

1 **Automatic Retinoblastoma Screening and Surveillance**

2 **Using Deep Learning**

3 Ruiheng Zhang, MD^{1, #}, Li Dong, MD^{1, #}, Ruyue Li, MSc^{2, #}, Kai Zhang, PhD^{3 #},
4 Yitong Li, MSc^{1, #}, Hongshu Zhao, MD⁴, Jitong Shi, MD⁴, Xin Ge, MD⁴, Xiaolin
5 Xu, MD⁴, Libin Jiang, MD⁴, Xuhan Shi, MD¹, Chuan Zhang, MD¹, Wenda Zhou,
6 MSc¹, Liangyuan Xu, MSc⁴, Haotian Wu, MSc¹, Heyan Li, MSc¹, Chuyao Yu, MSc
7 ¹, Jing Li⁴, Jianmin Ma, MD^{4, *}, Wenbin Wei, MD^{1, *}

8 ¹ Beijing Tongren Eye Center, Beijing Key Laboratory of Intraocular Tumor
9 Diagnosis and Treatment, Beijing Ophthalmology & Visual Sciences Key Lab,
10 Medical Artificial Intelligence Research and Verification Key Laboratory of the
11 Ministry of Industry and Information Technology, Beijing Tongren Hospital, Capital
12 Medical University, Beijing, China.

13 ² Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren
14 Hospital, Capital Medical University, Beijing, China.

15 ³ InferVision Healthcare Science and Technology Limited Company, Shanghai,
16 China.

17 ⁴ Beijing Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Eye
18 Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China.

19 # These authors contributed equally to the study.*Correspondence author: Jianmin Ma
20 (jmma@sina.com), Wenbin Wei (weiwenbintr@163.com).

21 **Keywords:** Retinoblastoma, Deep learning, Screening, Surveillance

22 **ABSTRACT**

23 Retinoblastoma is the most common intraocular malignancy in childhood. With the
24 advanced management strategy, the global salvage and overall survival have
25 significantly improved, which proposes subsequent challenges regarding long-term
26 surveillance and offspring screening. Here, we developed deep learning algorithm,
27 called Deep Learning Assistant for Retinoblastoma (DLA-RB) training on A total of
28 36623 images from 713 patients. We validated it in the prospectively collected dataset,
29 comprised of 1366 images form 139 eyes of 103 patients. In identifying active
30 retinoblastoma tumors (treatment required) from all clinical-suspected patients, the
31 area under the receiver operating characteristic curve (AUC) of DLA-RB reached
32 0.991 (95% CI 0.970-1.000). In identifying active retinoblastoma from stable
33 retinoblastoma patients (treatment is not required), AUC of DLA-RB reached 0.962
34 (95% CI 0.915-1.000), respectively. Cost-utility analysis revealed that DLA-RB based
35 diagnosis mode is more cost-effective in both retinoblastoma diagnosis and
36 retinoblastoma activity surveillance. The DLA-RB achieved high accuracy and
37 sensitivity in identifying active retinoblastoma from the normal and stable
38 retinoblastoma fundus. It can be incorporated within telemedicine programs in the
39 future.
40

41 **Introduction**

42 Retinoblastoma is the most common intraocular malignancy in childhood. It is
43 estimated to affect one case per 16,000–18,000 live births worldwide.¹ China, India,
44 and other populous countries in Asia and Africa have the largest amount of newly
45 diagnosed retinoblastoma patients. With advanced screening techniques and
46 multidisciplinary management, most retinoblastoma patients achieve life-saving,
47 globe salvage, and even useful vision.² In a recent national cohort study in China, the
48 overall survival rates of RB patients were 81%, 83%, and 91% in patients diagnosed
49 in 1989-2008, 2009-2013, 2014-2017, respectively.³ For globe salvage, almost all
50 retinoblastoma classified by the International Intraocular Retinoblastoma
51 Classification (IIRC)⁴ as group A-C can achieve globe salvage through systemic
52 chemotherapy with adjuvant laser therapy and cryotherapy.⁵ Group D retinoblastoma
53 can also achieve 40% (95%CI:31–51%) globe salvage rate through chemotherapy
54 alone⁶ and a significantly higher globe salvage rate when receiving intra-ocular
55 artery chemotherapy.⁷

56 Improved overall survival and successful local control propose subsequent challenges
57 regarding long-term surveillance after tumor control and screening for the offspring
58 and relatives of retinoblastoma survivors. The current expert consensus recommends
59 every 2–6 months for 4 years after tumor control.¹ In real-world practice, group D
60 retinoblastoma received an average of 21 examinations under anesthesia (EUAs)
61 during 5-year follow-up.⁸ Furthermore, offspring of retinoblastoma survivors have 50%

62 chance of inheriting the mutant RB1 from the affected parent, which would result in a
63 97% risk of developing retinoblastoma.⁹ Global preserving approach thus leads to
64 additional disease burden comprised of health care, socioeconomic, and mental
65 aspects. Furthermore, in most developing countries, only experienced medical
66 ophthalmologists at tertiary eye centers can perform EUAs. This unbalanced
67 distribution of health care resources may delay the diagnosis and proper management
68 during the referring process.

