Deep Learning–Based Prediction of Refractive Error Using Photorefraction Images Captured by a Smartphone: Model Development and Validation Study

Jaehyeong Chun1*, BSc, MSc; Youngjun Kim2*, MD; Kyoung Yoon Shin2, MD; Sun Hyup Han2, MD; Sei Yeul Oh2, MD, PhD; Tae-Young Chung2, MD, PhD; Kyung-Ah Park2, MD, PhD; Dong Hui Lim2,3, MD, PhD

1Department of Industrial and System Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
2Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
3Department of Digital Health, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Republic of Korea
*these authors contributed equally

Abstract

Background: Accurately predicting refractive error in children is crucial for detecting amblyopia, which can lead to permanent visual impairment, but is potentially curable if detected early. Various tools have been adopted to more easily screen a large number of patients for amblyopia risk.

Objective: For efficient screening, easy access to screening tools and an accurate prediction algorithm are the most important factors. In this study, we developed an automated deep learning–based system to predict the range of refractive error in children (mean age 4.32 years, SD 1.87 years) using 305 eccentric photorefraction images captured with a smartphone.

Methods: Photorefraction images were divided into seven classes according to their spherical values as measured by cycloplegic refraction.

Results: The trained deep learning model had an overall accuracy of 81.6%, with the following accuracies for each refractive error class: 80.0% for ≤−5.0 diopters (D), 77.8% for >−5.0 D and ≤−3.0 D, 82.0% for >−3.0 D and ≤−0.5 D, 82.8% for >−0.5 D and <+0.5 D, 83.3% for ≥+0.5 D and <+3.0 D, 79.3% for ≥+3.0 D and <+5.0 D, and 75.0% for ≥+5.0 D. These results indicate that our deep learning–based system performed sufficiently accurately.

Conclusions: This study demonstrated the potential of precise smartphone-based prediction systems for refractive error using deep learning and further yielded a robust collection of pediatric photorefraction images.

Introduction

Amblyopia is the most common cause of permanent visual impairment in children, and its worldwide prevalence is estimated to be approximately 1.6%-5% [1,2]. Refractive error is one of the leading causes of pediatric amblyopia [3]. Early detection of refractive error in children plays an important role in visual prognosis [4,5], and therefore, early pediatric screening...
is recommended by the American Academy of Pediatrics, American Academy of Pediatric Ophthalmology and Strabismus (AAPOS), and European Strabismological Association and Societies [6,7].

Cycloplegic retinoscopic refraction is the standard technique for measuring refractive error. However, this method has some limitations. It is difficult to get young children to cooperate during the procedure, and advanced clinical ophthalmologic training is required to perform the test (user dependent) [2,8].

Previously, autorefractors were developed for faster and easier refraction in children. However, autorefraction presents several difficulties, including maintaining the proper position for testing and maintaining visual fixation on the target for a sufficient duration [9,10]. Photorefraction data can confirm the presence of myopia, hyperopia, astigmatism, and anisometropia by evaluating the refraction type and the position of eccentric crescent images on the pupil after projecting a light source onto the retina [11,12]. Photorefraction is simple and fast, making it convenient for use in children with poor cooperation ability, and it is suitable for screening large populations [13,14]. Several tools have been developed to meet the growing demand to perform photorefraction in clinical settings [2,15,16]. Easy availability of these tools and accurate prediction algorithms are the most important factors for ensuring efficient screening by photorefraction. Recently, deep learning algorithms have yielded innovative results in the field of medical imaging diagnostics [17]. In particular, deep convolutional neural networks [18] have been widely applied to extract essential features directly from images without human input. In ophthalmology, deep convolutional neural networks showed remarkable performance for detecting various diseases, including diabetic retinopathy [19-21], glaucoma [22,23], and retinopathy of prematurity [24]. Deep learning can also capture biological signs that are difficult for even human experts to detect, such as retinal findings from fundus images associated with cardiovascular risk [25]. However, little research has been done on the application of deep learning to refractive error prediction among children, using photorefraction images. A previous study attempted to predict the refractive error from retinal fundus images using deep learning [26], but the application was limited because the average participant age was 55 years and a specialized device was required to obtain the fundus images.

The purpose of this study was to develop an automated deep learning–based prediction system for refractive error using eccentric photorefraction images of pediatric patients captured by a smartphone. We trained our deep convolutional neural network with photorefraction images to identify various refractive error ranges. Thereafter, we comparatively evaluated its performance on our network with conventional cycloplegic retinoscopic refraction.

**Methods**

### Study Approval

This study was performed at a single center according to the tenets of the Declaration of Helsinki. The Institutional Review Board of Samsung Medical Center (Seoul, Republic of Korea) approved this study (SMC 2017-11-114).

### Participants

Patients aged 6 months to 8 years who visited the outpatient clinic for a routine ocular examination were requested to participate in this study. Written informed consent was provided by parents prior to participation. All screening tests were conducted at Samsung Medical Center between June and September 2018. The exclusion criteria were diseases that could affect light reflection, such as congenital cataracts and corneal opacity, diseases involving visual pathways or extraocular muscles, a medical history of previous ophthalmic surgery (eg, strabismus, congenital cataract, and congenital glaucoma), limited cycloplegia, and poor cooperation during study activities.

### Data Collection

A total of 305 photorefraction images (191 images from 101 girls and 114 images from 63 boys) were obtained (mean age 4.32 years, SD 1.87 years). All patients underwent a complete ophthalmologic examination, including visual acuity, motility evaluation, and anterior segment evaluation. Eccentric photorefraction images were obtained using a smartphone with a 16-megapixel camera (LG-X800K; LG Electronics Inc, Seoul, Korea) at a 1-meter distance from the front of the patient in a dark room (<15 lux). The smartphone was placed straight forward to the face of the children without angulation. All photorefraction images were acquired in the same setting (in a dark room and before the cycloplegic procedure). The smartphone’s built-in flash, present next to the camera lens, was used as the light source for eccentric photorefraction, wherein light was refracted and reached the retinal surface and was then magnified and reflected. When optimal reflection was achieved, a characteristic crescent-shaped reflection appeared in the eye. A photograph of the crescent reflection was captured through LED control [13]. After acquisition of photorefraction images, 0.5% tropicamide and 0.5% phenylephrine (Tropherine; Hanmi Pharm, Seoul, Korea) were administered three times at 5-minute intervals. Cycloplegic retinoscopy and fundus examination to obtain spherical, cylindrical, cylindrical axis, and spherical equivalent values were performed between 30 and 60 minutes following the first instillation of cycloplegics, when the pupillary light reflex was eliminated. Both photorefraction and cycloplegic refraction were performed sequentially, and the ground truth for images acquired by photorefraction was labelled according to the values of cycloplegic refraction. Consequently, the result of cycloplegic refraction was provided as the ground truth for machine learning of photorefraction images.

The acquired eccentric photorefraction images were divided into the following seven classes according to the spherical values measured by cycloplegic refraction: \( \leq -5.0 \text{ diopter (D)} \), \( >-5.0 \text{ D and } \leq -3.0 \text{ D} \), \( >-3.0 \text{ D and } \leq -0.5 \text{ D} \), \( >-0.5 \text{ D and } \leq 0.5 \text{ D} \), \( >0.5 \text{ D and } \leq 3.0 \text{ D} \), \( >3.0 \text{ D and } \leq 5.0 \text{ D} \), and \( >5.0 \text{ D} \). The cutoff values of the seven classes for refractive errors were determined clinically. Among myopic refraction (minus values), \( -5.0 \text{ D}, -3.0 \text{ D}, \) and \( -0.5 \text{ D} \) were considered as thresholds of high, moderate, and mild myopia, respectively. In other words, refractive errors \( \leq -5.0 \text{ D} \) indicated high myopia, refractive errors \( -5.0 \text{ D}, \) and \( -0.5 \text{ D} \) were considered as thresholds of high, moderate, and mild myopia, respectively. In other words, refractive errors \( \leq -5.0 \text{ D} \) indicated high myopia, refractive errors \( -5.0 \text{ D}, \) and \( -0.5 \text{ D} \) were considered as thresholds of high, moderate, and mild myopia, respectively. In other words, refractive errors \( \leq -5.0 \text{ D} \) indicated high myopia, refractive errors \( -5.0 \text{ D}, \) and \( -0.5 \text{ D} \) were considered as thresholds of high, moderate, and mild myopia, respectively. In other words, refractive errors \( \leq -5.0 \text{ D} \) indicated high myopia, refractive errors \( -5.0 \text{ D}, \) and \( -0.5 \text{ D} \) were considered as thresholds of high, moderate, and mild myopia, respectively.
errors $> -5.0$ D and $\leq -3.0$ D indicated moderate myopia, and refractive errors $> -3.0$ D and $\leq -0.5$ D indicated mild myopia. Similarly, $+0.5$ D, $+3.0$ D, and $+5.0$ D were thresholds of mild, moderate, and high hyperopia, respectively, among plus values.

**Image Data Preparation for Training, Validation, and Testing**

Photorefraction images were processed for training our deep convolutional neural network. Initially, the images were cropped to capture the pupil. The images were resized to 224×224 pixels, and the pixel values were scaled from 0 to 1. To overcome an overfitting issue caused by an insufficiently sized training dataset, data augmentation was performed by altering brightness, saturation, hue, and contrast; adding Gaussian noise; and blurring images using Gaussian kernels. Thereafter, the image pixel values were normalized by subtracting the mean and dividing by the SD to ensure that each image had a similar data distribution and would converge faster during the training procedure.

For training, validation, and testing, we used the five-fold cross-validation approach to build a reliable deep learning model with a limited dataset. Initially, all the data were subdivided into five equal-sized folds with the same proportion of different classes in each fold. Four of the five folds were for training and validation (3.5 folds for training and 0.5 folds for validation), and one fold was for testing. After five repetitions of this process, we were able to evaluate the performance of the entire dataset because the test folds were independent of each other, and we confirmed the stability of our model for the entire dataset using the confusion matrix.

**Deep Convolutional Neural Network and Training**

We used a deep convolutional neural network to classify photorefraction images into the most probable class of refractive error. Among the various types of convolutional neural networks, we developed Residual Network (ResNet-18) [27] to avoid problems that occur when deep neural network depth increases, such as vanishing or exploding gradients and accuracy degradation. Residual Network addresses these issues using identity mapping with shortcut connections. The shortcut connections allow networks to skip over layers and also enable speed training. Figure 1 illustrates the overall structure of the deep learning approach we propose in this work. The basic block consists of two 3×3 convolutional layers, and the shortcut connection enables the network to learn identity mapping (Figure 2).

Because we did not have a sufficiently large training dataset, we performed transfer learning to capture low-level features, such as edge and color, without wasting image data [28]. Accordingly, pretrained parameters of Residual Network on the ImageNet [29] datasets were reused as starting points for our model. The pretrained Residual Network was available on Pytorch [30]. We then replaced the last fully connected layer to output seven predicted probabilities for each refractive error class ($\leq -5.0$ D, $> -5.0$ D and $\leq -3.0$ D, $> -3.0$ D and $\leq -0.5$ D, $> -0.5$ D and $\leq +0.5$ D, $> +0.5$ D and $\leq +3.0$ D, $> +3.0$ D and $\leq +5.0$ D, and $> +5.0$ D). During the training process, the first layer was frozen, and the learning rates for the subsequent layers were increased from $1e^{-10}$ to $1e^{-5}$ to finetune our network for preventing an overfitting issue. Furthermore, we designed the loss function as a weighted sum of cross-entropy by class, wherein the weight for each class was the reciprocal of the proportion of that class’s images in the training dataset. This technique was useful to achieve balanced accuracy for all classes, despite having an imbalanced training dataset. For convergence of network training, the learning rate was decayed by a factor of 0.95 every 10 epoch, and we trained the parameters of networks using stochastic gradient descent [31] with 0.9 momentum. We set the maximum training epoch as 500 and the minibatch size of training images as 16. All codes were implemented using Pytorch 1.2.0 [30]. Details of the network structure are shown in Table 1.

![Figure 1](https://medinform.jmir.org/2020/5/e16225)
Figure 2. Structure of the basic block and the shortcut connection. The basic block consists of two 3x3 convolutional layers, two Batch Normalization layers, and a Rectified Linear Unit (ReLU) activation function. The shortcut connection adds the input vector of the basic block to the output of the basic block.
Table 1. Configuration of the deep convolutional network.

| Layer type, feature map | Filters | Kernel | Stride | Padding | Learning rate |
|-------------------------|---------|--------|--------|---------|--------------|
| **Input**               |         |        |        |         | 0.0 (freeze) |
| 224x224x3               | a       |        |        |         |              |
| **Convolutional**       |         |        |        |         | 0.0 (freeze) |
| 112x112x64              | 64      | 7x7x3  | 2      | 3       |              |
| **Batch normalization** |         |        |        |         | 0.0 (freeze) |
| 112x112x64              |         |        |        |         |              |
| **Max pooling**         |         |        |        |         | 0.0 (freeze) |
| 56x56x64                | 1       | 3x3    | 2      | 1       |              |
| **Layer 1**             |         |        |        |         |              |
| Basic block 1-1         |         |        |        |         | 0.0 (freeze) |
| 56x56x64                | 64      | 3x3x64 | 1      | 1       |              |
| Basic block 1-2         |         |        |        |         | 0.0 (freeze) |
| 56x56x64                | 64      | 3x3x64 | 1      | 1       |              |
| **Layer 2**             |         |        |        |         |              |
| Basic block 2-1         |         |        |        |         | 1e-10        |
| 28x28x128               | 128     | 3x3x64 | 2      | 1       |              |
| Basic block 2-2         |         |        |        |         | 1e-10        |
| 28x28x128               | 128     | 3x3x128| 1      | 1       |              |
| 28x28x128               | 128     | 1x1x64 | 2      | 0       |              |
| **Layer 3**             |         |        |        |         |              |
| Basic block 3-1         |         |        |        |         | 1e-8         |
| 14x14x256               | 256     | 3x3x128| 2      | 1       |              |
| Basic block 3-2         |         |        |        |         | 1e-8         |
| 14x14x256               | 256     | 3x3x256| 1      | 1       |              |
| Basic block 3-2         |         |        |        |         | 1e-8         |
| 14x14x256               | 256     | 1x1x128| 2      | 0       |              |
| **Layer 4**             |         |        |        |         |              |
| Basic block 4-1         |         |        |        |         | 1e-6         |
| 7x7x512                 | 512     | 3x3x256| 2      | 1       |              |
| Basic block 4-2         |         |        |        |         | 1e-6         |
| 7x7x512                 | 512     | 3x3x512| 1      | 1       |              |
| **Average pooling**     |         |        |        |         |              |
| 1x1x512                 | 1       | 7x7    | 7      | 0       |              |
| **Fully connected layer** |     |        |        |         |              |
Results

Image Dataset Demographics

A total of 305 photorefraction images from 191 girls and 114 boys were acquired. The mean age was 4.32 years (SD 1.87 years), and the median age was 4 years (range 0-8 years). The mean spherical equivalent was 0.13 D (SD 2.27 D; range −5.50 to 6.75 D), and the mean astigmatism was −1.50 D (SD 1.38 D; range −6.50 to 0 D), according to cycloplegic refraction.

According to cycloplegic refraction results, 25 photorefraction images had a refractive error $\leq$ −5.0 D, 18 had an error >−5.0 D and $\leq$ −3.0 D, 50 had an error >−3.0 D and $\leq$ −0.5 D, 84 had an error >−0.5 D and $<$+0.5 D, 87 had an error $\geq$+0.5 D and $<$+3.0 D, 29 had an error $\geq$+3.0 D and $<$+5.0 D, and 12 had an error $\geq$+5.0 D. Table 2 summarizes patient demographics in detail, and examples of photorefraction images according to the refractive error class are shown in Figure 3.

Figure 3. Examples of photorefraction images from the seven different refractor error classes. A bright crescent appears in the pupillary reflex, and its size and shape indicate the diopter (D) value.

Table 2. Dataset participant demographics.

| Characteristic | Value |
|----------------|-------|
| Total images, n | 305   |
| Girls, n (%)    | 191 (62.6) |
| Age, mean (SD)  | 4.32 (1.87) |

$^a$Not applicable.
Performance of the Proposed Deep Convolutional Neural Network

We used five-fold cross-validation to evaluate our network’s performance. Training, validation, and testing were independently iterated five times. In each iteration, there were 213 training images, 31 validation images, and 61 testing images. We chose the network with the highest validation accuracy when loss of training was saturated. Thereafter, we measured the classification accuracy of the network in the test fold. All five networks, which were established in the training phase, had an accuracy of more than 80% for each validation set. Similarly, the performances of the five testing folds were 83.6%, 80.3%, 82.0%, 78.7%, and 83.6% (Table 3).

