**Maryam Zamanian** , **Iraj Abedi**

Department of Medical Physics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

# Convolutional neural networks in auto-segmentation of nasopharyngeal carcinoma tumor — a systematic review and meta-analysis

**Address for correspondence:**

Dr. Iraj Abedi

Department of Medical Physics,

School of Medicine, Isfahan University

of Medical Sciences

Hezar Jarib Street, Postal code:

8174673461, Isfahan, Iran

e-mail: I.abedi@med.mui.ac.ir

**ABSTRACT**

**Introduction.** Segmentation is one of the main stages of the treatment planning system (TPS), especially in nasopharyngeal carcinoma (NPC), because it is very heterogeneous and penetrates the skull bone tissue. An automated method to reduce the workload and human error caused by the lack of expertise and perspective would be very helpful. This meta-analysis evaluated the ability of convolutional neural networks (CNNs) to plan auto-segmentation computed tomography (CT) and magnetic resonance imaging (MRI) modalities.

**Material and methods.** Articles published in PubMed, Scholar, and Cochrane databases were examined. The risk of bias was evaluated by the QUADAS-2 tool. The dice similarity coefficient (DSC) as the effect size and standard error (SE) as the precision index were analyzed by random effects. To calculate the degree of heterogeneity and its agent, we used ($I^2$ and $\tau^2$) and meta-regression analysis ($p < 0.05$). A funnel plot was used to check for publication bias.

**Results.** In general, eight studies on CT and 12 on MRI modalities were selected from 3601 studies. The heterogeneity based on ($I^2$ and $\tau^2$) and DSC values (with a 95% confidence interval) for CT and MRI modalities were 88.7% ($\tau^2 = 0.011$), 0.67 (0.62–0.72), and 81.42% ($\tau^2 = 0.01$), 0.76 (0.72–0.80), respectively.

**Conclusions.** CNNs' ability to segment both CT and MRI modalities is at a medium level, and its improvement can make it more suitable for clinical use.

**Keywords**: convolutional neural network, computed tomography, magnetic resonance imaging, nasopharyngeal carcinoma, segmentation

## Introduction

Nasopharyngeal carcinoma (NPC) is the most common type of otolaryngological cancer that grows on the walls of the nasopharyngeal cavity. It has a heterogeneous distribution in different geographical regions, with the highest prevalence observed in Southeast Asia and moderate prevalence in South Asia and North Africa [1]. As the tumor grows and its grade increases to T4, it gradually spreads to the skeletal structure of the skull, even to the intracranial area [2].

The location of the tumor in the head and neck region, surrounding vital organs, and high sensitivity to radiation are reasons for choosing radiotherapy as

the best treatment method [3]. The most important step in the treatment planning system (TPS), which is performed by an experienced radiation oncology specialist, is contouring of the tumor tissue (PTV) and the organs at risk (OARs) before starting the treatment, which can involve two imaging modalities: computed tomography (CT), magnetic resonance imaging (MRI), or both of them [4].

Magnetic resonance imaging can provide the best soft tissue contrast for NPC, and it is a painless and non-invasive method that does not require ionizing radiation, which makes it possible to repeat it to take different sequences (such as T1W, T2W, and T1C). It can also show the shape and location of the lesion well [5]. Because of NPC location, its spread to bone tissues, and the advantage of CT images, including the quality of imaging with better contrast in bony areas and high speed, CT is the best choice. In addition, CT scans take less time than MRI and are cost-effective and available. Centers may use either of these two modalities according to the patient's condition [6, 7].

Image segmentation is a time-consuming person-dependent task that requires the rendering skill of the oncologist; therefore, its correct execution creates a large workload, and the smallest error in segmentation affects the treatment plan [8]. In addition, segmentation of NPC tumors is more difficult due to their greater diversity and heterogeneous intensity compared to other tumors. One of the other challenges and problems of NPC segmentation is its metamorphic form, and each stage of treatment may require re-segmentation. For this reason, an automatic and accurate method to implement segmentation would be of great help [2, 9].

Among alternative methods that have been tested in recent years is the use of artificial intelligence for the automatic and accurate implementation of all TPS parts in various tumors [10]. In recent studies, Convolutional Neural Networks (CNNs) are evaluated rapidly in image auto-segmentation [11–13]. Therefore, in this study, we decided to comprehensively analyze the available literature on CNN ability to automatically perform NPC tumor segmentation in CT and MRI modalities.

## Material and methods

We launched a comprehensive and systematic search of reliable sources to learn whether CNNs have sufficient ability to perform accurate segmentation. The study was registered at the beginning of its conceptualization in PROSPERO, the international open-access Prospective Register of Systematic Reviews (CRD42022379228).

### Search strategy

We searched electronic databases, including MEDLINE (through PubMed) and Cochrane Library. In addition, a Google Scholar search of gray literature and publications in the arXiv database was conducted. There were no limitations regarding study language. Considering that the investigation of CNNs does not have a long history and has been evaluated only in recent years, no time limit was set for the search (in the year 2022). The terms used for the search strategy included ("Nasopharyngeal carcinoma") AND ("Segmentation" OR "U-Net" OR "U-Res-Net" OR "Res-UNet") AND ("Computed tomography" OR "CT" OR "Magnetic resonance imaging" OR "MRI"). PubMed was searched using the restriction of placing the [Title/Abstract] fields in all terms, and no field restriction was placed in Scholar.

After searching, Endnote software was used to collect articles. First, duplicate articles were excluded from the study. The screening of the studies was carried out in three steps: title, abstract, and full text. The search and screening of articles were performed by two researchers. Our assessment overlapped in 95% of cases, and in the remaining cases, we resolved differences of opinion based on the eligibility criteria.

### Study exclusion criteria

All the selected studies investigated the power of all CNNs in relation to the NPC tumor segment, and the examination of OAR segments was excluded from the study. In terms of the investigated indicators, studies that reported the dice similarity coefficient (DSC) index were included. Studies in which the size of the network training samples was under 15 and studies that combined positron emission tomography (PET) images with CT and MRI were excluded from this analysis. All study reviews, case reports, editorials, and letters were excluded from the study.