69 The application of deep learning (DL) techniques proposes reliable and low-cost
70 methods for screening and diagnosing retinal diseases. Previous studies showed that
71 the DL algorithm based on fundus images achieves close-to-expert performance for
72 the automatic detection of retinal diseases, including diabetic retinopathy¹⁰,
73 age-related macular degeneration (AMD),¹¹ glaucoma,¹² myopic maculopathy,¹³ and
74 papilledema.¹⁴ Recently, researchers have applied DL in detecting plus diseases in
75 retinopathy of prematurity.^{15,16} The DL algorithm exceeds 6 of 8 experts in identifying
76 plus diseases, which indicates it can reach robust performance in images obtained by
77 EUAs.

78 Thus, in the current study, we developed Deep Learning Assistant for Retinoblastoma
79 (DLA-RB) and validated it in the prospectively collected dataset. The DLA-RB aims
80 to assist the fundus surveillance after local control and provide referral advice. The
81 DLA-RB also provides automatic surveillance of the contralateral eye of
82 retinoblastoma patients and the offspring of retinoblastoma survivors.

83

84 **Results**

85 Between March 2018 and January 2022, 47503 images from 713 patients were
86 retrospectively collected from the anonymous data center in Beijing Tongren hospital
87 (**Figure 1**). After manually excluding low-quality images due to non-fundus image,
88 halation, blurs, and defocus (**Supplementary Figure S1**), a total of 36623 images
89 from 713 patients were finally included for DLA-RB development, with 19045
90 (52.0%), 2918 (8.0%), and 14660 (40.0%) images were normal, stable and active
91 retinoblastoma (**Supplementary Table S1**). At patient levels, 713 patients were
92 randomly allocated for DLA-RB development and five-fold cross-validation.
93 Considering that it is more complicated to distinguish between stable and active
94 retinoblastoma, we first trained ResNet-50 and InceptionResnetV3 for this two-class
95 task to compare the performance of these two architectures. ROC indicated slight
96 better performance of ResNet-50 in five-fold cross-validation (0.940 [95%CI
97 0.851-0.996] vs. 0.934 [95%CI 0.735-0.946], **Supplementary Figure S2**). Thus,
98 ResNet-50 was used to establish DLA-RB. To distinguish between the normal fundus
99 and active lesion, DLA-RB achieved an AUC of 0.9982 (95%CI 0.986-1.000) in the
100 development dataset (**Supplementary Figure S3**).

101 From February 2022 to June 2022, 103 patients with clinical-suspected
102 retinoblastoma and treated retinoblastoma patients first visited Beijing Tongren
103 hospitals. Fundus images of these patients were not used in developing DLA-RB. If

104 more than one time of EUA, only the first EUA images were collected. In total, 139
105 eyes of 103 patients were included for prospective validation (**Table 1**). 69 eyes were
106 clinically diagnosed with retinoblastoma. According to international intraocular
107 retinoblastoma classification (IIRC),¹⁷ 3 (4.3%), 7 (10.1%), 8 (11.6%), 40 (58.0%),
108 and 11 (15.9%) eye was classified as A-E stage, respectively. 43.5% of
109 retinoblastoma was endophytic, while 56.5% were exophytic. For each EUA, the
110 highest probability among all images was assigned as an eye-level probability.

111 DLA-RB could accurately identify active retinoblastoma tumors from all
112 clinical-suspected retinoblastoma (**Figure 2A-B, Table 2**). The AUC, sensitivity, and
113 specificity reached 0.991 (95%CI 0.970-1.000), 0.979 (95%CI 0.927-1.000), and
114 1.000 (95%CI 1.000-1.000), respectively. Compared with the human
115 ophthalmologists, the DLA-RB reached significantly higher sensitivity and specificity
116 in identifying new retinoblastoma tumors (**Figure 2A-B, Supplementary Table**
117 **S2-3**). The only misclassified case of DLA-RB was a Group A retinoblastoma located
118 at superior-nasal ora Serrata (**Supplementary Figure S1**).

119 DLA-RB could also accurately distinguish active retinoblastoma from stable
120 retinoblastoma (**Figure 2C-D, Table 2**). Among all referred and treated
121 retinoblastoma patients, the AUC, sensitivity, and specificity reached 0.962 (95%CI
122 0.915-1.000), 0.978 (95%CI 0.932-1.000), and 0.800 (95%CI 0.556-1.000)
123 respectively. Compared with competent ophthalmologists, the DLA-RB reached
124 superior sensitivity yet inferior specificity in identifying active retinoblastoma from

125 stable retinoblastoma (**Figure 2C-D, Supplementary Table S2-3**). Of all 4
126 misclassified cases in this task, 3 cases were false-positive, whereas only 1 case was
127 false-negative (**Table 4**).

128 Heatmap visualization revealed that the DLA-RB mainly focused on the tumor of the
129 newly-diagnosed retinoblastoma (**Figure 3A**). The DLA-RB focused on the relapse
130 site of retinoblastoma on fundus images (**Figure 3B**).

