

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |                                     |                                                                                                                                                                                                                                                                                     |
|-------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| n/a                                 | Confirmed                                                                                                                                                                                                                                                                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested                                                                                                                                                                                                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated                                                                                                                                                         |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data collection | No Software was used to collect the data for this study.                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Data analysis   | Our code is publicly available at <a href="https://github.com/mic-dkz/nnunet">github.com/mic-dkz/nnunet</a> . The following open source libraries have been utilized as part of our data analysis pipeline (all of them are python (python version 3.8.5) packages to be installed via "pip install <package_name>"): torch=1.6, tqdm=4.48.2, dicom2nifti=2.2.10, scikit-image=0.17.2, medpy=0.4.0, scipy=1.5.2, batchgenerators=0.21, numpy=1.19.1, scikit-learn=0.23.2, SimpleITK=1.2.4, pandas=1.1.1. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All 23 datasets used in this study are publicly available and can be accessed via their respective challenge website: D1-D10 Medical Segmentation Decathlon: <http://medicaldecathlon.com/>; D11 BCV-Abdomen: <https://www.synapse.org/#!Synapse:syn3193805/wiki/>; D12 PROMISE12: <https://promise12.grand-challenge.org/>; D13 ACDC: <https://acdc.creatis.insa-lyon.fr/>; D14 LiTS: <https://competitions.codalab.org/competitions/17094>; D15 MSLes: <https://smart-stats-tools.org/lesion-challenge>; D16 CHAOS: <https://chaos.grand-challenge.org/>; D17 KiTS: <https://kits19.grand-challenge.org/>; D18 SegTHOR: <https://competitions.codalab.org/competitions/21145>; D19 CREMI: <https://cremi.org/>; D20-23 Cell Tracking Challenge: <http://celltrackingchallenge.net/>

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample size     | We evaluate our algorithm on 23 publicly available datasets. We did not alter the sample size of utilized datasets, however, we experimentally confirm the importance of evaluation on a sufficient number of datasets and training cases in order to draw reliable conclusions.                                                                                                                                                                                                         |
| Data exclusions | No data was excluded from the analysis.                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Replication     | All reported results are based on test sets from external challenges. Thus, scientific best practice dictates to perform only one submission (i.e. one replication) to avoid overfitting. During algorithm training we have not observed failed trainings or configurations, i.e. all attempted trainings were successful. The ablation studies in Figure 6 are reported as cross-validation results (1 run) from which 1000 bootstrap subsets were sampled to ensure ranking stability. |
| Randomization   | All experiments are international segmentation competitions with pre-defined random splits between training data and test data.                                                                                                                                                                                                                                                                                                                                                          |
| Blinding        | Experimenters were blinded with respect to the data collection process as well as the data splitting process and did not have access to labels of the corresponding test data.                                                                                                                                                                                                                                                                                                           |

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involved in the study                                |
|-------------------------------------|------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data               |

### Methods

| n/a                                 | Involved in the study                           |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |