

# BMJ Open Artificial intelligence method to classify ophthalmic emergency severity based on symptoms: a validation study

Hyunmin Ahn 

**To cite:** Ahn H. Artificial intelligence method to classify ophthalmic emergency severity based on symptoms: a validation study. *BMJ Open* 2020;**10**:e037161. doi:10.1136/bmjopen-2020-037161

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-037161>).

Received 27 January 2020

Revised 01 May 2020

Accepted 29 May 2020

## ABSTRACT

**Objectives** We investigated the usefulness of machine learning artificial intelligence (AI) in classifying the severity of ophthalmic emergency for timely hospital visits.

**Study design** This retrospective study analysed the patients who first visited the Armed Forces Daegu Hospital between May and December 2019. General patient information, events and symptoms were input variables. Events, symptoms, diagnoses and treatments were output variables. The output variables were classified into four classes (red, orange, yellow and green, indicating immediate to no emergency cases). About 200 cases of the class-balanced validation data set were randomly selected before all training procedures. An ensemble AI model using combinations of fully connected neural networks with the synthetic minority oversampling technique algorithm was adopted.

**Participants** A total of 1681 patients were included.

**Major outcomes** Model performance was evaluated using accuracy, precision, recall and F1 scores.

**Results** The accuracy of the model was 99.05%. The precision of each class (red, orange, yellow and green) was 100%, 98.10%, 92.73% and 100%. The recalls of each class were 100%, 100%, 98.08% and 95.33%. The F1 scores of each class were 100%, 99.04%, 95.33% and 96.00%.

**Conclusions** We provided support for an AI method to classify ophthalmic emergency severity based on symptoms.

## INTRODUCTION

South Korea has a population of about 50 million and is considered to have an advanced economy. In 2018, the gross domestic product per capita was about 35 000 dollars.<sup>1</sup> The key feature of the Korean medical system is the National Health Insurance System, and virtually all citizens have joined this insurance system. In 2016, the performance of South Korea in the healthcare access and quality (HAQ) index was ranked 25th out of 195 countries. Aside from tuberculosis, stroke and non-melanoma skin cancer, most of the diseases received very high HAQ index scores, with an average of 90 out of 100. Except for Japan (ranked 12th),

## Strengths and limitations of this study

- This artificial intelligence with various available mathematical and technical methods was a model to improve performance by overcoming the imbalance training and validation sample.
- It included a bigger data set than previous studies and reflects more detailed symptom presentations.
- The model was trained to reflect the final diagnoses and treatments, not just the known emergency symptoms, and the model training was intended to reflect both symptoms and also diagnoses and treatments.
- The utilisation of the AI model could be further improved if more various and bigger data set was obtained.

South Korea ranked second highest among Asian countries.<sup>2</sup>

To our knowledge, there has been no study or report on HAQ index scores for eye diseases in South Korea. According to a statistical report by Statistics Korea (KOSTAT, the national statistical office), there are 4.1 eye clinics per 100 000 people, and eye clinics are mainly concentrated around the Seoul capital area and provincial metropolitan cities (1088 of 2057). Unfortunately, 20 of 185 cities are without an ophthalmologist.<sup>3</sup> Also, there is no specific study on the Korean population regarding accessibility, management quality and prognostic differences in ophthalmic diseases. However, medical access and patient quality management for other diseases, such as acute myocardial infarction and cancer, have been studied.<sup>4,5</sup> The lack of ophthalmologists may cause difficulty in the management of eye diseases in some regions. In addition, ophthalmologic diagnoses and treatments by non-ophthalmologists in the primary healthcare system are very limited. The lack of an ophthalmologic educational programme in the medical curriculum and hospital training for general practitioners and the specificity of diagnoses and treatment of eye diseases underlies these problems.<sup>6,7</sup>



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Ophthalmology, Armed Forces Daegu Hospital, Daegu, Korea (the Republic of)

### Correspondence to

Hyunmin Ahn;  
overhyun31@gmail.com

Although South Korea has good primary HAQ across the board, there are blind spots, such as missing the proper management time and opportunity in eye healthcare for patients.