131 DLA-DR has proven to be sensitive and accuracy in identifying active retinoblastoma
132 from the normal fundus and stable retinoblastoma fundus. Yet, the cost-effectiveness
133 of a DLA-RB based approach to screen and surveillance retinoblastoma is unknown.

134 **Table 4** shows the base-case results of the cost-utility analysis. For identifying active
135 retinoblastoma tumors from all clinical-suspected retinoblastoma, the cumulative
136 costs for traditional ophthalmologic centers-based diagnosis mode and DLA-RB
137 based diagnosis mode were \$4882 and \$3519, respectively, and cumulative
138 quality-adjusted life-year (QALYs) for both diagnosis modes were 2.87769 and
139 2.8724 respectively. Compared with DLA-RB based diagnosis mode, the traditional
140 ophthalmologic centers-based diagnosis mode was not cost-effective since it resulted
141 in a gain of 1 QALY at the cost of \$257645. For distinguishing active retinoblastoma
142 from stable retinoblastoma patients, the cumulative costs and QALYs for traditional
143 ophthalmologic centers-based diagnosis mode were \$7614 and 5.46436, and the
144 figures for DLA-RB-based diagnosis mode were \$6981 and 5.46345. Traditional
145 ophthalmologic centers-based diagnosis mode was still unsatisfied with the

146 cost-effectiveness threshold since it produced an ICUR of \$694323. Therefore,
147 DLA-RB based diagnosis mode is cost-effective option in both RB diagnosis and
148 active lesion identification progress.

149 The tornado diagram shows the parameters that had the greatest impacts on base-case
150 outcomes and how much they influenced the outcomes (**Supplementary Figure S5**).
151 Probabilistic sensitivity analyses were conducted by taking 10000 random draws
152 surrounding the basic values (**Supplementary Figure S6**). According to
153 cost-effectiveness acceptability curves, at the current cost-effectiveness threshold,
154 DLA-RB based diagnosis mode had a 100% probability of being the more
155 cost-effective option in the process of RB diagnosis from all clinical-suspected
156 retinoblastoma, and it had a 59% probability to be cost-effective for identification of
157 active retinoblastoma from all treated retinoblastoma patients.

158

159 **Discussion**

160 In the current study, we developed Deep Learning Assistant for Retinoblastoma
161 (DLA-RB) that could accurately identify active retinoblastoma from the normal
162 fundus and stable retinoblastoma fundus. The DLA-RB further exhibited superior
163 sensitivity in identifying newly onset retinoblastoma and similar sensitivity in
164 identifying relapse retinoblastoma. The diagnosis accuracy and sensitivity were
165 superior/non-inferior than competent with about 2-5 years of experience in EUAs

166 examination. Compared with referral procedures to ophthalmologic centers, DLR-RB
167 based approach is cost-effective and readily implemented in the regional hospital.
168 According to the consensus from the American Association of Ophthalmic
169 Oncologists and Pathologists, children with a family history of the retinoblastoma are
170 highly recommended to take sufficient retina examination until the age of 7.¹⁸ Among
171 all family members of retinoblastoma patients, offspring of bilateral have near 50%
172 risk. In contrast, other siblings and nieces/nephews have less than 2.5% risk of
173 harboring RB1 mutation. EUA is strongly recommended for any child unable to
174 participate in an office examination sufficiently to allow a thorough examination of
175 the retina.¹⁸ However, the frequent screening schedule is often limited by the scarcity
176 of ophthalmological resources, especially in underdeveloped regions.¹⁹ Here, we
177 proposed a novel screening solution for offspring and family members of patients.
178 DLA-RB showed high efficiency with 0.991 of AUC and 0.979 sensitivity in the
179 prospective validation dataset. The diagnostic accuracy and sensitivity of DLA-RB
180 surpassed competent ophthalmologists. In real-world practice, the DLA-RB could
181 provide automatic and real-time feedback on retinoblastoma screening, especially
182 among the high-risk population.
183 Nowadays, although advanced multidisciplinary management concepts have been
184 adopted in developed countries, less significant improvement in globe salvage rate
185 was seen in the lower-middle-income country.²⁰ DLA-RB could also be used as a
186 screening technique for the contralateral eye of unilateral retinoblastoma patients who

187 received primary enucleation. In a retrospective analysis comprising
188 unilateral-presenting Group D retinoblastoma patients, during a median time of
189 approximately 5 years follow-up, only 1 out of 55 developed a new tumor in the
190 contralateral eye.⁸ Using DLA-RB at the local hospital could save time and financial
191 burden when ophthalmologic centers are not near at hand.

192 Through multidisciplinary management, the globe salvage rate in the high-income
193 country has tremendously improved from 34% in 1980–89 to 70% in 2010–20.²⁰

194 Based on a meta-analysis, retinoblastoma recurrence occurs in about 15% of patients.⁷

195 Advanced retinoblastoma (IIRC Group D-E) has a significantly higher risk of
196 recurrence.²¹ Extensive recurrence of subretinal or vitreous seeds was the most

197 common cause of treatment failure and final enucleation. Thus, there are increasing

198 burden of tumor surveillance after intraocular local control. Here, DLA-RB showed

199 high sensitivity to detect active retinoblastoma in the background of stable

200 retinoblastoma appearance (calcification, atrophy, laser and cryotherapy scar, and

201 fibrosis). In real-world practice, DLA-RB can provide automatic assistance in

202 treatment decision-making.