### Data extraction

The results were classified into two subgroups: CT and MRI modalities. The data extracted from the studies included the name of the first author, country and year of publication, network architecture, sample size and classification for training, external validation and testing, tumor staging, epochs number, learning rate, batch size, type of datasets, network dimension, CT contrast type, MRI sequence, feature extraction software, and processor characterization.

Furthermore, the indices of network performance included the DSC index and Hausdorff distance (HD) extracted from the studies. Meta-analysis results were

reported using the 2020-PRISMA criteria, and the study protocol was written accordingly (Supplementary Tab. S1).

## Quality assessment (risk of bias)

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to evaluate the quality of m eta-analyses and the risk of bias. This tool evaluates the quality of diagnostic studies and includes four key domains: (1) patient selection random sampling, (2) index test (assessment blinded for and independent of reference test), (3) reference standard (valid reference test, assessment independent of index test), and (4) flow and timing (sufficient time between index and reference, all data points included in the analysis). The set of questions for each domain had answers on three scores including "yes-(1) score", "no-(0) score", and "unclear-(0) score". This step was implemented by two persons.

## Statistical analysis

Stata software (version 17.0; College Station, TX 77845, US) was used to perform all statistical calculations. Excel software (Microsoft 2016) was used to extract primary information from the articles and perform some basic calculations. One of the most important indices for evaluating CNN segmentation results is the DSC index, which is used as effect size. The heterogeneity studies were calculated by a random effect model, $I^2$, $\tau^2$, and a level higher than 0.7 ($I^2 > 0$) was considered an indicator of heterogeneity. To predict and investigate the effect of a variable on the obvious change in the results, the regression method was used, and the existence of possible publication bias was evaluated using a funnel plot.

## Results

### Study selection

Among the 3625 studies that were obtained by searching PubMed, Scholar, and Cochrane databases, 20 studies met eligibility criteria. A flow diagram of the study selection process is shown in Figure 1.

### Study characteristics and quality assessment

The reviewed studies on both modalities were conducted in China in the years 2018–2022. Different CNNs included 2D–2.5D-3D UNet [14–26], modified UNet
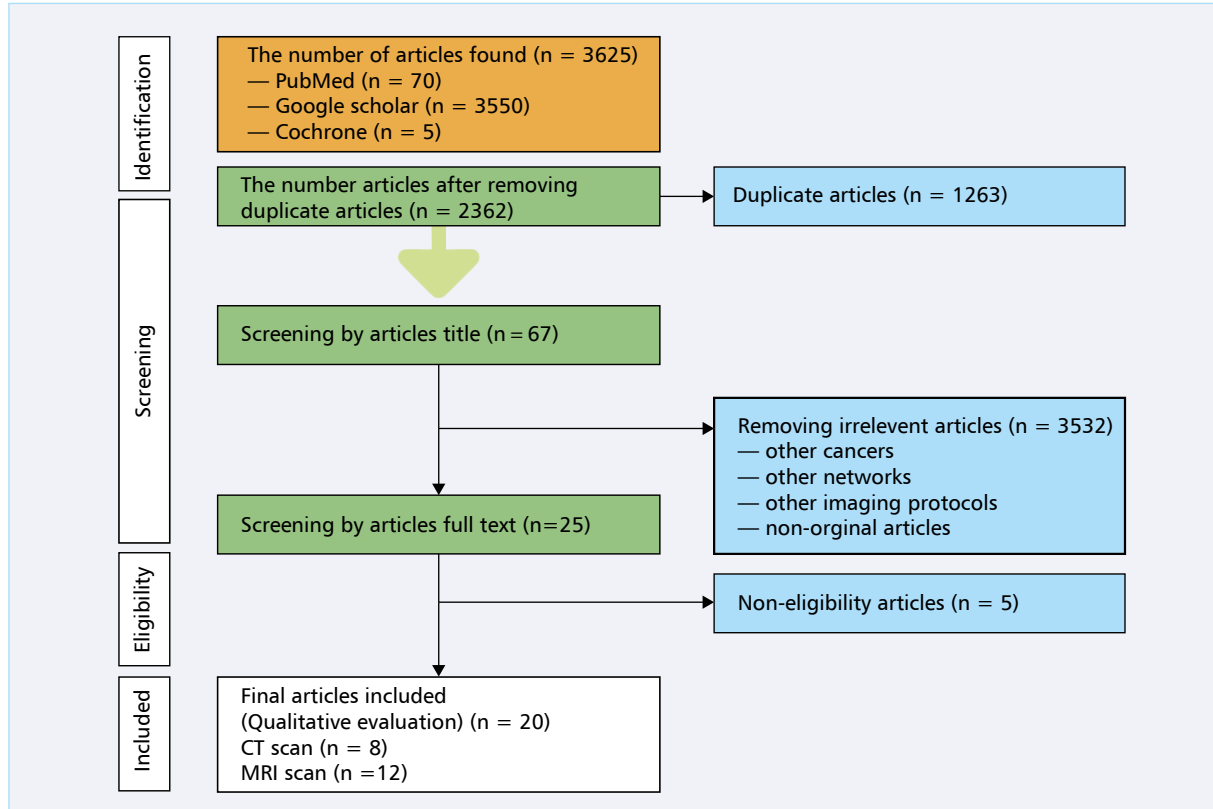


**Figure 1.** PRISMA flow diagram for study selection; CT — computed tomography; MRI — magnetic resonance imagining

**Figure 2.** Quality assessment of the studies. Part (**A**) is related to the quality assessment of computed tomography (CT) studies and part (**B**) is related to magnetic resonance imaging (MRI) studies

[17], 3D Res-UNet [17, 27, 28], modified 3D Res-UNet [14, 17], a mix of 2D and 3D Res-UNet [29, 30], 3D Vnet [15], 3D SI-UNet [18], 3D Nested UNet [14, 19, 31], 3D AttR2-UNet [14, 21, 31], 3D LW-UNet [32], and 3D DE-UNet [33].