Artificial intelligence (AI) is used in various fields, including medical, and ophthalmology is no exception. Especially, research on fundus image analysis using AI has been actively conducted.<sup>8,9</sup> Considering technological developments of AI in other fields, such as autonomous vehicles or climate forecasting, the applicability of AI in medical, in particular, ophthalmologic fields can be more versatile.<sup>10,11</sup> One of the primary applications of AI within emergency medicine is in triage.<sup>12</sup> There is possibility that AI systems could improve the emergency decision-making process in triage by supporting patient flow, wait times, resource utilisation and risk stratification.<sup>12</sup> However, to our knowledge, there is no available AI for ophthalmologic emergency but only some symptom checkers with poor accuracy are utilised.<sup>13-16</sup>

This study was aimed to determine the possibility for AI to guide patients for eye treatment, who do not know the urgent severity of the symptoms, due to lack of awareness of eye diseases, to an ophthalmologist. Specifically, we focused on whether AI can distinguish actual urgent patients.

## METHODS

### Study population and data set

All 1681 patients who first visited the Armed Forces Daegu Hospital, in Daegu, South Korea between June and December 2019 were retrospectively analysed. Prior to any data processing sequence, patient data were randomly divided into two data sets, that was, the training set and validation set. The target ratio of the training set and validation set was about 7 to 1, and about 200 cases were required to be included in a validation set. The validation set consisted of about 50 cases in each of the four emergency code classes (more details below), by random extraction to form a balanced validation data set.

### Emergency severity classification

Emergency severity was classified into four classes (red, orange, yellow and green classes, indicating immediate to no emergency cases, respectively) according to the articles with modification based on events, symptoms, diagnoses and treatments.<sup>17,18</sup> Two labelling methods were used. In the first labelling method (ECode I), only diagnoses and treatments were used. In the second labelling method (ECode II), events, symptoms, diagnoses and treatments were used. The corresponding criteria for classes were as follows (table 1).

**Table 1** Emergency severity classification

	Red (immediate)	Orange (within 24 hours)	Yellow (within 7 days)	Green (not emergency)
Diagnosis	Acute glaucoma Corneal laceration CRAO, CRVO<24 hour Giant cell arteritis (visual disturbance) Globe perforation Iris prolapse Intraocular FB Orbital cellulitis third CN palsy (pupil involve)	Corneal abrasion Corneal FB Corneal ulcer Corneal opacities (with pain) FB in conjunctival sac Hyphoema Iridocyclitis Orbital wall fracture Retinal detachment Retinal tear Vitreous haemorrhage	AMD, CSCR, or other retinal disorders (with fluid collection) Zoster ophthalmicus Episcleritis (not manage) Scleritis PVD Bell's palsy Optic neuritis Orbital cellulitis (preseptal) Severe infectious conjunctivitis Vein occlusions Proliferative DR	Allergy conjunctivitis Blepharitis Chalazion Dry eye Ectropion Watery eye Subconjunctival haemorrhage Non-proliferative DR Cataract
Symptom	Sudden continuous visual disturbance Moderate to severe continuous eyeball pain or photophobia	Marked one side redness (sudden/recent onset)	Diplopia or other visual disturbance	Squint (long-lasting)
Event	Pain in postop eye (less than 2 months) Stab/penetrating trauma Chemical burn	Arc eye Blunt trauma Contact lens/FB problems Corneal graft		
Treatment	Urgent operation	Continuous monitoring for 24–72 hours		

AMD, age-related macular degeneration; CN, cranial nerve; CRAO, central retinal artery occlusion; CRVO, central retinal vein occlusion; CSCR, central serous chorioretinopathy; DR, diabetic retinopathy; FB, foreign body; PVD, posterior vitreous detachment.

