

Review

Recent Advances in Deep Learning and Medical Imaging for Head and Neck Cancer Treatment: MRI, CT, and PET Scans

Mathew Illimoottil¹ and Daniel Ginat^{2,*} 

¹ School of Medicine, University of Missouri-Kansas City, Kansas City, MO 64018, USA

² Department of Radiology, The University of Chicago, Chicago, IL 60637, USA

* Correspondence: dtg1@uchicago.edu

Simple Summary: Deep learning techniques have significant potential in head and neck cancer imaging, particularly in tumor detection, segmentation, and outcome prediction using magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) scans. Advanced deep learning methods, such as convolutional autoencoders, generative adversarial networks (GANs), and transformer models, have further enhanced imaging capabilities. Comparing deep learning and traditional techniques and the advantages and limits of each reveals their complementary roles in cancer management. Integrating radiogenomics with deep learning models promises further advancements in personalized care. However, challenges such as standardization, data quality, model overfitting, and computational requirements persist. Addressing these issues, integrating multimodal and temporal information, enhancing explainability, and conducting clinical validation are crucial for implementing deep learning models in head and neck cancer diagnosis and treatment. Overcoming these obstacles will pave the way for improved patient outcomes and personalized treatment strategies in head and neck cancer management.

Abstract: Deep learning techniques have been developed for analyzing head and neck cancer imaging. This review covers deep learning applications in cancer imaging, emphasizing tumor detection, segmentation, classification, and response prediction. In particular, advanced deep learning techniques, such as convolutional autoencoders, generative adversarial networks (GANs), and transformer models, as well as the limitations of traditional imaging and the complementary roles of deep learning and traditional techniques in cancer management are discussed. Integration of radiomics, radiogenomics, and deep learning enables predictive models that aid in clinical decision-making. Challenges include standardization, algorithm interpretability, and clinical validation. Key gaps and controversies involve model generalizability across different imaging modalities and tumor types and the role of human expertise in the AI era. This review seeks to encourage advancements in deep learning applications for head and neck cancer management, ultimately enhancing patient care and outcomes.

Keywords: magnetic resonance imaging (MRI); computed tomography (CT); positron emission tomography (PET); deep learning; radiomics; radiogenomics; convolutional neural networks (CNNs); artificial intelligence (AI); generative adversarial networks (GANs); head and neck; cancer; treatment



Citation: Illimoottil, M.; Ginat, D. Recent Advances in Deep Learning and Medical Imaging for Head and Neck Cancer Treatment: MRI, CT, and PET Scans. *Cancers* **2023**, *15*, 3267. <https://doi.org/10.3390/cancers15133267>

Academic Editors: Qiong Wang, Teng Huang and Yan Pang

Received: 17 April 2023

Revised: 25 May 2023

Accepted: 27 May 2023

Published: 21 June 2023



Copyright: © 2023 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/>).

1. Introduction

Head and neck cancers represent a heterogeneous group of malignancies that arise from various structures in the head and neck region, including the oral cavity, pharynx, larynx, nasal cavity, paranasal sinuses, and salivary glands [1]. These cancers are often associated with significant morbidity and mortality, and their management requires a multidisciplinary approach involving surgery, radiation therapy, and chemotherapy [2]. Accurate diagnosis, staging, and monitoring of head and neck cancers are crucial for determining the optimal treatment strategy, predicting patient outcomes, and assessing treatment response.

Medical imaging techniques, such as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) scans, have been indispensable in the assessment of head and neck cancers [3]. However, the interpretation of these images can be challenging due to the complex anatomy, diverse tumor biology, and overlapping features of malignant and benign lesions [4]. Moreover, traditional imaging techniques often rely on human expertise, which may be subjective and prone to variability.

Deep learning algorithms, particularly convolutional neural networks (CNNs), have shown potential in improving the accuracy, efficiency, and consistency of medical image analysis [5]. These advanced computational techniques can automatically learn hierarchical feature representations from raw imaging data, enabling the extraction of clinically relevant information and potentially surpassing human-level performance in certain tasks.

In this review, we delve into the limitations of traditional imaging techniques in head and neck cancer imaging and discuss advanced deep learning techniques, such as convolutional autoencoders for data compression, image denoising, and feature extraction; generative adversarial networks (GANs) for super-resolution; and transformer models such as the Vision Transformer (ViT). We also compare deep learning and traditional imaging techniques in head and neck cancer imaging, emphasizing their complementary roles and the potential benefits of integrating these methods.

In recent years, deep learning applications have expanded to include tumor detection, segmentation, classification, and response prediction, as well as the integration of multi-modal imaging data and the development of radiomics-based predictive models [6]. We explore the potential of deep learning to advance precision medicine in head and neck cancer management by enabling personalized treatment planning and outcome predictions. Additionally, we examine the emerging field of radiogenomics, which utilizes information about the relationship between imaging features and genomic information to create predictive models that can better guide medical treatment. We identify gaps in current research and propose future directions that may further enhance the impact of deep learning on the diagnosis and treatment of head and neck cancers.

This review aims to inform researchers and clinicians about the progress, potential, and obstacles in the application of deep learning techniques in head and neck cancer imaging, ultimately guiding future research toward improving patient outcomes and personalized treatment strategies. By incorporating the latest advancements and a comprehensive comparison of traditional and deep learning techniques, we provide a concise but informative understanding of the current landscape and future prospects in head and neck cancer imaging.

2. Limitations of Traditional Imaging Techniques in Head and Neck Cancer Imaging

Traditional imaging techniques, such as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET) scans, play crucial role in the diagnosis, staging, and monitoring of head and neck cancers [7]. However, these methods often rely on human expertise for image interpretation, which may be subjective and prone to variability [8]. In this section, we discuss the limitations of traditional imaging techniques in head and neck cancer imaging, focusing on the challenges associated with human expertise and the potential consequences for clinical decision-making.

2.1. Subjectivity and Interobserver Variability

Radiologists may assess medical images based on qualitative criteria, such as lesion shape, size, or intensity, which can lead to inconsistencies in diagnosis and staging among different observers [9]. For example, in the case of head and neck cancers, the complex anatomy and overlapping features of malignant and benign lesions can make accurate interpretation of imaging data challenging, even for experienced radiologists [10]. Additionally, manual delineation of tumor boundaries in imaging data can be time-consuming, and interobserver variability can affect the accuracy and reliability of the results [11].