Table 3. Results for five-fold cross-validation.

| Iterationa | Validation accuracy (%) (N=31) | Test accuracy (%) (N=61) |
|------------|--------------------------------|-------------------------|
| First iteration | 87.1                          | 83.6                    |
| Second iteration | 80.6                          | 80.3                    |
| Third iteration | 80.6                          | 82.0                    |
| Fourth iteration | 83.9                          | 78.7                    |
| Fifth iteration | 83.9                          | 83.6                    |
| Average        | 83.2                          | 81.6                    |

*aIn each iteration, our network was trained using the rest of the validation and test dataset (213 training images).

In the five-fold test, our network had the following accuracies: 80.0% for class $\leq -5.0$ D, 77.8% for class $> -5.0$ D and $\leq -3.0$ D, 82.0% for class $> -3.0$ D and $\leq -0.5$ D, 83.3% for class $> -0.5$ D and $+0.5$ D, 82.8% for class $+0.5$ D and $-3.0$ D and $+3.0$ D, 79.3% for class $> +3.0$ D and $+5.0$ D, and 75% for class $\geq +5.0$ D (Table 4). Despite the imbalanced dataset, our model achieved consistent performance for all classes.

Table 4. Performance of our deep convolutional neural network with the overall test dataset.

| Class                  | Number | Accuracy (%) |
|------------------------|--------|--------------|
| $\leq -5.0$ D          | 25     | 80.0         |
| $>-5.0$ D and $\leq -3.0$ D | 18     | 77.8         |
| $>-3.0$ D and $\leq -0.5$ D | 50     | 82.0         |
| $>-0.5$ D and $+0.5$ D | 84     | 83.3         |
| $\geq +0.5$ D and $+3.0$ D | 87     | 82.8         |
| $\geq +3.0$ D and $+5.0$ D | 29     | 79.3         |
| $\geq +5.0$ D         | 12     | 75.0         |
| Total                  | 305    | 81.6         |

*aD: diopter.

For performance comparison, we developed the following five baseline models and calculated the performances: (1) pretrained VGG-11 [32]; (2) pretrained squeezeNet [33]; (3) Support Vector Machine (SVM) [34]; (4) Random Forest [35]; and (5) simple convolutional neural network. VGG-11 and squeezeNet were pretrained on the ImageNet [29] datasets, and their parameters were frozen, except the last four convolutional layers during training. Moreover, we designed the following two traditional machine learning approaches: SVM and Random Forest. SVM has a radial basis function kernel, 1.0 regularization parameter, and three degrees of the kernel function. Random Forests has 500 trees, the Gini index criterion, and two samples required to split an internal node. Lastly, the simple convolutional neural network has three convolutional layers with six kernels ($8\times8$ size, two strides), 16 kernels ($5\times5$ size, two strides), and 24 kernels ($3\times3$ size, one stride), respectively; a max-pooling layer ($2\times2$ size and two strides) after each convolutional layer; and three fully connected layers with 120, 84, and 7 hidden units, respectively, in a row at the end of the network. We evaluated the performances of the five baseline models using five-fold cross-validation, and the results of performance comparison are shown in Table 6. We confirmed that the proposed deep convolutional neural network outperformed all baseline models.
Table 5. Confusion matrix for refractive error classification of our deep convolutional neural network.

| Ground truth | Predictive value | Accuracy (%) |
|--------------|------------------|--------------|
|              | ≤−5.0 D          |              |
|              | >−5.0 D and ≤−3.0 D |              |
| ≤−5.0 D      | 20^b             |              |
| >−5.0 D and ≤−3.0 D | 3 |              |
| >−3.0 D and ≤−0.5 D | 2 |              |
| >−0.5 D and ≤+0.5 D | 0 |              |
| ≥+0.5 D and ≤+3.0 D | 0 |              |
| ≥+3.0 D and ≤+5.0 D | 0 |              |
| ≥+5.0 D      | 80.00000        |              |
| ≥−5.0 D      | 77.80010        |              |
| >−5.0 D and ≤−3.0 D | 82.000044        |
| >−3.0 D and ≤−0.5 D | 83.301870        |
| >−0.5 D and ≤+0.5 D | 82.80472        |
| ≥+0.5 D and ≤+3.0 D | 79.3123        |
| ≥+3.0 D and ≤+5.0 D | 75.09        |
| ≥+5.0 D      | 81.6——————— |              |
| Overall accuracy (%) | — — — — — — — — | 81.6       |

^aD: diopter.
^bNumber of correct predictions of our deep convolutional neural network.
^cNot applicable.