### Red (immediate)

Events (pain in postop eye (less than 2 months), stab/penetrated trauma and chemical burn), symptoms (sudden continuous visual disturbance, moderate-to-severe continuous eyeball pain or photophobia), diagnoses (acute glaucoma, corneal laceration, central retinal artery/vein occlusion under 24 hours, giant cell arteritis with visual disturbance, globe perforation, iris prolapse, intraocular foreign body, orbital cellulitis and third cranial nerve palsy with pupil involvement) and treatment (urgent operation).

### Orange (within 24 hours, except for red)

Events (arc eye, blunt trauma, contact lens-related injury, foreign body without globe injury and corneal graft patient), symptoms (sudden/recent onset with marked one-side redness), diagnoses (corneal abrasion, corneal foreign body, corneal ulcer, corneal opacities with pain, foreign body in conjunctival sac, hyphoema, iridocyclitis, orbital wall fracture, retinal detachment, retinal tear and vitreous haemorrhage) and treatments (needing continuous monitoring for 24–72 hours).

### Yellow (within 7 days, except for red and orange)

No event, symptom (diplopia or other visual disturbance), diagnosis (aggravating retinal disease such as age-related macular degeneration, central serous chorioretinopathy with fluid collection, herpes zoster ophthalmicus, non-managed episcleritis, scleritis, posterior vitreous detachment, Bell's palsy, optic neuritis, preseptal orbital cellulitis, severe infectious conjunctivitis, retinal vein occlusion over 24 hours and proliferative diabetic retinopathy) and no treatment.

### Green (not emergency)

No event, symptom (long-lasting squint), diagnosis (allergy conjunctivitis, blepharitis, chalazion, dry eye, ectropion, watery eye, subconjunctival haemorrhage, non-proliferative diabetic retinopathy, no macular oedema and senile cataract) and no specific treatment.

### Reliability analysis of emergency severity classification

We used an intraclass correlation coefficient (ICC,  $p$ -value < 0.05, 95% CI) for reliability of the emergency severity classification by the responsible author. ECode II was compared with two other results from one general practitioner and one other ophthalmologist.

### Clinical variables for training and validation

Three categories of variables (general patient information, event and symptoms) were used as predictors. For general patient information, seven variables (age, sex, systemic disease, allergy, eye disease/trauma, eye operation and operation elapsed time) were used. Systemic disease, eye disease, eye operation and operation elapsed time (if there was a multiple surgical history) were explored with multiple choice questions. In the event category, a trauma event (stab/penetrating and blunt), other event (lens-related/foreign body-related,

