Journal of Big Data

RESEARCH Open Access

Noninvasive identification of Benign and malignant eyelid tumors using clinical images via deep learning system

Shiqi Hui[1†], Li Dong[2†], Kai Zhang[3], Zihan Nie[1], Xue Jiang[1], Heyan Li[1], Zhijia Hou[1], Jingwen Ding[1], Yue Wang[1] and Dongmei Li[1,4*]

[†]Shiqi Hui and Li Dong contributed equally to the study

*Correspondence: ldmlily1@163.com

[1] Beijing Ophthalmology & Visual Science Key Laboratory, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China
[2] Beijing Key Laboratory of Intraocular Tumor Diagnosis and Treatment, Beijing Ophthalmology & Visual Sciences Key Lab, Medical Artificial Intelligence Research and Verification Key Laboratory of the Ministry of Industry and Information Technology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China
[3] InferVision Healthcare Science and Technology Limited Company, Shanghai, China
[4] 1 Dong Jiao Min Lane, Beijing, 100730, China

Abstract

Eyelid tumors accounts for 5–10% of skin tumors. It is important but difficult to identify malignant eyelid tumors from benign lesions in a cost-effective way. Traditional screening methods for malignancy in eyelid tumors require laborious and time-consuming histopathological process. Therefore, we aimed to develop a deep learning (DL)-based image analysis system for automatic identification of benign and malignant eyelid tumors. Using a common digital camera, we collected clinical images from patients who were histopathologically diagnosed with eyelid tumors. We trained 8 convolutional neural network (CNN) models to identify benign and malignant eyelid tumors, including ResNet-50, ResNet-101, InceptionV3, and InceptionResNetV2. Another group of patients with eyelid tumors were also collected as the prospective validation dataset. Performance of DL models and human clinicians in prospective validation dataset were evaluated and compared. A total of 309 images from 209 patients were used for training DL system, all eight models reached an average accuracy greater than 0.958 in the internal cross-validation. 36 images from 36 patients were included for the prospective validation, the models reached the best performance in accuracy, sensitivity, specificity, and area under curve (AUC) of 0.889 (95% CI 0.747–0.956), 0.933 (95% CI 0.702–0.988), 0.857 (95% CI 0.654–0.950), and 0.966 (95% CI 0.850–0.993), respectively. DL system had a similar performance as the senior ophthalmologists, and outreached the performance of junior ophthalmologists and medical students. DL system can identify benign and malignant tumors through common clinical images, with a better performance than most ophthalmologists. Combining DL system with smartphone may enable patients' self-monitoring for eyelid tumors and assist in doctors' clinical decision making.

Keywords: Deep learning, Eyelid tumor, Clinical image

Introduction

Eyelid tumors accounts for 5–10% of skin tumors [1, 2], which can be divided into benign and malignant lesions, according to their tissue or cell of origin. In general, 80–90% eyelid lesions appear as benign tumors [3–6]. However, malignant eyelid lesions can be life threatening, with up to 30% 5-year mortality in sebaceous cell carcinoma patients [7].

Malignant cases include basal cell carcinoma (80–90%), squamous cell carcinoma (5%), sebaceous cell carcinoma (1.0–5.5%), and melanoma (< 1%) [8]. The prevalence and outcomes of various eyelid tumor subtypes vary significantly, due to the different geographic location, genetic background, socio-economic status, and healthcare policies [9]. A complete surgical excision with intraoperative margin control is the standard treatment for malignant lesions, which can reduce the rate of recurrence [8]. Therefore, it is necessary to accurately differentiate malignant lesions from benign at treatment onset for the reduction in the mortality and complications. The diagnosis of eyelid lesions requires a specific expertise and biological or pathological process, which is laborious, time consuming and subject to the experience of the pathologists. There is also a concern that the lack of access to pathology expertise may result in foregoing intraoperative frozen section diagnosis and gross anatomical features for lesion identification. Thus, new efficient approaches are needed to address current limitations.

Over the past few years, artificial intelligence (AI), especially deep learning (DL) [10], has played a major role in the field of medicine, including image recognition [11], auxiliary diagnosis [12], drug development [13], and health care management [14]. In fact, DL based systems have been developed to detect eyelid melanoma and basal cell carcinoma using dermoscopic images [15, 16] or pathological images [17, 18]. However, they still require pathological examination or dermatoscopical process which are not convenient for screening malignancies among non-experts. Pathological examination is invasive and dermatoscopical process need professional equipment.

Digital photograph is the most commonly used approach to analyze facial data due to the conveniency and intuition. The application of DL based on digital photograph has achieved physician-equivalent classification accuracy in lid position [19–22] and skin cancer [23, 24]. Moreover, importing photograph into a smartphone can achieve portable and convenient telemedicine-technology [25] to determine whether an emergency medical treatment is required. We meant to provide a noninvasive identification in differentiate malignant eyelid tumors from benign ones without pathological or dermatoscopical process, makes it possible for patients to monitor eyelid tumors and identify malignant ones at early stage. The present study aimed to establish DL models to automatically differentiate benign and malignant eyelid tumors using common digital photographs. Then we compared the performance between DL system with different levels of ophthalmologists.

Methods

Study design and participants

This study was a single centre, diagnostic research with prospective validation. From 1 to 2017 to 30 September 2020, eyelid tumors patients who underwent basic ophthalmic examinations in Beijing Tongren Hospital (Beijing, China) were retrospectively collected. The pictures of eyelid tumor were captured using a digital camera (DSC-F828, Sony, Japan) at the first visit. Patients were asked to look horizontally, and camera positioned in the frontal plane at pupil height, one meter away from the patients. For tumors that cannot expose completely in primary position, we used a medical cotton swab for auxiliary exposure. Photographs were taken in outpatient clinics and inpatient wards, hence the lighting and background of the images were not uniform, indicating

the richness and diversity of our datasets. Those who finally underwent tumor resection surgery and had histopathological diagnosis were included in this study. These data were used as developmental dataset, and were randomly divided into two independent datasets with four-fold cross-validation as subject independent manner, and the best model can be chosen with repeated four rounds of training with a development dataset and testing with a validation dataset [26–28]. The tumors' images of a patient will not be split into training and testing datasets at the same time.

To further validate the performance of the DL system, another group of patients were prospectively collected in Beijing Tongren Hospital between 1 and 2020 and 30 June 2021 as the prospective validation dataset. All patients underwent surgical resection and received histopathological diagnosis. The pictures of those eyelid tumors were also captured using the same digital camera. The flow chart of data selection was showed in Fig. 1.

In this study, all procedures were conducted in accordance with the Declaration of Helsinki. The Ethics Committee of Beijing Tongren Hospital approved the study. Written informed consent was obtained from each subject.

Image preprocessing and quality control

To improve the DL analysis, we resized the images to a resolution of 256×256 pixels before developing the algorithm. In the quality control process, we assessed the image quality and filtered out unqualified images after mask removal. The pixel values of the selected images applied to a linear mapping with a pixel value ranging from (0, 255) to (0, 1). Based on several arguments, such as the readable region ratio, illumination, blurriness, and image contents, pictures with poor quality were also excluded. The tumors on patients' faces were encircled using polygon tool and annotated based on histopathological diagnosis, and then the regions of tumors were cropped for training the algorithms. The annotation tool is LabelMe (https://github.com/wkentaro/labelme).

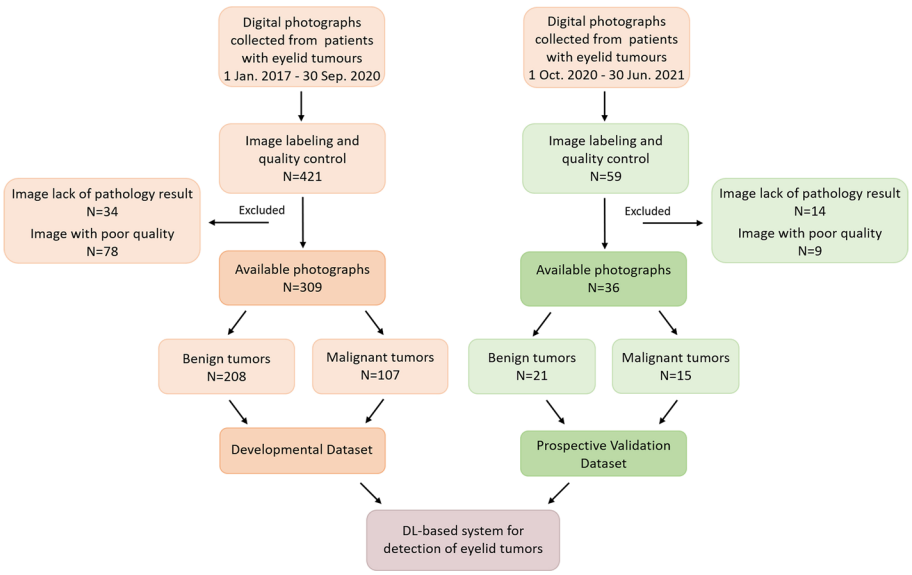

Fig. 1 Schematic diagram of this study

Algorithm development

We applied several convolutional neural networks to automatically detect whether the eyelid tumor was benign or malignant. Histopathological diagnosis was used as ground truth. We first compared the performance of some architectures including ResNet-50, ResNet-101, InceptionV3, and InceptionResnetV2 [29–31]. Based on the GPU (Graphical Processing Unit) memory and generalization ability, we chose these four types of CNNs mentioned above. We adopted four-fold cross-validation to develop the models and selected the optimal one. To further test the performance of the DL models, we then used the prospective validation datasets. The overview of the deep convolutional neural network-based model training pipeline was illustrated in Fig. 2. All models were developed with Tensorflow 1.10.0 and Keras 2.2.4 on the server with three NVIDIA 1080 GPUs, and were pretrained with imagenet dataset [32]. We fine-tuned the weight of CNNs from the pretrained models which were trained with imagenet dataset, instead of training from scratch. We used several data augmentation methods to enrich the dataset in the training stage, including horizonal flipping, vertical flipping and rotation up to 90°, which could reduce the possibility of overfitting. Because we thought the colour and shape features were the most important characteristics, some other data augmentation methods which might modify the pixel values and appearance were not adopted. The appearance images of eyes were used as the input to discern whether this tumor was benign or malignant. The samples were shown in Fig. 3. The optimization algorithm was SGD (Stochastic Gradient Descent) [33], the default hyperparameters in Keras 2.2.4 were used, at the same time, batch size was 15. Besides, class weight was used to trade off the effect of imbalanced distribution of two classes. Based on the repeated experiments, different epochs were also applied to train the models without underfitting. Early stopping was applied, and if the validation loss did not improve over 10 consecutive epochs [34].

Comparison between human and DL system

Three senior ophthalmologists (with more than 15 years clinical experience), two junior ophthalmologists (with more than 5 years clinical experience), and two medical students

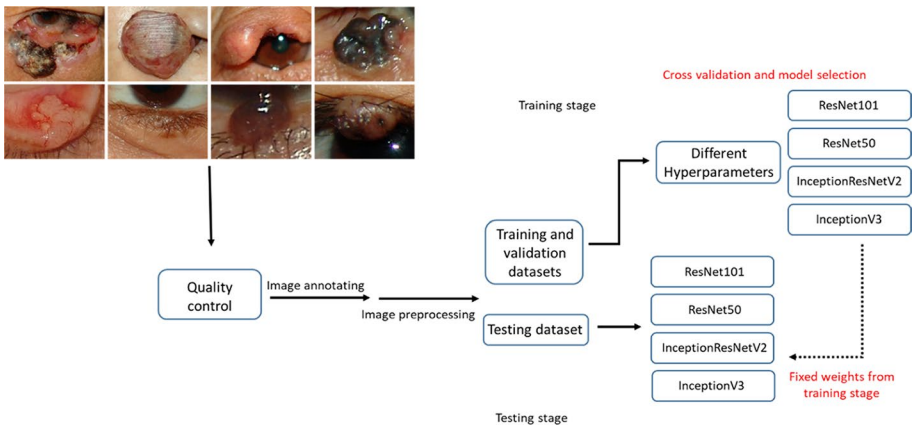

Fig. 2 Overview of the deep learning-based system to automatically predict eyelid tumors from digital clinical images

Fig. 3 Input samples of benign and malignant tumors

were invited to independently diagnose the tumors in the prospective validation dataset. The results of the DL system and histopathological information were not available to any human doctors. We compared the performance between these human ophthalmologists with the DL system.