Table 6. Performance comparison of the proposed model and baseline models.

| Model                                      | Accuracy (%) |
|--------------------------------------------|--------------|
| The proposed deep convolutional neural network | 81.6         |
| Pretrained VGG-11                          | 70.8         |
| Pretrained SqueezeNet                      | 77.4         |
| Support Vector Machine                     | 65.2         |
| Random Forest                              | 62.9         |
| Simple convolutional neural network        | 70.8         |

Additionally, we produced heatmaps using gradient-weighted class activation mapping (Grad-CAM) [36] to provide visual explanations for each screening decision. This technique is crucial for interpreting network output and validating whether the network learned meaningful features. The activation map visualizes where the network considered the critical locations to be within photorefraction images for detecting refractive error. Figure 4 shows the activated regions from four layers in the photorefraction images. Notably, we observed the heatmap from the fourth layer, which captured important features for classifying refractive error, particularly the region of the crescent in the pupil.
Figure 4. Examples of photorefraction images correctly classified by deep neural networks. (A), (B), (C) were identified as ≥+0.5 D and <+3.0 D, ≥+3.0 D and <+5.0 D, and ≥+5.0 D, respectively. The first layers captured low-level features, such as edge and color. With deeper layers, the network focused on high-level features that were regarded as important aspects for classification.

Discussion

The primary purpose of refractive error screening is the early detection of a refractive error to allow interventions that can reduce the risk of amblyopia. Early detection and treatment of refractive error can lead to better visual outcomes and reduce the prevalence and severity of amblyopia in children [4,37]. The cycloplegic refraction test has been an essential tool to accurately measure refractive error, because pediatric patients are more accommodating than adults [38]. However, young children tend not to cooperate well during the refraction test, and the test requires a skilled ophthalmic practitioner [2,8]. Additionally, the eye drops used during cycloplegia can cause side effects, such as flushing, fever, drowsiness, and red eye [39]. For these reasons, cycloplegic refraction is not suitable for large screening of refractive error and amblyopia [12]. Currently, smartphones are ubiquitous devices that allow physicians and other related medical professionals to overcome common diagnostic barriers in many clinical settings [40]. A photorefraction screening test using a smartphone is an easy and effective way to screen most young children. The photorefractive method is simple and takes no longer than a second to test both eyes simultaneously. The test requires minimal space (just a meter of distance between the subject and the testing device) and removes the need for cycloplegia, thereby greatly reducing side effects and testing time. Moreover, it does not require expert knowledge or experience to perform [6]. These advantages make the photorefractive method ideal for measuring refractive error, especially for poorly cooperative young children.

Several studies have compared the accuracy of photoscreeners for detecting various amblyopia risk factors [40-42]. One study evaluated a smartphone photoscreening application (GoCheckKids) and reported 76% sensitivity and 67.2% specificity [15] for detecting amblyopia risk factors using the 2013 AAPOS guidelines. Because we evaluated the accuracy of predicting refractive errors and not amblyopia risk factors, we were limited in our ability to directly compare the performance of our method against that of GoCheckKids. Instead, our deep convolutional neural network achieved satisfactory accuracy for predicting categories of refractive error using only a small image dataset. The results showed the potential for developing precise smartphone-based prediction systems for refractive error using deep learning. With further collection of pediatric photorefraction image data, more precise prediction of refractive error and effective detection of amblyopia would be possible.