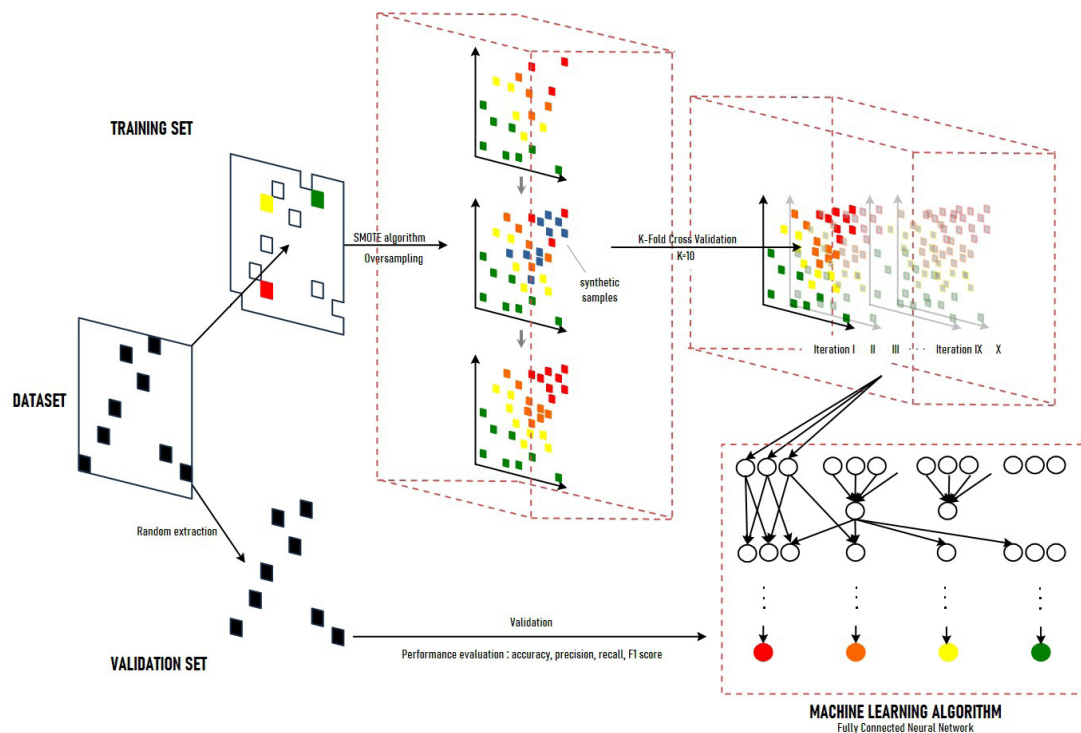
chemical burn and arc injury) and common cold were included. Corrected visual acuity determined with a near vision card was also checked. Symptom categories were divided into three subcategories such as vision, pain and other symptoms for emphasising vision and pain themselves and the temporal relationship of vision, pain and other symptoms. Each symptom subcategory included the aspect of time, aspect of symptoms and number of symptomatic eyes. The aspect of time consisted of duration, persistence (even with behaviours such as blinking, artificial tear usage or daytime changes) and worsening (in the disease period). In the vision subcategory, nine visual problems (poor vision, visual field defect, metamorphopsia, glare, photopsia, colorosis and nyctalopia) were used. In the pain subcategory, the aspect of pain was classified into eye pain (sharp, dull and aggressive pain (patient usually claimed as 'My eyes feel like they're going to pop out')) and eyelid pain. Visual Analogue Scale scores for self-reported pain degrees in both eyes and eyelid pain were also included. In the other symptoms subcategory, diplopia, lazy eye, nystagmus, eyelid mass, eyelid swelling, ptosis, enophthalmos, exophthalmos, white eyeball pigmentation/abnormalities, conjunctival swelling, red eye, dryness, stinging, burning sense, fatigue, itching, foreign body sense, discharge (sleep), tear and neurologic signs (nausea, vomiting and headache) were included. The symptoms were classified by referring to the American Association of Ophthalmology (AAO) homepage (<https://www.aao.org/eye-health/symptoms-list>) and were selected based on usual patient complaints. To establish symptom variables that patients complained of, prior screening was conducted. If there was an exceptional symptom that was difficult to classify using the AAO list because of regional and national characteristics, attempts were made to consistently comply with the AAO list. A total of 57 questions (32 symptoms) were used (table 2).

**Table 2** The clinical variables for training set and validation set

#### Category (number of variables)

General patient informations	Others
Age	Corrected visual acuity (with NVC)
Sex	Symptoms
Systemic disease	Visual symptoms (9)
Allergy	Pain symptoms (5)
Eye disease or trauma	Aspects (4)
Eye operation	VAS scores
Elapsed time	Other symptoms (20)
Events	Characteristics (in each symptom category)
Stab/penetrating trauma	Number of impact eyes
Blunt trauma	Duration
Lens/foreign body related	Persistence
Chemical burn	Worsening
Arc injury	
Common cold	

NVC, near vision card; VAS, Visual Analogue Scale.



**Figure 1** Data set, sampling and the model architecture. Prior to any data processing sequence, patient data were randomly divided into two data sets, that was, training set and validation set (ratio 7:1). The number of each class validation samples were extracted similarly. In training set, unbalanced data set was verified using the synthetic minority oversampling technique (SMOTE) algorithm. The model included the K-fold cross-validation method (K=10) for increasing the model performance. In machine learning algorithm, vision and pain variables were processed by each multilayer perceptron and concatenated with general patient information and other symptom variables. Overfitting was controlled by 50% dropout method, L2 regulation and the early stopping technique. The prediction performance was measured by accuracy of whole data set and precision, recall and F1 score of each class.