Statistical analyses

All statistical analyses were performed using Python 3.7.3 (Wilmington, DE, USA) and MATLAB R2016a (https://www.mathworks.com/). We used the accuracy, sensitivity, specificity, and receiver-operating characteristic curve to assess the performance of the DL model. The area under curve (AUC) with 95% confidence interval (CI) was calculated.

Results

A total of 309 pictures from 229 patients with eyelid tumors were retrospectively gathered for the training, tuning, and internal validation of the DL system (Table 1). The mean age (standard deviation, SD) was 49.3 ± 17.5 years old, and 63.76% patients were female. 157 subjects were histopathologically diagnosed with benign tumors, while 72 patients were diagnosed with malignant tumors. The most common malignant eyelid tumor in our datasets is basal cell carcinoma (60/122, 49.18%), followed by sebaceous adenocarcinoma eyelid (32/122, 26.23%), squamous cell carcinoma (12/122, 9.84%), and eyelid melanoma (9/122, 7.38%). The most common benign eyelid tumor is nevus (149/223, 66.82%), followed by cyst (14/223, 6.28%), seborrheic keratosis (13/223, 5.83%), and xanthelasma (10/223, 4.48%). The top two malignant eyelid tumors in our prospective validation datasets are basal cell carcinoma (5/15, 33.33%) and eyelid melanoma (3/15, 20%). The top two malignant eyelid tumors are nevus (7/21, 33.33%) and seborrheic keratosis (2/21, 9.52%). Patients with malignant eyelid tumors were older than those with benign tumors. Seven tumors existed in bilateral eyelids, and twenty-five tumors were located in both upper and lower eyelids. Another thirty sixty pictures of eyelid tumors images from 36 patients were prospectively collected as the prospective validation dataset. Similar age, sex distribution, and tumor location were recognised in the two datasets.

Table 1 Components of the developmental dataset and the prospective validation set

	Developmental set			Prospective validation set		
	Malignant tumor	Benign tumor	Total	Malignant tumor	Benign tumor	Total
Patients	72	157	229	15	21	36
Photographs	107	202	309	15	21	36
Age (SD)	61.5 (12.3)	43.6 (16.6)	49.3 (17.5)	64.9 (7.0)	39.6 (18.0)	50.1 (19.1)
Female (%)	58.33	66.24	63.76	46.67	80.95	66.67
Laterality						
Right eye	61	100	161	6	11	17
Left eye	46	95	141	9	10	19
Both	0	7	7	0	0	0
Location of tumor						
Upper eyelid	33	97	130	4	10	14
Lower eyelid	62	92	154	11	10	21
Both	12	13	25	0	1	1

SD standard deviation

Table 2 Performance of models in the internal validation dataset

Model number	Parameters	Architecture	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
1	Class weight = [0.1, 25] Epoch = 80	ResNet101	0.961 (0.930–0.991)	0.793 (0.607–0.978)	0.968 (0.946–0.991)	0.969 (0.942–0.995)
2		ResNet50	0.962 (0.950–0.974)	0.779 (0.674–0.884)	0.973 (0.954–0.991)	0.973 (0.957–0.990)
3		Inception-ResNetV2	0.960 (0.942–0.978)	0.769 (0.494-1.000)	0.970 (0.941-1.000)	0.946 (0.879-1.000)
4		InceptionV3	0.954 (0.933–0.976)	0.756 (0.664–0.848)	0.966 (0.948–0.985)	0.958 (0.931–0.986)
5	Class weight = [0.1, 30] Epoch = 60	ResNet101	0.956 (0.923–0.989)	0.769 (0.692–0.846)	0.967 (0.939–0.994)	0.958 (0.952–0.965)
6		ResNet50	0.963 (0.925-1.000)	0.883 (0.777–0.988)	0.965 (0.912-1.000)	0.972 (0.952–0.992)
7		Inception-ResNetV2	0.967 (0.942–0.992)	0.809 (0.686–0.933)	0.967 (0.919-1.000)	0.963 (0.941–0.986)
8		InceptionV3	0.939 (0.865-1.000)	0.833 (0.581-1.000)	0.945 (0.853-1.000)	0.944 (0.900-0.988)

CI confidence interval, *AUC* area under curve

Table 2 showed the performance of different DL models for the detection of eyelid tumors. All eight models reached an average accuracy greater than 0.958 in the internal cross-validation. The average sensitivity and specificity were greater than 0.795 and 0.965, respectively, and the mean AUCs were greater than 0.960. Table 3 showed the performance of these models in prospective validation dataset, the best model reached the accuracy, sensitivity, specificity, and AUC of 0.889 (95% CI 0.747–0.956), 0.933 (95% CI 0.702–0.988), 0.857 (95% CI 0.654–0.950), and 0.966 (95% CI 0.850–0.993), respectively. The ROC and PR curves of these eight models were shown in Fig. 4.