Magnetic resonance imagining modality studies used hospital data [21–24, 26, 28, 29, 31–35], and CT studies often used the 2019 MICCAI StructSeg data [14, 15, 17, 19] and hospital data [16, 18, 27, 30]. Two articles were conference papers [22, 30], and one was from the arXiv database [26].

In studies where CT images were analyzed, types of considered images were included without contrast (CT), and with contrast (CE-CT). Also, MRI images were the collection of different sequences of T1-Weighted (T1W), T2-Weighted (T2W), T1-Contrast (T1C), and multi-sequence (MS). In most studies of both modalities, full details of the task were not given; however, the epoch size was between 40 and 600, batch size was 1–8, and the learning rate was 0.01–0.001.

### Result of risk of bias evaluation

The quality of the articles in the CT and MRI modality groups was evaluated using the QUADAS-2 tool, as presented in Figure 2.

### Result of meta-analysis

The descriptive characteristics and some performance results of NPC segmentation studies on CT scan MRI modalities are listed in Tables 1 [15–19, 27, 30, 36] and 2 [21–24, 26, 28, 29, 31–33, 35, 37], respectively.

### NPC CT scan segmentation evaluation

Meta-analysis results of NPC segmentation studies of CT scan modality are presented as a forest plot in Figure 3. The pooled DSC was 0.67 (CI 95%, 0.62 to 0.72; $I^2 = 88.07\%$, $\tau^2 = 0.011$) (p = 0.00) for CT scan segmentation.

**Table 1. Details of the studies on convolutional neural networks segmentation of computed tomography (CT) images**

| The first author (publication year and country) | Sample Size | Training [number] | External. validation | Testing N | Dataset | Tumor staging | Image type | Architecture | Epoch | Batch size | Learning Rate | DSC (mean) | HD (mean) | Processor specifications |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Li et al. [16] (2019; China) | 502 | 302 | 100 | 100 | West China Hospital | T1–T4 | CT | 2D UNet | 40 | NM | 0.01 | 74 | 32.10 | Dual. Intel Xeon E5-2643 v4 (3.4 GHz) and dual. NVIDIA Tesla K40m graphics card |
| Xue et al. [18] (2020; China) | 150 | 120 | 15 | 15 | Hospital. of USTC | T1–T4 | CT | 3D UNet<br>3D SI-UNet | 200 | NM | 0.0001 | 84<br>74 | 9.7<br>8.7 | One Intel Xeon Processor E5-2695 CPU and an NVIDIA Tesla P100 GPU memory |
| Wang et al. [27] (2020; China) | 205 | NM | NM | NM | NM | T1–T4 | CE-CT<br>CT | 3D Res –UNet | NM | 1 | 0.03 | 73 | 4.96 | One NVIDIA GeForce RTX 2080Ti with 11 GB GPU memory |
| Bai et al. [17] (2021; China) | 60 | 50 | No | 10 | StructSeg 2019 Challenge | T1–T4 | CT | 2D UNet<br>2D PUNet<br>3D UNet<br>3D PUNet<br>3D Res -UNet<br>3D PRes-UNet | NM | 8 | 0.0005 | 57.01<br>60.59<br>59.71<br>59.8<br>58.97<br>62.88 | 8.12<br>6.75<br>14.52<br>11.94<br>7.41<br>6.07 | One NVIDIA RTX 2080Ti GPU and 32 GB memory |
| Liu et al. [19] (2021; China) | 140 | 60 | No | 40 | 2019 MICCAI StructSeg+ Sichuan Provincial. Cancer Hospital | T1–T4 | CT | 3D UNet<br>3D Nested-UNet | 300 | 4 | 0.001 | 33.9<br>25.6 | 13.2<br>13.7 | NM |
| Mei et al. [15] (2021; China) | 50 | 40 | No | 10 | 2019 MICCAI StructSeg | T1–T4 | CT | 3D UNet<br>2.5 UNet<br>3D VNet | NM | 16 | 0.0001 | 59.91<br>62.16<br>61.02 | NM | Two NVIDIA GTX 1080 Ti GPU memory |
| Jin et al. [30] (2021; China) | 90 | 63 | 18 | 9 | Sichuan Cancer Hospital. & Institute | T1–T4 | CT | 3D PUNet<br>3D ResSE-UNet | 200 | 8 | 0.0001 | 75<br>79 | 8.59<br>7.64 | NM |
| Yang et al. [36] (2022; China) | 257 | 205 | No | 52 | 2019 MICCAI StructSeg | T1–T4 | CE-CT<br>CT | 3D UNet<br>3D PRes-UNet<br>3D AttR2-UNet<br>3D Nested-UNet | 120 | 2 | 0.01 | 73.67<br>74.49<br>73.54<br>73.87 | 6.32<br>5.06<br>6.74<br>5.17 | One NVIDIA GeForce RTX 2080Ti with 11 GB GPU memory |

DSC — dice similarity coefficient; HD — Hausdorff distance; N — number; NM — not mentioned