### Model development

An ensemble AI model using combinations of fully connected neural networks was adopted (figure 1). Input variables included multiselect questions, and they were preprocessed by dummy processing. In the training set, unbalanced data set was verified using the synthetic minority oversampling technique (SMOTE) algorithm, which is the most widely used algorithm for generating artificial training samples.<sup>19</sup> The model included the K-fold cross-validation method (K=10) for increasing the model performance in relatively few data volumes.<sup>20</sup> This model is unique as it was configured to separate vision and pain variables from other variables, and the variables were processed primarily by each multilayer perceptron method. After the process, pain and vision variables were concatenated with respective general patient information, event and other symptom variables. Overfitting was controlled by the dropout method (50%), L2 regulation and the early stopping technique. Prediction performance was measured by the accuracy of the whole data set and precision, recall and F1 score of each class. Accuracy itself was not useful due to bias in the number of each class, which is common in medical reports.<sup>21 22</sup> The accuracy was high because of the high proportion of the mild severity class, which was the green class in this paper; therefore, precision, recall

and the F1 score were also calculated. Additionally, an 'acceptable prediction percentage' was confirmed. An acceptable prediction percentage was defined as an AI prediction that is determined to be 'true' when the predicted label is the same as or more urgent than the actual class. For example, when AI predicted a yellow-class patient as a red-class or yellow-class patient, the prediction was considered 'true'. The earlier hospital visit was considered acceptable in view of the risk of delayed treatment. A clinically dangerous situation was when an urgent patient visited the hospital late.

### The performance comparison between the model with ECode I and the model with ECode II

Two steps were used to analyse the difference between the model with ECode I and ECode II. First, for the correlation of labels between the model with ECode I and ECode II in the whole data set, Spearman's correlation analysis was used. The relationships among each class in the model with ECode I and ECode II were also analysed. In the validation set, statistically significant differences between the model oversampled by SMOTE algorithm with ECode I (ECcSMOTE I) accuracy and with ECode II (ECcSMOTE II) accuracy were analysed with  $p < 0.05$  indicating significance.



**Table 3** The proportion of ECode I and ECode II

Total 1681 (cases)	Red	Orange	Yellow	Green
<b>ECode I</b>				
Training set (1456 to 100%)	49 (3.39%)	63 (4.30%)	115 (7.92%)	1229 (84.39%)
Validation set (225 to 100%)	61 (27.11%)	61 (27.11%)	52 (23.11%)	51 (22.67%)
Total (100%)	110 (6.54%)	124 (7.38%)	167 (9.93%)	1280 (76.15%)
<b>ECode II</b>				
Training set (1471 to 100%)	77 (5.20%)	110 (7.47%)	120 (8.14%)	1164 (79.19%)
Validation set (210 to 100%)	54 (25.72%)	52 (24.76%)	52 (24.76%)	52 (24.76%)
Total (100%)	131 (7.79%)	162 (9.64%)	172 (10.23%)	1216 (72.34%)

Diagnoses and treatments, except symptoms such as the red flag sign and events, were only used for the first classification (ECode I). Another classification consisted of symptoms, events, diagnoses and treatment (ECode II).

### Patient and public involvement

No patient was involved in developing the research question, outcome measurement and design of the study. We are unable to disseminate the findings of the research directly to the study participants.

## RESULTS

### Reliability of emergency severity classification

The ICC score of ECode II by the general practitioner, other ophthalmologist and responsible authors was 0.979 ( $p < 0.001$ , 95% CI 0.975 to 0.982).