When comparing the performance between human ophthalmologists and the DL system, we found that DL system reached a similar, and even better diagnostic performance

Table 3 Performance of models in the prospective validation dataset

Model number	Parameters	Architecture	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
1	Class weight = [0.1, 25] Epoch = 80	ResNet101	0.718 (0.556–0.838)	0.800 (0.548–0.930)	0.667 (0.454–0.828)	0.880 (0.736–0.951)
2		ResNet50	0.833 (0.681–0.921)	0.733 (0.481–0.891)	0.905 (0.711–0.974)	0.930 (0.799–0.978)
3		Inception-ResNetV2	0.806 (0.650–0.903)	0.933 (0.702–0.988)	0.714 (0.500-0.862)	0.867 (0.720–0.943)
4		InceptionV3	0.778 (0.619–0.883)	0.800 (0.548–0.930)	0.762 (0.549–0.894)	0.903 (0.765–0.964)
5	Class weight = [0.1, 30] Epoch = 60	ResNet101	0.889 (0.747–0.956)	0.933 (0.702–0.988)	0.857 (0.654–0.950)	0.966 (0.850–0.993)
6		ResNet50	0.750 (0.589–0.863)	0.867 (0.621–0.963)	0.667 (0.454–0.828)	0.872 (0.726–0.946)
7		Inception-ResNetV2	0.833 (0.681–0.921)	0.933 (0.702–0.988)	0.762 (0.549–0.894)	0.954 (0.832–0.989)
8		InceptionV3	0.778 (0.619–0.883)	1.000 (0.796-1.000)	0.619 (0.409–0.793)	0.871 (0.726–0.946)

CI confidence interval, *AUC* area under curve

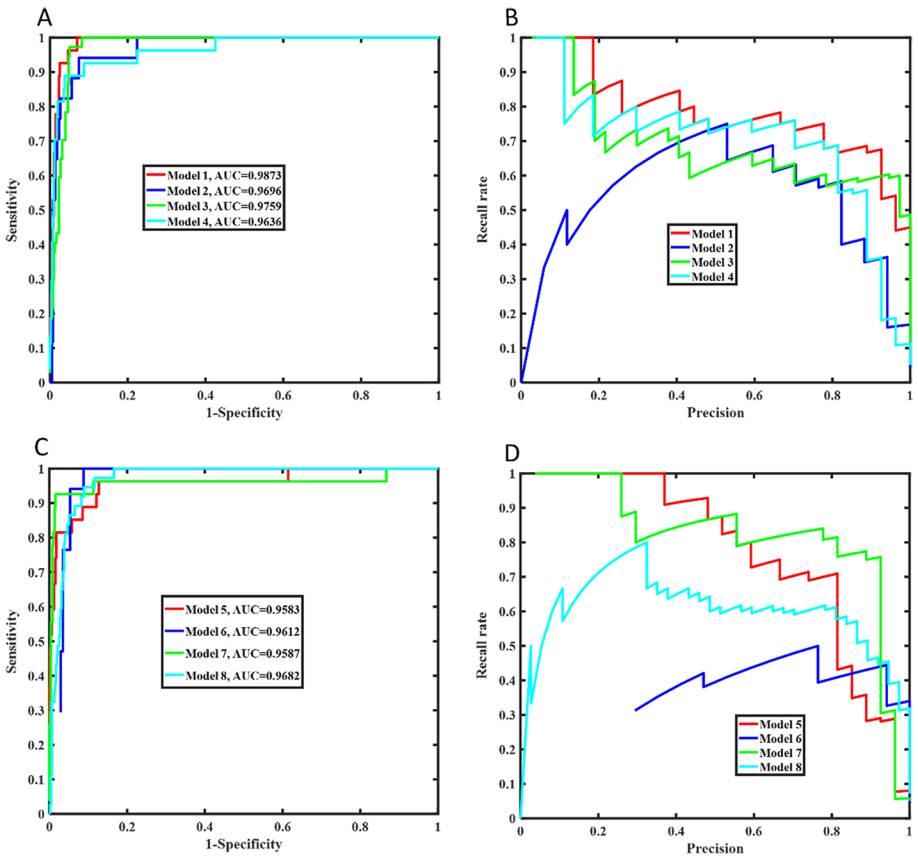

Fig. 4 Performance of models in cross validation. **A**, **B** Epoch = 80. **C**, **D**: Epoch = 60. Model 1: ResNet101; Model 2: ResNet50; Model 3: InceptionResNetV2; Model 4: InceptionV3; Model 5: ResNet101; Model 6: ResNet50; Model 7: InceptionResNetV2; Model 8: InceptionV3

