for adapting the weights of the last layer for the given task.²⁶ We used the *ViT base 16 model* with 12 attention heads (with 16 × 16 patches) together with transfer learning. The model was pre-trained on IMAGENET1K_V1 weights.²⁵ We then examined the ViT base 16 model's ability to classify MR images as originating from a control or an irradiated mouse.

Computations were conducted using a NVIDIA RTX 4090 GPU with 16 GB graphics memory. We used a batch size of 32 with the Adam optimizer²⁷ with a learning rate of 0.000009 and weight decay of 0.0006. Training for 10 epochs

Table 1. Mean Test Set Performance Metrics ($\pm$ SEM) for the ViT Base Model. the Model was Trained and Tested on Five Different Training and Test Sets Obtained by Random Splitting.

	Accuracy	Precision	Recall	F1
Control	0.69 ± 0.04	0.70 ± 0.06	0.65 ± 0.03	0.66 ± 0.04
Irradiated	0.69 ± 0.04	0.68 ± 0.03	0.72 ± 0.08	0.69 ± 0.05

was sufficient to achieve the model's maximum accuracy. Only the final layer of the transfer learning model was modified and trained, while the other layers (pre-trained on IMAGENET1K_V1) were frozen. The model output was classification probability p_c , and for the binary classification task, the control class was labeled as "1" ($p_c \geq 0.5$) and the irradiated class as "0" ($p_c < 0.5$). The 12 attention maps generated for the central MR image for each mouse were extracted to identify patterns that could explain what the model highlighted in the images. Here, the maps were split at the coronal midline, approximately coinciding with the radiation field border. Furthermore, a region of interest (ROI) including the mouth and throat (irradiated part in treated animals) was defined inferiorly to the midline, while a ROI including the brain and nape (unirradiated part) was defined superiorly to the midline. The group mean attention values for each attention head and ROI was then calculated.

Performance and Evaluation Metrics

The accuracy in the classification is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

True positive (TP) is the number of positive samples that were also correctly predicted to be positive, false positive (FP) is the number of samples incorrectly predicted as positive, true negative (TN) is the number of samples correctly predicted as negative and false negative (FN) is the number of samples incorrectly classified as negative. Recall (also called sensitivity), precision and F1-score are given by²⁸:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Precision gives the confidence of the classifier to correctly classify images to their true class. Recall, often known as 'true positive rate' provides information about the fraction of samples the classifier has correctly classified to the true class out of all the samples available in that class. Therefore, precision provides an indication as to how precise the classifier is while recall measures the completeness of

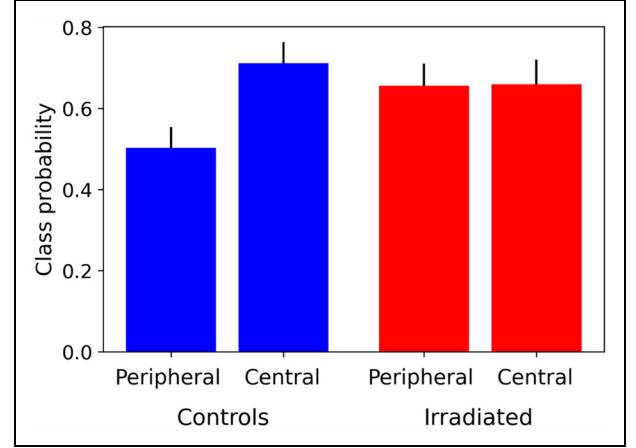


Figure 2. Mean Predicted Class Probability With Error Bars Indicating Sem For Control And Irradiated Mice Where Each Group Was Separated Into Central And Peripheral Mr Images.

classifying true samples. The F1 score combines the recall and precision using their harmonic mean.

Statistics

Group mean and standard error of the mean (SEM) are reported. Student's t-test was used to compare the mean values of the two groups, with a significance level of $p < 0.05$. Pearson's correlation coefficient r was calculated to assess associations between two variables.

Results

As seen in Table 1, the ViT Base 16 model showed good performance across the five splits of the data into training and test sets. The mean accuracy was 69% and was identical for irradiated and control animals. The control class had slightly higher mean precision than the irradiated class, though with higher standard error, whereas the irradiated class had higher recall and F1-score. However, none of these differences were statistically significant ($p > 0.05$).