203 In the past two decades, telemedicine technology has enabled the centralization of

204 care at a single center in developing countries to achieve patient outcomes comparable

205 to those of developed countries.²² Within the domain of telemedicine, the DLA-RB

206 can largely contribute to lowering the burden in diagnosis and follow-up and

207 concentrate the limited healthcare resource in making multidisciplinary management.

208 DLA-RB can also help to promote the availability of retinoblastoma screening and
209 surveillance by improving the early diagnosis of population at-risk and long-term
210 follow-up for retinoblastoma patients. Such an approach has successfully managed
211 retinopathy of prematurity, a leading cause of childhood blindness worldwide.^{15,16,23}
212 Despite these remarkable results, some inherent limitations of DLA-RB are worth
213 highlighting. First, the development and prospective validation of DLA-RB were
214 conducted in a single institution (Beijing Tongren Hospital). The generalizability of
215 diagnostic performance needs further validation. Yet, in the current study, fundus
216 images in the present study were captured by five ophthalmological experts and two
217 different commercially available cameras, which fully reflected the variability in
218 operators' experience, image quality, resolution, different camera systems,
219 illumination, and field of view. Thus, the result from the perspective validation
220 dataset can represent the performance of DLA-RB in real-world practice. Secondly,
221 retinoblastoma begins as a round, translucent, gray to white tumor in the retina,
222 similar to the halation caused by microbubble in couplant. Careful image quality
223 control was required for ophthalmologists. An artificial intelligence-based quality
224 control module is needed to incorporate with DLA-RB, which warrants further
225 development. Third, accurately identifying active retinoblastoma also rely on a
226 thorough examination of the posterior pole and peripheral retina, which depends on
227 the experience of the operator. A deep-learning-based operation indicator may need to
228 cooperate with DLA-RB to lower the risk of false-negative chance. Fourth, because of

229 the low incidence of retinoblastoma, only 103 patients were included in prospective
230 validation, which limited subgroup analysis. It warrants further study to explore the
231 diagnosis accuracy of DLA-RB in different IIRC group retinoblastoma.

232 In conclusion, DLA-RB achieved high accuracy and sensitivity in identifying active
233 retinoblastoma from the normal fundus and stable retinoblastoma fundus. Compared
234 with referral procedures to ophthalmologic centers, DLA-RB based automatic
235 screening and activity surveillance is also cost-effective. In the future, DLA-RB can
236 incorporate telemedicine programs to reduce the burden in diagnosis and follow-up
237 and concentrate the limited health-care resource in making multidisciplinary
238 management.

239

240 **Methods**

241 **Clinical trial registration and ethics approval**

242 This single-center diagnostic study was done in Beijing Tongren hospitals in China.
243 The methods were performed in accordance with relevant guidelines and regulations
244 and approved by the Medical Ethics Committee of Beijing Tongren Hospital. Because
245 individually identifiable information was removed during retrospective collection,
246 written informed consent was exempted in the retrospective collected dataset. In the
247 prospectively collected validation dataset, written informed consent was obtained
248 from all caregivers prior to their inclusion. This study was registered on
249 ClinicalTrials.gov (NCT05308043) on February 2022. Because this study was a

250 diagnostic study that did not assign designative interventions to participants, clinical
251 trial protocol is not applicable.

252 **Dataset collection**

253 Training and internal test dataset were retrospectively collected between March 2018
254 and January 2022. All EUAs were performed for screening or pretreatment
255 examination in daily clinical practice. Five ophthalmological experts (HS Zhao, X Ge,
256 XL Xu, LB Jiang, JS, and JM Ma) obtained images using a commercially available
257 camera (RetCam3; Natus Medical Incorporated) and the standard imaging protocol. In
258 brief, posterior pole and 12 clocks of peripheral retina images were obtained during
259 EUAs. Repeatedly images were taken if necessary. We collected all EUAs images
260 from patients clinically diagnosed with retinoblastoma. For patients with several
261 EUAs, we included all images in the retrospective dataset. Patients who were
262 considered with Coats' disease, persistent fetal vasculature and other retinal diseases
263 were excluded.

264 The validation dataset was prospectively collected from February 2022 to June 2022.
265 Clinically suspected retinoblastoma and treated retinoblastoma patients that first
266 visited Beijing Tongren hospitals were standard EUAs. Patients involved in algorithm
267 development and considered to have other retinal diseases were excluded. During
268 prospective collection, a new generation of a commercially available camera
269 (RetCam3) was deployed in Beijing Tongren hospital with updated illumination and

270 imaging capture system. Thus, the prospective validation dataset was further divided
271 into two subsets (Equipment 1 and 2) to examine the reliability of new equipment.