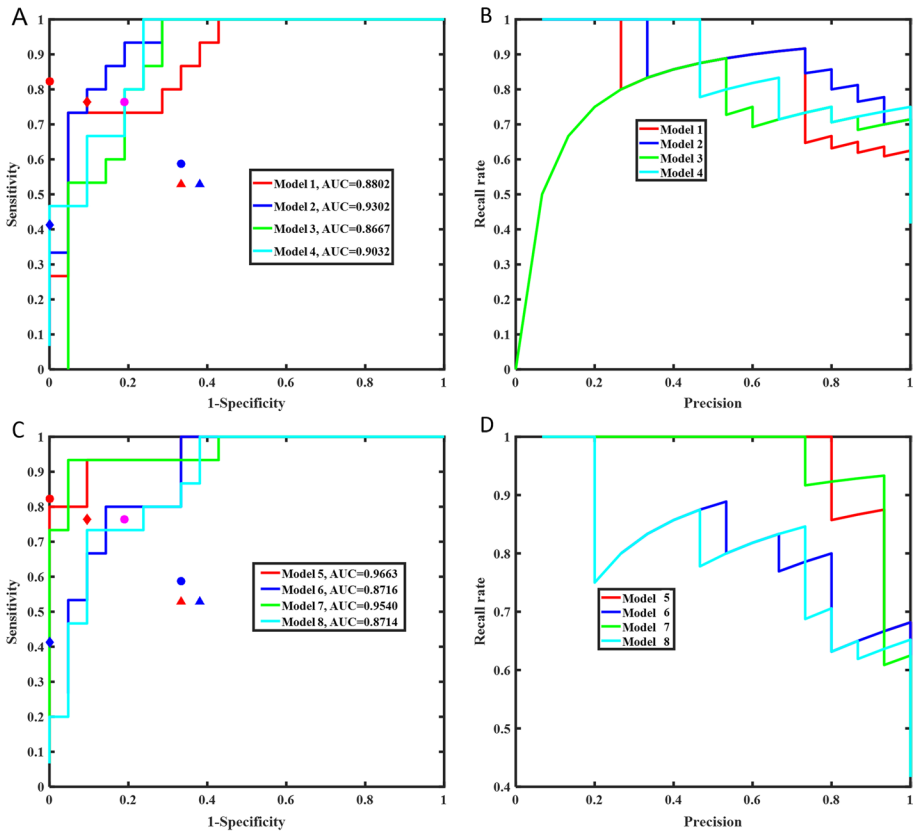

Fig. 5 Performance of models in prospective validation dataset and comparison with human ophthalmologists. **A**, **B** Epoch = 80. **C**, **D** Epoch = 60. Model 1: ResNet101; Model 2: ResNet50; Model 3: InceptionResNetV2; Model 4: InceptionV3; Model 5: ResNet101; Model 6: ResNet50; Model 7: InceptionResNetV2; Model 8: InceptionV3. Red filled rhombus: Senior Ophthalmologist 1, blue filled rhombus: Senior Ophthalmologist 2, red filled circle: Junior Ophthalmologist 1, blue filled circle: Junior Ophthalmologist 2, pink filled circle: Junior Ophthalmologist 3, red filled triangle: Medical student 1, blue filled triangle: Medical student 2

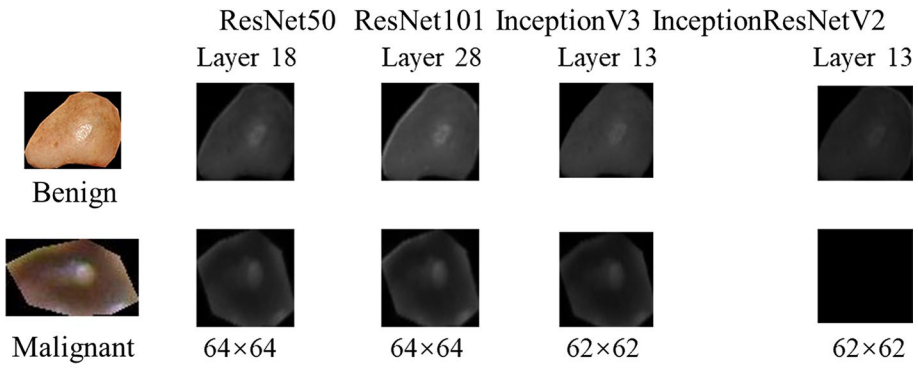

Fig. 6 The feature maps in four types of CNNs

than senior ophthalmologists. In general, DL system performed much better than junior ophthalmologists and medical students (Fig. 5). The features maps of four types of CNNs were shown in Fig. 6, which showed the contour and pixel values were more important.

Discussion

In this study, we successfully trained DL models that could automatically identify benign and malignant eyelid tumors from clinical images. Even with a rather small database in the training set, our CNN algorithms had more accurate diagnosis than junior ophthalmologists and medical students, reaching an 88.89% accuracy, 85.71% sensitivity, and 93.33% specificity in the detection of eyelid tumors. The DL system showed a comparable performance with senior experts.

Eyelid tumor is common seen, but it is important but difficult to distinguish benign and malignant tumors, as they sometimes have overlapping features of irregular shapes, irregular pigmentation, and telangiectasia to malignant ones. Benign tumor is most commonly described as a well-demarcated, waxy, pigmented lesion, and developed at a younger age. Malignant tumor may diffuse infiltration to surrounding tissues, damage the orbit and intraorbital tissues, result in loss of lashes, central ulceration and/or destruction of eyelid architecture, distant metastasis may also occur. Our model combines with smartphone may help patients to monitor the malignant eyelid tumors themselves and assist in doctors' clinical decision making. Lord et al. [35] first proposed the novel use of smartphone in ophthalmic imaging. Detailed use of smartphone-based image applications in ophthalmology was described later by various researchers [36–38]. Smartphone-based ophthalmic imaging techniques can be adopted by any clinician to obtain opinions from experts, and portable image documenting.

There are numerous imaging devices in ophthalmic examination, most of which are sophisticated and specialized for specific regions of the eye, which requires close interaction of the patient and the clinician. Therefore, a simple, portable alternative high-quality imaging tool for routine examination is needed. Clinical images are of more convenient, and intuitive compared to dermoscopic images [17, 39] or histopathological images [18, 40, 41] reported previously. Clinical images have been used to evaluate eyelid disorders [22, 25, 42, 43], ocular motility and strabismus [44–47].

Recently, Adamopoulos et al. [48] and Li et al. [49] also developed AI system to detect eyelid tumors. The details of comparison with these studies have been listed in Table 4. In brief, Adamopoulos et al. [48] used photographic images with small sample size to distinguish basal cell carcinoma. However, basal cell carcinoma is not the only malignant eyelid tumor so that this model may fail to identify other malignant mass. In addition, this study did not provide some important evaluation metrics including sensitivity, specificity, and AUC. Li [49] et al. also developed DL model to identify malignant eyelid tumors from benign ones with bigger sample size. Before identifying the characteristics of the tumors, they trained model to locate the tumor first with an average precision of 76.2%, which meant about a quarter of mass was wrongly located. Therefore, human ophthalmologists in our study were assigned to precisely delineate the tumors before the implementation of DL in the development stage and real deployment scenario, so that our model showed better performance and this approach might be more appropriate for decision-making in clinic. Also, as an attempt, we used another prospectively collected dataset to compare the performance between model and human ophthalmologists, which proved the model has surpassed most ophthalmologists.