### Model performance

In ECcSMOTE I, 1456 cases (49 (3.39%), 63 (4.30%), 115 (7.92%) and 1229 (84.39%) in the red, orange, yellow and green classes, respectively) were enrolled in the training set, and 225 cases (61 (27.11%), 61 (27.11%), 52 (23.11%) and 51 (22.67%) in each red, orange, yellow and green class) were enrolled in the validation set. In ECcSMOTE II, 1471 cases (77 (5.20%), 110 (7.47%), 120 (8.14%) and 1164 (79.19%), respectively) were enrolled in the training set and 210 cases (54 (25.72%), 52 (24.76%), 52 (24.76%) and 52 (24.76%), respectively) were enrolled in the validation set (table 3). The correlation coefficient between ECode I and ECode II was 0.837 ( $p = 0.000$ ). The relationships among each class between ECode I and ECode II are shown in table 4. The accuracy

of ECcSMOTE I model, calculated with the validation set, was 98.88%. The accuracy of ECcSMOTE II was 99.05%. There was no statistically significant difference between ECcSMOTE I and ECcSMOTE II accuracy ( $p = 0.825$ ). The precision of each class was 100%, 99.04%, 97.14% and 100% in ECcSMOTE I. The recalls were 100%, 100%, 99.04% and 97.16%. The F1 scores were 100%, 99.52%, 98.08% and 98.56%, respectively. The acceptable prediction percentage of ECcSMOTE I was 100%. The precision of each class was 100%, 98.10%, 92.73% and 100%, respectively, with ECcSMOTE II. The recalls were 100%, 100%, 98.08% and 95.33%, respectively. The F1 scores were 100%, 99.04%, 95.33% and 96.00%, respectively. The acceptable prediction percentage of ECcSMOTE II was also 100% (table 5).

## DISCUSSION

### Previous researches

Currently, mobile phone applications and internet sites that diagnose diseases and classify emergency severity based on symptoms are being used to decide whether patients require a hospital visit.<sup>13–16</sup> However, unfortunately, previous studies reported negative results for these utilities. The studies showed that the accuracy of disease diagnosis was less than 50% and that of emergency severity classification was about 80%. Also, a recent research study

**Table 4** The relationships of ECode I and ECode II

		ECode II			
		Red (%)	Orange (%)	Yellow (%)	Green (%)
ECode II	Red	65.2	13.0	–	21.8
	Orange	–	48.5	21.2	30.3
	Yellow	–	–	77.8	22.2
	Green	–	–	–	100

The correlations coefficient between ECode I and ECode II was 0.837 ( $p = 0.0000$ ).

**Table 5** The model performance of ECcSMOTE II

	Total (%)	Red (%)	Orange (%)	Yellow (%)	Green (%)
Accuracy	99.05	–	–	–	–
Precision	–	100	98.10	92.73	100
Recall	–	100	100	98.08	95.33
F1 score	–	100	99.04	95.33	96.00

ECcSMOTE II was the model labelled by ECode II with SMOTE algorithm and their input variables included general patient information. SMOTE, synthetic minority oversampling technique.

in the field of ophthalmology was performed, although the sample size was insufficient.<sup>23</sup> In this study, the overall accuracy of disease diagnosis was 26% (95% CI: 12% to 40%). The accuracy of emergency triage was 39% (95% CI: 14% to 64%) in an emergency situation and 88% (95% CI: 73% to 100%) in a non-emergency situation.

### Strengths and limitations

The results of our study differed from those of preceding studies. There were three main reasons for this. First, the biggest difference was the sample size.<sup>24</sup> In the present study, 1681 cases were studied in which the sample size was relatively large compared with previous studies that predicted about 50–99 diagnoses and 3–5 triage classifications with around 20–30 symptoms. If {(the number of input variables)/(sample size)} increased, the strength for each variable was lowered, and the variable whose actual strength was high did not have enough strength. If {(the number of output variables)/(sample size)} increased, there may be very few or no samples for each output class, which means that there was no significant difference between each class.<sup>25–28</sup> Second, more detailed input variables were employed in this study. Fifty-seven questions were used as input variables. This was clearly different from using 20–30 questions as in previous studies. The variables, such as persistence, aggravation and degree, were used in addition to the presence or absence of symptoms. The aspect refinement allowed building a model with completely different expressions for the same symptom variables. Third, emergency severity prediction, not diagnostic prediction, was assessed. The reason why diagnostic prediction is more difficult than emergency severity prediction was mentioned above. Also, different diseases are often expressed with the same symptoms, and it is very difficult to distinguish diseases clinically by symptoms without examination in ophthalmology. For example, wet-type age-related macular degeneration and central serous chorioretinopathy were labelled as the same ‘yellow’ class in this study, although they were diagnosed differently. The symptoms of both diseases were only persistent visual disturbances.