**Table 2. Details of the studies on convolutional neural networks segmentation of magnetic resonance imaging (MRI)**

| The first author (publication year and country) | Sample Size | Training [number] | External validation | Testing N | Dataset | Tumor staging | MRI sequence | Architecture | Epoch | Batch size | Learning Rate | DSC (mean) | HD (mean) | Processor specifications |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| He et al. [22] (2018; China) | 19 | 18 | No | 1 | NM | T1–T4 | T1W | 3D UNet | NM | NM | 0.0001 | 74.8 | NM | Ubuntu 14.04 with Tesla K80 at 3.6GHz and 11.18GB GPU memory |
| Wang et al. [23] (2018; China) | 15 | 11 | No | 4 | West China Hospital | T1–T4 | T1W | 3D UNet | NM | NM | NM | 79 | NM | NM |
| Chen et al. [26] (2019; China) | 149 | NM | NM | NM | Shandong Cancer Hospital | T1–T4 | T1W T2W T1C MS | 2D UNet 3D UNet | 100 | 8 | 0.001 | 57.97 64.33 | 84.66 21.02 | NVIDIA Titan Xp GPU with 12GB GPU memory |
| Lin et al. [28] (2019; China) | 1021 | 715 | 103 | 203 | Sun Yat-sen University Cancer Center | T1–T4 | T1W T2W T1C | 3D Res-UNet | NM | NM | NM | 79 | NM | NM |
| Ye et al. [33] (2020; China) | 44 | NM | NM | NM | Panyu Central. Hospital | T1–T4 | T2W T1W | 2D DE-UNet | 200 | 1 | 0.0001 | 66.1 | NM | NVIDIA Geforce GTX 1080 TI with 11 GB GPU memory |
| Guo et al. [24] (2020; China) | 120 | 96 | 14 | 10 | NM | T1–T4 | MRI | 3D UNet | 500 | 1 | 0.0001 | 73.7 | NM | NM |
| Wang et al. [29] (2021; China) | 45 | NM | NM | NM | West China Hospital | T1–T4 | T1W T2W T1C MS | 2D+3D Res-UNet | 50 | 5 | 0.05 | 89.6 | 5.07 | NM |
| Wong et al. [37] (2021; China) | 201 | 136 | No | 65 | NM | T1–T4 | T2W T1W | 2D UNet | 75 | 4 | 0.005 | 71 | NM | NM |
| Cai et al. [21] (2021; China) | 251 | 241 | No | 10 | Shanghai Cancer Center | T1–T4 | T1W T2W T1C | 3D UNet 3D AttR2-UNet | 600 | NM | 0.0001 | 81.1 81.5 | NM | Two NVIDIA Geforce GTX 1080 Ti GPU memory |
| Qi et al. [35] (2021; China) | 130 | NM | NM | NM | Shandong Cancer Hospital | T1–T4 | T1W T2W T1C | 3D UNet | NM | NM | NM | 88.2 | NM | NM |
| Zhang et al. [31] (2022; China) | 93 | 73 | 10 | 10 | NM | T1–T4 | T1W T1C | 2D AttR2-UNet 2D Nested-UNet 2D SE-UNet | 100 | 3 | 0.001 | 73.8 79 78.7 | NM | NM |
| Liu et al. [32] (2022; China) | 92 | 72 | 10 | 10 | NM | T1–T4 | T1W T1C | 2D LW-UNet | NM | 1 | NM | 81.3 | NM | NM |

DSC — dice similarity coefficient; HD — Hausdorff distance; N — number; NM — not mentioned
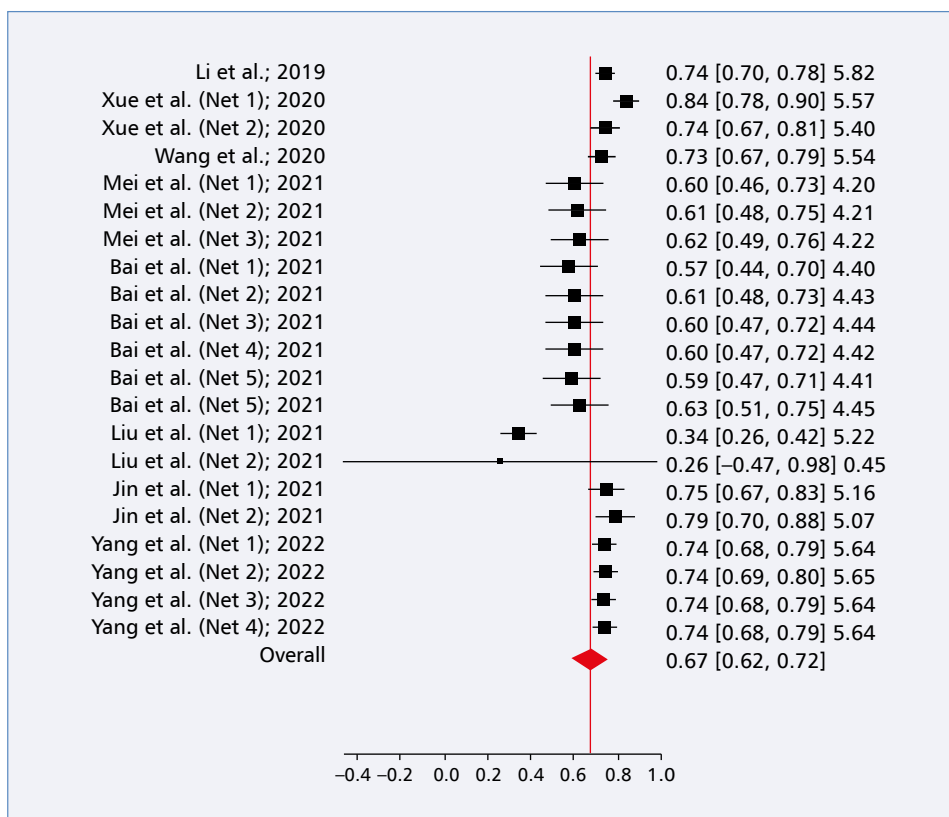
**Figure 3.** Forest plot of computed tomography (CT) modality segmentation studies. The pooled-dice similarity coefficient (DSC) value [calculated with the 95% confidence interval (CI) with range for each study is reported]. Studies are sorted by year, and all Network type values are indicated (Net: Network type).

## NPC MRI scan segmentation

The meta-analyses result of NPC segmentation on the MRI modality showed that the pooled DSC was 0.76 (95% CI 0.72 to 0.80; $I^2$ = 81.42%) (p = 0.01), and its forest plot is presented in Figure 4.

