

RESEARCH ARTICLE

Multimodal Deep Learning Model of Predicting Future Visual Field for Glaucoma Patients

QUANG T. M. PHAM¹, JONG CHUL HAN^{2,3}, DO YOUNG PARK⁴,
AND JITAE SHIN¹, (Member, IEEE)

¹Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, Republic of Korea

²Department of Ophthalmology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 03181, Republic of Korea

³Department of Medical Device, Management and Research, SAIHST, Sungkyunkwan University, Seoul 03181, Republic of Korea

⁴Department of Ophthalmology, Yeungnam University Hospital, Yeungnam University College of Medicine, Daegu 42415, Republic of Korea

Corresponding authors: Jitae Shin (jtsin@skku.edu) and Jong Chul Han (heartmedic79@gmail.com)

This study was supported in part by SMC-SKKU Future Convergence Research Program Grant and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2020R1F1A1065626).

ABSTRACT Glaucoma is one of the most common reasons for blindness worldwide, especially in elderly people. Glaucoma can be monitored using visual field (VF) tests. Therefore, predicting the future VF to monitor progression of glaucoma is important. In this paper, we proposed a deep learning model to predict future VF based on previous VF and optical coherence tomography (OCT) images (including thickness map, vertical tomogram, and horizontal tomogram). The image data were analyzed using a ResNet-50 model. Image features and previous VFs were combined, and a long short-term memory (LSTM) network was used to predict future VF. A weighted method was used to detect noisy data. The proposed method was improved when applying weighted loss. The mean absolute error (MAE) was 3.31 ± 1.37 , and the root mean square error (RMSE) was 4.58 ± 1.77 . The model showed high performance when combining VF data and OCT image data. Furthermore, the model was useful for detecting and re-weighting noisy data.

INDEX TERMS Deep learning, glaucoma, OCT, visual field.

I. INTRODUCTION

Glaucoma is a progressive optic neuropathy that can lead to blindness. The visual field (VF) test is one of the important methods for monitoring the progression of glaucoma. Based on the VF report, early treatment can be given to reduce the damage and decrease the progression rate [1]. VF defects in glaucoma patients usually occur in the peripheral VF, although defects can present in the central VF in some glaucoma patients. Because central VF is important to maintain visual acuity and quality of life in glaucoma patients, detecting and predicting the VF defect pattern is very important [2].

Studies about predicting glaucoma progression and VFs have been conducted to help doctors in monitoring the disease. Since the VF test can reveal the functional defects directly. Many of them utilize the serial of previous VFs to predict the next VFs [3], [4], [5], [6]. However, collecting serial VF data can be difficult because a long period of time

might be required, and inconsistent results can be obtained based on the condition of the patients. Thus, unreliable VF reports and noisy samples can be obtained. Therefore, it is essential to detect noisy samples during the training of deep learning models.

In addition to the VF test, optical coherence tomography (OCT) can provide structural information regarding optic nerve head damage associated with glaucoma. Compared with the VF test, OCT is faster and more dependable. The OCT device offers various types of reports regarding the optic nerve head such as thickness maps and tomograms (horizontal and vertical). The thickness map in the OCT usually provides information on the thickness of the retinal nerve fiber layer (RNFL) around the optic nerve head that is directly associated with glaucomatous damage, and the tomogram can provide information on structural features in the optic nerve head as well as the extent of glaucomatous damage (Fig. 1) [7].

In this paper, we proposed a deep learning model to predict future VF defect patterns in glaucoma patients using previous VF test data and OCT image data such as thickness

The associate editor coordinating the review of this manuscript and approving it for publication was Ahsan Khandoker¹.

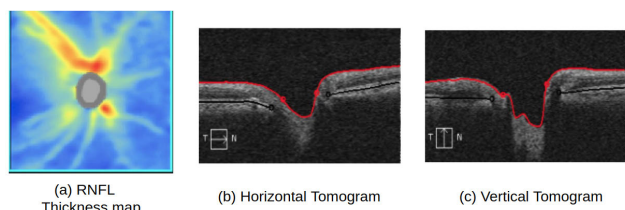


FIGURE 1. The image data from OCT test: (a) thickness map, (b) horizontal tomogram, and (c) vertical tomogram.

maps and tomograms. We exploited both CNN and RNN to extract features from image data and serial data. Furthermore, to manage the noisy data, an efficient training mechanism is applied. A regression model is made to detect noise and obtain the weight of samples. Then, the main model is trained with a new weighted loss. To the best of our knowledge, this is the first study attempting to deal with noisy data directly for the future VF prediction task. Based on quantitative results, these contributions can improve the performance of our proposed model.