To investigate the dependence of the classification accuracy on the MR slice location, we separated the image series of each mouse into central and peripheral images, where the former depicts submandibular glands and the vertebral column while the latter does not. The mean predicted class probability for central and peripheral images for control and irradiated animals is shown in Figure 2. For control mice, there was a significant difference ($p = 0.01$) between the central and peripheral images, but for the irradiated mice there was not.

The 12 attention maps corresponding to the central MR slice was extracted, visualized, and analyzed for all mice. Figure 3 shows an MR image illustrating the ROIs for the mouth and throat (irradiated part in treated animals) and the brain and nape (unirradiated part) along with attention maps for head # 9, averaged over each treatment group. For every attention map, the mean attention

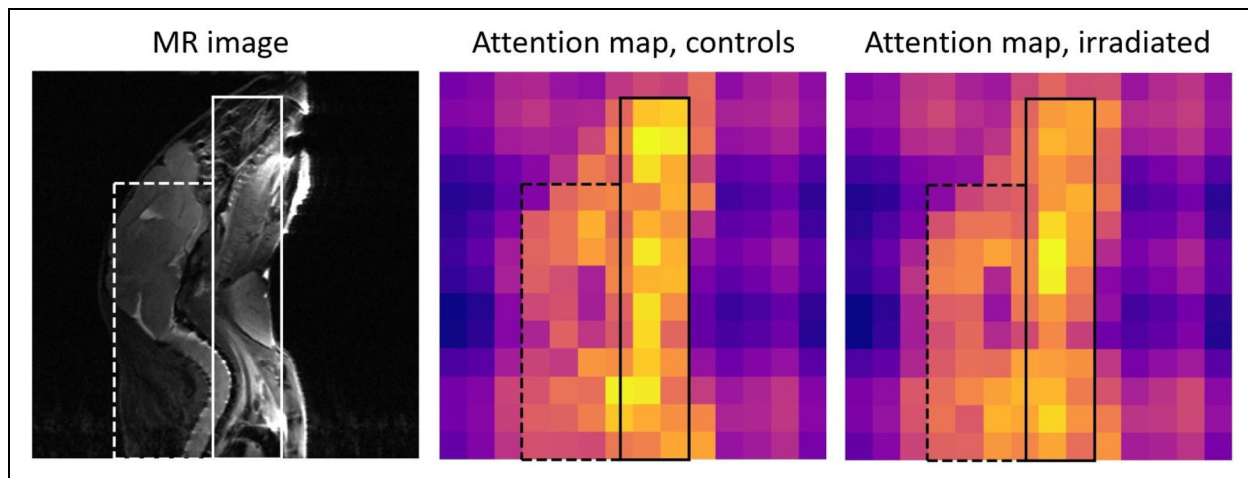


Figure 3. Central MR Image Of A Mouse (Left) And Mean Attention Maps For Head #9 For Control (Middle) And Irradiated Mice (Right). The Solid Line Represents The Roi Encompassing The Mouth And Throat (Irradiated Part In Treated Animals), While The Stippled Line Represents The Brain And Nape (Unirradiated Part).

value within each ROI was calculated for each treatment group. As seen in Figure 4, attention head #1, #4, #9, #10, and #12 had significant differences between the two ROIs for controls. Only head #9 had significant differences between the two ROIs in both controls and irradiated mice. As seen in Figure 3, the attention map for head #9 highlights the mouth and throat, corresponding to the irradiated part in the treated animals. There were generally small differences between the mean attention maps for the two treatment groups. Attention maps for heads #1 and #4 were quite noisy and predominantly highlighted air-tissue interfaces, while for heads #10 and #12 they appeared similar to #9 but with smaller differences between the ROIs (data not shown).

To investigate the relation between our image-based ViT model predictions and late toxicity in terms of reduced saliva production, we computed the mean predicted class probability over all images (including misclassified ones) for each mouse. Figure 5 shows the mean probability of an animal belonging to the control group plotted against the saliva production. As seen, there was a marked association, with a Pearson's correlation coefficient of 0.65 ($p=0.01$). A probability cutoff at 0.5 (also used in the image classification) and a saliva cutoff at 125 μL separated the two treatment groups with a classification accuracy of 0.86.

Discussion

We have shown that Vision Transformers, specifically the ViT Base 16, achieved a mean accuracy of approximately 0.69 in classifying MR images of control and irradiated mice. The transformer's predictions, based on images acquired very early after treatment, were associated with late toxicity in terms of reduced saliva production. Notably, one of the attention maps highlighted the mouth and throat, the irradiated area, demonstrating the model's ability to provide interpretable results. To our knowledge, these findings are novel and hold potential for future clinical applications.

