

Application of Artificial Intelligence for Nasopharyngeal Carcinoma Management – A Systematic Review

Wai Tong Ng,^{1,2} Barton But,²
Horace CW Choi,³
Remco de Bree,⁴ Anne WM Lee,^{1,2}
Victor HF Lee,^{1,2}
Fernando López,^{5,6}
Antti A Mäkitie,⁷⁻⁹
Juan P Rodrigo,^{5,6} Nabil F Saba,¹⁰
Raymond KY Tsang,¹¹
Alfio Ferlito¹²

¹Clinical Oncology Center, The University of Hong Kong-Shenzhen Hospital, Shenzhen, People's Republic of China; ²Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; ³Department of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; ⁴Department of Head and Neck Surgical Oncology, University Medical Center Utrecht, Utrecht, the Netherlands; ⁵Department of Otolaryngology, Hospital Universitario Central de Asturias (HUCA), Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Instituto Universitario de Oncología del Principado de Asturias (IUOPA), University of Oviedo, Oviedo, 33011, Spain; ⁶Spanish Biomedical Research Network Centre in Oncology, CIBERONC, Madrid, 28029, Spain; ⁷Department of Otorhinolaryngology - Head and Neck Surgery, HUS Helsinki University Hospital and University of Helsinki, Helsinki, Finland; ⁸Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland; ⁹Division of Ear, Nose and Throat Diseases, Department of Clinical Sciences, Intervention and Technology, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden; ¹⁰Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, GA, USA; ¹¹Division of Otorhinolaryngology, Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, People's Republic of China; ¹²Coordinator of the International Head and Neck Scientific Group, Padua, Italy

Correspondence: Barton But
Department of Clinical Oncology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, People's Republic of China
Tel +852 2255 4352
Fax +852 2872 6426
Email bbut@hku.hk

Introduction: Nasopharyngeal carcinoma (NPC) is endemic to Eastern and South-Eastern Asia, and, in 2020, 77% of global cases were diagnosed in these regions. Apart from its distinct epidemiology, the natural behavior, treatment, and prognosis are different from other head and neck cancers. With the growing trend of artificial intelligence (AI), especially deep learning (DL), in head and neck cancer care, we sought to explore the unique clinical application and implementation direction of AI in the management of NPC.

Methods: The search protocol was performed to collect publications using AI, machine learning (ML) and DL in NPC management from PubMed, Scopus and Embase. The articles were filtered using inclusion and exclusion criteria, and the quality of the papers was assessed. Data were extracted from the finalized articles.

Results: A total of 78 articles were reviewed after removing duplicates and papers that did not meet the inclusion and exclusion criteria. After quality assessment, 60 papers were included in the current study. There were four main types of applications, which were auto-contouring, diagnosis, prognosis, and miscellaneous applications (especially on radiotherapy planning). The different forms of convolutional neural networks (CNNs) accounted for the majority of DL algorithms used, while the artificial neural network (ANN) was the most frequent ML model implemented.

Conclusion: There is an overall positive impact identified from AI implementation in the management of NPC. With improving AI algorithms, we envisage AI will be available as a routine application in a clinical setting soon.

Keywords: machine learning, neural network, deep learning, prognosis, diagnosis, auto contouring

Introduction

According to the International Agency for Research on Cancer, nasopharyngeal carcinoma (NPC) is the twenty-third most common cancer worldwide. The global number of new cases and deaths in 2020 were 133,354 and 80,008, respectively.^{1,2} Although it is not uncommon, it has a distinct geographical distribution where it is most prevalent in Eastern and South-Eastern Asia, accounting for 76.9% of global cases. It was also found that almost half of the new cases occurred in China.² Because of its late symptoms and anatomical location, it makes it difficult to be detected in the early stages. Radiotherapy is the primary treatment modality, and concomitant/adjunctive chemotherapy is often needed for advanced locoregional disease.³ Furthermore, there are many organs-at-risk (OARs) nearby that are

sensitive to radiation; these include the salivary glands, brainstem, optic nerves, temporal lobes and the cochlea.⁴ Hence, it is of interest whether the use of artificial intelligence (AI) can help improve the diagnosis, treatment process and prediction of outcomes for NPC.

With the advances of AI over the past decade, it has become pervasive in many industries playing both major and minor roles. This includes cancer treatment, where medical professionals search for methods to utilize it to improve treatment quality. AI refers to any method that allows algorithms to mimic intelligent behavior. It has two subsets, which are machine learning (ML) and deep learning (DL). ML uses statistical methods to allow the algorithm to learn and improve its performance, such as random forest and support vector machine. Artificial neural network (ANN) is an example of ML and is also a core part of DL.⁵ DL can be defined as a learning algorithm that can automatically update its parameters through multiple layers of ANN. Deep neural networks such as convolutional neural network (CNN) and recurrent neural network are all DL architectures.

Besides histological, clinical and demographic information, a wide range of data ranging from genomics, proteomics, immunohistochemistry, and imaging must be integrated by physicians when developing personalized treatment plans for patients. This has led to an interest in developing computational approaches to improve medical management by providing insights that will enhance patient outcomes and workflow throughout a patient's journey.

Given the increased use of AI in cancer care, in this systematic literature review, papers on AI applications for NPC management were compiled and studied in order to provide an overview of the current trend. Furthermore, possible limitations discussed within the articles were explored.

Methods

Search Protocol

A systematic literature search was conducted to retrieve all studies that used AI or its subfields in NPC management. Keywords were developed and combined using boolean logic to produce the resulting search phrase: (“artificial intelligence” OR “machine learning” OR “deep learning” OR “Neural Network”) AND (“nasopharyngeal carcinoma” OR “nasopharyngeal cancer”). Using the search phrase, a search of research articles from the past 15

years to March 2021 was performed on PubMed, Scopus and Embase. The results from the three databases were consolidated, and duplicates were removed. The Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) was followed where possible, and the PRISMA flow diagram and checklist were used as a guidelines to consider the key aspects of a systematic literature review.⁶

Eligibility

Exclusion and inclusion criteria were determined to assess the eligibility of the retrieved publications. The articles were first checked to remove those that were not within the exclusion criteria. These included book chapters, conference reports, literature reviews, editorials, letters to the editors and case reports. In addition, articles in languages other than English or Chinese and papers with inaccessible full-texts were also excluded.

The remaining studies were then filtered by reading the title and abstract to remove any articles that were not within the inclusion criteria (applications of AI or its subfield and experiments on NPC). A full-text review was further performed to confirm the eligibility of the articles based on both these criteria. The process was conducted by two independent reviewers (B.B & H.C.).

Data Extraction

Essential information from each article was extracted and placed in a data extraction table (Table 1). These included the author(s), year of publication, country, sample type, sample size, AI algorithm used, application type, study aim, performance metrics reported, results, conclusion, and limitations. The AI model with the best performance metrics from each study was selected and included. Moreover, the performance results of models trained with the training cohort were obtained from evaluating the test cohort instead of the training cohort. This was to prevent overfitting by avoiding to train and test models with the same dataset.

Quality Assessment

The selected articles were assessed for risk of bias and applicability using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS)-2 tool in Table 2.⁷ Studies with more than one section rated “high” or “unclear” were eliminated. Further quality assessment was also completed to ensure the papers meet the required standard. This was performed using the guidelines for developing and

Table 1 Data Extraction Table

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Wang et al (2010) ¹⁰ (Hong Kong SAR, China)	NPC and other type 54 (Protein data)	Machine learning (Diagnosis)	1. Classification: Non-linear regulatory network (hopfield-like network)	To find relationships between protein biomarkers and classify different disease groups	1. Classification performance - Sensitivity: 0.95 - Specificity: 0.80 - Accuracy: 0.93	The developed regulatory network out-performed Fisher linear discriminant, KNN, linear SVM and radial basis function SVM in classification performance.	The proposed technique has promise in assisting disease diagnosis by finding protein regulation relationships.	N/A
Aussem et al (2012) ¹¹ (France)	NPC 1298 (Risk factors)	Machine learning (Miscellaneous applications - Risk factor identification)	1. Feature selection: Markov boundary discovery algorithm 2. Classification: Bayesian network	To extract relevant dietary, social and environmental risk factors related to increasing risk of NPC	1. Identification of potential risk factors - Odd ratio ^a 2. Predictive performance - AUC: 0.743 3. Empirical performance - Accuracy: 0.654 - Euclidean distance ^a	The proposed model had a better performance in recognizing risk factors associated with NPC than other algorithms.	The proposed techniques can integrate experts' knowledge and information extracted from data to analyze epidemiologic data	N/A
Kumdee, Bhongmakapat, and Ritthipravat (2012) ¹² (Thailand)	NPC 1257 (Clinicopathological and serology data)	Machine learning (Prognosis)	1. Prediction: Generalized neural network-type SIRM (G-NN-SIRM)	To predict NPC recurrence	1. Classification performance - AUC: 0.8809 - Sensitivity: 0.8003 - Specificity: 0.8486 - Accuracy: 0.8245 2. Predictive performance - Mean square error: 0.595360	G-NN-SIRM had a significantly higher performance than the other techniques in NPC recurrence prediction	The G-NN-SIRM can be applied to NPC recurrence prediction	N/A
Ritthipravat, Kumdee, and Bhongmakapat (2013) ¹³ (Thailand)	NPC 495 (Clinicopathological and serology data)	Machine learning (Prognosis)	1. Missing data technique to complete data for model training: Complete-case analysis, Mean imputation, expectation-maximization imputation and KNN imputation 2. Prognosis: Single-point ANN model, Multiple-point ANN model and Sequential neural network	To predict NPC recurrence	1. Predictive performance - AUC (5 years): 0.7300 - Chi-square statistics of goodness of fit test (5 years recurrence-free survival): 4.30	The closest performance to the Kaplan-Meier model were the expectation-maximization imputation technique models, particularly with sequential neural network.	Missing data technique cross-combined with ANNs were investigated for predicting NPC recurrence.	N/A

(Continued)

Table 1 (Continued).

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Zhu, and Kan (2014) ⁴⁴ (China)	NPC 312 (microRNA expressions data)	Machine learning (Prognosis)	1. Data transformation, data integration, or prediction output: ANN 2. Death risk assessment: Neural network cascade	To assess cancer prognosis	1. Risk prediction performance - AUC: 0.951 - Accuracy: 0.83	The neural network cascade out-performed both the transformed and untransformed neural network models.	The study proposed a potential method for constructing a microRNA biomarker selection and prediction model	N/A
Jiang et al. (2016) ¹⁴ (China)	NPC 453 (Clinicopathological and serology data)	Machine learning (Prognosis)	1. Feature Selection: Recursive feature elimination procedure based on SVM 2. Classification: SVM	To predict the survival of NPC patients with synchronous metastases	1. Prognostic performance - AUC: 0.633 - Sensitivity: 0.713 - Specificity: 0.807	The ML model had a better prognostic performance than classifiers using clinical indexes alone or with haematological markers	The model has the potential to help clinicians choose the most appropriate treatment strategy for metastatic NPC patients	1. Not all clinical indexes and haematological markers were included in the study. 2. The effectiveness of the combination treatment is uncertain due to differences in being treated in different institutions
Liu et al (2016) ⁴³ (China)	NPC 53 (MRI images)	Machine learning (Prognosis)	1. Classification: KNN & ANN	To predict NPC response to chemoradiotherapy	1. Classification performance - Specificity: 1.000 - Accuracy: 0.909 - Precision: 1.000 - F-score: 0.875 - Matthews correlation coefficient: 0.810 - No information rate: 0.727	The model using parameters extracted from T1 sequence had a better classification performance than the other two models	Integrating texture parameters to ML algorithms can act as imaging biomarkers for NPC tumor response to chemoradiotherapy	1. Relatively small sample size
Wang et al (2016) ⁶³ (Taiwan and China)	NPC 335 (MRI and PET/CT images)	Machine learning (Diagnosis)	1. Classification: ANN	To assess the diagnostic accuracy of adding additional nodal parameters and PET/CT	1. Diagnostic efficacy - Sensitivity: 0.822 - Specificity: 0.952 - Youden's index ² - Accuracy: 0.888 - PPV: 0.943 - NPV: 0.846	ANN demonstrated that combining three (and four) of the proposed parameters yielded good results.	Additional parameters in MRI and PET/CT were found which can improve prediction accuracy.	1. Possible diagnostic errors from not using histopathology. 2. Not all factors useful in the diagnosis were included due to limitations of the facility in the center.