High accuracy and efficiency rates are the main advantages in the application of DL system in medical diagnosis, since the developed algorithms can capture and integrate

Table 4 Comparison of the current study and previous methods

	Our study	Adamopoulos et al. [48]	Li et al. [49]
What was solved	• DL system can identify benign and malignant tumors through common clinical images, with a better performance than most ophthalmologists • Enable patients' self-monitoring and assist in doctors' clinical decision making when combining with smartphone App	• Provide full support that DL ANN can count as a powerful pattern recognition and classification tool and can be widely applied on tasks related to medical diagnosis on BCC	• The potential for promoting early detection and treatment of malignant eyelid tumors at screening stage • A roust performance in identifying malignant eyelid tumors from benign ones
Methods and details	CNN for distinguishing malignant tumors Class weight and different epochs of 60 and 80 were applied to train the models without underfitting	ANN and CNN for distinguishing BCC	Faster-RCNN (Faster Region based CNN) for localizing tumors and CNN for distinguishing malignant tumors Each algorithm was trained for 80 epochs
Outcomes (95% CI)			
Accuracy	0.889 (0.747–0.956)	Approximately 0.500–1.000	0.818 (0.773–0.862)
Sensitivity	0.933 (0.702–0.988)	–	0.915 (0.844–0.986)
Specificity	0.857 (0.654–0.950)	–	0.792 (0.739–0.845)
AUC	0.966 (0.850–0.993)	–	0.899 (0.854–0.934)
Dataset	339 photographic images from 255 patients, collected from Beijing Tongren Hospital, Beijing, China	143 photographic images collected from the Clinic of Ophthalmology of the University Hospital at Heraklion, Crete, Greece	1417 photographic images from 851 patients, collected from Ningbo Eye Hospital, Jiangdong Eye Hospital, and Zunyi First People's Hospital
Limitations	Cannot provide a specific subtype diagnosis based on images	May not be employed for other malignant eyelid tumors other than BCC	Cannot provide a specific subtype diagnosis based on images

DL deep learning, *ANN* artificial neural network, *CNN* convolutional neural network, *CI* confidential interval, *AUC* area under curve

information in ways in a fraction of time that the human brain cannot perform. Different diagnosis can be found for the same eyelid tumor, due to the dependence of patients' collaboration and clinician's experience. DL system enable automatic risk stratification of tumors and can be used as a triaging tool before clinician assessment, which may reduce unnecessary biopsies [50]. Its successful implementation could reduce human error, providing early diagnosis and consequent cost reduction in eyelid tumor treatments [51].

The majority of eyelid tumors represent as benign proliferative ones. Skin tumors involve the eyelid region in up to 10% of cases, with BCCs being the most prevalent among Chinese population [6]. Various factors affecting the incidence of benign and malignant eyelid tumors, including race, geography, and genetics. Suspicions of malignancy should be aroused when clinical signs arise, such as loss of lashes, central ulceration, infiltration, gradual enlargement, loss of sensation, induration, irregular or 'pearly' borders, destruction of eyelid structure, telangiectasia. Risk factors such as smoking, history of previous skin cancer, excessive sun exposure, previous radiation and immunosuppression also contribute to the occurrence of eyelid tumors [52–54].

In our study, 35.36% eyelid tumors are malignant tumors. According to epidemiological investigations, Sendul et al. [54] found that 87.1% were benign tumor, the left (12.9%) composed of malignant tumors. Xu et al. [55] revealed that 86.2% were benign tumors with data from a same medical center. The difference may be due to the different sample size and that difficult cases are more likely to visit our medical center. Supporting the literatures, malignant eyelid tumors developed at an elder age than the benign tumors [2, 54, 55]. Similar to some previous studies, we also observed a lower eyelid predominance for the malignant eyelid tumors [9, 56]. Prolonged exposure to sunlight seems to be an important predisposing factor of this predominance [57]. No preference of benign tumors in the upper eyelid was observed as described before [6, 53]. Of seven patients with bilateral eyelid tumors in our study, six of them were xanthoma. The similarities and differences in the incidence rates of benign tumors among studies describing eyelid tumors can mainly be attributed to racial and regional factors rather than data bias.

This application available to general public can be used for patient self-examination and in community outreach programs when applying on smartphones, and it also provides support for junior clinicians while documenting cases and following up in a better way. Thereby, it can help to reduce unnecessary biopsies, minimize over diagnosis and other potential harms associated with screening, as well as to improve clinician workload and timely access to specialist care for people requiring urgent attention.

Although the performance of DL system in date analysis is promising, and several studies have also reported CNN algorithms have surpassed the classification efficacy of physician, the real performance of DL system still remains unclear. Future work could focus on the differentiation on subtypes of eyelid tumors, assuring the best classification outcome. Rigorously tests should be performed before implementation, and monitored after the utilization of this technology.

There are some limitations in our study. First, in comparison with skin tumors, eyelashes, eyebrows, and the background color of pupil, sclera, and iris could be the possible confounding factors associated with bias. Second, the classification in this study included only benign and malignant tumors, rather than specific types of tumors. We intend to enrich our dataset of each subtype tumors in order to develop a system with

better outcome employed to make specific diagnosis. Third, lesions were identified only from 2-dimensional photographs without any additional clinical information. Combining available data with the algorithm for classification such as clinical images, dermoscopic images or pathological images was proved to have a higher accuracy than single CNN model.

Conclusions

This study proves DL system is a convenient way that can be used in the identification of benign and malignant tumors through common clinical images. Our system has achieved a medical application of AI, with a better performance than most ophthalmologists. Compared with related researches, our study avoids the object detection procedure and reached better classification performance, which is appropriate for clinical use. In the future, combining DL system with smartphone may further enable patients' self-monitoring for malignancy in eyelid tumors and assist in doctors' clinical decision making.

Abbreviations

DL	Deep learning
CNNs	Convolutional neural networks
AI	Artificial intelligence
GPU	Graphical processing unit
CI	Confidential interval
ROC curve	Receiver operatingcharacteristic curve
SD	Standard deviation
AUC	Area under the curve
PR curve	Precision recall curve
BCC	Basal cell carcinoma
Faster-RCNN	Faster region based CNN

Acknowledgements
Not applicable.