272 **Image quality control and labeling**

273 All EUAs images were stored in jpeg format in the imaging databases. Poor-quality
274 images resulting from halation, blur, and defocus, as well as non-EUAs images, were
275 excluded manually. In addition, blurred images due to leukoplakia and secondary
276 cataract in advanced retinoblastoma were also excluded. A multidisciplinary
277 management team consisting of ophthalmology, pediatric and radiologist experts
278 made personal treatment strategies after EUAs. The personal treatment strategies
279 comprised a combination of a systemic chemotherapy regimen of carboplatin,
280 vincristine, etoposide, and focal consolidation therapy. Each chemotherapy cycle was
281 repeated every 3 to 4 weeks for 6 to 8 cycles, according to condition and tumor status.
282 Follow-up by EUAs was undertaken before each cycle of chemotherapy and every 3
283 to 4 weeks thereafter, during which the adjuvant laser therapy and cryotherapy were
284 applied as needed. In accordance with treatment strategies, five ophthalmological
285 experts (HS Zhao, X Ge, XL Xu, LB Jiang, JS, and JM Ma) with a minimum of 15
286 years of experience labeled all images as “normal fundus”, “stable retinoblastoma”
287 that specific treatment is not required, and “active retinoblastoma” that specific
288 treatment is required.

289 **Development of DLA-RB**

290 The patients from the retrospective dataset were randomly split into training and the
291 internal validation dataset as five-fold cross-validation for developing and evaluating
292 the performance of DLA-RB, respectively. To automatically distinguish “normal
293 fundus”, “stable retinoblastoma”, and “active retinoblastoma”, we first compared the
294 performance of some architectures including ResNet-50 and InceptionV3.²⁴⁻²⁶
295 ResNet-50 which did not occupy much computational resource achieved better
296 performance and was chosen to complete the task. The input of the model was an
297 image from EUAs. The output of the first model is a binary output determined
298 whether the input image contained active retinoblastoma was among normal fundus
299 and active retinoblastoma. The second binary classification task for determining
300 among stable retinoblastoma and active retinoblastoma, whether the input image
301 contained active retinoblastoma. In addition, the training dataset is imbalanced, so we
302 adopted a class weight policy in the training process. All models were developed with
303 Tensorflow 1.10.0 and Keras 2.2.4 on the server with three NVIDIA 1080 GPUs
304 (Graphical Processing Units).²⁷ All images were resized to 256×256 and then fed into
305 models to train or test. The optimization algorithm was SGD (Stochastic Gradient
306 Descent)²⁸, the default hyperparameters in Keras 2.2.4 were used, and at the same
307 time, the batch size was 32. Besides, class weight was used for trading off the effect
308 of the imbalanced distribution of two classes. Based on the repeated experiments,
309 different epochs were also applied to train the models without underfitting. Moreover,
310 the best models in terms of validation accuracy is saved as the final deployed model.

311 **Validation of DLA-RB**

312 We first validated the performance of DLA-RB in identifying retinoblastoma activity
313 in patients using an internal validation dataset and a prospective validation dataset
314 from Beijing Tongren Hospital. For further performance evaluation, two varying
315 degrees of expertise (competent and trainee), masked to the patients' demographics
316 and final multidisciplinary management strategy, were asked to classify every patient
317 in the prospective dataset independently, and their results were compared with those
318 of DLA-RB. The competent ophthalmologists attended doctors with about 2-5 years
319 of experience in general ophthalmology and EUAs examination. The trainee was a
320 resident who had finished EUAs training.

321 **Cost-utility analysis**

322 Two Markov models were built using TreeAge Pro (TreeAge Software; Williamstown,
323 MA, USA) to compare the cost-effectiveness between traditional ophthalmologic
324 centers-based diagnosis mode and DLA-RB based diagnosis mode from a societal
325 perspective. For the first binary model, we simulated a hypothetical cohort of a
326 newborn with RB through 5 1-year Markov cycles. With the growth of age, children
327 could be found to have eye symptoms by their parents according to specific
328 probabilities and then go to the hospital. Diagnosed patients received routine
329 treatment and management, whereas undiagnosed patients could still be examined in
330 the next cycle. According to the Management Guidelines for Childhood Screening for
331 Retinoblastoma Families, we determined the screening intervals for undiagnosed
332 patients.¹⁸ For the second binary model, simulated a hypothetical cohort of
333 2-year-old RB patients who have received regular treatment and were in an inactive

334 stage through a total of 3 1-year Markov cycles. They were routinely reviewed every
335 three months, and those who had active lesions received further treatment.

336 Parameters used in Markov models were collected from our study and previous
337 studies. Examination and treatment costs were collected in Chinese yuan from Beijing
338 Tongren Hospital and converted to US dollars at an exchange rate of 6.45 yuan per
339 dollar (**Supplementary Table S4**). Both direct and indirect costs were included, and
340 the specific composition of costs and annual costs of traditional ophthalmologic
341 centers-based diagnosis mode and DLA-RB-based diagnosis mode was shown in
342 **Supplementary Table S5-6**. Under traditional ophthalmologic centers-based
343 diagnosis mode, the patient and one accompanying family member spent more time in
344 tertiary hospitals or eye hospitals, resulting in transportation, accommodation, and
345 food costs. In DLA-RB-based diagnosis mode, we charged an additional \$15.5 for
346 each use. Only one accompanying family member's wage loss was counted.