2.2. Factors Affecting Human Expertise

Human interpretation of medical images may be influenced by factors such as fatigue, experience, and cognitive biases, which can contribute to discrepancies in decision-making [12]. For instance, studies have shown that radiologists' diagnostic performance can be affected by factors such as workload, time constraints, and confirmation bias, leading to potential errors in cancer detection and staging [13,14]. Furthermore, the lack of standardized guidelines and protocols for image interpretation in head and neck cancer imaging can exacerbate these issues, resulting in suboptimal clinical decision-making and patient management.

2.3. Need for Improved Imaging Techniques

Given the limitations of traditional imaging techniques and the potential impact on patient outcomes, there is a growing need for more accurate, efficient, and consistent methods for head and neck cancer imaging. Deep learning algorithms, particularly convolutional neural networks (CNNs), offer a promising solution to these challenges since they can automatically learn hierarchical feature representations from raw imaging data and potentially surpass human-level performance in certain tasks [15,16]. In the following sections, we discuss the recent advances in deep learning applications for head and neck cancer treatment, focusing on the use of MRI, CT, and PET scans.

3. Deep Learning in Medical Imaging

3.1. Deep Learning: A Brief Overview

Deep learning is a subfield of machine learning that involves the use of artificial neural networks with multiple layers, also known as deep neural networks, to automatically learn hierarchical representations of input data [15]. This advanced computational approach has shown remarkable success in various domains, including natural language processing, speech recognition, and computer vision. Convolutional neural networks (CNNs) are a type of deep learning architecture specifically designed for image analysis. They have demonstrated exceptional performance in various computer vision tasks, such as image classification, object detection, and segmentation. Figure 1 shows a lymph node metastasis segmented by a neural network algorithm. CNNs typically consist of multiple layers, including convolutional, pooling, and fully connected layers, that work together to automatically learn features and representations from raw image data, making them particularly suitable for medical image analysis, as shown in Figure 2. Some CNNs also include additional layers such as upsampling, which can be useful for image segmentation tasks.

3.2. Deep Learning in Medical Image Analysis: Performance Improvements over Traditional Methods

The application of deep learning techniques to medical image analysis has shown promising results in various domains, leading to significant advancements in the detection and diagnosis of diseases, the segmentation of anatomical structures, and the prediction of treatment outcomes [16]. Deep learning algorithms can learn complex patterns from medical images, generalizing well to new data and achieving human-level or even superior performance in many tasks [17]. This ability offers the potential to greatly improve the accuracy, efficiency, and consistency of medical image analysis, ultimately benefiting patient care and outcomes.

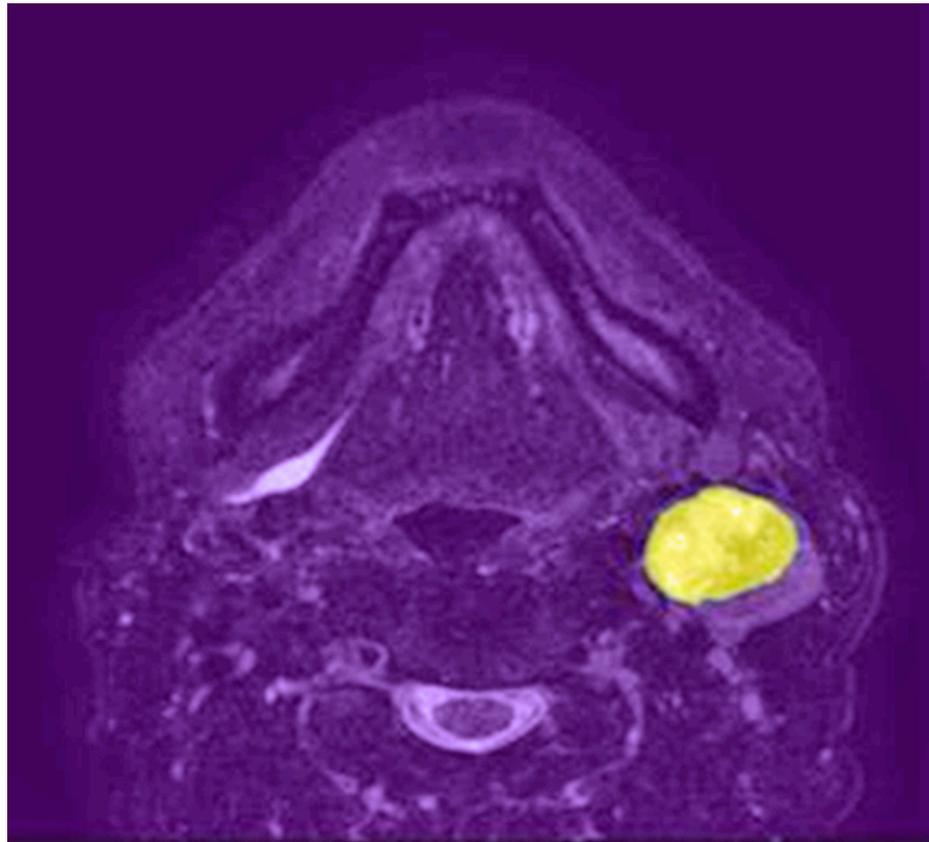


Figure 1. Axial T2-weighted MRI shows a left neck lymph node metastasis segmented by a neural network algorithm (highlighted in yellow).

The Basic Structure and Operation of CNNs

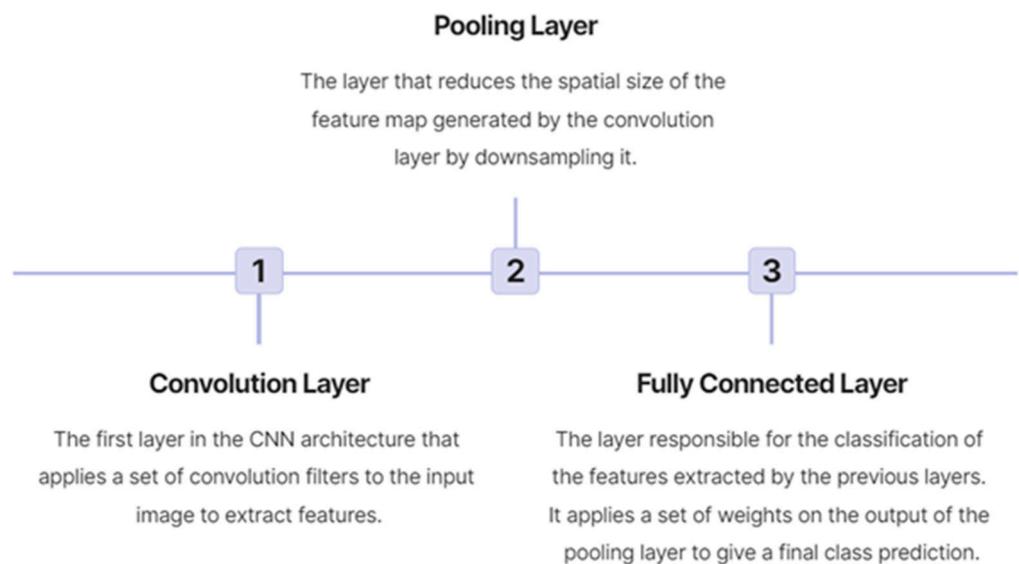


Figure 2. On the basic structure and operation of convolutional neural networks (CNNs).

Quantitative comparisons between deep learning models and traditional pre-deep learning imaging techniques have demonstrated substantial improvements in performance. For instance, in the diagnosis of diabetic retinopathy from retinal images, deep learning models have achieved sensitivity of 96.8% and specificity of 87.0%, significantly surpassing traditional methods with sensitivities of 49.3–85.5% and specificities of 71.0–93.4% [18]. Similarly, deep learning-based detection of pulmonary nodules on CT scans has exhibited higher accuracy (94.2%) than conventional computer-aided detection methods (79.1%) [19]. In the segmentation of brain tumors from MRI scans, deep learning models have achieved a Dice similarity coefficient of 0.88, outperforming traditional methods with coefficients ranging from 0.65 to 0.85 [20].

Moreover, deep learning has enabled the development of models that can integrate multi-modal and multi-scale imaging data, as well as clinical and demographic information, to generate more accurate and comprehensive predictions for patient outcomes and treatment responses [17]. This ability has contributed to the growing field of radiomics, which aims to extract and analyze high-dimensional quantitative features from medical images to build predictive models for personalized medicine [21]. As a result, deep learning is playing an increasingly important role in advancing precision medicine and improving patient care across a range of diseases and medical conditions. Figure 3 demonstrates the process by which CNNs are trained and fine-tuned.

Training and Fine-Tuning of CNNs

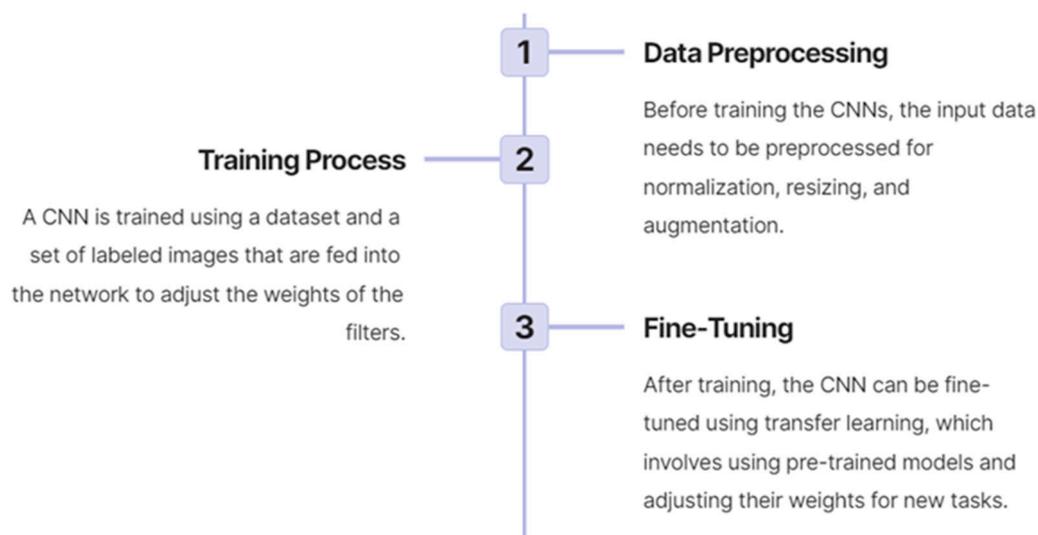


Figure 3. On the training and fine-tuning of convolutional neural networks (CNNs).

4. Deep Learning in Head and Neck Cancer Imaging

4.1. MRI

Magnetic resonance imaging (MRI) is a non-invasive imaging modality that provides excellent soft tissue contrast and high spatial resolution, making it a valuable tool for the assessment of head and neck cancers [22]. Deep learning techniques have been applied to various tasks in head and neck MRI, ranging from tumor detection and segmentation to treatment response prediction and prognosis assessment.

4.1.1. Tumor Detection and Segmentation

Convolutional neural networks (CNNs) have been employed for the automated detection and segmentation of head and neck tumors on MRI scans [23]. Recent studies have reported promising results in segmenting primary tumors and lymph nodes in patients with oropharyngeal cancer [24]. For instance, Men et al. (2017) reported a Dice similarity

coefficient (DSC) of 0.85 for primary tumor segmentation using a 3D CNN, compared to a DSC of 0.74 for an atlas-based approach [25]. Advanced CNN architectures, such as U-Net and its variants, have been particularly successful in these segmentation tasks [26].

However, challenges remain in segmenting tumors with irregular shapes and heterogeneous intensity profiles, which may require the development of more sophisticated deep learning models or the incorporation of additional information, such as clinical or demographic data, to improve segmentation performance [5].

4.1.2. Treatment Response Prediction and Prognosis Assessment

Deep learning models have also shown potential in predicting treatment response in head and neck cancer patients using MRI scans. Recent studies have demonstrated the ability of CNNs to predict outcomes, such as tumor regression, local control, and overall survival, from pre- and post-treatment MRI scans for various forms of cancer [27,28]. Some research in other forms of cancer has explored the use of multi-parametric MRI, combining multiple imaging sequences to improve the accuracy of treatment response prediction [29].