## Subgroup analysis

The type of networks and their dimensions were evaluated in the following subgroups:

— CT scan: based on the number of network types, subgroups were divided into 12 categories. The number of six Network types was reported without a meta-analysis evaluation (including one study). The DSC index for 2.5D UNet, 2D UNet, 3D UNet, 2D P-UNet, 3D P-UNet, 3D AttR2-UNet, 3D Nested UNet, 3D Res-UNet, 3D P-Res-UNet, 3D ResSE-UNet, 3D SI-UNet, and 3D VNet was 0.62 (0.49 to 0.76), 0.67 (95% CI 0.50 to 0.83; $I^2$ = 84.52%), 0.62 (0.95% CI 0.46 to 0.79; $I^2$ = 95.64%), 0.61 (95% CI 0.48 to 0.73), 0.68 (0.95% CI 0.53 to 0.83; $I^2$ = 74.68%), 0.74 (0.68 to 0.79),

0.64 (0.95% CI 0.25 to 1.02; $I^2$ = 41.30%), 0.67 (95% CI 0.53 to 0.81; $I^2$ = 74.63%), 0.70 (95% CI 0.59 to 0.81; $I^2$ = 74.68%), 0.79 (0.70 to 0.88), 0.74 (0.67 to 0.81), and 0.61 (0.48 to 0.75), respectively. Furthermore, the pooled DSC values for Network dimensions including 2D, 2.5D, and 3D were 0.65 (95% CI 0.54 to 0.76; $I^2$ = 75.96%), 0.62 (0.49 to 0.76), and 0.68 (0.95% CI 0.62 to 0.74; $I^2$ = 89.20%), respectively;

— MRI scan: in this modality, network types were divided into ten categories (10 network types) which nine categories were reported without a meta-analysis evaluation (including one study). The DSC index on 2D UNet, 3D UNet, 2D AttR2-UNet, 3D AttR2-UNet, 2D Nested-UNet, 2D SE-UNet, 2D+3D Res-UNet, 3D Res-UNet, 3D DE-UNet, and 3D LW-UNet was 0.64 (0.57 to 0.72), 0.76 (95% CI 0.68 to 0.84; $I^2$ = 87.20%), 0.78 (0.70 to 0.87), 0.81 (0.77 to 0.86), 0.79 (0.71 to 0.87), 0.79 (0.70 to 0.87), 0.79 (0.67 to 0.91), 0.79 (0.77 to 0.81), 0.66 (0.52 to 0.80), and 0.81 (0.73 to 0.89), respectively.

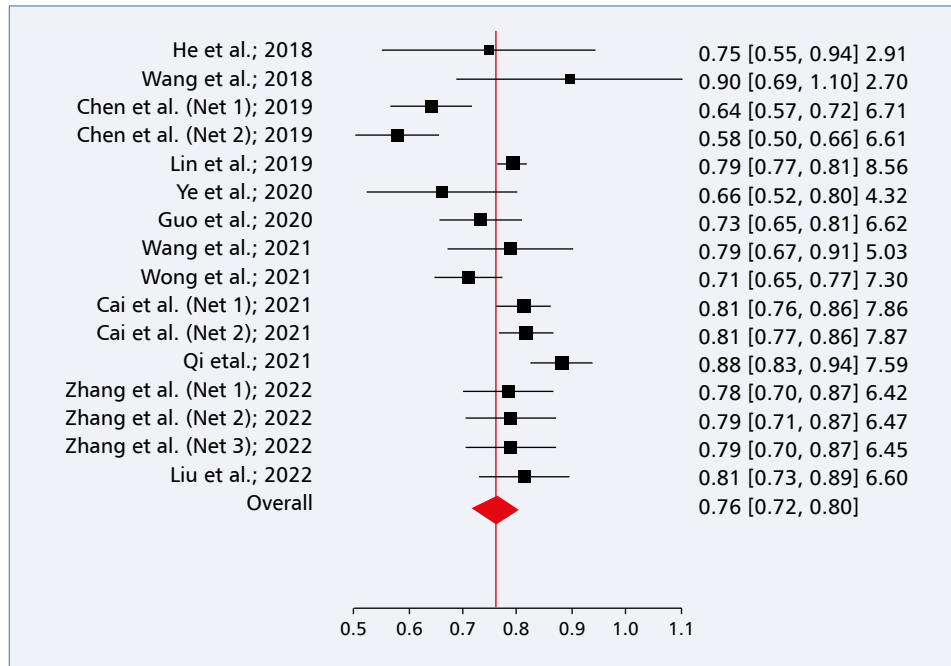The pooled DSC analysis for the subgroups of Network Dimensions including 2D, 2D + 3D, and

**Figure 4.** Forest plot of magnetic resonance imagining modality segmentation studies. The pooled dice similarity coefficient (DSC) value (calculated with a 95% confidence interval) with range for each study is reported. Studies are sorted by year, and all Network type values are indicated (Net: Network type)
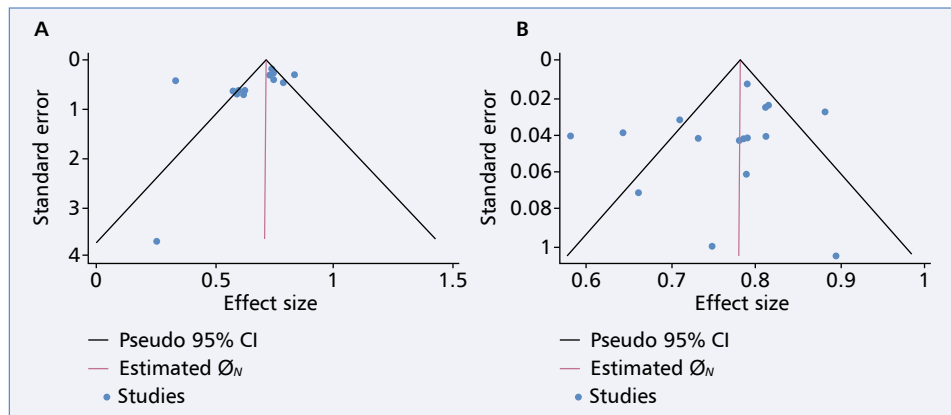


**Figure 5.** Funnel plot on computed tomography (**A**) and magnetic resonance imagining (**B**) modalities for evaluation of publication bias. The dice similarity coefficient (DSC) index was calculated as the effect size [95% confidence interval (CI)]

3D achieved 0.75 (95% CI 0.68 to 0.82; $I^2 = 67.92\%$), 0.79 (0.67 to 0.91) and 0.77 (95% CI 0.71 to 0.82; $I^2 = 86.87\%$), respectively.