This study compared refractive error estimation with precycloplegic photorefraction images and cycloplegic refraction. The results showed consistent measurements between the two methods. Dubious results regarding estimation of refractive error using photorefractors have been uncovered by previous studies [12,14,42]. Erdurmus et al reported that noncycloplegic photorefraction (Plusoptix CR03; PlusoptiX GMBH, Nurnberg, Germany) tended to overestimate negative refraction in children, resulting in overdiagnosis of myopia (−0.70 D) [12]. Lim et al reported similar results and showed that refractive error measured by a photorefractor without cycloplegia (Plusoptix S09; PlusoptiX GmbH) tended to be more myopic compared with cycloplegic refractive error [42]. On the other hand, Schimizek et al claimed that noncycloplegic refraction using a photorefractometer (Power Refractor; Plusoptix GmbH) resulted in underestimation of spherical equivalents owing to uncontrolled accommodation [14]. Another study showed that cycloplegic refraction results and photorefractor Plusoptix S08 (Plusoptix GmbH, Nurnberg, Germany) results were similar [2]. In this study, photorefraction results without cycloplegia showed reasonable agreement with cycloplegic refraction, suggesting that our deep learning–based system achieved considerably accurate performance under noncycloplegic conditions.
This study has several limitations. First, manifest refraction was not performed in all subjects. Since photorefractive refraction tests were performed without the use of a cycloplegic agent, useful information might have been obtained if the number of manifest refraction results without cycloplegia were enough to compare with photorefraction data in the same patient. Second, the number of photorefraction images was relatively small and the model could only predict a range of refractive errors (not a specific value). Third, all children involved in the study were Korean. Thus, a trained model using the eyes of Korean children may not be applicable to the eyes of pediatric patients having different ethnicities [43,44]. Future studies with more patients of multiple ethnicities and a greater range of refractive errors would be beneficial for providing a more precise clinical perspective.

In conclusion, this study showed that our deep learning–based system successfully yielded accurate and precise refractive measurements. This further demonstrates the potential for developing simplified smartphone-based prediction systems for refractive error using deep learning with large-scale collection of pediatric photorefraction images from patients with various ages and refractive errors.

Acknowledgments
This research was supported by a National Research Foundation of Korea grant funded by the Government of Korea’s Ministry of Education (NRF-2018R1D1A1A02045884; Seoul, Korea), which was received by Dong Hui Lim, and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI19C0577), which was received by Dong Hui Lim.

Authors’ Contributions
DHL designed the study. JC, YK, KYS, DHL, and K-AP analyzed and interpreted the clinical data. JC and YK wrote the submitted manuscript draft. TYC, SYO, SHH, DHL, and KAP reviewed the design, the results, and the submitted draft. JC and YK contributed equally to the work as co-first authors. DHL and KAP are the corresponding authors for this study. All authors read and approved the final manuscript.

Conflicts of Interest
None declared.

References
1. Simons K. Amblyopia characterization, treatment, and prophylaxis. Surv Ophthalmol 2005 Mar;50(2):123-166. [doi: 10.1016/j.survophthal.2004.12.005] [Medline: 15749306]
2. Demirci G, Arslan B, Özstütçü M, Eliyaçık M, Gulkilik G. Comparison of photorefraction, autorefractometry and retinoscopy in children. Int Ophthalmol 2014 Aug 10;34(4):739-746. [doi: 10.1007/s10792-013-9864-x] [Medline: 24114503]
3. Miller JM, Dobson V, Harvey EM, Sherrill DL. Cost-efficient vision screening for astigmatism in native american preschool children. Invest Ophthalmol Vis Sci 2003 Sep 01;44(9):3756-3763. [doi: 10.1167/iovs.02-0970] [Medline: 12939288]
4. Cotter SA, Pediatric Eye Disease Investigator Group, Edwards AR, Wallace DK, Beck RW, Arnold RW, et al. Treatment of anisometropic amblyopia in children with refractive correction. Ophthalmology 2006 Jun;113(6):895-903 [FREE Full text] [doi: 10.1016/j.ophtha.2006.01.068] [Medline: 16751032]
5. U.S. Preventive Services Task Force. Screening for visual impairment in children younger than age 5 years: recommendation statement. Ann Fam Med 2004 May 01;2(3):263-266 [FREE Full text] [doi: 10.1370/afm.193] [Medline: 15209205]
6. Schimitzek T, Haase W. Efficiency of a video-autorefractometer used as a screening device for amblyogenic factors. Graefes Arch Clin Exp Ophthalmol 2002 Sep 27;240(9):710-716. [doi: 10.1007/s00417-002-0524-5] [Medline: 12271366]
7. American Academy of Ophthalmology Pediatric Ophthalmology/Strabismus Panel. Preferred Practice Pattern Guidelines. Pediatric Eye Evaluations. American Academy of Ophthalmology 2007.
8. Safir A. Retinoscopy. Int Ophthalmol Clin 1971;11(1):115-129. [doi: 10.1097/000004397-19711110-00008] [Medline: 5129703]
9. Prabakaran S, Dirani M, Chia A, Gazzard G, Fan Q, Leo SW, et al. Cycloplegic refraction in preschool children: comparisons between the hand-held autorefractor, table-mounted autorefractor and retinoscopy. Ophthalmic Physiol Opt 2009 Jul;29(4):422-426. [doi: 10.1111/j.1475-1313.2008.00616.x] [Medline: 19523087]
10. La TY, Oh JR. Reliability of Refractive Measurement by Hand-held Autorefractor. J Korean Ophthalmol Soc 2002:2241-2245.
11. Donauhe SP, Arnold RW, Ruben JB. Preschool vision screening: what should we be detecting and how should we report it? Uniform guidelines for reporting results of preschool vision screening studies. Journal of American Association for Pediatric Ophthalmology and Strabismus 2003 Oct;7(5):314-316. [doi: 10.1016/s1091-8531(03)00182-4]
12. Erdurmus M, Yagci R, Karadag R, Durmus M. A comparison of photorefraction and retinoscopy in children. J AAPOS 2007 Dec;11(6):606-611. [doi: 10.1016/j.aaapos.2007.04.006] [Medline: 17588794]
13. Cole TD. Multimeridian Photorefraction: A technique for the detection of visual defects in infants and preverbal children. Johns Hopkins APL Technical Digest 1991:166-175.
37. Scheiman MM, Hertle RW, Beck RW, Edwards AR, Birch E, Cotter SA, Pediatric Eye Disease Investigator Group. Randomized trial of treatment of amblyopia in children aged 7 to 17 years. Arch Ophthalmol 2005 Apr 01;123(4):437-447. [doi: 10.1001/archopht.123.4.437] [Medline: 15824215]