Men et al (2017) ⁵⁵ (China)	NPC 230 (CT images)	Deep learning (Auto-contouring)	1. Segmentation: Deep deconvolutional neural network	To assess the segmentation ability of the developed model	1. Segmentation performance - DSC: 0.753±0.113 - HD: 12.6mm±11.5mm	The deep deconvolutional neural network had a much better performance when compared with VGG-16 model.	The proposed model has promise in enhancing the consistency of delineation and streamline the radiotherapy workflows.	1. The model may be hard to converge due to having N0 and N+ patients in both the training and testing set. 2. Interobserver variability may cause bias 3. The segmentation performance may be affected by the inclusion of both stage I and II patients as the target may have diverse contrast, shape and sizes.
Mohammed et al (2017) ⁵⁴ (Malaysia and Iraq)	NPC 149 (Microscopic images)	Machine learning (Auto-contouring/ Diagnosis)	1. Segmentation: K-means clustering algorithm and ANN 2. Classification: K-means clustering algorithm and ANN	To evaluate the segmentation and identification performance of the developed models	1. Classification performance - Sensitivity: 0.9324 - Specificity: 0.9054 - Accuracy: 0.911 2. Segmentation performance - Accuracy: 0.883	The proposed method had an improved classification and segmentation performance over SVM.	Texture features can assist in differentiating benign and malignant tumors. Thus, the fully automated proposed model can help with doctors' diagnosis and support them.	1. Disproportionate distribution between benign and malignant tumors in the dataset. 2. Sample size can be larger to improve reliability 3. Combining different techniques is complicated.
Zhang et al (2017) ⁴² (China)	NPC 110 (MRI images)	Machine learning (Prognosis)	1. Classification: LR, Kernel SVM, Linear SVM, AdaBoost, RF, ANN, KNN, Linear discriminant analysis and Naïve bayes 2. Feature Selection: LR, Logistic SVM, RF, Distance correlation, Elastic net LR & Sure independence screening	To predict local and distant failure of advanced NPC patients' prior treatment	1. Prognostic performance - AUC: 0.8464±0.0069 - Test error: 0.3135±0.0088	Using RF for both feature selection and classification had the best prognostic performance	The most optimal ML methods for local and distant failure prediction in advanced NPC can improve precision oncology and clinical practice	N/A
Zhang et al (2017) ⁴¹ (China)	NPC 118 (MRI images)	Machine learning (Prognosis)	1. Analysis Clustering (with radiomics)	Individualized progression-free survival evaluation of advanced NPC patients' prior treatment	1. Prognostic performance - C-index: 0.776 [CI: 0.678–0.873]	Integrating radiomics signature with other factors within a nomogram such as TNM staging system or clinical data improved its performance.	The use of quantitative radiomics models can be useful in precision medicine and assist with the treatment strategies for NPC patients	1. The analysis did not consider two-way or higher order interactions between features. 2. The generalizability of the model cannot be determined as the validation cohort was taken from the same institution. 3. The training phase of the model may be affected from selection bias due to stringent inclusion criteria.

(Continued)

Table 1 (Continued).

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Li et al (2018) ¹⁵ (China)	NPC 30396 (Endoscopic images)	Deep learning (Auto-contouring/ Diagnosis)	1. Detection: Fully CNN	To evaluate the performance of the developed model to segment and detect nasopharyngeal malignancies in endoscopic images	1. Detection performance - AUC: 0.930- Sensitivity: 0.902 [CI: 0.878–0.922] - Specificity: 0.855 [CI: 0.827–0.880] - Accuracy: 0.880 [CI: 0.861–0.896] - PPV: 0.869 [CI: 0.843–0.892] - NPV: 0.892 [CI: 0.865–0.914] - Time taken: 0.67 min 2. Segmentation performance - DSC: 0.75±0.26	The developed model had a better performance than oncologists in nasopharyngeal mass differentiation. It was also able to automatically segment malignant areas from the background of nasopharyngeal endoscopic images.	The proposed method has potential in guided biopsy for nasopharyngeal malignancies.	1. Limited diversity due to all samples being acquired from the same institution, which leads to over-fitting.
Mohammed et al (2018) ⁶² (Malaysia, Iraq and India)	NPC 381 (Endoscopic images)	Artificial intelligence and machine learning (Diagnosis)	1. Feature selection: Genetic algorithm 2. Classification: ANN	To evaluate the proposed method in detecting NPC from endoscopic images	1. Classification performance - Sensitivity: 0.9535 - Specificity: 0.9455 - Accuracy: 0.9622	The detection performance of the trained ANN was close to that done manually by ear, nose and throat specialists.	The study demonstrated the feasibility of using ANNs for NPC identification in endoscopic images.	N/A
Mohammed et al (2018) ¹⁶ (Malaysia, Iraq and India)	NPC 381 (Endoscopic images)	Artificial intelligence and machine learning (Auto-contouring/ Diagnosis)	1. Feature selection: Genetic algorithm 2. Classification: ANN & SVM	To evaluate the proposed model in diagnosing NPC from endoscopic images	1. Segmentation performance - Accuracy: 0.9265 2. Classification performance - Sensitivity: 0.9480 - Specificity: 0.9520 - Precision: 0.9515	The developed model yielded similar results to that of ENT specialists in segmentation performance. The classification performance achieved high results but the training dataset had a better performance.	The study demonstrated the effectiveness and accuracy of the proposed method.	N/A
Du et al (2019) ⁶⁰ (Hong Kong SAR, China)	NPC 277 (MRI images and clinicopathological data)	Machine learning (Prognosis)	1. Feature selection: Cluster analysis 2. Classification: SVM	To predict early progression of nonmetastatic NPC	1. Model performance - AUC: 0.80 [CI: 0.73–0.89] - Sensitivity: 0.92 - Specificity: 0.52 - PPV: 0.32 - NPV: 0.96 2. Model calibration - Brier score: 0.150	The proposed model trained with five clinical features and radiomic features had the best performance over the other models. Tumor shape sphericity, first-order mean absolute deviation, T stage, and overall stage were important factors affecting 3-year disease progression.	The use of radiomics can be used for tumor diagnosis and risk assessment. Shapley additive explanations helped to find relationship between features in the model.	1. The association between Epstein-Barr virus and progression-free survival was not explored. 2. Test-retest or time-dependent variability of radiomic features were not examined. 3. The image resolutions of the patient samples were diverse.
Jiao et al (2019) ⁶⁶ (China)	NPC 106 (IMRT plans)	Machine learning (Miscellaneous applications - radiotherapy planning)	1. Prediction: Distance-to-target histogram (DTH) general regression neural network (GRNN) and DTH +Conformal plan DVH GRNN	To predict dose-volume histograms of OARs from IMRT plan	1. DVH prediction accuracy (OARs) - Dosimetric results ^b - Average R squared: 0.95 - Average MAE: 3.67	The addition of dosimetric information improved the DVH prediction of the developed model.	The study showed the prediction capability of the model when patient dosimetric information was added to geometric information.	N/A

Jing et al (2019) ¹⁷ (China)	NPC and other types 6449 (Clinicopathological and plasma EBV DNA data)	Deep learning (Prognosis)	I. Prediction: RankDeepSurv	Compare RankDeepSurv with other survival models and clinical experts in the analysis and prognosis of four public medical datasets and NPC	I. Predictive performance of survival analysis - C-index: 0.681 [0.678–0.684]	The RankDeepSurv had a better performance than the other three referenced methods in four public medical clinical datasets and in the NPC dataset versus clinical experts	The proposed model can assist clinicians in providing more accurate predictions for NPC recurrence	N/A
Li et al (2019) ⁵³ (China)	NPC 502 (CT images)	Deep learning (Auto-contouring)	I. Segmentation: U-net	To assess the developed model's accuracy in delineating CT images	I. Model performance - DSC: 0.74 - HD: 12.85mm	The modified U-net had a higher consistency and performance than manual contouring, while using less time per patient.	The developed model has the potential to help lighten clinicians' workload and improve NPC treatment outcomes.	N/A
Liang et al (2019) ⁵² (China)	NPC 185 (CT images)	Deep learning (Auto-contouring/ Diagnosis)	I. Recognition and classification: OAR detection and segmentation network (ODS net)	To assess the performance of the model in detecting and segmenting OARs	I. Detection performance - Sensitivity: 1.000 [CI: 0.994–1.000] - Specificity: 0.999 [CI: 0.997–1.000] 2. Segmentation performance - DSC: 0.934±0.04	The ODS net had good result in both detection and segmentation performances.	The fully automatic model may help to facilitate therapy planning.	1. The manual segmentation of images by a radiologist may not be consistent and may not be the true standard of reference. 2. The method was only examined by one type of CT scan. Investigation with other CT scanners was needed. 3. Contrast CT and MRI images were not applied in the model, affecting its performance.
Lin et al (2019) ⁵¹ (China, Hong Kong SAR, China, and Singapore)	NPC 1021 (MRI images)	Deep learning (Auto-contouring)	I. Feature extraction: 3D CNN	To evaluate the developed model in auto-contouring of primary gross target volume	I. Model performance - DSC: 0.79 [CI: 0.78–0.79] - ASSD: 2.0mm [CI: 1.9–2.1]	The model yielded good accuracy. It also helped improve the contouring accuracy and time of practitioners in the study.	The model has the potential to help tumor control and patient survival by enhancing the delineation accuracy and lower contouring variation by different practitioners and the time required.	1. Low statistical power due to small number of events. 2. There may be memory bias during evaluation. 3. Poor contouring of oropharynx, hypopharynx and intracranial regions caused lower accuracy at the cranial-caudal edges.
Liu et al (2019) ⁶⁵ (China)	NPC 190 (Helical tomotherapy plans)	Deep learning (Miscellaneous applications - radiotherapy planning)	I. Prediction: U-ResNet-D	To predict the three-dimensional dose distribution of helical tomotherapy	I. Predictive performance a) Dose difference - Average: 0.0%±2.4% - Average MAE: 1.9%±1.8% b) Dosimetric index (DI) comparison - Prediction accuracy (Dmean): 2.4% - Prediction accuracy (Dmax): 4.2% c) DSC: 0.95–1	The U-ResNet-D model yielded good results in predicting 3D dose distribution.	The developed method has the potential to increase the quality and consistency of treatment plans.	1. Can only predict one type of dose distribution. 2. The model was unable to apply the predicted 3D dose distribution into a clinical plan.
Ma et al (2019) ¹⁸ (China and USA)	NPC 90 (CT and MRI images)	Deep learning (Auto-contouring)	I. Segmentation: Multi-modality CNN & Combined CNN	To use MRI and CT in NPC segmentation with the proposed models	I. Segmentation performance - Sensitivity: 0.718±0.121 - PPV: 0.797±0.109 - DSC: 0.752±0.043 - ASSD: 1.062mm±0.298mm	The proposed model out-performed the same model but without multi-modal information fusion and other existing CNN-based multi-modality segmentation models.	The proposed model was the first CNN-based method to solve the challenge of performing multi-modality tumor contouring on NPCs.	N/A

(Continued)

Table I (Continued).

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Peng et al (2019) ¹⁹ (China)	NPC 707 (PET/CT images)	Deep learning (Prognosis)	I. Feature extraction: Nomogram using Deep CNN	To assess risk and guide induction chemotherapy for patients	I. Prognostic performance - C-index: 0.722 [CI: 0.652–0.792] - AUC (5 years): 0.735	The DL-based radiomics nomogram out-performed the EBV DNA-based model in risk stratification and induction chemotherapy guiding	The DL-based radiomics nomogram can be used for individualized treatment strategies.	1. The follow-up period was too short 2. Data sample were only from one center. 3. There may be patient selection biases confounded with radiomics signatures and outcomes may be present as induction chemotherapy was not randomly assigned to patients.
Rehioui et al (2019) ²⁰ (Morocco)	NPC 90 (Risk factors)	Machine learning (Miscellaneous applications - Risk factor identification)	I. Prediction: Clustering algorithms (K means, Expectation Maximization algorithm, density-based algorithms (DENCLUE and its variants)	Compare the prognosis performance of different algorithms	I. Model performance - Dunn Index: 0.614 - Davies-Bouldin Index: 2.110 - Compactness Index: 0.781 - Accuracy: 0.592 - Normalized Mutual Information: 0.245 - Entropy: 0.806	The density-based algorithms (DENCLUE and its variants) obtained a better result than partitioning or statistical models	Familial history of cancer, living conditions and tobacco consumption are all associated with advanced stage of NPC.	N/A
Zhong et al (2019) ⁵⁰ (China)	NPC 140 (CT images)	Deep learning (Auto-contouring)	I. Segmentation: Boosting Resnet & Voxelwise Resnet	To assess the proposed model in delineating OARs	I. Segmentation performance - DSC: 0.9188±0.0351 -95HD: 3.15mm±0.58mm - Volume overlap error: 14.83% ±5.26%	The proposed cascaded method gave a significantly better performance than other existing single network architecture or segmentation algorithms.	The study showed the effectiveness of the developed model when auto-contouring OARs and the benefits of using the cascaded DL structure.	N/A
Zou et al (2019) ²¹ (China and United Kingdom)	NPC 99 (CT and MRI images)	Deep learning (Miscellaneous applications - Image registration)	I. Image registration: Full convolution network, CNN & Random sample consensus	To develop a model for image registration	I. Image registration - Precision ¹ - Recall ¹ - Target registration error (TRE) ^a	The proposed method with additional use of transfer learning and fine-tuning out-performed both the proposed method and scale invariant feature transform (SIFT).	The use of transfer learning and fine tuning in the proposed model is promising in improving image registration.	N/A
Abd Ghani et al (2020) ⁶¹ (Malaysia, Iraq and India)	NPC 381 (Endoscopic images)	Machine learning (Diagnosis)	I. Classification: KNN, Linear SVM & ANN	To develop a model with endoscopic images for NPC identification	I. Classification performance - Sensitivity: 0.925 - Specificity: 0.937 - Accuracy: 0.947	The majority rule for decision-based fusion technique had a significantly lower performance than using a single best performing feature scheme for the SVM classifier, which uses pair-wise fusion of only two features.	A fully automated NPC detection model with good accuracy was developed. Although the proposed method had a high accuracy, the single best performing feature scheme for the SVM classifier outperforms it.	N/A