347 Primary results were incremental cost-utility ratios (ICURs), which were calculated
348 using the following formula:

$$349 \quad ICURs = \text{incremental cost} / \text{quality adjusted life years gained}.$$

350 China's per capita gross domestic product (GDP) in 2021 was \$12551. According to
351 the WHO definition, intervention costing between 1 to 3 times the per capita GDP
352 was cost-effective, costing less than the per capita GDP was highly cost-effective, and
353 costing more than three times the per capita GDP was not cost-effective.²⁴ Both
354 1-way deterministic and simulated probabilistic sensitivity analyses were performed
355 to test the sensitivity and robustness of base-case values, the uncertainty ranges of
356 parameters were presented in **Supplementary Table S7**.

357 **Visualization and statistical analysis**

358 To visualize the decision ways of the model used, we applied the Grad-CAM to
359 generate heatmaps.²⁵ The performance of DLA-RB was estimated by accuracy,
360 sensitivity, specificity, positive predictive value, negative predictive value, and F1
361 score for the identification of each category. We used the receiver operating
362 characteristic (ROC) curve and precision-recall curve to show the diagnostic
363 performance of the DLA-RB in discriminating binary classification tasks. N-out-of-N
364 Bootstrapping with 1000 replicates was used to estimate 95% confidence intervals (95%
365 CI) of the performance metrics at eye level. Human ophthalmologists' performance
366 was compared with 95% CI of DLA-RB.

367 All statistical analysis was performed using R Statistical Software (version 4.1.1; R
368 Foundation for Statistical Computing, Vienna, Austria), Stata (17.0, StataCorp LLC,
369 College Station, TX).

370 **Data availability**

371 Python scripts enabling the main steps of the analysis are available from the
372 corresponding author on reasonable request. The data and materials in this study are
373 available from the corresponding author on reasonable request.

374 **Role of the funding source**

375 The funders had no role in the study design, data collection, data analysis, data
376 interpretation, or in the writing of the manuscript.

377

378 **Acknowledgments:** The authors thanks Beijing Hospitals Authority' Ascent Plan
379 (DFL20190201); National Natural Science Foundation of China (82141128); The
380 Capital Health Research and Development of Special (2020-1-2052);
381 Science & Technology Project of Beijing Municipal Science & Technology Commissi
382 on (Z201100005520045, Z181100001818003) for supporting this study.

383 **Contributors:** JMM, RHZ and DL contributed to the concept of the study. JMM,
384 WBW critically reviewed the manuscript. RHZ, LD, RYL,KZ, and YTL designed the
385 study and did the literature search. RHZ, LD, RYL,KZ,YTL, HSZ, JTS, XG, XLX,
386 LBJ, XHS, CZ, Wdz, LYX, HTW, HYL, CYY, and JL collected the data. RHZ, LD,
387 RYL,KZ, andYTL contributed to the design of the statistical analysis plan. RHZ, LD,
388 and KZ did the data analysis and data interpretation. RHZ and RYL drafted the
389 manuscript. JMM and WBW provided research funding, coordinated the research, and
390 oversaw the project. All authors had access to all the raw datasets and the
391 corresponding authors (JMM and WBW) has verified the data and had final decision
392 to submit for publication. All authors reviewed and approved the final manuscript

393 **Competing Interests:** All authors declare that the research was conducted in the
394 absence of any commercial or financial relationships that could be construed as a
395 potential conflict of interest.

396 **References**

- 397 1 Dimaras, H. *et al.* Retinoblastoma. *Nat Rev Dis Primers* **1**, 15021,
398 doi:10.1038/nrdp.2015.21 (2015).
- 399 2 Fabian, I. D. *et al.* The management of retinoblastoma. *Oncogene* **37**, 1551-1560,
400 doi:10.1038/s41388-017-0050-x (2018).
- 401 3 Luo, Y. *et al.* Contemporary Update of Retinoblastoma in China: Three-Decade
402 Changes in Epidemiology, Clinical Features, Treatments, and Outcomes. *Am J*
403 *Ophthalmol* **236**, 193-203, doi:10.1016/j.ajo.2021.09.026 (2022).
- 404 4 Linn Murphree, A. Intraocular retinoblastoma: the case for a new group classification.
405 *Ophthalmol Clin North Am* **18**, 41-53, viii, doi:10.1016/j.ohc.2004.11.003 (2005).
- 406 5 Stathopoulos, C. *et al.* Current Indications of Secondary Enucleation in
407 Retinoblastoma Management: A Position Paper on Behalf of the European
408 Retinoblastoma Group (EURbG). *Cancers (Basel)* **13**, doi:10.3390/cancers13143392
409 (2021).
- 410 6 Daniels, A. B. *et al.* Effect of Intravenous Chemotherapy Regimen on Globe Salvage
411 Success Rates for Retinoblastoma Based on Disease Class-A Meta-Analysis.
412 *Cancers (Basel)* **13**, doi:10.3390/cancers13092216 (2021).
- 413 7 Chen, Q. *et al.* Comparison between intravenous chemotherapy and intra-arterial
414 chemotherapy for retinoblastoma: a meta-analysis. *BMC Cancer* **18**, 486,
415 doi:10.1186/s12885-018-4406-6 (2018).
- 416 8 Fabian, I. D. *et al.* Primary enucleation for group D retinoblastoma in the era of