Author contributions
All authors had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: SH, LD, DL. Acquisition, analysis, or interpretation of data: SH, LD, KZ, ZN, XJ, HL, ZH, JD, YW, WW, DL. Drafting of the manuscript: SH, LD, DL. Critical revision of the manuscript for important intellectual content: All authors. Administrative, technical, or material support: KZ. All authors read and approved the final manuscript.

Funding
This study was supported by the National Natural Science Foundation of China (82071005); The Special Fund of the Pediatric Medical Coordinated Development Center of Beijing Hospitals Authority (XTCX201824).

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

Declarations

Ethics approval and consent to participate
The Ethics Committee of Beijing Tongren Hospital approved the study. Written informed consent was obtained from each subject.

Consent for publication
Not applicable.

Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 January 2022 Accepted: 5 June 2022
Published online: 22 June 2022

References

1. Cook BE, Bartley GB. Treatment options and future prospects for the management of eyelid malignancies: an evidence-based update. Ophthalmology. 2001;108:2088–98.
2. Depressed M, Uffer S. Clinicopathological features of eyelid skin tumors. A retrospective study of 5504 cases and review of literature. Am J Dermatopathol. 2009;31:256–62.
3. Welch RB, Duke JR. Lesions of the lids: a statistical note. Am J Ophthalmol. 1958;45:415–6.
4. Aurora AL, Blodi FC. Lesions of the eyelids: clinicopathological study. Surv Ophthalmol. 1970;15:94–104.
5. Tesluk GC. Eyelid lesions: incidence and comparison of benign and malignant lesions. Ann Ophthalmol. 1985;17:704–7.
6. Wang L, Shan Y, Dai X, You N, Shao J, Pan X, et al. Clinicopathological analysis of 5146 eyelid tumors and tumor-like lesions in an eye centre in South China, 2000–2018: a retrospective cohort study. BMJ Open. 2021;11:e041854.
7. Shields JA, Demirci H, Marr BP, Eagle RC, Shields CL. Sebaceous carcinoma of the ocular region: a review. Surv Ophthalmol. 2005;50:103–22.
8. Sun MT, Huang S, Huilgol SC, Selva D. Eyelid lesions in general practice. Aust J Gen Pract. 2019;48:509–14.
9. Yu SS, Zhao Y, Zhao H, Lin JY, Tang X. A retrospective study of 2228 cases with eyelid tumors. Int J Ophthalmol. 2018;11:1835–41.
10. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Farhan L, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data. 2021;8:53.
11. Lin C, Song X, Li L, Jiang M, Sun R, et al. Detection of active and inactive phases of thyroid-associated ophthalmopathy using deep convolutional neural network. BMC Ophthalmol. 2021;21:39.
12. Bi S, Chen R, Zhang K, Xiang Y, Wang R, Lin H, et al. Differentiate cavernous hemangioma from schwannoma with artificial intelligence (AI). Ann Transl Med. 2020;8:710.
13. Liang G, Fan W, Luo H, Zhu X. The emerging roles of artificial intelligence in cancer drug development and precision therapy. Biomed Pharmacother. 2020;128:110255.
14. Gunasekeran DV, Ting DSW, Tan GSW, Wong TY. Artificial intelligence for diabetic retinopathy screening, prediction and management. Curr Opin Ophthalmol. 2020;31:357–65.
15. Cui X, Wei R, Gong L, Qi R, Zhao Z, Chen H, et al. Assessing the effectiveness of artificial intelligence methods for melanoma: a retrospective review. J Am Acad Dermatol. 2019;81:1176–80.
16. Maron RC, Utikal JS, Hekler A, Hauschild A, Sattler E, Sondermann W, et al. Artificial Intelligence and its effect on dermatologists' accuracy in dermoscopic melanoma image classification: web-based survey study. J Med Internet Res. 2020;22:e18091.
17. Wang L, Ding L, Liu Z, Sun L, Chen L, Jia R, et al. Automated identification of malignancy in whole-slide pathological images: identification of eyelid malignant melanoma in gigapixel pathological slides using deep learning. Br J Ophthalmol. 2020;104:318–23.
18. Jiang YQ, Xiong JH, Li HY, Yang XH, Yu WT, GAO M, et al. Recognizing basal cell carcinoma on smartphone-captured digital histopathology images with a deep neural network. Br J Dermatol. 2020;182:754–62.
19. Hung JY, Perera C, Chen KW, Myung D, Chiu HK, Fuh CS, et al. A deep learning approach to identify blepharoptosis by convolutional neural networks. Int J Med Inform. 2021;148:104402.
20. Thomas PBM, Gunasekera CD, Kang S, Baltrusaitis T. An artificial intelligence approach to the assessment of abnormal lid position. Plast Reconstr Surg Glob Open. 2020;8:e3089.
21. Chen HC, Tzeng SS, Hsiao YC, Chen RF, Hung EC, Lee OK, et al. Smartphone-based artificial intelligence-assisted prediction for eyelid measurements: algorithm development and observational validation study. JMIR Mhealth Uhealth. 2021;9:e32444.
22. Lou L, Cao J, Wang Y, Gao Z, Jin K, Xu Z, et al. Deep learning-based image analysis for automated measurement of eyelid morphology before and after blepharoptosis surgery. Ann Med. 2021;53:2278–85.
23. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115–8.
24. Maintz L, Welchowski T, Herrmann N, Brauer J, Kläschen AS, Fimmers R, et al. Machine learning-based deep phenotyping of atopic dermatitis: severity-associated factors in adolescent and adult patients. JAMA Dermatol. 2021;157:1414–24.
25. Ahuja AA, Kohli P, Lomte S. Novel technique of smartphone-based high magnification imaging of the eyelid lesions. Indian J Ophthalmol. 2017;65:1015–16.
26. Zhang K, Li X, He L, Guo C, Yang Y, Dong Z, et al. A human-in-the-loop deep learning paradigm for synergic visual evaluation in children. Neural Netw. 2020;122:163–73.
27. Yang J, Zhang K, Fan H, Huang Z, Xiang Y, Yang J, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. Commun Biol. 2019;2:390.
28. Zhang K, Liu X, Jiang J, Li W, Wang S, Liu L, et al. Prediction of postoperative complications of pediatric cataract patients using data mining. J Transl Med. 2019;17:2.
29. Pan Q, Zhang K, He L, Dong Z, Zhang L, Wu X, et al. Automatically diagnosing disk bulge and disk herniation with lumbar magnetic resonance images by using deep convolutional neural networks: method development study. JMIR Med Inform. 2021;9:e14755.
30. Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, et al. An interpretable and expandable deep learning diagnostic system for multiple ocular diseases: qualitative study. J Med Internet Res. 2018;20:e11144.
31. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. Deep learning from "passive feeding" to "selective eating" of real-world data. NPJ Digit Med. 2020;3:143.