Bai et al (2020) ⁶⁴ (China and USA)	NPC 140 (IMRT plans)	Machine learning (Miscellaneous applications - radiotherapy planning)	I. Prediction: ANN	To explore viability of a model for knowledge-based automated intensity-modulated radiation therapy planning	I. Plan quality - DVH ^a - Dose distribution ^b - Monitor unit: 685.04±59.63 - Planning duration: 9.85min±1.13min	The proposed model had a similar performance but with a higher efficiency in treatment planning when compared with manual planning.	The proposed technique can significantly reduce the treatment planning time while maintaining the same plan quality.	N/A
Chen et al (2020) ²² (China)	NPC 149 (MRI images)	Deep learning (Auto-contouring)	I. Segmentation: Multi-modality MRI fusion network	To evaluate the segmentation performance of a model which uses T1, T2 and contrast-enhanced T1 MRI	I. Segmentation performance - DSC: 0.7238±0.1099 - HD: 18.31mm±16.73mm - ASSD: 2.07mm±2.32mm	In comparison with other existing DL algorithms, the proposed model had the best segmentation performance.	Multi-modality MRI is useful to the proposed model for NPC delineation.	N/A
Chen et al (2020) ⁶⁸ (China)	NPC 99 (VMAT plans)	Machine learning (Miscellaneous applications - radiotherapy planning)	I. Prediction: ANN	To develop models for radiotherapy planning for planning quality control	I. Plan quality a) DVH ^a b) Dose distribution ^b	The developed ANN model had a lower capability than the junior physicist in designing radiotherapy plans.	The proposed model enhancing the quality and stability of individualized radiotherapy planning.	N/A
Chuang et al (2020) ²³ (Taiwan)	NPC 726 (WSIs)	Deep learning (Diagnosis)	I. Classification: Patch-level deep CNN & Slide-level deep CNN	To assess proposed model in detecting NPC in biopsies	I. Diagnostic performance - AUC: 0.9900±0.004 - Sensitivity ^a - Specificity ^a	The slide-level model had a better performance than pathology residents. However, its diagnostic ability is slightly worse than both attending pathologists and the chief resident.	The study demonstrated for the first time that DL algorithms can identify NPC in biopsies.	N/A
Cui et al (2020) ³⁹ (China)	NPC 792 (MRI images)	Machine learning and deep learning (Prognosis)	I. Feature selection: Generalized linear model (ridge/lasso), XRT, Gradient boosting machine, RF & DL (Unknown)	To assign prediction scores to NPC patients and compare with the current clinical staging system	I. Prognostic performance a) AUC - OS: 0.796 (s.d.=0.044) - DMFS: 0.752 (s.d.=0.042) - LRFs: 0.721 (s.d.=0.052) b) Specificity - OS: 0.721 (s.d.=0.061) - DMFS: 0.576 (s.d.=0.114) - LRFs: 0.540 (s.d.=0.153) c) Test error - OS: 0.208 (s.d.=0.037) - DMFS: 0.271 (s.d.=0.052) - LRFs: 0.287 (s.d.=0.050)	The new scoring system had a better prognostic performance than the TNM/AJCC system in predicting treatment outcome for NPC	The new scoring system has the potential to improve image data-based clinical predictions and precision oncology	1. The time to event was not considered. 2. No testing set. 3. Only data from MRI images were used for the scoring system
Diao et al (2020) ⁶⁰ (China)	NPC 1970 (WSIs)	Deep learning (Diagnosis)	I. Classification: Inception-v3	To assess the pathologic diagnosis of NPC with the proposed model	I. Diagnostic performance - AUC: 0.930 - Sensitivity: 0.929 - Specificity: 0.801 - Accuracy: 0.905 - Jaccard index: 0.879 - Euclidean distance: 0.242 - Kappa factor: 0.842	Inception-v3 performed better than the junior and intermediate pathologists, but was worse than the senior pathologist in accuracy, specificity, sensitivity, AUC and consistency.	The proposed model can be used to support pathologists in clinical diagnosis by acting as a diagnostic reference.	1. Improvement in the model's design is required. 2. The model only identifies if the tumour was cancerous but not its subtype. 3. Limited sample size.

(Continued)

Table I (Continued).

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Du et al (2020) ⁵⁹ (China, USA and Canada)	NPC 76 (PET/CT images)	Machine learning (Diagnosis)	I. Classifications: Decision tree, KNN, Linear discriminant analysis, LR, Naive bayes, RF & SVM with radial basis function kernel SVM	To evaluate and compare different machine learning methods in differentiating local recurrence and inflammation	I. Diagnostic performance - AUC: 0.883 [CI: 0.675–0.979] - Sensitivity: 0.833 - Specificity: 1.000 - Reliability (test error): 0.091 [CI: 0.001–0.244]	The combination of fisher score with KNN, FSCR with support vector machines with RBF-SVM, fisher score with RF, and minimum redundancy maximum relevance with RBF-SVM had significantly better performance in accuracy, sensitivity, specificity and reliability than other combination of techniques.	Several methods to integrate ML algorithms with radiomics have the potential to improve NPC diagnostics.	1. Limited by the retrospective nature and small sample size from one source. 2. Only common feature selection and classification techniques were chosen. No parameter tuning was performed. 3. Clinical parameters and genomic data were not included.
Guo et al (2020) ²⁴ (China)	NPC 120 (3D MRI images)	Deep learning (Auto-contouring)	I. Segmentation: 3D CNN with multi-scale pyramid network	To evaluate the segmentation performance of a model	I. Segmentation performance - DSC: 0.7370 - ASSD: 1.214mm - F1-score: 0.7540	The developed model out-performed the other DL models. Furthermore, the Jaccard loss function improved the segmentation performance of all models substantially.	The Jaccard loss function solved the issue of extreme foreground and background imbalance in image segmentation. However, further validation is required.	N/A
Jing et al (2020) ²⁵ (China)	NPC 1846 (3D MRI images and clinicopathological data)	Deep learning (Prognosis)	I. Prediction: End-to-end multi-modality deep survival network (MDSN)	To predict and categorize the risk scores of NPC patients	I. Model performance - C-index: 0.651	The end-to-end MDSN had a better performance than the other four survival methods. The integration of clinical stages into the MDSN model further improves its performance.	MDSN has the potential to support clinicians in making treatment decisions.	N/A
Ke et al (2020) ⁴⁹ (China)	NPC 4100 (3D MRI images)	Deep learning (Auto-contouring/ Diagnosis)	I. Classification and segmentation: Self-constrained 3D DenseNet model	To assess the detection and segmentation ability of the developed model	I. Diagnostic performance - AUC: 0.976 [CI: 0.966–0.987] - Sensitivity: 0.9968 [CI: 0.9792–0.9998] - Specificity: 0.9167 [CI: 0.8377–0.9607] - Accuracy: 0.9777 [CI: 0.9566–0.9891] - PPV: 0.9909 [CI: 0.9861–0.9941] - NPV: 0.9637 [CI: 0.9473–0.9753] 2. Segmentation performance - DSC: 0.77±0.07	The model had encouraging segmentation ability and the diagnostic performance of the proposed model obtained a better result than that of experienced radiologists.	The developed model may be able to improve the diagnostic efficiency and assist in clinical practice.	1. No external validation. 2. The model only trained with MRI images. Other clinical factors were not considered. 3. Patients diagnosed with benign hyperplasia were not confirmed with histopathology.
Liu et al (2020) ³⁸ (China)	NPC 1055 (WSIs)	Deep learning (Prognosis)	I. Risk score calculation: DeepSurv	To assess the survival risk of NPC patients in order to make treatment decisions	I. Survival risk assessment - C-index: 0.723	DeepSurv to analyze the pathological microscopic features was a higher independent prognostic risk factor than EBV DNA copies and N stage	DeepSurv to analyze the pathological microscopic features can be used as a reliable tool for assessing survival risk in NPC patients.	1. Decreased generalizability when applied to other centers or populations. 2. Samples only consisted of undifferentiated non-keratinization NPC in the endemic region. 3. The principle in which pathological microfeatures assist in guiding treatment is unknown

Men et al (2020) ⁴⁷ (China)	NPC 600 (CT images)	Deep learning (Auto-contouring)	1. Segmentation: CNN 2. Classification: Resnet-101	To assess the proposed method to improve segmentation constantly with less labelling effort	1. Classification Performance - AUC: 0.91 - Sensitivity: 0.92 - Specificity: 0.90 - Accuracy: 0.92 2. Segmentation performance - DSC: 0.86±0.02	The proposed method could improve segmentation performance, while reducing the amount of labelling required.	The developed model decreased the amount of labelling and boosted segmentation performance by constantly obtaining, fine-tuning and transferring knowledge over long periods of time.	1. The effect of the number of locked layers were not investigated. 2. The study did not use 3D segmentation model and counted the labelling reduction based on slices, resulting in poor automatic segmentation in several slices for nearly all patients
Mohammed et al (2020) ²⁶ (Malaysia, Iraq and India)	NPC 381 (Endoscopic images)	Machine learning (Diagnosis)	1. Classification: Multilayer perceptron ANN	To detect NPC from endoscopic images	1. Classification performance - AUC: 0.931±0.017 - Sensitivity: 0.9543±0.0165 - Specificity: 0.9578±0.0221 - Accuracy: 0.9566±0.0175 - PPV: 0.9455±0.0433	The developed models yielded good results and ANN,50–50-A, had the best performance.	The study was the first to consolidate diverse features into one fully automated classifier.	1. Insufficient sample size and limited changeability. 2. Possible misidentification of endoscopy image data by experts.
Wang et al (2020) ²⁷ (China, USA and Thailand)	NPC 186 (CT and MRI images)	Machine learning (Radiation-induced injury diagnosis)	1. Prediction: XGBoost	To assess the feasibility in developing a model for predicting radiation-related fibrosis	1. Predictive performance - AUC: 0.69 - Sensitivity: 0.0215 - Specificity: 0.9866 - Accuracy: 0.65	The proposed model trained with CT images had a better diagnostic accuracy than when using MRI features.	The proposed technique can be used to perform patient specific treatments by adjusting the administered dose on the neck, which can minimize the side effects.	1. There is subject bias in fibrosis grading. 2. The radiomic analysis protocol is impractical to be used in daily clinical practice. 3. The investigation does not differentiate between radiation-induced fibrosis and residual or recurrent tumour.
Wang et al (2020) ⁴⁸ (China)	NPC 205 (CT images)	Deep learning (Auto-contouring)	1. Feature extraction: Modified 3D U-Net based on Res-block and SE-block	To develop a model for automatic delineation of NPC in computed tomography	1. Delineation accuracy - Precision: 0.7538 - Sensitivity: 0.7634 - DSC: 0.7372 - HD95: 4.96mm - ASSD: 1.47mm	The proposed model out-performed the other methods in the experiment. In addition, using CT combined with contrast-enhanced-CT instead of CT alone improves the performance of all models.	The study showed that the proposed fully automated model has promise in helping clinicians in 3D delineation of tumour during radiotherapy planning by minimize delineation variability.	1. The patient samples were all from one medical center. 2. The model could only automatically delineate the nasopharynx gross tumour volume
Xie et al (2020) ³⁷ (Hong Kong SAR, China)	NPC 166 (PET/CT images)	Machine learning (Prognosis)	1. Prediction: LR, SVM, RF & XGBoost	To investigate the effect of re-sampling technique and machine learning classifiers on radiomics-based model	1. Predictive performance - AUC: 0.66 - Geometric mean score: 0.65 - Precision: 0.90 - Recall: 0.74 - F-measures: 0.81	The combination of adaptive synthetic re-sampling technique and SVM classifier gave the best performance	Re-sampling technique significantly improved the prediction performance of imbalanced datasets	1. The relatively small number of instances and features in the retrospective dataset may reduce the generalizability to other kinds of cancer. 2. The paper mainly focused on the data level approach using re-sampling technique
Xue et al (2020) ⁶⁹ (China)	NPC 150 (Combined CT and MRI images)	Deep learning (Auto-contouring)	1. Segmentation: Deeplabv3+ 2. Feature extraction: MobileNetV2	To evaluate the performance of the proposed model in segmenting high risk tumors	1. Segmentation accuracy - DSC: 0.76±0.11 - HD: 10.9mm±8.6mm - ASSD: 3.4mm±2.0mm - Jaccard index: 0.63±0.13	The developed model had a better performance when compared with the U-net model. Its results were closer to manual contouring.	The developed model has promise in increasing the effectiveness and consistency of primary tumour gross target volume delineation for NPC patients.	1. Insufficient training data. 2. Delineation variability between clinicians. 3. MRI images were not used for training the model.

(Continued)

Table I (Continued).