One of the aims in this study was to investigate the attention maps and to determine which region(s) in the MR images that were highlighted. As recently reviewed,²⁹ explainability in ViTs is challenging because there is no clear standard for what constitutes an appropriate explanation. However, in the current work, one of the extracted attention maps predominantly pointed to the irradiated region, with focus on the submandibular glands, oral cavity, and throat. This substantiates that the ViT model used MR information from the irradiated region in the classification. However, the per-image classification varied and was poorer towards the periphery of the field of view for control images. This could be due to noise and little tissue-specific information, but it is not clear why this was not the case for irradiated animals. It may be that the irradiation caused pronounced effects throughout the animal, making it easier for the vision transformer to extract relevant image information. In future studies, it could be relevant to investigate more recent and sophisticated transformer architectures like Swin or Diet³⁰ in an effort to increase the model accuracy. Furthermore, it would be interesting to apply ViTs to MR images of patients with head and neck cancer and explore their role in a clinical setting, as these applications can potentially aid in detecting early post-irradiation effects.

Previously, it has been shown that the treatment protocol used in this study gives both acute and late toxicities in all the irradiated animals.^{18,24,31,32} Thus, the current classification task was in essence a prediction of the probability of developing toxicity. Indeed, we found a correlation between per-animal prediction probability, as based on MR images acquired early after irradiation, and late saliva production, indicating that the ViT classifier could potentially be used for toxicity prediction. The advantage of our approach is that radiation-induced toxicity may be detected early, allowing appropriate action to be taken to determine the best course of therapy. Furthermore, prompt information regarding the possible causes of any side effects that may have