- 417 systemic and targeted chemotherapy: the price of retaining an eye. *Br J Ophthalmol*
- 418 **102**, 265-271, doi:10.1136/bjophthalmol-2017-310624 (2018).
- 419 9 Dimaras, H. *et al.* Retinoblastoma. *Lancet* **379**, 1436-1446,
- 420 doi:10.1016/S0140-6736(11)61137-9 (2012).
- 421 10 Gulshan, V. *et al.* Development and Validation of a Deep Learning Algorithm for
- 422 Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA* **316**,
- 423 2402-2410, doi:10.1001/jama.2016.17216 (2016).
- 424 11 Peng, Y. *et al.* DeepSeeNet: A Deep Learning Model for Automated Classification of
- 425 Patient-based Age-related Macular Degeneration Severity from Color Fundus
- 426 Photographs. *Ophthalmology* **126**, 565-575, doi:10.1016/j.ophtha.2018.11.015 (2019).
- 427 12 Li, Z. *et al.* Efficacy of a Deep Learning System for Detecting Glaucomatous Optic
- 428 Neuropathy Based on Color Fundus Photographs. *Ophthalmology* **125**, 1199-1206,
- 429 doi:10.1016/j.ophtha.2018.01.023 (2018).
- 430 13 Tan, T. E. *et al.* Retinal photograph-based deep learning algorithms for myopia and a
- 431 blockchain platform to facilitate artificial intelligence medical research: a retrospective
- 432 multicohort study. *The Lancet. Digital health* **3**, e317-e329,
- 433 doi:10.1016/S2589-7500(21)00055-8 (2021).
- 434 14 Milea, D. *et al.* Artificial Intelligence to Detect Papilledema from Ocular Fundus
- 435 Photographs. *N Engl J Med* **382**, 1687-1695, doi:10.1056/NEJMoa1917130 (2020).
- 436 15 Brown, J. M. *et al.* Automated Diagnosis of Plus Disease in Retinopathy of Prematurity
- 437 Using Deep Convolutional Neural Networks. *JAMA Ophthalmol* **136**, 803-810,

- 438 doi:10.1001/jamaophthalmol.2018.1934 (2018).
- 439 16 Redd, T. K. *et al.* Evaluation of a deep learning image assessment system for
440 detecting severe retinopathy of prematurity. *Br J Ophthalmol*,
441 doi:10.1136/bjophthalmol-2018-313156 (2018).
- 442 17 Shields, C. L. *et al.* The International Classification of Retinoblastoma predicts
443 chemoreduction success. *Ophthalmology* **113**, 2276-2280,
444 doi:10.1016/j.ophtha.2006.06.018 (2006).
- 445 18 Skalet, A. H. *et al.* Screening Children at Risk for Retinoblastoma: Consensus Report
446 from the American Association of Ophthalmic Oncologists and Pathologists.
447 *Ophthalmology* **125**, 453-458, doi:10.1016/j.ophtha.2017.09.001 (2018).
- 448 19 Ramirez-Ortiz, M. A. *et al.* Diagnostic delay and sociodemographic predictors of stage
449 at diagnosis and mortality in unilateral and bilateral retinoblastoma. *Cancer Epidemiol*
450 *Biomarkers Prev* **23**, 784-792, doi:10.1158/1055-9965.EPI-13-1069 (2014).
- 451 20 Wong, E. S. *et al.* Global retinoblastoma survival and globe preservation: a systematic
452 review and meta-analysis of associations with socioeconomic and health-care factors.
453 *Lancet Glob Health* **10**, e380-e389, doi:10.1016/S2214-109X(21)00555-6 (2022).
- 454 21 Munier, F. L. *et al.* Conservative management of retinoblastoma: Challenging
455 orthodoxy without compromising the state of metastatic grace. "Alive, with good vision
456 and no comorbidity". *Progress in retinal and eye research* **73**, 100764,
457 doi:10.1016/j.preteyeres.2019.05.005 (2019).
- 458 22 Yousef, Y. A. *et al.* How Telemedicine and Centralized Care Changed the Natural

- 459 History of Retinoblastoma in a Developing Country: Analysis of 478 Patients.
460 *Ophthalmology* **128**, 130-137, doi:10.1016/j.ophtha.2020.07.026 (2021).
- 461 23 Morrison, S. L., Dukhovny, D., Chan, R. V. P., Chiang, M. F. & Campbell, J. P.
462 Cost-effectiveness of Artificial Intelligence-Based Retinopathy of Prematurity
463 Screening. *JAMA Ophthalmol* **140**, 401-409, doi:10.1001/jamaophthalmol.2022.0223
464 (2022).
- 465 24 Hutubessy, R., Chisholm, D. & Edejer, T. T.-T. Generalized cost-effectiveness
466 analysis for national-level priority-setting in the health sector. *Cost Eff Resour Alloc* **1**,
467 8 (2003).
- 468 25 Selvaraju, R. R. *et al.* in *Proceedings of the IEEE international conference on*
469 *computer vision*. 618-626.
- 470
- 471

472 **Figure legends**

473 **Figure 1.** Workflow diagram for the development and evaluation of DLA-RB.