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Xue et al (2020) ⁴⁶ (China)	NPC 150 (Combined CT and MRI images)	Deep learning (Auto-contouring)	I. Segmentation: SI-net	To assess the model's ability to segment high-risk tumors	I. Segmentation performance - DSC: 0.84±0.04 - HD: 8.7mm±2.5mm - ASSD: 2.8mm±1.0mm - Jaccard index: 0.74±0.05	The SI-Net model had a better segmentation performance than the U-net model. The mean contouring time of the model is also less than when performed manually.	The proposed model has the potential to help with treatment planning by improving the efficiency and consistency of CTVp1 segmentation.	1. Insufficient training data. 2. Delineation variability between practitioners. 3. MRI images were not used for training the model.
Yang et al (2020) ²⁸ (China)	NPC 1138 (MRI images)	Deep learning (Prognosis)	I. Prediction: Resnet	To evaluate an automatic T staging system that requires no additional annotation	I. Prognostic performance a. AUC: 0.943 b. Sensitivity: 0.882 c. Specificity: 0.880 d. Accuracy: 0.7559 [CI: 0.7151–0.7967] e. C-index - OS: 0.652 [CI: 0.567–0.737] - PFS: 0.612 [CI: 0.537–0.686]	The proposed model had a similar performance to the TNM staging system	The model had a good prognostic performance in fully automated T staging of NPC.	1. Some imaging information may be ignored as contrast-enhanced-T1 weighted images in the coronal plane and T1 weighted images in the sagittal plane were not included in the model construction. 2. The generalizability was unknown as the model was not externally verified.
Yang et al (2020) ⁶⁷ (China)	NPC 147 (CT images)	Deep learning (Auto-contouring)	I. Feature extraction: Based on 3D U-net	To investigate the segmentation accuracy of OARs	I. Segmentation performance - DSC: 0.62±0.02 - HD: 3.4mm±1.0mm	There was no statistical significance between the results obtained from the proposed model and manual contouring of the OARs except for the optic nerves and chiasm.	The developed model can be used for auto-contouring of OARs.	N/A
Zhang et al (2020) ⁵⁸ (China)	NPC 242 (MRI images)	Machine learning (Radiation-induced injury diagnosis)	I. Radiomic analysis: RF	To develop a model for early detection of radiation-induced temporal lobe injury	I. Predictive performance - AUC: 0.830 [CI: 0.823–0.837]	The use of texture features in feature selection improved the performance of the prediction model.	The developed models have the potential to support in providing early detection and taking preventive measures against radiation-induced temporal lobe injury.	1. Insufficient sample size. 2. 3D-conformal radiotherapy was performed on NPC patients instead of IMRT. 3. Limited generalizability as it was a single institution study.
Zhang et al (2020) ³⁶ (China)	NPC 220 (WSIs, MRI images and clinicopathological data)	Deep learning (Prognosis)	I. Prediction: Resnet-18 and DeepSurv	To explore the use of magnetic resonance imaging and microscopic whole-slide images to improve the prognosis of model	I. Prognostic performance - C-index: 0.834 [0.736–0.932]	The established nomogram had a much higher performance compared to the clinical model.	The developed multi-scale nomogram has the potential to be a non-invasive, cost-effective tool for assisting in individualized treatment and decision making on NPC.	1. The study was retrospective and the sample size was relatively small 2. A molecular profile was not used in the model. 3. The subjects in the cohorts were all Chinese, hence, the generalizability needs to be verified

Zhao et al (2020) ³⁵ (China)	NPC 123 (MRI images, clinicopathological and plasma EBV DNA data)	Machine learning (Prognosis)	I. Prediction: SVM	To investigate an MRI-based radiomics nomogram in predicting induction chemotherapy response and survival	I. Prediction performance - AUC: 0.8725 - Accuracy: 0.8696 - PPV: 71.43% - NPV: 93.75%	The proposed nomogram had a better performance than the clinical nomogram.	The constructed nomogram could be used for personalized risk stratification and for treating NPC patients that received induction chemotherapy.	1. Small sample size due to the strict inclusion criteria. 2. Limited generalizability due to the single institutional nature of the study 3. The induction chemotherapy response evaluation had a lower accuracy as the assessment was based only on anatomical MRI imaging.
Zhong et al (2020) ²⁹ (China)	NPC 638 (MRI images, clinicopathological and plasma EBV DNA data)	Deep learning (Prognosis)	I. Prediction: ResNeXt	To predict the survival of stage T3N1M0 NPC patients treated with induction chemotherapy and concurrent chemoradiotherapy	I. Model performance - C-index: 0.788 [CI: 0.695–0.882]	The DL-based radiomics model had a higher predictive performance than the clinical model.	It has the potential to be a useful non-invasive tool for risk stratification and prognostic prediction	1. Only the basilar region was used for analysis, while nasopharyngeal and other regions were not considered. 2. Only T3N1M0 patients were considered
Bai et al (2021) ³⁰ (China)	NPC 60 (CT images)	Deep learning (Auto-contouring)	I. Segmentation: ResNeXt U-net	To use computed tomography for the segmentation of NPC	I. Segmentation performance - DSC: 0.6288±0.0812 - 95HD: 6.07mm±2.53mm - FI-score: 0.6615	The developed DL algorithm had a significantly better performance than three existing DL models	An NPC-seg algorithm was developed and won 9th place on the StructSeg 2019 Challenge leader-board	N/A
Cai et al (2021) ³¹ (China)	NPC 251 (MRI images and clinicopathological data)	Deep learning (Auto-contouring)	I. Segmentation: Attention U-net with T-channel module	To use image and T-staging information to improve NPC tumor delineation accuracy	I. Segmentation performance - DSC: 0.845 [CI: 0.791–0.879] - ASSD: 0.533mm [CI: 0.174–1.254]	Having the attention module and T-channel improved the effectiveness of the model. The proposed model had the best performance over four other state-of-the-art methods.	Integrating both the attention and the T-channel module can improve the delineation performance of a model substantially	1. Small batch size due to GPU memory limitation 2. Smaller number of epochs to reduce the training time. 3. The proportion of T-channel against the input volume is small.
Tang et al (2021) ³² (China and Australia)	NPC 95 (MRI images)	Deep learning (Auto-contouring)	I. Segmentation: DA-DSU-net	To develop a model for NPC segmentation using MRI	I. Segmentation performance - DSC: 0.8050 - ASSD: 0.8021mm - Prevent match: 0.8026 - Correspondence ratio: 0.7065	The developed network had a higher performance than three other segmentation methods	The proposed model can help clinicians by delineating the tumor in order to provide accurate staging and radiotherapy planning of NPC.	1. Insufficient training data. 2. The model processes in 2D form. 3. Multi-modality input was not used.
Wen et al (2021) ⁵⁷ (China)	NPC 8194 (Clinicopathological and dosimetric data)	Machine learning (Radiation-induced injury diagnosis)	I. Dosimetric factors selection: LASSO, RF, Stochastic gradient boosting and SVM	To predict temporal lobe injury after intensity-modulated radiotherapy in NPC	I. Identification of dosimetric factors associated with temporal lobe injury incidence - AUC: 0.818 - C-index: 0.775 [CI: 0.751–0.799] - Spearman correlation matrix ³	The nomogram that included dosimetric and clinical factors had a better prediction performance than the nomogram with only DVH. D0.5cc was considered the most important dosimetric factor by LASSO, Stochastic gradient boosting and SVM.	The proposed method was able to predict temporal lobe injury accurately and can be used to help provide individualized follow-up management.	1. Selection bias due to the retrospective nature of the study. 2. No external validation due to the single institutional nature of the study. 3. TLI were not grouped by the severity.

(Continued)

Table I (Continued).

Authors, Year and Country	Site, No. of Cases (Data Type)	AI Subfield (Application)	Artificial Intelligence Methods and its Application	Study Aim	Performance Metric (s)	Results	Conclusion	Limitations
Wong et al (2021) ⁴⁵ (Hong Kong SAR, China)	NPC 412 (MRI images)	Deep learning (Diagnosis)	1. Classification: Residual Attention Network	To differentiate early stage NPC from benign hyperplasia using T2-weighted MRI	1. Diagnostic performance - AUC: 0.96 [CI: 0.94–0.98] - Sensitivity: 0.924 [CI: 0.858–0.959] - Specificity: 0.906 [CI: 0.728–0.951] - Accuracy: 0.915 - PPV: 0.905 - NPV: 0.924	The CNN obtained a good result in discriminating NPC and benign hyperplasia.	The proposed fully automatic network model demonstrated the prospect of CNN in identifying NPC at an early stage.	1. There is limited generalizability as only MRI scans of the head and neck region with the field of view centered on the nasopharynx can be used. 2. No external validation. 3. No association between CNN score and having a nasopharyngeal biopsy. 4. Possible chance to include undetected NPC in patients with benign hyperplasia.
Wong et al (2021) ⁵⁶ (Hong Kong SAR, China)	NPC 201 (MRI images)	Deep learning (Auto-contouring)	1. Delineation: U-net	To evaluate the delineation performance of a model using non-contrast-enhanced MRI	1. Delineation performance - DSC: 0.71±0.09 - ASSD: 2.1mm±4.8mm - Δ Primary tumor volume: 1.0 ±12.2cm ³	The performance of CNN using fs-T2W images was similar to that of CNNs using contrast-enhanced-T1 weighted and contrast-enhanced-fat-suppressed-T1 weighted images.	Although using contrast-enhanced sequence for head and neck MRI is still recommended, when avoiding use of contrast agent is preferred, CNN is a potential future option.	1. Limited generalizability to other CNN architectures due to variations in tissue contrasts. 2. Only considered slice-based algorithms and no other ones. 3. The study did not take into account whether CNN delineation performance could be affected by different fat-suppression techniques.
Wu et al (2021) ³⁴ (China)	NPC and other types 233 (MRI images)	Machine learning and deep learning (Prognosis)	1. Classification: Resnet18 2. Feature selection LASSO	To assess the predicted value of peritumoral regions and explore the effects of different peritumoral sizes in learning models	1. Model performance a) AUC: 0.660 [CI: 0.484–0.837] b) Sensitivity: 0.344 [CI: 0.179–0.508] c) Specificity: 0.800 [CI: 0.598–1.000]	Radiomics is more suitable than DL for modelling peritumors	The peritumoral models, and ML and DL helped improved the prediction performance.	1. Datasets were small. 2. Maximum slice of the tumor was used for analysis directly, but multi-slices of the tumor were not.
Zhang et al (2021) ³³ (China)	NPC 252 (MRI images, clinicopathological and plasma EBV DNA data)	Machine learning and deep learning (Prognosis)	1. Prediction: Residual network and LR analysis 2. Feature selection: Minimum redundancy-maximum relevance, LASSO & Akaike information criterion algorithms	To predict DMFS and to investigate the influence of additional chemotherapy to concurrent chemoradiotherapy for different risk groups.	1. Prediction performance of Distant metastasis-free survival a) AUC - DMFS: 0.808 [CI: 0.654–0.962]	By integrating DL signature with N stage, EBV DNA and treatment regimen, the MRI-based combined model had a better predictive performance than the DL signature-based, radiomic signature-based and clinical-based model	The MRI-based combined model could be used as a complementary tool for making treatment decisions by assessing the risk of DMFS in locoregionally advanced NPC patients	1. The value of the deep learning model and the collected information were limited. 2. The repeatability of radiomic signatures was poor. 3. The generalizability of the model is affected due to difference in scan protocols between institutions.

Notes: ^aIndicates performance metric presented in graph and not as a numerical value. ^bValues found in publication.

Abbreviations: NPC, nasopharyngeal carcinoma; MRI, magnetic resonance imaging; SVM, support vector machines; KNN, k-nearest neighbor; ANN, artificial neural network; AUC, area under the receiver operating characteristic curve; ML, machine learning; PET, positron emission tomography; CT, computed tomography; PPV, positive predictive values; NPV, negative predictive values; HD, Hausdorff distance; LR, logistic regression; RF, random forest; C-index, concordance index; CNN, convolutional neural network; IMRT, intensity-modulated radiation therapy; DVH, dose-volume histogram; MAE, mean absolute error; OAR, organ-at-risk; EBV DNA, Epstein-Barr Virus DNA; DL, deep learning; VMAT, volumetric modulated arc therapy; WSI, whole slide image; LASSO, least absolute shrinkage and selection operator; OS, Overall survival; DMFS, distant metastasis-free survival; LRFS, local-region relapse-free survival; AJCC, American Joint Committee on Cancer; PFS, progression-free survival.

Table 2 Quality Assessment via the QUADAS-2 Tool

Authors Publication Year	Risk of Bias ^a				Applicability Concerns ^a			At Risk ^b	
	Patients Selection	Index Test	Reference Standard	Flow and Timing	Patients Selection	Index Test	Reference Standard	Risk of Bias	Applicability
Wang et al (2010) ¹⁰	?	✓	✓	✓	✓	✓	✓	Yes	No
Aussem et al (2012) ¹¹	?	✓	✓	✓	✓	✓	✓	Yes	No
Kumdee, Bhongmakapat and Ritthipravit (2012) ¹²	✓	✓	✓	✓	✓	✓	✓	No	No
Ritthipravit, Kumdee, and Bhongmakapat (2013) ¹³	✓	✓	✓	✓	✓	✓	✓	No	No
Zhu and Kan (2014) ⁴⁴	✓	✓	✓	✓	✓	✓	✓	No	No
Jiang et al (2016) ¹⁴	?	✓	✓	✓	✓	✓	✓	Yes	No
Liu et al (2016) ⁴³	✓	✓	✓	?	✓	✓	✓	Yes	No
Wang et al (2016) ⁶³	?	✓	✓	✓	✓	✓	✓	Yes	No
Men et al (2017) ⁵⁵	✓	✓	✓	?	✓	✓	✓	Yes	No
Mohammed et al (2017) ⁵⁴	?	✓	✓	✓	✓	✓	✓	Yes	No
Zhang et al (2017) ⁴²	✓	✓	✓	✓	✓	✓	✓	No	No
Zhang et al (2017) ⁴¹	✓	✓	✓	?	✓	✓	✓	Yes	No
Li et al (2018) ¹⁵	✓	✓	✓	✓	✓	✓	✓	No	No
Mohammed et al (2018) ⁶²	?	✓	✓	✓	✓	✓	✓	Yes	No
Mohammed et al (2018) ¹⁶	?	✓	✓	✓	✓	✓	✓	Yes	No
Du et al (2019) ⁴⁰	✓	✓	✓	✓	✓	✓	✓	No	No
Jiao et al (2019) ⁶⁶	✓	✓	?	✓	✓	✓	✓	Yes	No
Jing et al (2019) ¹⁷	✓	✓	✓	✓	✓	✓	✓	No	No
Li et al (2019) ⁵³	✓	✓	✓	✓	✓	✓	✓	No	No
Liang et al (2019) ⁵²	✓	✓	✓	✓	✓	✓	✓	No	No
Lin et al (2019) ⁵¹	✓	✓	✓	✓	✓	✓	✓	No	No
Liu et al (2019) ⁶⁵	✓	✓	?	✓	✓	✓	✓	Yes	No
Ma et al (2019) ¹⁸	?	✓	✓	✓	✓	✓	✓	Yes	No
Peng et al (2019) ¹⁹	✓	✓	✓	✓	✓	✓	✓	No	No
Rehioui et al (2019) ²⁰	✓	✓	✓	?	✓	✓	✓	Yes	No
Zhong et al (2019) ⁵⁰	?	✓	✓	✓	✓	✓	✓	Yes	No
Zou et al (2019) ²¹	?	✓	✓	✓	✓	✓	✓	Yes	No
Abd Ghani et al (2020) ⁶¹	?	✓	✓	✓	✓	✓	✓	Yes	No
Bai et al (2020) ⁶⁴	✓	✓	✓	✓	✓	✓	✓	No	No
Chen et al (2020) ²²	?	✓	✓	✓	✓	✓	✓	Yes	No

(Continued)

Table 2 (Continued).