### Evaluation of possible causes of heterogeneity

In regression evaluation, coefficients of variables caused heterogeneity for CT studies based on the training number, external validation, and epoch number, which were (0.00073, p = 0.014), (–0.13648, p = 0.008), (–0.00109, p = 0.041), respectively, and for MRI studies based on batch size (–0.02199, p = 0.010).

### Publication bias

We used a funnel plot to evaluate the publication bias in the studies that evaluated CNNs in image segmentation of both CT and MRI modalities (Fig. 5).

## Discussion

The automatic system for the segmentation of heterogeneous NPC tumors is very valuable because it reduces the workload and speeds up diagnosis and treatment. It is necessary to know how successful deep learning networks have been so far, thus the results of this study will be very helpful in decision-making. The DSC index value was selected as the effect size parameter, and a meta-analysis was performed along with SE.

Convolutional neural networks as a subgroup of deep learning were initially tested as 2D in 2018 for CT scans and then in 2019 for MRI. After introducing innovative 3D networks, more studies have been devoted to these networks (Tab. 1, 2). However, 3D networks require a higher volume of calculations and more complex hardware for processing [39]. Recently, the expansion of network layers to improve network performance has been considered. AttR2-UNet and Nested-UNet are examples of such networks [40, 41].

Overall, considering the classification of the DSC index into three levels: good ($0.8 \leq DSC \leq 1$), medium ($0.6 \leq DSC < 0.8$), and poor ($0 \leq DSC < 0.6$)] [38], both MRI image (0.76) and CT image (0.67) segmentation networks achieved medium results, while MRI studies obtained better results than NPC CT image segmentation studies. However, due to the different characteristics of the networks and the heterogeneous distribution of the studies in the two categories, it is not possible to draw definitive conclusions in this regard. The included studies were performed in the past five years, which indicates that this field is very new and will gain more success with further investigations.

In addition, the pooled DSC of both CT and MRI modalities for different dimensions of networks (2D–2.5D–3D), reported almost similar values (~0.02 difference). In detail, the highest value of the DSC index for CT and MRI modalities was observed in 3D-ResSe-UNet (0.79), AttR2-UNet, and LW-UNet (0.81), respectively.

The limitation of analysis based on the results of the evaluated networks was the difference in details and performance of networks, such as the used loss function and epoch number even in similar networks. In addition, there was heterogeneity regarding the training of the networks using CT (with and without contrast) and MRI in different sequences. Due to the dependence of deep learning on the dataset, the heterogeneous distribution of patients, and the small number of patients in some geographical areas, may have affected the results of studies. Almost half of the CT scan segmentation studies used the same dataset presented in the 2019 MICCAI StructSeg, which reduces the impact of data type on the results and makes their comparison more valid. What is characteristic of these studies is that external validation was not performed in more than half of both modalities. Overall, we were able to reduce the heterogeneity analysis of the dimensions and type of the network subgroups.

Notably, all eligible studies were conducted in China, and on the other hand, the highest prevalence of NPC cancer was reported in China (~80%) [42]. Probably, the number of appropriate datasets, compared to other countries, and prioritizing this cancer in research facilitated the implementation of studies.

Since it is not easy to determine the margin of small tumors in magnetic resonance (MR) images, it may affect the ability of the network [43]. Therefore, more empowerment of networks to segmentation of MR images should be given more attention in future studies. Due to the impossibility of using contrast agents for patients with renal failure and the possibility of long-term complications [44], using images with contrast is likely to be used less in the future, thus it is better to enable networks to use non-contrast images.

## Conclusions

The medium capability level of CNNs was observed in both CT and MRI modalities, while this capability was better in MRI segmentation. By improving CNNs, their clinical application can be made more practical.

## Article Information and Declarations

### Ethics statement

This article does not involve any studies with human participants or animals performed by any of the authors.

### Author contributions

I.A. and M.Z. were equally involved in the design, literature review, and analysis of the study.

### Conflict of interest

The authors declare that they have no conflict of interest.

### Supplementary material

Supplementary Table S1.

# References

1. Chang ET, Ye W, Zeng YX, et al. The Evolving Epidemiology of Nasopharyngeal Carcinoma. Cancer Epidemiol Biomarkers Prev. 2021; 30(6): 1035–1047, doi: 10.1158/1055-9965.EPI-20-1702, indexed in Pubmed: 33849968.

2. Guo R, Mao YP, Tang LL, et al. The evolution of nasopharyngeal carcinoma staging. Br J Radiol. 2019; 92(1102): 20190244, doi: 10.1259/bjr.20190244, indexed in Pubmed: 31298937.

3. Blanchard P, Biau J, Huguet F, et al. Radiotherapy for nasopharyngeal cancer. Cancer Radiother. 2022; 26(1-2): 168–173, doi: 10.1016/j.canrad.2021.08.009, indexed in Pubmed: 34953699.

4. Minniti G, Goldsmith C, Brada M. Radiotherapy. Handb Clin Neurol. 2012; 104: 215–228, doi: 10.1016/B978-0-444-52138-5.00016-5, indexed in Pubmed: 22230446.

5. King AD. MR Imaging of Nasopharyngeal Carcinoma. Magn Reson Imaging Clin N Am. 2022; 30(1): 19–33, doi: 10.1016/j.mric.2021.06.015, indexed in Pubmed: 34802578.

6. Choopani MR, Abedi I, Dalvand F. Quality Assessment of Computed Tomography Images using a Channelized Hoteling Observer: Optimization of Protocols in Clinical Practice. Adv Biomed Res. 2023; 12: 8, doi: 10.4103/abr.abr_353_21, indexed in Pubmed: 36926443.

7. Patel PR, De Jesus O. CT Scan. In: De Je. ed. StatPearls. StatPearls Publishing LLC, Treasure Island (FL) 2022.

8. Schaue D, McBride WH. Opportunities and challenges of radiotherapy for treating cancer. Nat Rev Clin Oncol. 2015; 12(9): 527–540, doi: 10.1038/nrclinonc.2015.120, indexed in Pubmed: 26122185.