38. Morgan IG, Iribarren R, Fotouhi A, Grzybowski A. Cycloplegic refraction is the gold standard for epidemiological studies. Acta Ophthalmol 2015 Sep 18;93(6):581-585 [FREE Full text] [doi: 10.1111/aos.12642] [Medline: 25597549]

39. Wakayama A, Nishina S, Miki A, Utsumi T, Sugawara J, Hayashi T, et al. Incidence of side effects of topical atropine sulfate and cyclopentolate hydrochloride for cycloplegia in Japanese children: a multicenter study. Jpn J Ophthalmol 2018 Sep 25;62(5):531-536. [doi: 10.1007/s10384-018-0612-7] [Medline: 30046935]

40. 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. OPTH 2018 Aug;12:1533-1537. [doi: 10.2147/opth.s171935]

41. Arnold RW, Armitage MD. Performance of four new photoscreeners on pediatric patients with high risk amblyopia. J Pediatr Ophthalmol Strabismus 2014 Jan 03;51(1):46-52. [doi: 10.3928/01913913-20131223-02] [Medline: 24369683]

42. Lim JH, Bae GH, Shin SJ. Reliability and Usefulness of Refractive Measurements by PlusoptiX S09 in Children. J Korean Ophthalmol Soc 2014;55(7):1071. [doi: 10.3341/jkos.2014.55.7.1071]

43. Sravani NG, Nilagiri VK, Bharadwaj SR. Photorefraction estimates of refractive power varies with the ethnic origin of human eyes. Sci Rep 2015 Jan 23;5(1):7976 [FREE Full text] [doi: 10.1038/srep07976] [Medline: 25613165]

44. Bharadwaj SR, Sravani NG, Little J, Narasaiyah A, Wong Y, Woodburn R, et al. Empirical variability in the calibration of slope-based eccentric photorefraction. J Opt Soc Am A 2013 Apr 19;30(5):923. [doi: 10.1364/josaa.30.000923]

Abbreviations

AAPOS: American Academy of Pediatric Ophthalmology and Strabismus

SVM: Support Vector Machine

©Jaehyeong Chun, Youngjun Kim, Kyoung Yoon Shin, Sun Hyup Han, Sei Yeul Oh, Tae-Young Chung, Kyung-Ah Park, Dong Hui Lim. Originally published in JMIR Medical Informatics (http://medinform.jmir.org), 05.05.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on http://medinform.jmir.org/, as well as this copyright and license information must be included.