Authors Publication Year	Risk of Bias ^a				Applicability Concerns ^a			At Risk ^b	
	Patients Selection	Index Test	Reference Standard	Flow and Timing	Patients Selection	Index Test	Reference Standard	Risk of Bias	Applicability
Chen et al (2020) ⁶⁸	✓	✓	✓	✓	✓	✓	✓	No	No
Chuang et al (2020) ²³	✓	✓	✓	?	✓	✓	✓	Yes	No
Cui et al (2020) ³⁹	✓	✓	✓	✓	✓	?	✓	No	Yes
Diao et al (2020) ⁶⁰	✓	✓	✓	?	✓	✓	✓	Yes	No
Du et al (2020) ⁵⁹	✓	✓	✓	✓	✓	✓	✓	No	No
Guo et al (2020) ²⁴	?	✓	✓	✓	✓	✓	✓	Yes	No
Jing et al (2020) ²⁵	✓	✓	✓	✓	✓	✓	✓	No	No
Ke et al (2020) ⁴⁹	✓	✓	✓	✓	✓	✓	✓	No	No
Liu et al (2020) ³⁸	?	✓	✓	✓	✓	✓	✓	Yes	No
Men et al (2020) ⁴⁷	?	✓	✓	✓	✓	✓	✓	Yes	No
Mohammed et al (2020) ²⁶	?	✓	✓	✓	✓	✓	✓	Yes	No
Wang et al (2020) ²⁷	✓	✓	✓	✓	✓	✓	✓	No	No
Wang et al (2020) ⁴⁸	✓	✓	✓	?	✓	✓	✓	Yes	No
Xie et al (2020) ³⁷	✓	✓	✓	✓	✓	✓	✓	No	No
Xue et al (2020) ⁶⁹	✓	✓	✓	✓	✓	✓	✓	No	No
Xue et al (2020) ⁴⁶	✓	✓	✓	?	✓	✓	✓	Yes	No
Yang et al (2020) ²⁸	✓	✓	✓	✓	✓	✓	✓	No	No
Yang et al (2020) ⁶⁷	?	✓	✓	✓	✓	✓	✓	Yes	No
Zhang et al (2020) ⁵⁸	✓	✓	✓	✓	✓	✓	✓	No	No
Zhang et al (2020) ³⁶	✓	✓	✓	✓	✓	✓	✓	No	No
Zhao et al (2020) ³⁵	✓	✓	✓	?	✓	✓	✓	Yes	No
Zhong et al (2020) ²⁹	✓	✓	✓	✓	✓	✓	✓	No	No
Bai et al (2021) ³⁰	✓	✓	✓	?	✓	✓	✓	Yes	No
Cai et al (2021) ³¹	✓	✓	X	✓	✓	✓	✓	Yes	No
Tang et al (2021) ³²	?	✓	✓	✓	✓	✓	✓	Yes	No
Wen et al (2021) ⁵⁷	✓	✓	?	✓	✓	✓	✓	Yes	No
Wong et al (2021) ⁴⁵	✓	✓	✓	✓	✓	✓	✓	No	No
Wong et al (2021) ⁵⁶	✓	✓	X	✓	✓	✓	✓	Yes	No
Wu et al (2021) ³⁴	✓	✓	✓	✓	✓	✓	✓	No	No
Zhang et al (2021) ³³	✓	✓	✓	?	✓	✓	✓	Yes	No

Notes: ^aA check mark (✓) refers to passing (ie, absence of risk) of the criteria; a cross mark (X) refers to not passing (ie, presence of risk) of the criteria; and a question mark (?) refers to missing information to assess the criteria. ^bThe domain “risk of bias” and “applicability” were considered as no bias (ie, “No”) if passing all of the corresponding criteria (ie, all ✓); and were considered as having bias (ie, “Yes”) if not passing any of the corresponding criteria (ie, at least one ? Or X).

Table 3 Quality Assessment Guidelines

Article Sections	Parameters	Explanation
Title and Abstract	Title (Nature of study)	Introduce predictive model
	Abstract (Structured summary)	Include background, objectives, data sources, performance metrics of predictive models and conclusion about model value
Introduction	Rationale	Define the clinical goal, and review the current practice and performance of existing models
	Objectives	Identify how the proposed method can benefit the clinical target
Method	Describe the setting	Describe the data source, sample size, year and duration of the data
	Define the prediction problem	Define the nature of the study (retrospective/prospective), model function (prognosis, diagnosis, etc.) and performance metrics
	Prepare data for model building	Describe the inclusion and exclusion criteria of the data, data pre-processing method, performance metrics for validation, and define the training and testing set. External validation is recommended
	Build the predictive model	Describe how the model was built including AI modelling techniques used (eg random forest, ANN, CNN)
Results	Report the final model and performance	Reports the performance of the final proposed model, comparison with other models and human performance. It is recommended to include confidence intervals
Discussion	Clinical implications	Discuss any significant findings
	Limitations of the model	Discuss any possible limitations found
Conclusion		Discuss the clinical benefit of the model and summarize the result and findings

Note: Data from the guideline of Luo et al.⁸

reporting ML predictive models from Luo et al and Alabi et al (Table 3).^{8,9} The guideline was summarised, and a mark was given for each guideline topic followed. The threshold was set at half of the maximum marks, and the score was presented in Table 4.

Results

Database Search

The selection process was performed using the PRISMA flow diagram in Figure 1. 304 papers were retrieved from the three databases. After 148 duplicates were removed, one inaccessible article was rejected. The papers not meeting the inclusion (n=59) and exclusion (n=20) criteria were also filtered out. Moreover, two additional studies found in literature reviews were included after removing one for being duplicated and another that did not meet the exclusion criteria. Finally, 78 papers were then assessed for quality (Figure 1).

Quality Assessment

18 papers failed due to having more than one section with a “high” or “unclear” rating, leaving 60 studies to be

further evaluated. The QUADAS-2 tool showed that 48.3% of articles showed an overall low risk of bias, while 98.3% of them had a low concern regarding applicability (Table 2).

An additional evaluation was performed based on Table 3, which was adapted from the guidelines by Luo et al and the modified version from Alabi et al.^{8,9} Of the 60 relevant studies, 52 of them scored greater than 70% (Table 4). It should also be noted that 23 papers included the evaluation criteria items but did not fully follow the structure of the proposed guidelines.^{10–32} However, this only affects the ease of reading and extracting information from the articles, but not the content and quality of the papers.

Characteristics of Relevant Studies

The characteristics of the 7 articles finally included in the current study were shown in Table 1. The articles were published in either English (n=57)^{10–66} or Chinese (n=3);^{67–69} 3 studies examined sites other than the NPC.^{10,17,34}

Table 4 Quality Scores of the Finalized Articles

Studies	Title	Abstract	Rationale	Objectives	Setting Description	Problem Definition	Data Preparation	Build Model	Report Performance	Clinical Implications	Limitations	Scores (%)
Wang et al (2010) ¹⁰	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	73%
Aussem et al (2012) ¹¹	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	73%
Kumdee, Bhongmakapat and Ritthipravat (2012) ¹²	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✗	82%
Ritthipravat, Kumdee and Bhongmakapat (2013) ¹³	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✗	73%
Zhu and Kan (2014) ⁴⁴	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	82%
Jiang et al (2016) ¹⁴	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	82%
Liu et al (2016) ⁴³	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Wang et al (2016) ⁶³	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	82%
Men et al (2017) ⁵⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Mohammed et al (2017) ⁵⁴	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✓	82%
Zhang et al (2017) ⁴²	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	91%
Zhang et al (2017) ⁴¹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Li et al (2018) ¹⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Mohammed et al (2018) ⁶²	✓	✗	✓	✓	✗	✗	✓	✓	✓	✗	✗	55%
Mohammed et al (2018) ¹⁶	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗	64%
Du et al (2019) ⁴⁰	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Jiao et al (2019) ⁶⁶	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	73%
Jing et al (2019) ¹⁷	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✗	55%
Li et al (2019) ⁵³	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	91%
Liang et al (2019) ⁵²	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Lin et al (2019) ⁵¹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Liu et al (2019) ⁶⁵	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	91%

Ma et al (2019) ¹⁸	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	73%
Peng et al (2019) ¹⁹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Rehioui et al (2019) ²⁰	✓	✗	✗	✗	✓	✓	✗	✓	✓	✓	✗	55%
Zhong et al (2019) ⁵⁰	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	✗	73%
Zou et al (2019) ²¹	✓	✗	✓	✓	✗	✗	✓	✓	✓	✓	✗	64%
Abd Ghani et al (2020) ⁶¹	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	✗	64%
Bai et al (2020) ⁶⁴	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	82%
Chen et al (2020) ²²	✓	✗	✓	✓	✗	✓	✓	✓	✓	✓	✗	73%
Chen et al (2020) ⁶⁸	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	91%
Chuang et al (2020) ²³	✓	✓	✗	✗	✓	✓	✓	✓	✓	✗	✗	64%
Cui et al (2020) ³⁹	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	91%
Diao et al (2020) ⁶⁰	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	91%
Du et al (2020) ⁵⁹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	91%
Guo et al (2020) ²⁴	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗	82%
Jing et al (2020) ²⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	91%
Ke et al (2020) ⁴⁹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Liu et al (2020) ³⁸	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	91%
Men et al (2020) ⁴⁷	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓	73%
Mohammed et al (2020) ²⁶	✓	✗	✓	✓	✗	✗	✓	✓	✗	✗	✓	55%
Wang et al (2020) ²⁷	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓	82%
Wang et al (2020) ⁴⁸	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Xie et al (2020) ³⁷	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Xue et al (2020) ⁶⁹	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	82%
Xue et al (2020) ⁴⁶	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	82%
Yang et al (2020) ²⁸	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%

(Continued)

Table 4 (Continued).

Studies	Title	Abstract	Rationale	Objectives	Setting Description	Problem Definition	Data Preparation	Build Model	Report Performance	Clinical Implications	Limitations	Scores (%)
Yang, et al (2020) ⁶⁷	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	82%
Zhang et al (2020) ⁵⁸	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Zhang et al (2020) ³⁶	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Zhao et al (2020) ³⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Zhong et al (2020) ²⁹	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Bai et al (2021) ³⁰	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	73%
Cai et al (2021) ³¹	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	91%
Tang et al (2021) ³²	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	91%
Wen et al (2021) ⁵⁷	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Wong et al (2021) ⁴⁵	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Wong et al (2021) ⁵⁶	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%
Wu et al (2021) ³⁴	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	91%
Zhang et al (2021) ³³	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100%

Notes: A check mark (✓) refers to passing of the criteria; and a cross mark (X) refers to not passing of the criteria. The score refers to the proportion of passed criteria for that publication. Assessment parameters based on the guideline of Luo et al.⁸