Radiomics, which involves the extraction of high-dimensional features from medical images, has been combined with deep learning for improved prediction of treatment response and prognosis [6]. This approach can lead to more personalized treatment plans and better monitoring of patient outcomes. However, the clinical utility of these models remains to be established through large-scale, prospective studies.

4.2. CT

Computed tomography (CT) scans are widely used in the diagnosis and staging of head and neck cancers due to their high spatial resolution, fast acquisition times, and ability to visualize bony structures [30]. Deep learning applications in head and neck cancer CT imaging include tumor detection and segmentation, outcome prediction, and treatment planning.

4.2.1. Tumor Detection and Segmentation

Several studies have reported the use of CNNs for the detection and segmentation of head and neck tumors in CT scans [31,32]. CNNs have demonstrated high accuracy in segmenting primary tumors and lymph nodes, as well as in delineating organs at risk, such as the spinal cord and parotid glands, for radiotherapy planning [16,25,31].

Techniques such as 3D CNNs and multi-scale learning have been employed to better capture the spatial relationships between structures and to improve segmentation performance [33].

However, the performance of these models can be affected by the presence of artifacts, such as metal implants and dental fillings, which are common on head and neck CT scans [34]. Future research may explore the use of deep learning techniques for artifact reduction or may develop models that are more robust to artifacts in the image data.

4.2.2. Outcome Prediction and Treatment Planning

Deep learning models have also been developed for predicting treatment outcomes in head and neck cancer patients using CT scans. A recent study demonstrated the ability of a CNN to predict overall survival and disease-free survival from pre-treatment CT scans in patients with nasopharyngeal carcinoma [35]. Moreover, deep learning models incorporating radiomic features have shown improved prediction of treatment response and survival [33].

In addition, deep learning has been applied to treatment planning, including the optimization of radiotherapy plans, for head and neck cancer patients. These models can automatically generate dose distribution predictions and identify optimal beam configurations to minimize radiation exposure to healthy tissues while maximizing tumor coverage [36]. However, the generalizability and clinical impact of these models warrant

further investigation, including multi-center studies and the evaluation of their performance in diverse patient populations.

4.3. PET

Positron emission tomography (PET) scans, which provide functional information about tumor metabolism and tissue perfusion, have gained increasing importance in the management of head and neck cancers [37]. Deep learning techniques have been applied to PET scans for tasks such as tumor detection, segmentation, and outcome prediction.

4.3.1. Tumor Detection and Segmentation

CNNs have shown promising results in the detection and segmentation of head and neck tumors on PET scans [24]. Methods such as transfer learning and the use of multi-modal data, combining PET with CT, or MRI, have been explored to improve the performance of deep learning models in this domain [38].

However, the performance of these models can be affected by the low spatial resolution and high noise levels inherent to PET imaging, as well as variations in image acquisition protocols and reconstruction algorithms [39]. Future research should focus on developing models that are more robust to these challenges and investigating strategies for harmonizing PET data from different sources.

4.3.2. Outcome Prediction and Treatment Monitoring

Deep learning models have also been explored for predicting treatment outcomes in head and neck cancer patients using PET scans. Studies have reported the potential of CNNs to predict tumor response, recurrence, and survival from pre- and post-treatment PET scans [40]. Some research has also examined the use of PET-derived radiomic features in combination with deep learning models for improved outcome prediction [41].

Furthermore, deep learning has been applied to monitor treatment response in head and neck cancer patients using PET scans. These models can potentially identify early metabolic changes that indicate treatment effectiveness, allowing for timely adjustments to treatment plans and improved patient outcomes [42]. However, further research is needed to validate these models in large, prospective cohorts and to establish their clinical utility.

5. Comparison of Deep Learning and Traditional Imaging Techniques in Head and Neck Cancer Imaging

While deep learning techniques have shown promise in various head and neck cancer imaging tasks, it is crucial to consider the specific situations in which traditional imaging techniques remain valuable and those in which deep learning tools are more appropriate.

5.1. Advantages of Traditional Imaging Techniques

1. **Interpretability:** Traditional imaging techniques provide interpretable results, as they are often based on well-established, handcrafted features and statistical methods [43]. This interpretability allows clinicians to better understand the rationale behind the decision-making process, which is essential for building trust in the results and ensuring appropriate clinical actions.
2. **Lower computational requirements:** Traditional imaging methods typically have lower computational demands compared to deep learning approaches [16], making them more accessible and easier to implement on standard workstations without the need for high-performance computing resources.
3. **Robustness to variations:** Traditional imaging techniques may be more robust to variations in imaging protocols and acquisition parameters [44] since they rely on well-established features that are less sensitive to changes in image quality and appearance.

5.2. Advantages of Deep Learning Techniques

1. Improved accuracy: Deep learning techniques, particularly CNNs, have demonstrated superior performance in many head and neck cancer imaging tasks, including tumor detection, segmentation, and outcome prediction, compared to traditional methods [25,27,45].
2. Automatic feature learning: Deep learning models can automatically learn relevant features from the data without the need for manual feature engineering [5], reducing the potential for human bias and enabling the discovery of novel imaging biomarkers that may not be apparent using traditional techniques.
3. Integration of multi-modal and multi-parametric data: Deep learning models can efficiently manage and integrate information from various imaging modalities (e.g., MRI, CT, and PET) and different image sequences [16], potentially providing a more comprehensive assessment of tumor characteristics and treatment response.

5.3. Guidance for Optimal Use

1. Preliminary analysis and simpler tasks: In situations in which a quick, preliminary analysis is required or for less complex tasks, traditional imaging techniques may be more suitable due to their lower computational demands and ease of implementation [16].
2. Interpretability and trust: When interpretability and trust in the decision-making process are critical, traditional methods may be more appropriate, as they provide more transparent and explainable results [46].
3. Resource-limited settings: In settings in which computational resources are limited, traditional imaging techniques may be more feasible since they generally have lower computational requirements [16].
4. Complex tasks and improved performance: For more complex tasks or when seeking improved accuracy and performance, deep learning techniques are more suitable [24,25,27,45]. This suitability includes for tasks such as tumor detection and segmentation, outcome prediction, and treatment planning in head and neck cancer imaging.

By carefully considering the specific situation and the desired outcomes, clinicians and researchers can select the most appropriate method for achieving accurate and reliable results in head and neck cancer imaging.