9. Claude L, Jouglar E, Duverge L, et al. Update in pediatric nasopharyngeal undifferentiated carcinoma. Br J Radiol. 2019; 92(1102): 20190107, doi: 10.1259/bjr.20190107, indexed in Pubmed: 31322911.

10. Wang C, Zhu X, Hong JC, et al. Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future. Technol Cancer Res Treat. 2019; 18: 1533033819873922, doi: 10.1177/1533033819873922, indexed in Pubmed: 31495281.

11. Shen G, Jin X, Sun C, et al. Artificial Intelligence Radiotherapy Planning: Automatic Segmentation of Human Organs in CT Images Based on a Modified Convolutional Neural Network. Front Public Health. 2022; 10: 813135, doi: 10.3389/fpubh.2022.813135, indexed in Pubmed: 35493368.

12. Liu Z, Liu F, Chen W, et al. Automatic Segmentation of Clinical Target Volume and Organs-at-Risk for Breast Conservative Radiotherapy Using a Convolutional Neural Network. Cancer Manag Res. 2021; 13: 8209–8217, doi: 10.2147/CMAR.S330249, indexed in Pubmed: 34754241.

13. 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, indexed in Pubmed: 30302589.

14. Yang G, Dai Z, Zhang Y, et al. Multiscale Local Enhancement Deep Convolutional Networks for the Automated 3D Segmentation of Gross Tumor Volumes in Nasopharyngeal Carcinoma: A Multi-Institutional Dataset Study. Front Oncol. 2022; 12: 827991, doi: 10.3389/fonc.2022.827991, indexed in Pubmed: 35387126.

15. Mei H, Lei W, Gu R, et al. Automatic segmentation of gross target volume of nasopharynx cancer using ensemble of multiscale deep neural networks with spatial attention. Neurocomputing. 2021; 438: 211–222, doi: 10.1016/j.neucom.2020.06.146.

16. Li S, Xiao J, He L, et al. 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, indexed in Pubmed: 31736433.

17. Bai X, Hu Y, Gong G, et al. A deep learning approach to segmentation of nasopharyngeal carcinoma using computed tomography. Biomedical Signal Processing and Control. 2021; 64: 102246, doi: 10.1016/j.bspc.2020.102246.

18. 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, indexed in Pubmed: 32793483.

19. Liu Y, Yuan X, Jiang X, et al. Dilated Adversarial U-Net Network for automatic gross tumor volume segmentation of nasopharyngeal carcinoma. Applied Soft Computing. 2021; 111: 107722, doi: 10.1016/j.asoc.2021.107722.

20. Wong L, Ai Qy, Mo F, et al. Non contrast-enhanced imaging as a replacement for contrast-enhanced imaging for MRI automatic delineation of nasopharyngeal carcinoma. medRxiv. 2020, doi: 10.1101/2020.07.09.20148817.

21. 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.

22. He Yu, Yu Xi, Liu C, et al. A 3D Dual Path U-Net of Cancer Segmentation Based on MRI. 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). 2018, doi: 10.1109/icivc.2018.8492781.

23. Wang Y, Zu C, Hu G, et al. Automatic Tumor Segmentation with Deep Convolutional Neural Networks for Radiotherapy Applications. Neural Processing Letters. 2018; 48(3): 1323–1334, doi: 10.1007/s11063-017-9759-3.

24. Guo F, Shi C, Li X, et al. Image segmentation of nasopharyngeal carcinoma using 3D CNN with long-range skip connection and multi-scale feature pyramid. Soft Computing. 2020; 24(16): 12671–12680, doi: 10.1007/s00500-020-04708-y.

25. Qi Y, Yin Y, Li T, et al. A Computer Aided System for Nasopharyngeal Carcinoma Segmentation and Visualization Based on CT Images. 2018 2nd International Conference on Robotics and Automation Sciences (ICRAS). 2018, doi: 10.1109/icras.2018.8443238.

26. 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.

27. 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. Journal of Radiation Research and Applied Sciences. 2020; 13(1): 568–577, doi: 10.1080/16878507.2020.1795565.

28. Lin Li, Dou Qi, Jin YM, 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, indexed in Pubmed: 30912722.

29. Wang D, Gong Z, Zhang Y, et al. Convolutional Neural Network Intelligent Segmentation Algorithm-Based Magnetic Resonance Imaging in Diagnosis of Nasopharyngeal Carcinoma Foci. Contrast Media Mol Imaging. 2021; 2021: 2033806, doi: 10.1155/2021/2033806, indexed in Pubmed: 34456649.

30. Jin Z, Li X, Shen L, et al. Automatic Primary Gross Tumor Volume Segmentation for Nasopharyngeal Carcinoma using ResSE-UNet. 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). 2020, doi: 10.1109/cbms49503.2020.00116.

31. Zhang J, Gu L, Han G, et al. AttR2U-Net: A Fully Automated Model for MRI Nasopharyngeal Carcinoma Segmentation Based on Spatial Attention and Residual Recurrent Convolution. Front Oncol. 2021; 11: 816672, doi: 10.3389/fonc.2021.816672, indexed in Pubmed: 35155206.

32. Liu Yi, Han G, Liu X. Lightweight Compound Scaling Network for Nasopharyngeal Carcinoma Segmentation from MR Images. Sensors (Basel). 2022; 22(15), doi: 10.3390/s22155875, indexed in Pubmed: 35957432.

33. Ye Y, Cai Z, Huang B, et al. Fully-Automated Segmentation of Nasopharyngeal Carcinoma on Dual-Sequence MRI Using Convolutional Neural Networks. Front Oncol. 2020; 10: 166, doi: 10.3389/fonc.2020.00166, indexed in Pubmed: 32154168.