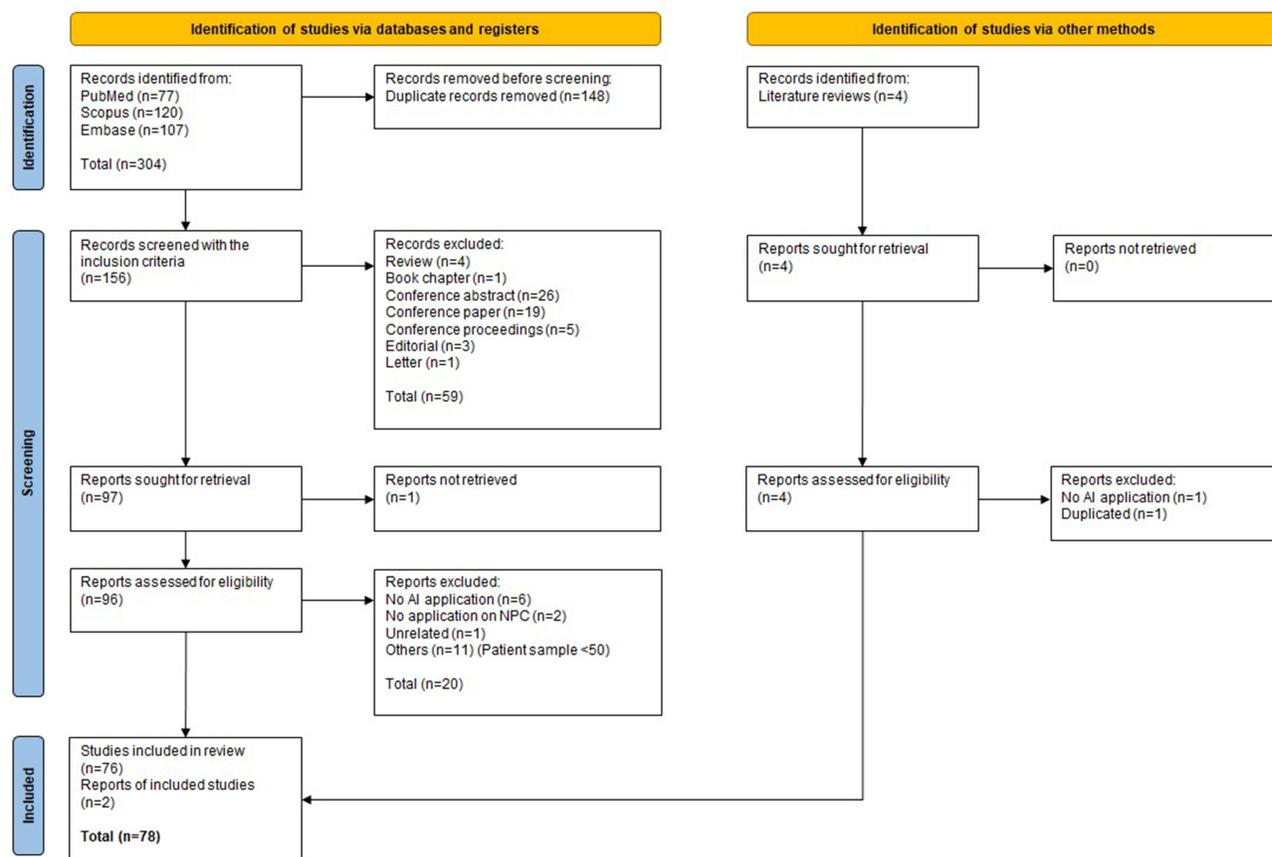


Figure 1 PRISMA flow diagram 2020.

Notes: Adapted from Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. Creative Commons license and disclaimer available from: <http://creativecommons.org/licenses/by/4.0/legalcode>.⁶

When observing the origins of the studies, 45 were published in Asia, while Morocco and France contributed one study each. Furthermore, 13 papers were collaborated work from multiple countries. The majority of the studies were from the endemic regions.

The articles used various types of data to train the models. 66.7% (n=40) only used imaging data such as magnetic resonance imaging, computed tomography or endoscopic images.^{15,16,18,19,21–24,26–28,30,32,34,37–39,41–43,45–56,58–63,67,69} There were also four studies that included clinicopathological data as well as images for training models,^{25,31,36,40} while three other studies developed models using images, clinicopathological data, and plasma Epstein-Barr virus (EBV) DNA.^{29,33,35} Furthermore, 4 studies used treatment plans,^{64–66,68} while proteins and microRNA expressions data were each extracted by one study.^{10,44} There were also four articles that trained with both clinicopathological and plasma EBV DNA/serology data,^{12–14,17} while one article trained its model with clinicopathological and dosimetric data.⁵⁷

Risk factors (n=2), such as demographic, medical history, familial cancer history, dietary, social and environmental factors, were also used to develop AI models.^{11,20}

The studies could be categorized into 4 domains, which were auto-contouring (n=21),^{15,16,18,22,24,30–32,45–55,67,69} diagnosis (n=17),^{10,15,16,23,26,27,49,52,54,56–63} prognosis (n=20)^{12–14,17,19,25,28,29,33–44} and miscellaneous applications (n=7),^{11,20,21,64–66,68} which included risk factor identification, image registration and radiotherapy planning (Figure 2A). Five studies examined both diagnosis and auto-contouring simultaneously.^{15,16,49,52,54}

Analyses on the purpose of the application showed that, only in auto-contouring, DL is the most heavily used (with 19 out of 22 instances). For the rest of the categories (NPC diagnosis, prognosis and miscellaneous applications), ML is the most common technique (more than half of the publications in each category) (Figure 2A). In addition, studies applying DL models selected in this literature review were published from 2017 to 2021, where

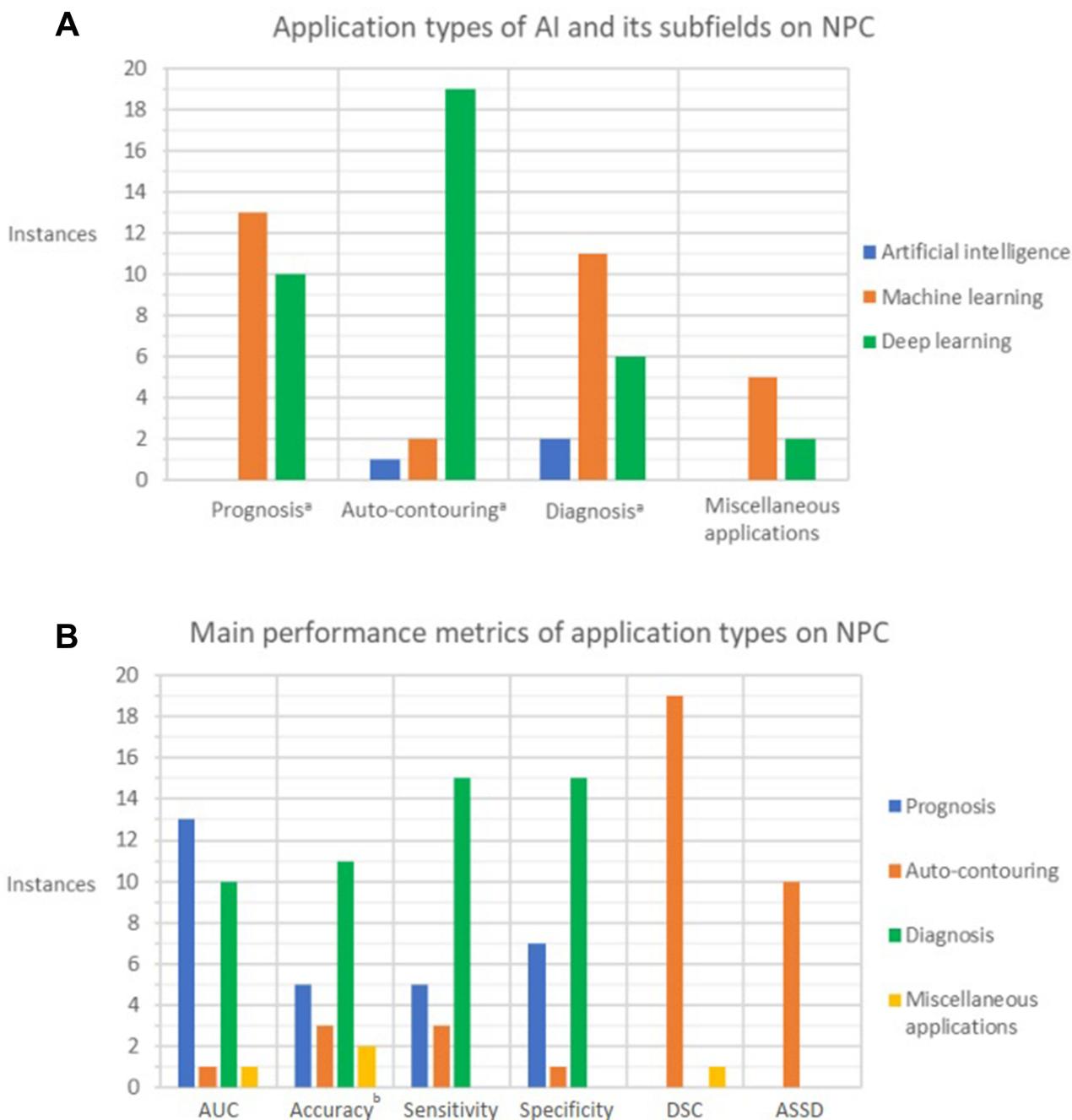


Figure 2 Comparison of studies on AI application for NPC management. **(A)** Application types of AI and its subfields on NPC; **(B)** Main performance metrics of application types on NPC.

Notes: ^aMore than one AI subfield (artificial intelligence, machine learning and deep learning) was used in the same study. ^bAuto-contouring and diagnosis accuracy values were found in the same study.⁵⁴

Abbreviations: AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; DSC, dice similarity coefficient; ASSD, average symmetric surface distance; NPC, nasopharyngeal carcinoma.

there was a heavier focus on experimenting with DL. It was observed that the majority of the papers applying DL models used various forms of CNN (n=30),^{15,18,19,21–24,28–34,36,45–53,55,56,60,65,67,69} while the main ML method used was ANN (n=12).^{13,16,26,42–44,54,61–64,68}

The primary metrics reported were the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, dice similarity coefficient (DSC) and average symmetric surface distance (ASSD), as shown in Figure 2B.

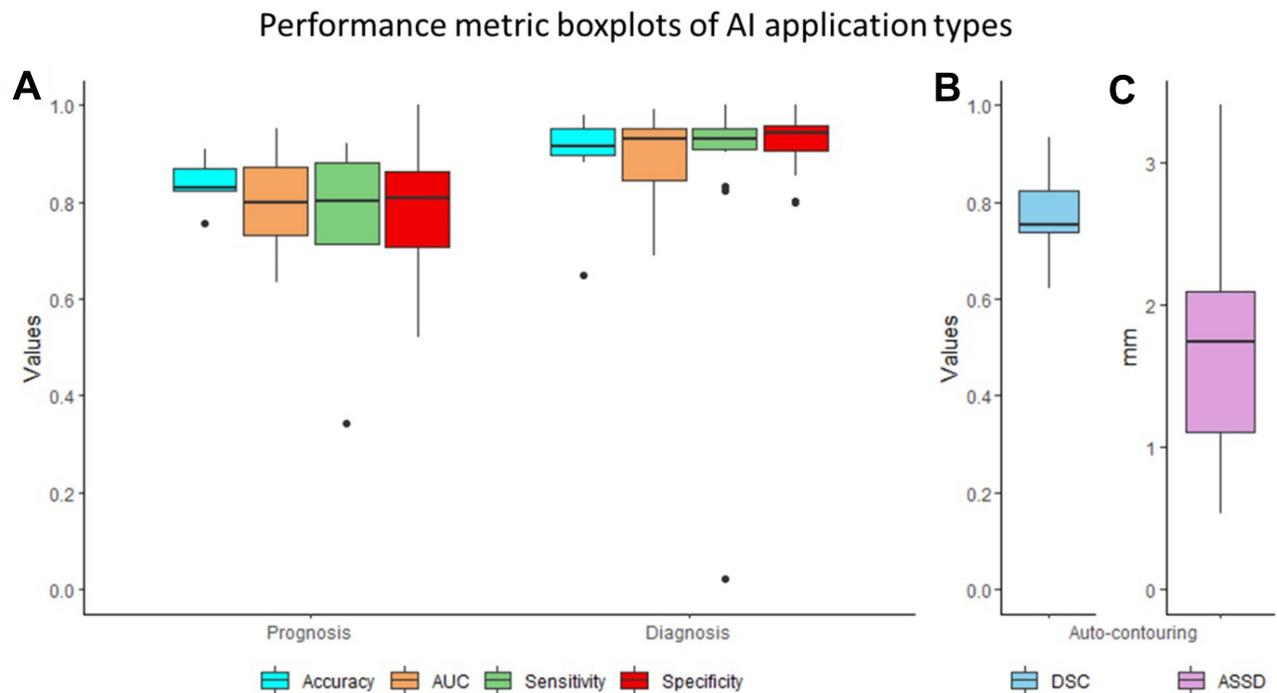


Figure 3 Performance metric boxplots of AI application types on NPC. (A) Prognosis and diagnosis: accuracy, AUC, sensitivity and specificity metric; (B) Auto-contouring: DSC metric; (C) Auto-contouring: ASSD metric.

Abbreviations: AI, artificial intelligence; ASSD, average symmetric surface distance; AUC, area under the receiver operating characteristic curve; DSC, dice similarity coefficient; NPC, nasopharyngeal carcinoma.

AUC was used to evaluate the models' capabilities in 25 papers, with the majority measuring the prognostic (n=13)^{12–14,19,28,33–35,37,39,40,42,44} and diagnostic abilities (n=10).^{15,23,26,27,49,56–60} Similarly, accuracy was the parameter most frequently reported in the diagnosis and prognosis application: 11 and 5 out of 20 articles respectively.^{10,12,15,26–28,35,43,44,49,54,56,60–63} Sensitivity was the most common studied parameter for diagnostic performance: 15 out of 23 papers.^{10,15,16,23,26,27,49,52,54,56,59–63} The specificity was only reported for prognosis (n=7)^{12,14,28,34,39,40,43} and diagnosis (n=15).^{10,15,16,23,26,27,49,52,54,56,59–63} In addition, the DSC (n=20)^{15,18,22,24,30–32,45–53,55,65,67,69} and ASSD (n=10)^{18,22,24,31,32,45,46,48,51,69} were the primary metrics reported in studies on auto-contouring (Figure 2B).