6. Gaps, Controversies, and Future Directions

6.1. Lack of Standardization and Benchmarking

One major challenge in the field of deep learning for head and neck cancer imaging is the lack of standardization and benchmarking. Diverse datasets, preprocessing techniques, model architectures, and training strategies have been used in different studies, making it difficult to compare the performance of various models and to assess their clinical utility [38]. Inconsistent reporting of model performance metrics, such as accuracy, sensitivity, and specificity, further complicates the comparison of results across studies. Future research should focus on establishing standardized datasets, evaluation metrics, and reporting guidelines to facilitate the benchmarking and comparison of deep learning models. In addition, promoting open science practices, such as sharing of data and code, can help to accelerate the development and validation of deep learning models for head and neck cancer imaging.

6.2. Integration of Multimodal, Temporal, and Clinical Information

Most deep learning studies in head and neck cancer imaging have focused on single-modality imaging data, such as MRI, CT, or PET scans. However, integrating multimodal information, such as combining functional and anatomical imaging data, can potentially improve the performance of deep learning models and provide a more comprehensive understanding of tumor characteristics [39]. Moreover, incorporating temporal informa-

tion from longitudinal imaging data can potentially enhance the prediction of treatment response, tumor recurrence, and patient outcomes [40]. In addition to imaging data, the integration of clinical information, such as patient demographics, tumor histology, and treatment details, can further improve the performance of deep learning models in head and neck cancer management. Future research should explore the development of deep learning models that can effectively integrate multimodal, temporal, and clinical information for personalized treatment planning and prognostication.

6.3. Explainability, Interpretability, and Trustworthiness

Deep learning models, particularly CNNs, are often considered “black boxes” due to their complex architectures and the lack of transparency in the decision-making process [47]. This lack of transparency can hinder the adoption of deep learning techniques in clinical practice since clinicians may be reluctant to trust a model’s predictions without understanding the underlying reasoning. Developing explainable and interpretable deep learning models is crucial to bridge this gap and promote their acceptance in the medical community [41]. Techniques such as attention mechanisms, layer-wise relevance propagation, and visualization of feature maps can help to elucidate the reasoning behind a model’s predictions and build trust among clinicians. Future research should focus on incorporating explainability and interpretability into the design of deep learning models for head and neck cancer imaging, as well as on developing methods to assess the trustworthiness and robustness of these models in the face of noisy, incomplete, or adversarial data.

6.4. Clinical Implementation, Validation, and Impact Assessment

The translation of deep learning models from research to clinical practice requires rigorous validation and assessment of their impact on patient outcomes [42]. Large-scale, prospective studies would be useful to establish the performance, generalizability, and clinical utility of deep learning models for head and neck cancer imaging [25]. These studies should involve diverse patient populations and imaging data from multiple centers to ensure the robustness of the models in real-world settings. Moreover, the integration of these models into existing clinical workflows, the assessment of their cost-effectiveness, and the evaluation of their impact on patient care, such as the reduction in diagnostic errors, optimization of treatment planning, and improvement of patient outcomes, are essential steps toward their successful implementation [48]. Future research should also explore the development of user-friendly, scalable, and secure software tools and platforms that can facilitate the deployment of deep learning models in clinical settings and enable their widespread adoption in head and neck cancer management.

7. Challenges and Limitations of Deep Learning Models in Head and Neck Cancer Imaging

7.1. Data Quality and Quantity

The performance of deep learning models heavily relies on the quality and quantity of the training data [5]. Obtaining large, diverse, and well-annotated datasets for head and neck cancer imaging can be challenging due to factors such as privacy concerns, data sharing restrictions, and the time-consuming nature of manual annotation by experts [49]. Additionally, variations in imaging protocols, scanner types, and image acquisition parameters across different institutions can lead to inconsistencies in the data, affecting model performance [50]. Future research should focus on developing methods to leverage smaller or imperfect datasets, explore data augmentation techniques, and promote multi-institutional collaboration to enhance the performance of deep learning models.

7.2. Model Overfitting

Deep learning models are prone to overfitting, especially when trained on small datasets, leading to poor generalization to new data [6]. Overfitting can result in models that perform well on the training data but fail to accurately predict outcomes for new

patients, limiting their clinical utility. Regularization techniques and model architectures that mitigate overfitting, such as dropout, batch normalization, and transfer learning, should be considered when developing deep learning models for head and neck cancer imaging [51]. Moreover, the use of cross-validation and external validation cohorts can help to evaluate and improve model generalizability.

7.3. Computation Requirements

The training and deployment of deep learning models often require substantial computational resources, such as graphics processing units (GPUs) and specialized hardware [5]. This need may pose a challenge for small clinical centers and researchers with limited access to high-performance computing facilities. Developing efficient model architectures, exploring strategies for model compression and acceleration, and utilizing cloud-based platforms for training and deployment can help to overcome these challenges and make deep learning models more accessible to a broader range of institutions and researchers.

7.4. Ethical Considerations

The use of deep learning models in medical imaging raises several ethical concerns, including those concerning data privacy, algorithmic fairness, and accountability [52]. Ensuring the protection of patient data is paramount, and researchers should adhere to data protection regulations and employ techniques such as data anonymization and secure data storage to safeguard patient privacy [53]. Moreover, biases in training data can lead to unfair treatment recommendations or misdiagnoses for certain patient populations, making it essential to address potential biases in the development and deployment of deep learning models. Establishing transparent and accountable practices, such as reporting model performance across diverse patient groups and involving stakeholders in the model development process can help to ensure ethical and responsible use of these technologies in head and neck cancer imaging.

8. Radiogenomics: A Promising Avenue for Deep Learning in Head and Neck Cancer Imaging

8.1. Radiogenomics: An Overview

Radiogenomics is an emerging interdisciplinary field that investigates the relationship between imaging features and genomic information, with the goal of developing predictive models that can guide personalized treatment decisions [10]. By integrating radiogenomic data with deep learning models, researchers and clinicians can gain a more comprehensive understanding of tumor biology, heterogeneity, and treatment response, ultimately leading to improved patient outcomes in head and neck cancer management.

8.2. Radiogenomic Features in Deep Learning Models

Several studies have explored the incorporation of radiogenomic features into deep learning models for head and neck cancer imaging [25,33]. Combining radiomic features extracted from MRI, CT, or PET scans with genomic data, such as gene expression profiles or mutational status, can improve the accuracy of tumor detection, segmentation, and outcome prediction [54]. Additionally, deep learning models can be designed to automatically learn radiogenomic representations from the imaging data, potentially revealing novel imaging-genomic associations that may have clinical implications [54]. This approach could lead to the identification of previously unknown biomarkers and therapeutic targets in head and neck cancers.