474 RB, retinoblastoma.

475 **Figure 2.** Receiver operating characteristic curves and precision-recall curve of

476 DLA-RB performance in prospective validation dataset.

477 Receiver operating characteristic and precision-recall curve of DLA-RB in identifying

478 normal-active lesion identification (A-B), and stable-active lesion identification

479 (C-D).

480 **Figure 3.** Heatmap Visualization of DLA-RB

481 Heatmap demonstrating representative lesions, shown in original fundus image (first

482 column), general heatmap (second column) for (A) identifying active retinoblastoma

483 tumors from all clinical-suspected retinoblastoma and (B) identifying active

484 retinoblastoma from all referred and treated retinoblastoma patients.

485

486 **Table 1.** Basic characteristics of patients in prospective validation dataset.

	Equipment 1	Equipment 2
Number of patients	64	39
Age at diagnosis		
<12month	14 (21.9%)	11 (28.2%)
12-23 month	19 (29.7%)	15 (38.5%)
24-35 month	14 (21.9%)	6 (15.4%)
≥36 month	17 (26.6%)	7 (17.9%)
Gender		
male	36 (56.3%)	23 (59.0%)
female	28 (43.8%)	16 (41.0%)
Laterality		
unilateral	44 (68.7%)	22 (56.4%)
bilateral	19 (29.7%)	16 (41.0%)
Number of eyes	88	51
IIRC Grade		
normal fundus	46 (52.3%)	24 (47.1%)
A	3 (3.4%)	0 (0.0%)
B	5 (5.7%)	2 (3.9%)
C	6 (6.8%)	2 (3.9%)
D	22 (25.0%)	18 (35.3%)
E	6 (6.8%)	5 (9.8%)
Number of eyes with retinoblastoma	42	27
Tumor focality		
unifocal	16 (38.1%)	4 (14.8%)
multifocal	26 (61.9%)	23 (85.2%)
Retinoblastoma growth pattern		
endophytic	24 (57.1%)	15 (55.6%)

exophytic	18 (42.9%)	12 (44.4%)
Fovea involvement		
not involved	14 (33.3%)	6 (22.2%)
subfoveal fluid	2 (4.8%)	4 (14.8%)
foveal tumor	26 (61.9%)	17 (63.0%)
Retinoblastoma seeds		
not seeds	17 (40.5%)	7 (25.9%)
subretinal	16 (38.1%)	10 (37.0%)
vitreous	9 (21.4%)	10 (37.0%)

487 IIRC, international intraocular retinoblastoma classification.

488

489 **Table 2.** Performance of DLA-RB in prospective validation dataset,

	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	F1 score (95% CI)
Normal-active retinoblastoma identification	0.992 [0.977-1.000]	0.979 [0.927-1.000]	1.000 [1.000-1.000]	1.000 [1.000-1.000]	0.988 [0.961-1.000]	0.989 [0.932-1.000]
Stable-active retinoblastoma identification	0.934 [0.869-0.984]	0.978 [0.932-1.000]	0.800 [0.556-1.000]	0.937 [0.868-1.000]	0.923 [0.727-1.000]	0.957 [0.930-1.000]

490

491

492 **Table 3.** Confusion matrix of DLA-RB in identifying active retinoblastoma among
493 treated patients.

		DLA-RB		
		stable	active	total
Gold standard	stable	12	3	15
	active	1	45	46
	total	13	48	61

494

495

496 **Table 4.** Base-case cost-utility results.

	Costs per person, \$	QALYs per person	Incremental costs per person, \$	Incremental QALYs per person	ICURs, \$
Diagnosis of RB					
DLA-RB-based diagnosis mode	3519	2.87240
Traditional ophthalmologic centers-based diagnosis mode	4882	2.87769	1363	0.00529	257645
Identification of active lesion in treated retinoblastoma patients					
DLA-RB-based diagnosis mode	6981	5.46345
Traditional ophthalmologic centers-based diagnosis mode	7614	5.46436	633	0.00091	694323

497 DLA-RB, Deep Learning Assistant for retinoblastoma; ICUR, incremental cost-utility ratio; QALY,

498 quality-adjusted life-year.

499 Costs are given in US dollars. Negative ICURs are regarded as dominating.

500

Development dataset

47503 images from 713 patients were retrospectively collected between March, 2018 and January, 2022

Prospective validation dataset

2032 images from 103 patients (139 eyes) were prospectively collected from February, 2022 to June, 2022

666 pictures were removed because of poor quality

10880 pictures were removed because of poor quality

Equipment 1
833 images from
64 patients (88 eyes)

Equipment 2
533 images from
39 patients (51 eyes)

36623 images from 713 patients
19045 images were normal
2918 images were stable RB
14660 images were active RB

1366 images from 103 patients
751 images were normal
132 images were stable RB
516 images were active RB

Deep Learning Assistant for Retinoblastoma Monitoring (DLA-RB) System