Performance metrics with five or more instances of each application method were presented in a boxplot (Figure 3). The median AUC, accuracy, sensitivity and specificity of prognosis were 0.8000, 0.8300, 0.8003 and 0.8070 respectively, while their range were 0.6330–0.9510, 0.7559–0.9090, 0.3440–0.9200 and 0.5200–1.000 respectively. For diagnosis, the AUC's median was 0.9300, while the median accuracy was 0.9150. In addition, the median sensitivity and specificity were 0.9307

and 0.9413, respectively. The range for diagnosis' AUC, accuracy, sensitivity and specificity were 0.6900–0.9900, 0.6500–0.9777, 0.0215–1.000 and 0.8000–1.000, respectively. The median DSC value for auto-contouring was 0.7530, while the range was 0.6200–0.9340. Furthermore, the median ASSD for auto-contouring was 1.7350 mm, and the minimum and maximum values found in the studies were 0.5330 mm and 3.4000 mm, respectively.

Auto-Contouring

Publications on auto-contouring experimented on segmenting gross tumor volumes, clinical target volume, OARs and primary tumor volumes. The target most delineated was the gross target volume (n=7),^{30,48,49,51,53,55,69} while the second most were the OARs (n=3).^{50,52,67} The clinical target volumes and the primary tumor volume were studied in two and one articles respectively.^{46,55,56} However, nine articles did not mention the specific target volume contoured.^{15,16,18,22,24,31,32,47,54} Two out of three articles reported that the DSC for delineating optic nerves was substantially lower than the other OARs.^{52,67} In contrast, for the remaining paper, although the segmentation of the optic nerve is not the worst, the paper reported that

the three OARs it tested, which included optic nerves, were specifically more challenging to contour.⁵⁰ This is because of the low soft-tissue contrast in computed tomography images and their diverse morphological characteristics. When analyzing the OARs, automatic delineation of the eyes yielded the best DSC. Furthermore, apart from the spinal cord, optic nerve and optic chiasm, the AI models have a DSC value greater than 0.8 when contouring OARs.^{50,52,67}

Diagnosis

As for the detection of NPC, six papers compared the performance of AI and humans. Two of them found that AIs had better diagnostic capabilities than humans (oncologists and experienced radiologists),^{15,49} while another two reported that AIs had similar performances to ear, nose and throat specialists.^{16,62} However, the last two papers found that it depends on the experience of the person. For example, senior-level clinicians performed better than the AI, while junior level ones were worse.^{23,60} This is because of the variations in possible sizes, shapes, locations, and image intensities of NPC, making it difficult to determine the diagnosis. These factors make it challenging for clinicians with less experience, and it showed that AI diagnostic tools could support junior-level clinicians.

On the other hand, within the 17 papers experimenting on the diagnostic application of AI, three articles analyzed radiation-induced injury diagnosis.^{27,57,58} Two of which were concerned with radiation-induced temporal lobe injury,^{57,58} while the remaining one predicted the fibrosis level of neck muscles after radiotherapy.²⁷ It was suggested that through early detection and prediction of radiation-induced injuries, preventive measures could be taken to minimize the side effects.

Prognosis

For studies on NPC prognosis, 11 out of 20 publications focused on predicting treatment outcomes, with the majority including disease-free survival as one of the study objectives.^{12,13,17,19,29,33,36,39–42} The rest studied treatment response prediction (n=2),^{35,43} predicting patients' risk of survival (n=5),^{14,25,37,38,44} T staging prediction and the prediction of distant metastasis (n=2).^{28,34} Therefore, the versatility of AI in different functionalities was demonstrated. The performances of the models were reported in (Table 1) and the main metric analyzed was AUC with 13 out of 25 articles (Figure 2B).

Miscellaneous Applications

In addition to the above aspects, AI was also used to study risk factor identification (n=2),^{11,20} image registration (n=1)²¹ and dose/dose-volume histogram (DVH) distribution (n=4).^{64–66,68} In particular, dose/DVH distribution prediction was frequently used for treatment planning. A better understanding of the doses given to the target and OARs can help clinicians give a more individualized treatment plan with better consistency and a lower planning duration. However, further development is required to obtain similar plan qualities as created by people. This is because one paper's model showed the same quality as manual planning by an experienced physicist,⁶⁴ but another study using a different model was unable to achieve a similar plan quality designed by even a junior physicist.⁶⁸

Discussion

As evident in this systematic review, there is an exponential growth in interest to apply AI for the clinical management of NPC. A large proportion of the articles collected were published from 2019 to 2021 (n=45) compared to that from 2010 to 2018 (n=15).

A heavier focus is also placed on specific fields of AIs, such as ML and DL. There are only three reports on AI, while there are 31 studies on ML and 37 on DL. The choice of AI subfield sometimes depends on the task. For example, 86% of the papers focused on DL for NPC auto-contouring (n=19), while although the majority of the studies in the other applications used ML, they were more evenly distributed (Figure 2A). The reason why there is such a significant difference in the type of AI used in auto-contouring may be due to the capability of the algorithms and the nature of the data. The medical images acquired have many factors affecting the auto-contouring quality; these include the varying tumour sizes and shapes, image resolution, contrast between regions, noise and lack of consistency during data acquisition being collected from different institutions.⁷⁰ Because of these challenges, ML-based algorithms have difficulty in performing automated segmentation on NPC as image processing before training is required, which is time-consuming. Furthermore, handcrafted features are necessary to precisely contour each organ or tumour as there are significant variations in size and shape for NPC. On the other hand, DL does not have this issue as they can

process the raw data directly without the need for hand-crafted features.⁷⁰

ANN is the backbone of DL, as DL algorithms are ANNs with multiple (2 or more) hidden layers. In the development of AI applications for NPC, 80% of the studied articles incorporated either ANN or DL technique in their models-^{12,13,15-19,21-26,28-34,36,38,39,42-56,60-69} because neural networks are generally better for image recognition. However, one study cautioned that ANNs were not necessarily better than other ML models in NPC identification.⁶¹ Hence, even though DL-based models and ANNs should be considered the primary development focus, other ML techniques should still not be neglected.

Based on the literature collected, the integration of AI applications in each category is beneficial to the practitioner. Automated contouring by AIs not only can make contouring less time-consuming for clinicians,^{46,51,53,64} it can also help to improve the user's accuracy.⁵¹ Similarly, AI can be used to reduce the treatment planning time for radiotherapy,⁶⁴ thus improve the efficiency and effectiveness of the radiotherapy planning process.

For some NPC studies, additional features from images and parameters were extracted to further improve the performances of models. However, it should be noted that not all features are suitable as some features have a more significant impact on the model's performance than others.^{40,57,58,61} Therefore, feature selection should be considered where possible.

At its current state, AI cannot yet replace humans to perform the most complex and time-consuming tasks. This is because multiple articles which compared the performance of their developed model with medical professionals showed conflicting results. The reason for this is that the experience of the clinician is an important factor that affects the resulting comparison. The models developed by Chuang et al and Diao et al performed better than junior-level professionals, but performed worse when compared to more experienced clinicians.^{23,60} One article even showed that an AI model had a lower capability than a junior physicist.⁶⁸ Furthermore, the quality of the training data and the experience of the AI developers are critical.

The review revealed that AI at its current state still has several limitations. The first concern was the uncertainty regarding the generalizability of the models, because datasets of many studies are retrospective and single institutional in nature.^{15,19,28,33,35-38,41,48,57-59} The dataset may not represent the true population and may only represent a

population subgroup or a region. Hence, this reduces the applicability of the models and affects their performance when applied to other datasets. Another reason was the difference in scan protocol between institutions. Variations in tissue contrasts or field of views may affect the performance as the model was not trained for the same condition.^{45,56} Therefore, consistency of scan protocols among different institutions is important to facilitate AI model training and validation.

Another limitation was the small amount of data used to train the models. 33% (n=20) of the articles chosen had ≤ 150 total samples for both training and testing the model. The reason for this was not only were the articles usually based on single-centre data, but also because NPC is less common compared to other cancers. This particularly affects DL-based models as they are more reliant on a much larger dataset to achieve their potential when compared to ML models; over-fitting will likely occur when there is only limited data; thus, data augmentation is often used to increase the dataset size. In addition, some studies had patient selection bias, while others had concerns about not implementing multi-modality inputs into the training model (Table 1).

Future work should address these issues when developing new models. Possible solutions include incorporating other datasets or cooperating with other institutions for external validation or to expand the dataset, which were lacking in most of the analysed papers in this review. The former suggestion can boost generalizability and avoid any patient selection bias, while the latter method can increase the capability of the AI models by providing more training samples. Other methods to expand dataset have also been explored, one of which is by using big data which can be done at a much larger scale. Big data can be defined as the vast data generated by technology and the internet of things, allowing easier access to information.⁷¹ In the healthcare sector, it will allow easier access to an abundance of medical data which will facilitate AI model training. However, with the large collection of data, privacy protection becomes a serious challenge. Therefore, future studies are required to investigate how to implement it.

The performances of the AI models could also be improved by increasing the amount of data and diversifying it with data augmentation techniques which were performed in some of the studies. However, it should be noted that with an increase in training samples, more data labelling will be required, making the process more time-consuming. Hence, one study proposed the use of continual

learning, which it found to boost the model's performance while reducing the labelling effort.⁴⁷ However, continual learning is susceptible to catastrophic forgetting, which is a long-standing and highly challenging issue.⁷² Thus, further investigation into methods to resolve this problem would be required to make it easier to implement in other research settings.

There are several limitations in this literature review. The metric performance results extracted from the publications were insufficient to perform a meta-analysis. Hence, the insight obtained from this review is not comprehensive enough. The quality of the included studies was also not consistent, which may affect the analysis performed.

Conclusion

There is growing evidence that AI can be applied in various situations, particularly as a supporting tool in prognostic, diagnostic and auto-contouring applications and to provide patients with a more individualized treatment plan. DL-based algorithm was found to be the most frequently used AI subfield and usually obtained good results when compared to other methods. However, limited dataset and generalizability are key challenges that need to be overcome to further improve the performances and accessibility of AI models. Nevertheless, studies on AI demonstrated highly promising potential in supporting medical professionals in the management of NPC; therefore, more concerted efforts in swift development is warranted.

Disclosure

Dr Nabil F Saba reports personal fees from Merck, GSK, Pfizer, Uptodate, and Springer, outside the submitted work; and Research funding from BMS and Exelixis. Professor Raymond KY Tsang reports non-financial support from Atos Medical Inc., outside the submitted work. The authors report no other conflicts of interest in this work.

References

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–249. doi:10.3322/caac.21660
- Ferlay J, Ervik M, Lam F, et al. Global cancer observatory: cancer today; 2020. Available from: <https://gco.iarc.fr/today>. Accessed June 4, 2021.
- Lee AWM, Ma BBY, Ng WT, Chan ATC. Management of nasopharyngeal carcinoma: current practice and future perspective. *J Clin Oncol.* 2015;33(29):3356–3364. doi:10.1200/JCO.2015.60.9347
- Chan JW, Parvathaneni U, Yom SS. Reducing radiation-related morbidity in the treatment of nasopharyngeal carcinoma. *Future Oncology.* 2017;13(5):425–431. doi:10.2217/fo-2016-0410
- Shimizu H, Nakayama KI. Artificial intelligence in oncology. *Cancer Sci.* 2020;111(5):1452–1460. doi:10.1111/cas.14377
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. doi:10.1136/bmj.n71
- Whiting PF, Rutjes AWS, Westwood ME. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529–536. doi:10.7326/0003-4819-155-8-201110180-00009
- Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J Med Internet Res.* 2016;18(12):e323. doi:10.2196/jmir.5870
- Alabi RO, Youssef O, Pirinen M, et al. Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future—A systematic review. *Artif Intell Med.* 2021;115:102060. doi:10.1016/j.artmed.2021.102060
- Wang HQ, Zhu HL, Cho WCS, Yip TTC, Ngan RKC, Law SCK. Method of regulatory network that can explore protein regulations for disease classification. *Artif Intell Med.* 2010;48(2):119–127. doi:10.1016/j.artmed.2009.07.011
- Aussem A, de Moraes SR, Corbex M. Analysis of nasopharyngeal carcinoma risk factors with Bayesian networks. *Artif Intell Med.* 2012;54(1):53–62. doi:10.1016/j.artmed.2011.09.002
- Kumdee O, Bhongmakapat T, Ritthipravat P. Prediction of nasopharyngeal carcinoma recurrence by neuro-fuzzy techniques. *Fuzzy Sets Syst.* 2012;203:95–111. doi:10.1016/j.fss.2012.03.004
- Ritthipravat P, Kumdee O, Bhongmakap T. Efficient missing data technique for prediction of nasopharyngeal carcinoma recurrence. *Inf Technol J.* 2013;12:1125–1133. doi:10.3923/itj.2013.1125.1133
- Jiang R, You R, Pei X-Q, et al. Development of a ten-signature classifier using a support vector machine integrated approach to subdivide the M1 stage into M1a and M1b stages of nasopharyngeal carcinoma with synchronous metastases to better predict patients' survival. *Oncotarget.* 2016;7(3):3645–3657. doi:10.18632/oncotarget.6436
- Li C, Jing B, Ke L, et al. Development and validation of an endoscopic images-based deep learning model for detection with nasopharyngeal malignancies. *Cancer Commun.* 2018;38(1):59. doi:10.1186/s40880-018-0325-9
- Mohammed MA, Abd Ghani MK, Arunkumar N, Mostafa SA, Abdullah MK, Burhanuddin MA. Trainable model for segmenting and identifying Nasopharyngeal carcinoma. *Comput Electr Eng.* 2018;71:372–387. doi:10.1016/j.compeleceng.2018.07.044
- Jing B, Zhang T, Wang Z, et al. A deep survival analysis method based on ranking. *Artif Intell Med.* 2019;98:1–9. doi:10.1016/j.artmed.2019.06.001
- Ma Z, Zhou S, Wu X, et al. Nasopharyngeal carcinoma segmentation based on enhanced convolutional neural networks using multi-modal metric learning. *Phys Med Biol.* 2019;64(2):025005. doi:10.1088/1361-6560/aaf5da
- Peng H, Dong D, Fang M-J, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res.* 2019;25(14):4271–4279. doi:10.1158/1078-0432.CCR-18-3065
- Rehioui H, Idrissi A. On the use of clustering algorithms in medical domain. *Int J Artif Intell.* 2019;17:236.