8.3. Challenges and Future Directions

Despite the promising potential of radiogenomics in deep learning for head and neck cancer imaging, several challenges need to be addressed. First, the acquisition of genomic data can be costly and invasive, limiting their availability in clinical settings [55]. Non-invasive methods, such as liquid biopsy, may offer a more accessible source of genomic

information in the future [56]. Second, the integration of imaging and genomic data presents computational challenges since these data types have different scales and dimensions [10]. Developing novel deep learning architectures and data fusion techniques that can effectively combine and process multimodal data is a crucial area for future research. Additionally, addressing the issues of data heterogeneity, missing data, and standardization of imaging and genomic data will be essential for the successful application of radiogenomics in deep learning models [57].

8.4. Clinical Implementation and Validation

As with any novel approach in medical imaging, the translation of radiogenomic-based deep learning models to clinical practice requires rigorous validation and assessment of their impact on patient outcomes. Large-scale, multi-institutional studies would be useful to establish the performance, generalizability, and clinical utility of these models in head and neck cancer management [16]. Furthermore, the integration of radiogenomic-based models into existing clinical workflows and the assessment of their cost-effectiveness are essential steps toward their successful implementation. As radiogenomics continues to evolve, it is expected that deep learning models incorporating radiogenomic features will play an increasingly important role in guiding personalized treatment strategies for head and neck cancer patients.

9. Advanced Deep Learning Techniques in Head and Neck Cancer Imaging

9.1. Convolutional Autoencoders for Data Compression, Image Denoising, and Feature Extraction

Convolutional autoencoders (CAEs) are a type of deep learning model that can be used for unsupervised feature learning, image denoising, and data compression [58]. They consist of an encoder network, which reduces the input image to a compact representation, and a decoder network, which reconstructs the image from this representation. CAEs have been applied in various medical imaging tasks, including head and neck cancer imaging, for improving image quality and reducing noise, as well as extracting relevant features that can be used in downstream analysis [59,60].

9.2. Generative Adversarial Networks for Super-Resolution

Generative adversarial networks (GANs) are a class of deep learning models that consist of two competing networks: a generator and a discriminator [61]. The generator creates synthetic images, while the discriminator evaluates the quality of these images, aiming to distinguish between real and generated images. GANs have been used for various medical image enhancement tasks, such as super-resolution, in which the goal is to generate high-resolution images from low-resolution input data. In the context of head and neck cancer imaging, GANs can be utilized to improve image resolution, potentially leading to more accurate tumor detection and segmentation [62,63].

9.3. Transformer Models: Vision Transformer (ViT)

Transformer models, initially introduced for natural language processing tasks, have been adapted for image analysis with the development of the Vision Transformer (ViT) [64]. ViT divides an image into non-overlapping patches, linearly embeds them into a sequence of fixed-size vectors, and processes these vectors using the transformer architecture. This approach allows the model to capture long-range dependencies and spatial relationships between image regions. ViT has shown competitive performance in various medical imaging tasks, including head and neck cancer detection and segmentation, and it can be considered an alternative to traditional CNN-based methods [65,66].

10. Conclusions

Deep learning techniques have made significant strides in the field of head and neck cancer imaging, demonstrating exceptional performance in critical tasks, such as tumor detection, segmentation, and outcome prediction, using various imaging modalities, such

as MRI, CT, and PET scans. Advanced deep learning techniques, including convolutional autoencoders, generative adversarial networks (GANs), and transformer models such as the Vision Transformer (ViT), have further enhanced the capabilities of these models, offering promising improvements in data compression, image denoising, super-resolution, and feature extraction. The incorporation of radiogenomics into deep learning models offers a promising avenue for further advancements, potentially enabling a more comprehensive understanding of tumor biology and heterogeneity and guiding personalized treatment strategies in head and neck cancer management.

Despite these encouraging results, a number of challenges must be addressed to facilitate the successful implementation of deep learning models in head and neck cancer diagnosis and treatment. One major obstacle is the lack of standardization in terms of datasets, preprocessing techniques, and model architectures, rendering it difficult to compare the performances of various models and assess their clinical utility. Additionally, acquiring large, diverse, and well-annotated datasets for head and neck cancer imaging poses a challenge due to privacy concerns, data sharing restrictions, and the labor-intensive nature of manual annotation by experts.

In light of the limitations of traditional imaging techniques, such as subjectivity and variability in human interpretation, the integration of deep learning and traditional imaging techniques could lead to more robust and accurate diagnosis and treatment planning. Recognizing the complementary roles of these approaches is essential for advancing head and neck cancer imaging and management.

Model overfitting, a common issue in deep learning, is another concern, particularly when models are trained on small datasets. This situation can lead to poor generalization from new data and reduced clinical utility. Moreover, the computational requirements for training and deploying deep learning models can be substantial, often necessitating specialized hardware and resources that may not be readily accessible to smaller clinical centers or researchers. Addressing these challenges, along with integrating multimodal and temporal information from different imaging modalities and timepoints, is crucial for further enhancing the performance of deep learning models. Developing explainable and interpretable models is also essential to bridge the gap between complex model architectures and clinical decision-making, fostering trust and acceptance among medical professionals.

Finally, conducting rigorous clinical validations through large-scale, prospective studies is necessary to establish the performance, generalizability, and clinical utility of deep learning models for head and neck cancer imaging. Overcoming these obstacles, while embracing the synergy between advanced deep learning techniques and traditional imaging approaches, will be instrumental in paving the way for improved patient outcomes and the realization of personalized treatment strategies in head and neck cancer management.