In this paper, emergency severity was classified using two methods. Diagnoses and treatments, except symptoms such as the red flag sign and events, were only used for the first classification (ECode I). Another classification consisted of symptoms, events, diagnoses and treatment (ECode II). Both ECode I and ECode II showed excellent accuracy with the SMOTE algorithm (ECcSMOTE I and II, 98.88% and 99.05%, respectively). Contrary to concern that the prediction of the ECcSMOTE I model was milder than the actual severity, it was equal or more urgent than the real values (allowable prediction percentage 100%). Also, the values for each class in ECode II were equal or higher than those in ECode I (table 5). Nevertheless, the accuracy of ECcSMOTE I was not lower than that of ECcSMOTE II ( $p=0.825$ ). It can be seen that labelling is also possible in emergency severity machine learning, even if the red flag sign and event is not considered.

This phenomenon was due to the input variation of the training set already including the patient’s symptoms and events, and the input variation was correlated with the output variation during the training process.

There were two major limitations in this study. First, there was a lack of data set samples. There were bigger sample sizes than in the previous reports, but it was relatively insufficient considering the number of input variables. The classification formula implemented in the model was different from simple linear regression, but the lack of samples was limited in machine learning based on statistic operations.<sup>29</sup> Additional studies with larger and multicentre validation sets are considered for clinical application. The second limitation was imbalance among the classes, which is a common problem in medical research. This problem suggests that the accuracy of minor classes becomes greatly degraded.<sup>21 22 30</sup> The accuracy of the ECsSMOTE II model was only 79.52%, which was almost similar to other studies by the limitation. To compensate for the limitation, the SMOTE algorithm, which is an oversampling method that mathematically emphasises class relationships was used in this study.<sup>20</sup> Considering the occurrence of bias with oversampling when evaluating model performance, the actual cases were randomly extracted from the entire data set before oversampling as a validation set. The SMOTE algorithm does not show superior performance compared with the undersampling method, but it can possibly be utilised for high-dimensional class-imbalanced data.<sup>31</sup> By further research that overcomes these limitations, the risk of clinical use of AI in ophthalmologic emergency should be supplemented.

### Conclusions

The accuracies of the models using the SMOTE algorithm (ECcSMOTE I and II) were close to 100% (98.88% and 99.05%, respectively). This preliminary study yielded excellent predictions, in contrast to previous reports, using various available mathematical and technical methods that aimed to compensate for the limitations, although studies in more diverse populations than in this study are ideal. In conclusion, we propose an AI method to classify ophthalmic emergency severity based on symptoms.

**Contributors** HA: substantial contributions to the conception, design of the work; the acquisition, analysis and interpretation of data for the work.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** Approval was obtained from the Medical Research Ethics Committee, the Armed Force Medical Command, Republic of Korea (protocol number AFMC-19126-IRB-19-096). All research protocols were conducted in accordance with the Declaration of Helsinki.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data relevant to the study are included in the article. Data are available in a public, open access repository.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Hyunmin Ahn <http://orcid.org/0000-0003-1529-5123>