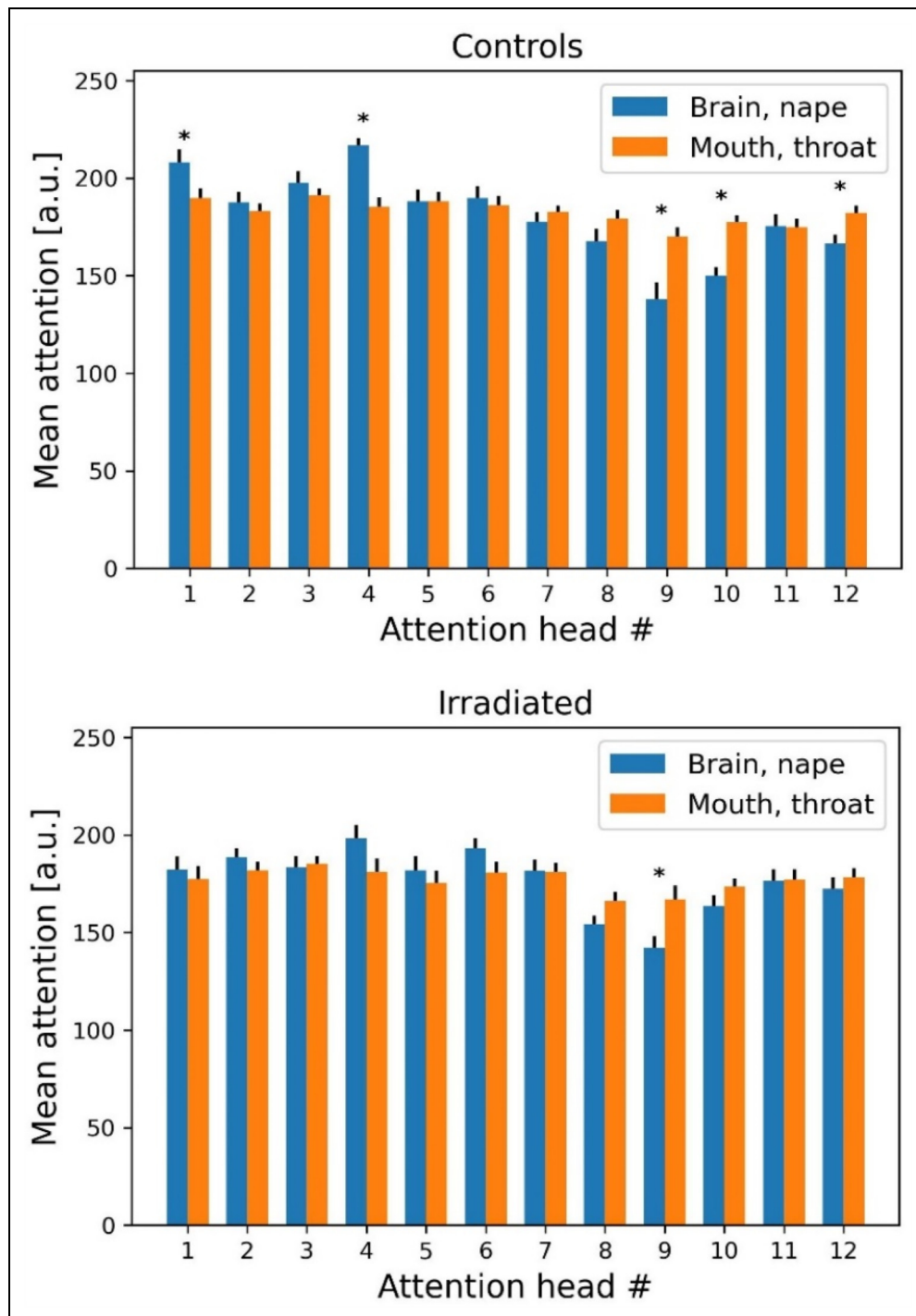


Figure 4. Mean Attention Value With Sem As Error Bar For The Mouth And Throat (Irradiated Part In Treated Animals) And Brain And Nape (Unirradiated Part) For Each Attention Head And Treatment Group. Data For Control And Irradiated Animals Are Given In The Upper And Lower Panel, Respectively. * $p < 0.05$.

resulted from radiation toxicity, and possible mitigation strategies, can be provided to the patient.

AI approaches such as deep learning on medical images have previously been used to predict toxicity in patients receiving radiotherapy (see^{11,33} for a review). Typically, these studies have employed pre-treatment, patient-specific radiation dose

distribution³³ and thus fail to account for inter-patient variations in radiosensitivity. Treatment monitoring by imaging and AI may potentially be able to identify radiosensitive individuals at an early stage after treatment. Also, attention mapping could point to regions in the patient having subtle radiation-induced changes that may progress to clinical toxicity at a later stage.

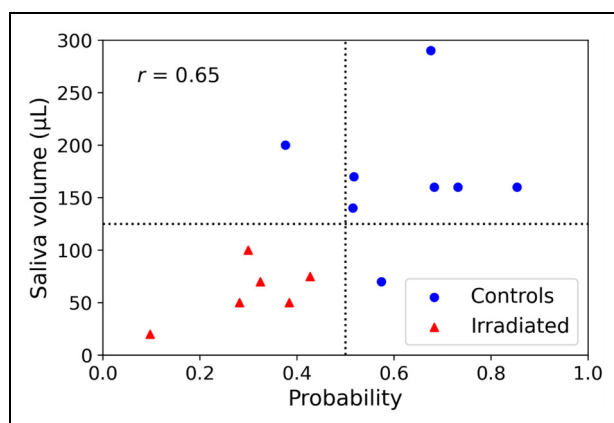


Figure 5. Late Saliva Production Per Mouse Versus the Image-Based ViT Probability of a Mouse Being Classified as a Control Mouse. The Dotted Lines Represent a Probability Cutoff at 0.5 and a Saliva Cutoff at 125 µL, Respectively, Used When Calculating Prediction Accuracy. Pearson's Correlation Coefficient r is Given.

To our knowledge, there are very few publications on the topic. Zhen et al used CNNs and unfolded dose maps for patients receiving radiotherapy of cervical cancer to highlight the possible rectum toxicity location.³⁴ Still, only pre-treatment dose maps were employed, disregarding any radiographic information that could indicate likely progression to toxicity in individual patients. In a very recent study, Kapoor et al used CNNs and integrated gradient techniques to identify regions in CT and dose images indicative of radiation pneumonitis for lung cancer patients receiving RT.³⁵ Still, the analysis employed CT images acquired at 3 and 6 months after RT, time points where clinical symptoms and radiographic CT changes are present and when it is too late for any early mitigation. Thus, there is a lack of methods that could indicate toxicity at an early stage during or immediately after RT.