34. Wong LM, Ai QiY, Mo FKF, et al. 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, indexed in Pubmed: 33544302.

35. Qi Y, Li J, Chen H, et al. Computer-aided diagnosis and regional segmentation of nasopharyngeal carcinoma based on multi-modality medical images. Int J Comput Assist Radiol Surg. 2021; 16(6): 871–882, doi: 10.1007/s11548-021-02351-y, indexed in Pubmed: 33782844.

36. Yang B, Chen X, Li J, et al. A feasible method to evaluate deformable image registration with deep learning-based segmentation. Phys Med. 2022; 95: 50–56, doi: 10.1016/j.ejmp.2022.01.006, indexed in Pubmed: 35091332.

37. Wong ML. Applications of Deep Learning in MRI of Nasopharyngeal Carcinoma. The Chinese University of Hong Kong, Hong Kong 2021.

38. Velker VM, Rodrigues GB, Dinniwell R, et al. Creation of RTOG compliant patient CT-atlases for automated atlas based contouring of local regional breast and high-risk prostate cancers. Radiat Oncol. 2013; 8: 188, doi: 10.1186/1748-717X-8-188, indexed in Pubmed: 23885662.

39. Nguyen D, Long T, Jia X, et al. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from

patient anatomy using deep learning. Sci Rep. 2019; 9(1): 1076, doi: 10.1038/s41598-018-37741-x, indexed in Pubmed: 30705354.

40. Zhu N, Liu C, Forsyth B, et al. Segmentation with Residual Attention U-Net and an Edge-Enhancement Approach Preserves Cell Shape Features. Annu Int Conf IEEE Eng Med Biol Soc. 2022; 2022: 2115–2118, doi: 10.1109/EMBC48229.2022.9871026, indexed in Pubmed: 36085725.

41. Kundu S, Karale V, Ghorai G, et al. Nested U-Net for Segmentation of Red Lesions in Retinal Fundus Images and Sub-image Classification for Removal of False Positives. J Digit Imaging. 2022; 35(5): 1111–1119, doi: 10.1007/s10278-022-00629-4, indexed in Pubmed: 35474556.

42. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016; 66(2): 115–132, doi: 10.3322/caac.21338, indexed in Pubmed: 26808342.

43. Spicer GJ, Kazim M, Glass LR, et al. Accuracy of MRI in defining tumor-free margin in optic nerve glioma surgery. Ophthalmic Plast Reconstr Surg. 2013; 29(4): 277–280, doi: 10.1097/IOP0b013e318291658e, indexed in Pubmed: 23715516.

44. Pasquini L, Napolitano A, Visconti E, et al. Gadolinium-Based Contrast Agent-Related Toxicities. CNS Drugs. 2018; 32(3): 229–240, doi: 10.1007/s40263-018-0500-1, indexed in Pubmed: 29508245.

## Supplementary material

**Table S1. PRISIMA 2020 checklist (based on: Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71)**

| Section and topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| **TITLE** | | | |
| Title | 1 | Identify the report as a systematic review | Page 1 |
| **ABSTRACT** | | | |
| Abstract | 2 | See the PRISMA 2020 for Abstracts checklist | Page 2, P1 |
| **INTRODUCTION** | | | |
| Rationale | 3 | Describe the rationale for the review in the context of existing knowledge | Page 3, P4,5,6 |
| Objectives | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses | Page 3, P7 |
| **METHODS** | | | |
| Eligibility criteria | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses | Page 4, P4 |
| Information sources | 6 | Specify all databases, registers, websites, organizations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted | Page 4, P2 |
| Search strategy | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used | Page 4, P1 |
| Selection process | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process | N/R |
| Data collection process | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process | Page 4, P1 |
| Data items | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect | Page 4, P6 |
| | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information | Page 4, P6 |
| Study risk of bias assessment | 11 | Specify the methods used to assess the risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process | Page 4, P5 |
| Effect measures | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results | Page 4, P6 |
| Synthesis methods | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5) | Page 5, P1 |
| | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions | Page 5, P1 |
| | 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses | Page 5, P1 |
| | 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used | Page 5, P1 |
| | 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta- regression) | Page 5, P1 |
| | 13f | Describe any sensitivity analyses conducted to assess the robustness of the synthesized results | Page 5, P1 |

→

**Table S1 cont. PRISIMA 2020 checklist (based on: Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021; 372: n71)**

| Section and Topic | Item # | Checklist item | Location where item is reported |
|---|---|---|---|
| Reporting bias assessment | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases) | Page 4, P5 |
| Certainty assessment | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome | Page 5, P1 |
| **RESULTS** | | | |
| Study selection | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram | Page 6, P3 |
| | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded | Page 6, P4, Fig 1 |
| Study characteristics | 17 | Cite each included study and present its characteristics | Page 6, P2-6 |
| Risk of bias in studies | 18 | Present assessments of risk of bias for each included study | Page 6, P1, Fig 2 |
| Results of individual studies | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots | Page 10-13 |
| Results of syntheses | 20a | For each synthesis, briefly summarize the characteristics and risk of bias among contributing studies | Page 6, P1 |
| | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect | Page 10-13 |
| | 20c | Present results of all investigations of possible causes of heterogeneity among study results | Page 10-13 |
| | 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results | Page 10-13 |
| Reporting biases | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed | Page 6, 14 |
| Certainty of evidence | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed | Page 10-13 |
| **DISCUSSION** | | | |
| Discussion | 23a | Provide a general interpretation of the results in the context of other evidence | Page 15, P6 |
| | 23b | Discuss any limitations of the evidence included in the review | Page 16, P5 |
| | 23c | Discuss any limitations of the review processes used | Page 16, P5 |
| | 23d | Discuss implications of the results for practice, policy, and future research | Page 16, P3 |
| **OTHER INFORMATION** | | | |
| Registration and protocol | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered | Page 3, P8 |
| | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared | Page 3, P8 |
| | 24c | Describe and explain any amendments to information provided at registration or in the protocol | Page 3, P8 |
| Support | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review | Page 16, P6 |
| Competing interests | 26 | Declare any competing interests of review authors | Page 16, P7 |
| Availability of data, code, and other materials | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review | N/R |