21. Zou M, Hu J, Zhang H, et al. Rigid medical image registration using learning-based interest points and features. *Comput Mater Continua*. 2019;60(2):511–525. doi:10.32604/cmc.2019.05912
22. Chen H, Qi Y, Yin Y, et al. MMFNet: a multi-modality MRI fusion network for segmentation of nasopharyngeal carcinoma. *Neurocomputing*. 2020;394:27–40. doi:10.1016/j.neucom.2020.02.002
23. Chuang W-Y, Chang S-H, Yu W-H, et al. Successful identification of nasopharyngeal carcinoma in nasopharyngeal biopsies using deep learning. *Cancers (Basel)*. 2020;12(2):507. doi:10.3390/cancers12020507
24. Guo F, Shi C, Li X, Wu X, Zhou J, Lv J. Image segmentation of nasopharyngeal carcinoma using 3D CNN with long-range skip connection and multi-scale feature pyramid. *Soft Comput*. 2020;24(16):12671–12680. doi:10.1007/s00500-020-04708-y
25. Jing B, Deng Y, Zhang T, et al. Deep learning for risk prediction in patients with nasopharyngeal carcinoma using multi-parametric MRIs. *Comput Methods Programs Biomed*. 2020;197:105684. doi:10.1016/j.cmpb.2020.105684
26. Mohammed MA, Abd Ghani MK, Arunkumar N, et al. Decision support system for nasopharyngeal carcinoma discrimination from endoscopic images using artificial neural network. *J Supercomput*. 2020;76(2):1086–1104. doi:10.1007/s11227-018-2587-z
27. Wang J, Liu R, Zhao Y, et al. A predictive model of radiation-related fibrosis based on the radiomic features of magnetic resonance imaging and computed tomography. *Transl Cancer Res*. 2020;9(8):4726–4738. doi:10.21037/tcr-20-751
28. Yang Q, Guo Y, Ou X, Wang J, Hu C. Automatic T staging using weakly supervised deep learning for nasopharyngeal carcinoma on MR images. *Journal of Magnetic Resonance Imaging*. 2020;52(4):1074–1082. doi:10.1002/jmri.27202
29. Zhong L-Z, Fang X-L, Dong D, et al. A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0. *Radiother Oncol*. 2020;151:1–9. doi:10.1016/j.radonc.2020.06.050
30. Bai X, Hu Y, Gong G, Yin Y, Xia Y. A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography. *Biomed Signal Process*. 2021;64:102246. doi:10.1016/j.bspc.2020.102246
31. Cai M, Wang J, Yang Q, et al. Combining images and t-staging information to improve the automatic segmentation of nasopharyngeal carcinoma tumors in MR images. *IEEE Access*. 2021;9:21323–21331. doi:10.1109/ACCESS.2021.3056130
32. Tang P, Zu C, Hong M, et al. DA-DSU-net: dual attention-based dense SU-net for automatic head-and-neck tumor segmentation in MRI images. *Neurocomputing*. 2021;435:103–113. doi:10.1016/j.neucom.2020.12.085
33. Zhang L, Wu X, Liu J, et al. MRI-based deep-learning model for distant metastasis-free survival in locoregionally advanced Nasopharyngeal carcinoma. *J Magn Reson Imaging*. 2021;53(1):167–178. doi:10.1002/jmri.27308
34. Wu X, Dong D, Zhang L, et al. Exploring the predictive value of additional peritumoral regions based on deep learning and radiomics: a multicenter study. *Med Phys*. 2021;48(5):2374–2385. doi:10.1002/mp.14767
35. Zhao L, Gong J, Xi Y, et al. MRI-based radiomics nomogram may predict the response to induction chemotherapy and survival in locally advanced nasopharyngeal carcinoma. *Eur Radiol*. 2020;30(1):537–546. doi:10.1007/s00330-019-06211-x
36. Zhang F, Zhong L-Z, Zhao X, et al. A deep-learning-based prognostic nomogram integrating microscopic digital pathology and macroscopic magnetic resonance images in nasopharyngeal carcinoma: a multi-cohort study. *Ther Adv Med Oncol*. 2020;12:175835920971416. doi:10.1177/1758835920971416
37. Xie C, Du R, Ho JWK, et al. Effect of machine learning re-sampling techniques for imbalanced datasets in 18F-FDG PET-based radiomics model on prognostication performance in cohorts of head and neck cancer patients. *Eur J Nucl Med Mol Imaging*. 2020;47(12):2826–2835. doi:10.1007/s00259-020-04756-4
38. Liu K, Xia W, Qiang M, et al. Deep learning pathological microscopic features in endemic nasopharyngeal cancer: prognostic value and protentional role for individual induction chemotherapy. *Cancer Med*. 2020;9(4):1298–1306. doi:10.1002/cam4.2802
39. Cui C, Wang S, Zhou J, et al. Machine learning analysis of image data based on detailed MR image reports for nasopharyngeal carcinoma prognosis. *Biomed Res Int*. 2020;2020:8068913. doi:10.1155/2020/8068913
40. Du R, Lee VH, Yuan H, et al. Radiomics model to predict early progression of nonmetastatic nasopharyngeal carcinoma after intensity modulation radiation therapy: a multicenter study. *Radiology*. 2019;1(4):e180075. doi:10.1148/ryai.2019180075
41. Zhang B, Tian J, Dong D, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin Cancer Res*. 2017;23(15):4259–4269. doi:10.1158/1078-0432.CCR-16-2910
42. Zhang B, He X, Ouyang F, et al. Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Lett*. 2017;403:21–27. doi:10.1016/j.canlet.2017.06.004
43. Liu J, Mao Y, Li Z, et al. Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma. *J Magn Reson Imaging*. 2016;44(2):445–455.
44. Zhu W, Kan X, Calogero RA. Neural network cascade optimizes MicroRNA biomarker selection for nasopharyngeal cancer prognosis. *PLoS One*. 2014;9(10):e110537. doi:10.1371/journal.pone.0110537
45. Wong LM, Ai QYH, Mo FKF, Poon DMC, King AD. Convolutional neural network in nasopharyngeal carcinoma: how good is automatic delineation for primary tumor on a non-contrast-enhanced fat-suppressed T2-weighted MRI? *Jpn J Radiol*. 2021;39(6):571–579. doi:10.1007/s11604-021-01092-x
46. Xue X, Qin N, Hao X, et al. Sequential and iterative auto-segmentation of high-risk clinical target volume for radiotherapy of nasopharyngeal carcinoma in planning CT images. *Front Oncol*. 2020;10:1134. doi:10.3389/fonc.2020.01134
47. Men K, Chen X, Zhu J, et al. Continual improvement of nasopharyngeal carcinoma segmentation with less labeling effort. *Phys Med*. 2020;80:347–351. doi:10.1016/j.ejmp.2020.11.005
48. Wang X, Yang G, Zhang Y, et al. Automated delineation of nasopharynx gross tumor volume for nasopharyngeal carcinoma by plain CT combining contrast-enhanced CT using deep learning. *J Radiat Res Appl Sci*. 2020;13(1):568–577. doi:10.1080/16878507.2020.1795565
49. Ke L, Deng Y, Xia W, et al. Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. *Oral Oncol*. 2020;110:104862. doi:10.1016/j.oraloncology.2020.104862
50. Zhong T, Huang X, Tang F, Liang S, Deng X, Zhang Y. Boosting-based cascaded convolutional neural networks for the segmentation of CT organs-at-risk in nasopharyngeal carcinoma. *Med Phys*. 2019;46(12):5602–5611. doi:10.1002/mp.13825
51. Lin L, Dou Q, Jin Y-M, et al. Deep learning for automated contouring of primary tumor volumes by MRI for nasopharyngeal carcinoma. *Radiology*. 2019;291(3):677–686. doi:10.1148/radiol.2019182012
52. Liang S, Tang F, Huang X, et al. Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning. *Eur Radiol*. 2019;29(4):1961–1967. doi:10.1007/s00330-018-5748-9
53. Li S, Xiao J, He L, Peng X, Yuan X. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods. *Technol Cancer Res Treat*. 2019;18:1533033819884561. doi:10.1177/1533033819884561
54. Mohammed MA, Abd Ghani MK, Hamed RI, Ibrahim DA, Abdullah MK. Artificial neural networks for automatic segmentation and identification of nasopharyngeal carcinoma. *J Comput Sci*. 2017;21:263–274.

55. Men K, Chen X, Zhang Y, et al. Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. *Front Oncol.* 2017;7:315. doi:10.3389/fonc.2017.00315
56. Wong LM, King AD, Ai QYH, et al. Convolutional neural network for discriminating nasopharyngeal carcinoma and benign hyperplasia on MRI. *Eur Radiol.* 2021;31(6):3856–3863. doi:10.1007/s00330-020-07451-y
57. Wen D-W, Lin L, Mao Y-P, et al. Normal tissue complication probability (NTCP) models for predicting temporal lobe injury after intensity-modulated radiotherapy in nasopharyngeal carcinoma: a large registry-based retrospective study from China. *Radiother Oncol.* 2021;157:99–105. doi:10.1016/j.radonc.2021.01.008
58. Zhang B, Lian Z, Zhong L, et al. Machine-learning based MRI radiomics models for early detection of radiation-induced brain injury in nasopharyngeal carcinoma. *BMC Cancer.* 2020;20(1):502. doi:10.1186/s12885-020-06957-4
59. Du D, Feng H, Lv W, et al. Machine learning methods for optimal radiomics-based differentiation between recurrence and inflammation: application to nasopharyngeal carcinoma post-therapy PET/CT images. *Mol Imaging Biol.* 2020;22(3):730–738. doi:10.1007/s11307-019-01411-9
60. Diao S, Hou J, Yu H, et al. Computer-aided pathologic diagnosis of nasopharyngeal carcinoma based on deep learning. *Am J Pathol.* 2020;190(8):1691–1700. doi:10.1016/j.ajpath.2020.04.008
61. Abd Ghani MK, Mohammed MA, Arunkumar N, et al. Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques. *Neural Comput Appl.* 2020;32(3):625–638. doi:10.1007/s00521-018-3882-6
62. Mohammed MA, Abd Ghani MK, Arunkumar N, Hamed RI, Abdullah MK, Burhanuddin MA. A real time computer aided object detection of nasopharyngeal carcinoma using genetic algorithm and artificial neural network based on Haar feature fear. *Future Gener Comput Syst.* 2018;89:539–547. doi:10.1016/j.future.2018.07.022
63. Wang Y-W, Wu C-S, Zhang G-Y, et al. Can parameters other than minimal axial diameter in MRI and PET/CT further improve diagnostic accuracy for equivocal retropharyngeal lymph nodes in nasopharyngeal carcinoma? *PLoS One.* 2016;11(10):e0163741–e0163741. doi:10.1371/journal.pone.0163741
64. Bai P, Weng X, Quan K, et al. A knowledge-based intensity-modulated radiation therapy treatment planning technique for locally advanced nasopharyngeal carcinoma radiotherapy. *Radiat Oncol.* 2020;15(1):188. doi:10.1186/s13014-020-01626-z
65. Liu Z, Fan J, Li M, et al. A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy. *Med Phys.* 2019;46(5):1972–1983. doi:10.1002/mp.13490
66. Jiao S-X, Chen L-X, Zhu J-H, Wang M-L, Liu X-W. Prediction of dose-volume histograms in nasopharyngeal cancer IMRT using geometric and dosimetric information. *Phys Med Biol.* 2019;64(23):23NT04. doi:10.1088/1361-6560/ab50eb
67. Yang X, Li X, Zhang X, Song F, Huang S, Xia Y. Segmentation of organs at risk in nasopharyngeal cancer for radiotherapy using a self-adaptive Unet network. *Nan Fang Yi Ke Da Xue Xue Bao.* 2020;40(11):1579–1586. doi:10.12122/j.issn.1673-4254.2020.11.07
68. Chen X, Yang J, Yi J, Dai J. Quality control of VMAT planning using artificial neural network models for nasopharyngeal carcinoma. *Chin J Radiol Med Prot.* 2020;40(2):99–105.
69. Xue X, Hao X, Shi J, Ding Y, Wei W, An H. Auto-segmentation of high-risk primary tumor gross target volume for the radiotherapy of nasopharyngeal carcinoma. *J Image Graph.* 2020;25(10):2151–2158.
70. Rizwan I, Haque I, Neubert J. Deep learning approaches to biomedical image segmentation. *Inf Med Unlocked.* 2020;18:100297. doi:10.1016/j.imu.2020.100297
71. Leclerc B, Cale J. *Big Data*. Milton, UK: Taylor & Francis Group; 2020.
72. Parisi GI, Kemker R, Part JL, Kanan C, Wermter S. Continual lifelong learning with neural networks: a review. *Neural Netw.* 2019;113:54–71. doi:10.1016/j.neunet.2019.01.012

Cancer Management and Research

Dovepress

Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient.

The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/cancer-management-and-research-journal>