The number of animals included in the study was a limitation, as small sample sizes can impact model robustness. However, we employed pre-training and transfer learning, which are expected to reduce the risk of overfitting. Additionally, by using random splitting with repeated training-test splits, we obtained a more reliable estimate of model performance compared to using a single training-test split. This approach helped minimize the likelihood of the model becoming overly tuned to specific animals. Moreover, we used MR images from only a single MR sequence (T2-weighted), and incorporating multiparametric MRI could have potentially enhanced model performance, although this would increase the risk of overfitting. Finally, when evaluating associations between the model prediction and saliva production, the saliva sampling procedure itself introduces inherent uncertainty, resulting in inter-animal variation of approximately 50%.¹⁸ Additionally, improving the interpretation of attention scores is important, as this could provide valuable insights into the underlying mechanisms of radiotherapy-induced toxicity. We also plan to investigate different vision transformer architectures, which may offer improved performance and

interpretation capabilities. Furthermore, we intend to test our methodology with a larger cohort of patients with H&N cancer, also incorporating multiparametric MRI. In conclusion, while our approach is still in its early stages, it is promising and lays the groundwork for future studies.

ORCID iDs

Manish Kakar <https://orcid.org/0009-0003-6951-6910>

Bao Ngoc Huynh <https://orcid.org/0000-0001-5210-132X>

Statements and Declarations

Ethical Considerations and Informed Consent Statements

All experiments were approved by the Norwegian Food Safety Authority (ID 27 931) and performed in accordance with directive 2010/63/EU on the protection of animals used for scientific purposes.

Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Author Contributions/CRediT

Manish Kakar: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original Draft

Bao Ngoc Huynh: Writing - Review & Editing, Formal analysis

Olga Zlygosteva: Writing - Review & Editing, Resources

Inga Solgård Juvkam: Writing - Review & Editing, Resources

Nina Edin: Writing - Review & Editing, Resources

Oliver Tomic: Writing - Review & Editing, Formal analysis

Cecilia Marie Futsaether: Writing - Review & Editing, Formal analysis

Eirik Malinen: Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing, Project administration.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability

The data that support the findings of this study are available from Prof. Eirik Malinen (eirik.malinen@fys.uio.no) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

References

1. Siddiqui F, Movsas B. Management of radiation toxicity in head and neck cancers. *Semin Radiat Oncol*. 2017;27(4):340-349.
2. Russi EG, Moretto F, Rampino M, et al. Acute skin toxicity management in head and neck cancer patients treated with radiotherapy and chemotherapy or EGFR inhibitors: Literature review and consensus. *Crit Rev Oncol Hematol*. 2015;96(1):167-182.
3. Shu Z, Zeng Z, Yu B, et al. Nutritional status and its association with radiation-induced oral mucositis in patients with