#### REFERENCES

- International Monetary Fund. *World economic outlook database*. 2018. Washington, DC: International Monetary Fund, 2019.
- Fullman N, Yearwood J, Abay SM, *et al*. Measuring performance on the healthcare access and quality index for 195 countries and territories and selected subnational locations: a systematic analysis from the global burden of disease study 2016. *Lancet* 2018;391:2236–71.
- Statistics Korea. *Statistics Korea outlook database*. 2019. Daejeon: Statics Korea, 2019.
- Dejardin O, Remontet L, Bouvier AM, *et al*. Socioeconomic and geographic determinants of survival of patients with digestive cancer in France. *Br J Cancer* 2006;95:944–9.
- Pierce RP, Williamson HA, Kruse RL. Distance, use of resources, and mortality among rural Missouri residents with acute myocardial infarction. *The Journal of Rural Health* 1998;14:28–35.
- McDonnell PJ. How do general practitioners manage eye disease in the community? *Br J Ophthalmol* 1988;72:733–6.
- Sheldrick JH, Wilson AD, Vernon SA, *et al*. Management of ophthalmic disease in general practice. *Br J Gen Pract* 1993;43:459–62.
- Ting DSW, Pasquale LR, Peng L, *et al*. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167–75.
- Du X-L, Li W-B, Hu B-J. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol* 2018;11:1555.
- Aytek A, Asce M, Alp M. An application of artificial intelligence for rainfall-runoff modeling. *Journal of Earth System Science* 2008;117:145–55.
- Meiring GAM, Myburgh HC. A review of intelligent driving style analysis systems and related artificial intelligence algorithms. *Sensors* 2015;15:30653–82.
- Grant K, McParland A. Applications of artificial intelligence in emergency medicine. *University of Toronto Medical Journal* 2019;96.
- Semigran HL, Linder JA, Gidengil C, *et al*. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ* 2015;351:h3480.
- Chambers D, Cantrell AJ, Johnson M, *et al*. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. *BMJ Open* 2019;9:e027743.
- Fine AM, Nizet V, Mandl KD. Participatory medicine: a home score for streptococcal pharyngitis enabled by real-time biosurveillance: a cohort study. *Ann Intern Med* 2013;159:577–83.
- Black P. *The dangers of using Google as a diagnostic aid*. London: MA Healthcare, 2009: 18. 1157.
- Heng LZ, Hamilton RD. Ocular emergencies. *Medicine* 2018;46:754–9.
- Kilduff C, Lois C. Red eyes and red-flags: improving ophthalmic assessment and referral in primary care. *BMJ Qual Improv Rep* 2016;5:u211608–4680.
- Chawla NV, Bowyer KW, Hall LO, *et al*. SMOTE: synthetic minority Over-sampling technique. *J Artif Intell Res* 2002;16:321–57.
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. in proceedings of Ijcai:1137–45.
- Rahman MM, Davis DN. Addressing the class imbalance problem in medical datasets. *Int J Mach Learn Comput* 2013;3:224–8.
- Li D-C, Liu C-W, Hu SC. A learning method for the class imbalance problem with medical data sets. *Comput Biol Med* 2010;40:509–18.
- Shen C, Nguyen M, Gregor A, *et al*. Accuracy of a popular online symptom Checker for ophthalmic diagnoses. *JAMA Ophthalmol* 2019;137:690–2.
- Sandelowski M. Sample size in qualitative research. *Res Nurs Health* 1995;18:179–83.
- Francis JJ, Johnston M, Robertson C, *et al*. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychol Health* 2010;25:1229–45.
- Guest G, Bunce A, Johnson L. How many interviews are enough? an experiment with data saturation and variability. *Field methods* 2006;18:59–82.
- Kotrlik J, Higgins C. Organizational research: determining appropriate sample size in survey research appropriate sample size in survey research. *Information technology, learning, and performance journal* 2001;19:43.
- Peduzzi P, Concato J, Kemper E, *et al*. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
- Byrd RH, Chin GM, Nocedal J, *et al*. Sample size selection in optimization methods for machine learning. *Math Program* 2012;134:127–55.
- Liu Y, Yu X, Huang JX, *et al*. Combining integrated sampling with SVM ensembles for learning from imbalanced datasets. *Inf Process Manag* 2011;47:617–31.
- Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics* 2013;14:106.