Author Contributions: Writing—review and editing, M.I. and D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This article was prepared in part using ChatGPT.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chow, L.Q.M. Head and neck cancer. *N. Engl. J. Med.* **2020**, *382*, 60–72. [[CrossRef](#)] [[PubMed](#)]
2. Marur, S.; Forastiere, A.A. Head and neck squamous cell carcinoma: Update on epidemiology, diagnosis, and treatment. *Mayo Clin. Proc.* **2016**, *91*, 386–396. [[CrossRef](#)]
3. Chiesa-Estomba, C.M.; Lequerica-Fernández, P.; Fakhry, N. Imaging in head and neck cancer: United in diversity. *Cancers* **2019**, *11*, 1077.
4. Wang, J.; Wu, C.J.; Bao, M.L.; Zhang, J.; Wang, X.N.; Zhang, Y.D. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur. Radiol.* **2017**, *27*, 4082–4090. [[CrossRef](#)]

5. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)] [[PubMed](#)]
6. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **2016**, *278*, 563–577. [[CrossRef](#)]
7. Pignon, J.P.; le Maître, A.; Maillard, E.; Bourhis, J. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. *Radiother. Oncol.* **2009**, *92*, 4–14. [[CrossRef](#)]
8. Alizadeh, M.; Safaei, A.; Dorri, G.; Zendehbad, S.A. Application of deep learning and machine learning methods to radiomics and texture analysis of medical images: A review. *J. Biomed. Phys. Eng.* **2018**, *8*, 405–424.
9. Petrick, N.; Sahiner, B.; Armato, I.I.S.G.; Bert, A.; Correale, L.; Delsanto, S.; Freedman, M.T.; Fryd, D.; Gur, D.; Hadjiiski, L.; et al. Evaluation of computer-aided detection and diagnosis systems. *Med. Phys.* **2013**, *40*, 087001. [[CrossRef](#)]
10. Lambin, P.; Rios-Velazquez, E.; Leijenaar, R.; Carvalho, S.; Van Stiphout, R.G.; Granton, P.; Zegers, C.M.; Gillies, R.; Boellard, R.; Dekker, A.; et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **2012**, *48*, 441–446. [[CrossRef](#)]
11. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [[CrossRef](#)]
12. Krupinski, E.A. Current perspectives in medical image perception. *Atten. Percept. Psychophys.* **2010**, *72*, 1205–1217. [[CrossRef](#)]
13. Drew, T.; Vo ML, H.; Olwal, A.; Jacobson, F.; Seltzer, S.E.; Wolfe, J.M. Scanners and drillers: Characterizing expert visual search through volumetric images. *J. Vis.* **2013**, *13*, 3. [[CrossRef](#)] [[PubMed](#)]
14. Berbaum, K.S.; Franken, E.A.; Dorfman, D.D.; Miller, E.M.; Caldwell, R.T.; Kuehn, D.M.; Berbaum, M.L. Role of faulty decision making in the satisfaction of search effect in chest radiography. *Acad. Radiol.* **2000**, *7*, 1098–1106. [[CrossRef](#)] [[PubMed](#)]
15. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
16. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [[CrossRef](#)]
17. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
18. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)]
19. Ciompi, F.; Chung, K.; Van Riel, S.J.; Setio, A.A.; Gerke, P.K.; Jacobs, C.; Scholten, E.T.; Schaefer-Prokop, C.; Wille, M.M.; Marchiano, A.; et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci. Rep.* **2017**, *7*, 46479. [[CrossRef](#)]
20. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [[CrossRef](#)]
21. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)]
22. Czernin, J.; Allen-Auerbach, M.; Schelbert, H.R. Improvements in cancer staging with PET/CT: Literature-based evidence as of September 2006. *J. Nucl. Med.* **2007**, *48*, 78S–88S.
23. Ibragimov, B.; Xing, L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med. Phys.* **2017**, *44*, 547–557. [[CrossRef](#)]
24. Zhong, Z.; Kim, Y.; Zhou, L.; Plichta, K. Deep-learning based, automated segmentation of head and neck cancer in PET/CT images. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 177–181.
25. Men, K.; Dai, J.; Li, Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med. Phys.* **2017**, *44*, 6377–6389. [[CrossRef](#)]
26. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 234–241.
27. Lao, J.; Chen, Y.; Li, Z.C.; Li, Q.; Zhang, J. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci. Rep.* **2017**, *7*, 10353. [[CrossRef](#)]
28. Wang, H.; Zhou, Z.; Li, Y.; Chen, Z.; Lu, P.; Wang, W.; Liu, W.; Yu, L. Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images. *EJNMMI Res.* **2019**, *9*, 15.
29. Hötter, A.M.; Tarlinton, L.; Mazaheri, Y.; Woo, K.M.; Gönen, M.; Saltz, L.B.; Goodman, K.A.; Garcia-Aguilar, J.; Gollub, M.J. Multiparametric MRI in the assessment of response of rectal cancer to neoadjuvant chemoradiotherapy: A comparison of morphological, volumetric and functional MRI parameters. *Eur. Radiol.* **2016**, *26*, 4303–4312. [[CrossRef](#)] [[PubMed](#)]
30. Som, P.M.; Curtin, H.D. *Head and Neck Imaging*, 4th ed.; Mosby: St. Louis, MO, USA, 2003.
31. Zhu, W.; Huang, Y.; Zeng, L.; Chen, X.; Liu, Y.; Qian, Z.; Du, N.; Fan, W.; Xie, X. AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med. Phys.* **2019**, *46*, 576–589. [[CrossRef](#)] [[PubMed](#)]
32. Haarbuerger, C.; Valentinitich, A.; Rienmüller, A.; Kuder, T.A.; Bickelhaupt, S. Automated detection and segmentation of head and neck carcinomas using a 3D U-Net on planning CT scans. *Eur. Radiol. Exp.* **2020**, *4*, 1–10.