- nasopharyngeal carcinoma during radiotherapy: A prospective study. *Front Oncol, Sec. Rad. Oncol.* 2020;10:594687.
4. Maria OM, Eliopoulos N, Muanza T. Radiation-Induced oral mucositis. *Front Oncol Sec Radiat Oncol.* 2017;7: 89-112.
 5. Sroussi HY, Epstein JB, Bensadoun RJ, et al. Common oral complications of head and neck cancer radiation therapy: Mucositis, infections, saliva change, fibrosis, sensory dysfunctions, dental caries, periodontal disease, and osteoradionecrosis. *Cancer Med.* 2017;6(12):2918-2931.
 6. Chibly AM, Aure MH, Patel VN, Hoffman MP. Salivary gland function, development, and regeneration. *Physiol Rev.* 2022; 102(3):1495-1552.
 7. Jensen SB, Pedersen AM, Vissink A, et al. A systematic review of salivary gland hypofunction and xerostomia induced by cancer therapies: Prevalence, severity and impact on quality of life. *Support Care Cancer.* 2010;18(8):1039-1060.
 8. Hajjo R, Sabbah DA, Bardaweel SK, Tropsha A. Identification of tumor-specific MRI biomarkers using machine learning (ML). *Diagnostics (Basel).* 2021;11(5):742.
 9. Robbins ME, Brunso-Bechtold JK, Peiffer AM, Tsien CI, Bailey JE, Marks LB. Imaging radiation-induced normal tissue injury. *Radiat Res.* 2012;177(4):449-466.
 10. Chakraborty C, Bhattacharya M, Pal S, Lee S-S. From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. *Curr Res Biotechnol Rev.* 2024;7:100164.
 11. Yuan X, Ma C, Hu M, et al. Machine learning in image-based outcome prediction after radiotherapy: A review. *J Appl Clin Med Phys.* 2025;26(1):e14559.
 12. Guo W, Liu J, Dong F, et al. Review of machine learning and deep learning models for toxicity prediction. *Exp Biol Med (Maywood).* 2023;248(21):1952-1973.
 13. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: A literature review. *BMC Med Imaging.* 2022;22(1):69.
 14. Longo L, Brcic M, Cabitza F, et al. Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Inf Fusion Mini Rev.* 2024;106: 102301.
 15. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16(16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:201011929.* 2020.
 16. What's the next word in large language models? *Nat Mach Intell.* 2023;5(4):331-332. doi:10.1038/s42256-023-00655-z
 17. Mauricio J, Domingues I, Bernardino J. Comparing vision transformers and convolutional neural networks for image classification: A literature review. *Appl Sci.* 2023;13(9):5521.
 18. Juvkam IS, Zlygosteva O, Arous D, et al. A preclinical model to investigate normal tissue damage following fractionated radiotherapy to the head and neck. *J Radiat Res.* 2023;64(1):44-52. doi:10.1093/jrr/trac066
 19. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Proceedings 31st Conference on Neural Information Processing Systems (NIPS' 2017), Curran Assoc. Inc., Red Hook, NY, USA, 2017: 6000-6010.
 20. Tutek M, Šnajder J. Toward practical usage of the attention mechanism as a tool for interpretability. vol. 10, pp. 47011-47030, 2022, doi: 10.1109/ACCESS.2022.3169772.
 21. Percie du Sert N, Hurst V, Ahluwalia A, et al. The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *BMC Vet Res.* 2020;16(1):242. doi:10.1186/s12917-020-02451-y
 22. National Research Council Committee for the Update of the Guide for the C, Use of Laboratory A. The National Academies Collection: Reports funded by National Institutes of Health. *Guide for the Care and Use of Laboratory Animals.* National Academies Press (US), Copyright © 2011, National Academy of Sciences; 2011.
 23. Bagavant H, Trzeciak M, Papinska J, et al. A method for the measurement of salivary gland function in mice. *JoVE (J Visual Exp).* 2018;131(131):e57203.
 24. Zlygosteva O, Juvkam IS, Aass HCD, et al. Cytokine levels in saliva are associated with salivary gland fibrosis and hyposalivation in mice after fractionated radiotherapy of the head and neck. *Int J Mol Sci.* 2023;24(20):15218.
 25. Pytorch. https://pytorch.org/vision/main/models/vision_transformer.html.
 26. Zhuang F, Qi Z, Duan K, et al. A comprehensive survey on transfer learning. *Proc IEEE.* 2021;109(1):43-76.
 27. Kingma DP, Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* 2014.
 28. Kent A, Berry MM, Luehrs Jr. FU, Perry JW. Machine literature searching VIII. Operational criteria for designing information retrieval systems. *Am Doc.* 1955;6(2):93-101. <https://doi.org/10.1002/asi.5090060209>
 29. Stassin S, Corduant V, Mahmoudi SA, Siebert X. Explainability and evaluation of vision transformers: An in-depth experimental study. *Electronics (Basel).* 2024;13(1):175.
 30. Islam K. Recent advances in vision transformer: A survey and outlook of recent work. *arXiv preprint arXiv:220301536.* 2022.
 31. Zlygosteva O, Juvkam IS, Arous D, et al. Acute normal tissue responses in a murine model following fractionated irradiation of the head and neck with protons or X-rays. *Acta Oncol (Madr).* 2023;62(11):1574-1580.
 32. Juvkam IS, Zlygosteva O, Sitarz M, et al. Proton- compared to X-irradiation leads to more acinar atrophy and greater hyposalivation accompanied by a differential cytokine response. *Sci Rep.* 2024;14(1):22311. doi:10.1038/s41598-024-73110-7
 33. Tan D, Nasir NM, Manan HA, Yahya N. Prediction of toxicity outcomes following radiotherapy using deep learning-based models: A systematic review. *Cancer/Radiothérapie.* 2023;27(5):398-406.
 34. Zhen X, Chen J, Zhong Z, et al. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: A feasibility study. *Phys Med Biol.* 2017;62(21):8246-8263.
 35. Kapoor R, Sleeman IVW, Palta J, Weiss E. 3D Deep convolution neural network for radiation pneumonitis prediction following stereotactic body radiotherapy. *J Appl Clin Med Phys.* 2023; 24(3):e13875.