

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) 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
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ 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)
☒	☐ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), 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	For whole slide image, the slide-level annotations are provided to us by the cytopathologists (excel table), and the images are captured by our compact microscope. The region-level annotations are supported by Qupath software (0.2.0).
Data analysis	The source code of the paper is available at ShenghuaCheng/high-risk-Cervical-Precancerous-Screening (github.com). We use Origin (2021) and python (3.8) with sklearn (0.23.2) to analysis the data.

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The original LRWSI and annotation data are private and are not publicly available since the protection of patients' privacy in cooperative hospitals. We released demonstration data, which can be checked at ShenghuaCheng/high-risk-Cervical-Precancerous-Screening (github.com). All data supporting the findings of this study

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	All participants included in this study are women, as only women undergo cervical cancer screening.
Reporting on race, ethnicity, or other socially relevant groupings	Owing to the study implementation hospitals for this study being located in China, the majority of the participants included in this research are Chinese, with the ethnicity being predominantly of the Asian subgroup.
Population characteristics	This retrospective study utilized 5,441 fully anonymized cervical slides (age range: 20–65 years) from routine cervical cancer screening and gynecological clinic visits. Consent Waiver: Granted (anonymized retrospective data) Sex: 100% female (cervical cancer cohort) Age Range: 20–65 years
Recruitment	For internal dataset, we first collect 3,899 LRWSIs from four different hospitals at different periods: A1-A5, B1-B2, C1-C2, and D1-D4. A and D are slides from the same hospital which is Tongji Hospital of Huazhong University of Science and Technology, but they differ in slide preparation and staining scheme. Slides made by Wuhan Union Hospital of Huazhong University of Science and Technology is referred as B. C is slides made by Maternal and Child Hospital of Hubei Province. We randomly separated the whole internal dataset to train, validate and test set by 8:1:1. For external dataset, we collect 972 LRWSIs. E1 and F1 are named External-A and External-B in paper respectively. E is from Wuhan Landing Institute for Artificial Intelligence Cancer Diagnosis Industry Development. F is slides from Duodao People's Hospital, Jingmen, Hubei. NewCohorts is composed by C3, D5 and D6 which is new cohorts from the same hospital of internal datasets. It is worth noting that E and F are sourced from enterprises and primary hospitals in resource-limited region. Compared to central hospitals with specialized equipment and experienced pathology experts, these data are more aligned with our application scenario. To demonstrate the performance of our method on HPV-positive samples, we additionally collected a dataset of 570 slides from HPV-positive cases. These slides were obtained using our improved compact microscope. All slides in the dataset were derived from cases with HPV-positive test results. Details of the dataset are presented in Table 1. The HPV-positive dataset is from Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.
Ethics oversight	This study was approved by the Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology (2020-S353). Informed consent was waived for using anonymized, retrospectively collected cervical slides solely for imaging and analysis. As a retrospective investigation utilizing routinely accumulated clinical pathology data, no participant compensation was provided.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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

Life sciences study design

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

Sample size	We totally used 5,441 whole slide images from six hospitals (of these, 3,899 cervical slides served as the internal training/testing dataset, 972 slides comprised the external test dataset, and an additional 570 slides formed the co-testing dataset with HPV assays). The origin and distribution of these samples are shown in Table 2. The samples were based on real-world clinical practice in these hospitals during multiple time periods, including central hospitals and primary hospitals in relatively resource-limited areas.
Data exclusions	We just removed samples from the raw data that were damaged, or the quality of slide was completely inappropriate, such as completely out of focus, or the number of fields of view was far from sufficient.
Replication	For internal data, the data is randomly divided into train, validation and test sets. For external data, the data is only used for testing. We release the test code and model weights of our proposed method. We also submit the raw test results for each sample of all experiments. All results can be repeated.
Randomization	For internal data, the data is randomly divided into train, validation and test sets. For external data, the data is only used for testing.
Blinding	For this retrospective study, as the slides were anonymized, it was not necessary to employ either double-blind or single-blind design.

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
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.