32. Deng J, Dong W, Socher R, Li L, Li K, Li F. Imagenet: a large-scale hierarchical image database. IEEE Conf Comput Vis Pattern Recognit. 2009. https://doi.org/10.1109/CVPR.2009.5206848.
33. Bottou L. Stochastic gradient descent tricks BT—neural networks: tricks of the trade. Berlin: Springer; 2012. p. 421–36.
34. Li Z, Guo C, Nie D, Lin D, Zhu Y, Chen C, et al. Deep learning for detecting retinal detachment and discerning macular status using ultra-widefield fundus images. Commun Biol. 2020;3:15.
35. Lord RK, Shah VA, San Filippo AN, Krishna R. Novel uses of smartphones in ophthalmology. Ophthalmology. 2010;117:1274–1274.e3.
36. Hogarty DT, Hogarty JP, Hewitt AW. Smartphone use in ophthalmology: what is their place in clinical practice? Surv Ophthalmol. 2020;65:250–62.
37. Chhablani J, Kaja S, Shah VA. Smartphones in ophthalmology. Indian J Ophthalmol. 2012;60:127–31.
38. Zvornicanin E, Zvornicanin J, Hadziefendic B. The use of smart phones in ophthalmology. Acta Inform Med. 2014;22:206–9.
39. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, et al. Deep neural networks are superior to dermatologists in melanoma image classification. Eur J Cancer. 2019;119:11–7.
40. Hekler A, Utikal JS, Enk AH, Solass W, Schmitt M, Klode J, et al. Deep learning outperformed 11 pathologists in the classification of histopathological melanoma images. Eur J Cancer. 2019;118:91–6.
41. Hekler A, Utikal JS, Enk AH, Berking C, Klode J, Schadendorf D, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. Eur J Cancer. 2019;115:79–83.
42. Sinha KR, Yeganeh A, Goldberg RA, Rootman DB. Assessing the accuracy of eyelid measurements utilizing the volk eye check system and clinical measurements. Ophthalmic Plast Reconstr Surg. 2018;34:346–50.
43. Godfrey KJ, Wilsen C, Satterfield K, Korn BS, Kikkawa DO. Analysis of spontaneous eyelid blink dynamics using a 240 frames per second smartphone camera. Ophthalmic Plast Reconstr Surg. 2019;35:503–5.
44. Pundlik S, Tomasi M, Liu R, Houston K, Luo G. Development and preliminary evaluation of a smartphone app for measuring eye alignment. Transl Vis Sci Technol. 2019;8:19.
45. Phanphruk W, Liu Y, Morley K, Gavin J, Shah AS, Hunter DG. Validation of StrabisPIX, a mobile application for home measurement of ocular alignment. Transl Vis Sci Technol. 2019;8:9.
46. Gupta R, Agrawal S, Srivastava RM, Singh V, Katiyar V. Smartphone photography for screening amblyogenic conditions in children. Indian J Ophthalmol. 2019;67:1560–3.
47. Arnold RW, O'Neil JW, Cooper KL, Silbert DI, Donahue SP. Evaluation of a smartphone photoscreening app to detect refractive amblyopia risk factors in children aged 1–6 years. Clin Ophthalmol. 2018;12:1533–7.
48. Adamopoulos A, Chatzopoulos EG, Anastassopoulos G, Detorakis E. Eyelid basal cell carcinoma classification using shallow and deep learning artificial neural networks. Evolving Systems. 2021;12: 583–90.
49. Li Z, Qiang W, Chen H, Pei M, Yu X, Wang L, et al. Artificial intelligence to detect malignant eyelid tumors from photographic images. NPJ Digit Med. 2022;5:23.
50. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human–computer collaboration for skin cancer recognition. Nat Med. 2020;26:1229–34.
51. Gui C, Chan V. Machine learning in medicine. Univ West Ont Med J. 2017;86:77–8.
52. Lear JT, Tan BB, Smith AG, Bowers W, Jones PW, Heagerty AH, et al. Risk factors for basal cell carcinoma in the UK: case-control study in 806 patients. J R Soc Med. 1997;90:371–4.
53. Leung C, Johnson D, Pang R, Kratky V. Identifying predictive morphologic features of malignancy in eyelid lesions. Can Fam Physician. 2015;61:e43-9.
54. Sendul SY, Akpolat C, Yilmaz Z, Eryilmaz OT, Guven D, Kabukcuoglu F. Clinical and pathological diagnosis and comparison of benign and malignant eyelid tumors. J Fr Ophtalmol. 2021;44:537–43.
55. Xu XL, Li B, Sun XL, Li LQ, Ren RJ, Gao F, et al. Eyelid neoplasms in the Beijing Tongren Eye Centre between 1997 and 2006. Ophthalmic Surg Lasers Imaging. 2008;39:367–72.
56. Burgic M, Iljazovic E, Vodencarevic AN, Burgic M, Rifatbegovic A, Mujkanovic A, et al. Clinical characteristics and outcome of malignant eyelid tumors: a five-year retrospective study. Med Arch. 2019;73:209–12.
57. Pieh S, Kuchar A, Novak P, Kunstfeld R, Nagel G, Steinkogler FJ. Long-term results after surgical basal cell carcinoma excision in the eyelid region. Br J Ophthalmol. 1999;83:85–8.

Publisher's Note