33. Wu, J.; Gensheimer, M.F.; Dong, X.; Rubin, D.L.; Napel, S.; Diehn, M. Robust intratumor partitioning to identify high-risk subregions in lung cancer: A pilot study. *Int. J. Radiat. Oncol. Biol. Phys.* **2017**, *98*, 647–656. [[CrossRef](#)]
34. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. [[CrossRef](#)] [[PubMed](#)]
35. Lu, L.; Ehmke, R.C.; Schwartz, L.H.; Zhao, B. Assessing agreement between radiomic features computed for multiple CT imaging settings. *PLoS ONE* **2015**, *10*, e0143848. [[CrossRef](#)] [[PubMed](#)]
36. Kearney, V.; Chan, J.W.; Haaf, S.; Descovich, M.; Solberg, T.D. DoseNet: A volumetric dose prediction algorithm using 3D fully convolutional networks. *Med. Phys.* **2018**, *45*, 6343–6353. [[CrossRef](#)] [[PubMed](#)]
37. Adams, M.C.; Turkington, T.G.; Wilson, J.M.; Wong, T.Z. A systematic review of the factors affecting accuracy of SUV measurements. *Am. J. Roentgenol.* **2010**, *195*, 310–320. [[CrossRef](#)] [[PubMed](#)]
38. Liu, S.; Liu, S.; Cai, W.; Che, H.; Pujol, S.; Kikinis, R.; Feng, D.; Fulham, M.J. Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer’s disease. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 1132–1140. [[CrossRef](#)]
39. Eiber, M.; Weirich, G.; Holzapfel, K.; Souvatzoglou, M.; Haller, B.; Rauscher, I.; Beer, A.J.; Wester, H.J.; Gschwend, J.; Schwaiger, M.; et al. Simultaneous 68Ga-PSMA HBED-CC PET/MRI Improves the Localization of Primary Prostate Cancer. *Eur. Urol.* **2016**, *70*, 829–836. [[CrossRef](#)]
40. Zhou, M.; Scott, J.; Chaudhury, B.; Hall, L.; Goldgof, D.; Yeom, K.W.; Iv, M.; Ou, Y.; Kalpathy-Cramer, J.; Napel, S.; et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *Am. J. Neuroradiol.* **2018**, *39*, 208–216. [[CrossRef](#)]
41. Vallières, M.; Freeman, C.R.; Skamene, S.R.; El Naqa, I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys. Med. Biol.* **2015**, *60*, 5471–5496. [[CrossRef](#)]
42. Vallières, M.; Kay-Rivest, E.; Perrin, L.J.; Liem, X.; Furstoss, C.; Aerts, H.J.; Khaouam, N.; Nguyen-Tan, P.F.; Wang, C.S.; Sultanem, K.; et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci. Rep.* **2017**, *7*, 10117. [[CrossRef](#)]
43. Bankier, A.A.; Levine, D.; Halpern, E.F.; Kressel, H.Y. Consensus interpretation in imaging research: Is there a better way? *Radiology* **2010**, *257*, 14–17. [[CrossRef](#)]
44. Fortin, J.P.; Parker, D.; Tunç, B.; Watanabe, T.; Elliott, M.A.; Ruparel, K.; Roalf, D.R.; Satterthwaite, T.D.; Gur, R.C.; Gur, R.E.; et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* **2017**, *161*, 149–170. [[CrossRef](#)]
45. Men, K.; Chen, X.; Zhang, Y.; Zhang, T.; Dai, J.; Yi, J.; Li, Y. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front. Oncol.* **2019**, *9*, 1200. [[CrossRef](#)] [[PubMed](#)]
46. Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; Pedreschi, D. A survey of methods for explaining black box models. *ACM Comput. Surv. (CSUR)* **2018**, *51*, 1–42. [[CrossRef](#)]
47. Castelvechi, D. Can we open the black box of AI? *Nat. News* **2016**, *538*, 20–23. [[CrossRef](#)]
48. Jha, S.; Topol, E.J. Adapting to artificial intelligence: Radiologists and pathologists as information specialists. *JAMA* **2016**, *316*, 2353–2354. [[CrossRef](#)]
49. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [[CrossRef](#)] [[PubMed](#)]
50. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)]
51. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
52. Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—addressing ethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981–983. [[CrossRef](#)]
53. Rodrigues, J.J.; de la Torre, I.; Fernández, G.; López-Coronado, M. Analysis of the security and privacy requirements of cloud-based electronic health records systems. *J. Med. Internet Res.* **2013**, *15*, e186. [[CrossRef](#)]
54. Aerts, H.J.; Velazquez, E.R.; Leijenaar, R.T.; Parmar, C.; Grossmann, P.; Carvalho, S.; Bussink, J.; Monshouwer, R.; Haibe-Kains, B.; Rietveld, D.; et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **2014**, *5*, 4006. [[CrossRef](#)]
55. Leek, J.T.; Scharpf, R.B.; Bravo, H.C.; Simcha, D.; Langmead, B.; Johnson, W.E.; Geman, D.; Baggerly, K.; Irizarry, R.A. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **2010**, *11*, 733–739. [[CrossRef](#)]
56. Siravegna, G.; Marsoni, S.; Siena, S.; Bardelli, A. Integrating liquid biopsies into the management of cancer. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 531–548. [[CrossRef](#)]
57. Sollini, M.; Antunovic, L.; Chiti, A.; Kirienko, M. Towards clinical application of image mining: A systematic review on artificial intelligence and radiomics. *Eur. J. Nucl. Med. Mol. Imaging* **2018**, *45*, 2076–2089. [[CrossRef](#)]
58. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. In Proceedings of the 2015 8th International Congress on Image and Signal Processing (CISP), Shenyang, China, 14–16 October 2015; pp. 393–398.
59. Chen, H.; Zhang, Y.; Zhang, W.; Liao, P.; Li, K.; Zhou, J.; Wang, G. Low-dose CT via convolutional neural network. *Biomed. Opt. Express* **2018**, *8*, 679–694. [[CrossRef](#)] [[PubMed](#)]

60. Zhou, Z.; Siddiquee MM, R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2020; pp. 3–11.
61. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014; pp. 2672–2680.
62. Wolterink, J.M.; Leiner, T.; Viergever, M.A.; Isgum, I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans. Med. Imaging* **2017**, *36*, 2536–2545. [[CrossRef](#)] [[PubMed](#)]
63. Son, J.; Park, S.J.; Jung, K.H. Retinal vessel segmentation in fundoscopic images with generative adversarial networks. *arXiv* **2018**, arXiv:1801.00863.
64. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
65. Chen, C.; Qin, C.; Qiu, H.; Tarroni, G.; Duan, J.; Bai, W.; Rueckert, D. Iterative residual refinement for joint optic-disc and cup segmentation with transformers. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Strasbourg, France, 27 September–1 October 2021; Springer: Cham, Switzerland, 2021; pp. 118–127.
66. Sharma, A.; Gupta, A.; Kumar, A.; Ray, A.; Sharma, S. Transformer-based deep learning model for multi-class segmentation of head and neck cancer. In *Proceedings of the International Conference on Image Analysis and Recognition*, Lecce, Italy, 23–27 May 2022; Springer: Cham, Switzerland, 2022; pp. 202–214.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.