II. RELATED WORK

A structure-function relationship exists between the VF map and the structure of the optic nerve head and can be captured on the OCT thickness map [8], [9], [10] (Fig. 2). Thus, OCT images can be used to analyze VF. Tsai et al. [11] proposed a parametric linear model to learn the structure-function relationship between the OCT thickness map and VF. They showed that the OCT thickness maps were associated with VF defect patterns in glaucoma patients.

Compared with the principal component analysis (PCA)-based coupled parametric model in a previous study [11], deep learning models are mainly applied to OCT images. For predicting VF from OCT images, convolutional neural network (CNN) models were mainly used in previous studies [12], [13]. To manage the image data, 3D CNN models were introduced in several studies [5], [14]. However, the structure-function relationships are different for each individual [9]. Consequently, without knowing the specific relationship for each patient, the prediction ability of the models is limited. Thus, a model with multi inputs (OCT images and VF tests) can help to learn the structure-function relationships better.

Previous VF data are commonly used to predict glaucoma because the VF usually progresses based on the initial defect pattern. Therefore, in many studies, previous VFs were used for learning the progression patterns [5], [6]. However, most of the studies only focus on the binary classification task, which detects whether there is a glaucoma progression or not [4], [6]. Yousefi et al. used a linear regression model to predict glaucoma progression based on previous VFs in [4] while a CNN model is applied in [15]. Kouros et al. exploited both VFs and thickness measurements from OCT for prediction [6].

To predict future VF, many researchers applied deep learning models [3], [14], [16]. A variational autoencoder

(VAE) was used to estimate future VF in [3]. To manage the time series data, a recurrent neural network (RNN) model was used in [17]. In order to improve the performance, a linear regularization method is introduced to combine both previous VFs and OCT images in the same latent space [14], [18]. Although those recent researchers are able to learn the glaucoma progression. However, there are no previous studies that can handle noisy data efficiently.

III. METHOD

A. DATASET AND DATA AVAILABILITY

A retrospective study was conducted on the patients in the Department of Ophthalmology at Samsung Medical Center between January 2018 and December 2019. All consecutive patients were selected for medical record review. This study was approved by the Institutional Review Board (IRB) at Samsung Medical Center (IRB No. 2020-08-040). Informed consent was waived due to the retrospective nature of the study, and the waiver was provided by the IRB. All study protocols adhered to the tenets of the Declaration of Helsinki.

Our private dataset included VF data and OCT image data from 266 patients. The glaucoma patients in the study were open-angle glaucoma. Each patient had 2–11 tests (including both VF and OCT tests, mean = 5.74). In our dataset, since glaucoma is a bilateral disease, one eye per patient was included. For the VF test, we extract the raw Humphrey VF (HVF) 30-2 from automated perimetry using a central 30-2 Humphrey Field Analyzer (HFA model 640; Humphrey Instruments, Inc., San Leandro, CA, USA) with the Swedish interactive threshold algorithm standard. For the OCT test, we extract the retinal nerve fiber layer (RNFL) thickness map, vertical tomogram, and horizontal tomogram from the Cirrus OCT (Carl Zeiss Meditec, Dublin, CA, USA) using the customized program. VF tests were repeated once or twice per year, depending on the patient's condition. Reliable VF analysis was defined as a false-negative rate of <15%, a false-positive rate of <15%, and a fixation loss of <20%. Unreliable VF test results were excluded from the analyses. If the error in the automated segmentation of OCT that might induce the wrong outcome in the analysis was confirmed, then the results were also excluded. Overall, 1,502 pairs of VF and OCT image samples were included in the dataset. Participant information is shown in Table 1. For more information, the source code is available at https://github.com/QuangBK/future_VF_prediction.

Furthermore, we also collect an external dataset for testing. The study for external validation was approved by the Institutional Review Board of Yeungnam University Hospital (IRB No. 2022-04-045). For the VF test, we use 30-2 HFA model 840 with the Swedish interactive threshold algorithm standard. For the OCT test, we extract the RNFL thickness map, vertical tomogram, and horizontal tomogram from the Cirrus OCT 500 (Carl Zeiss Meditec, Dublin, CA, USA). The external dataset included 50 patients with both eyes. Each patient had an average of 2.3 tests (including both VF and OCT tests).

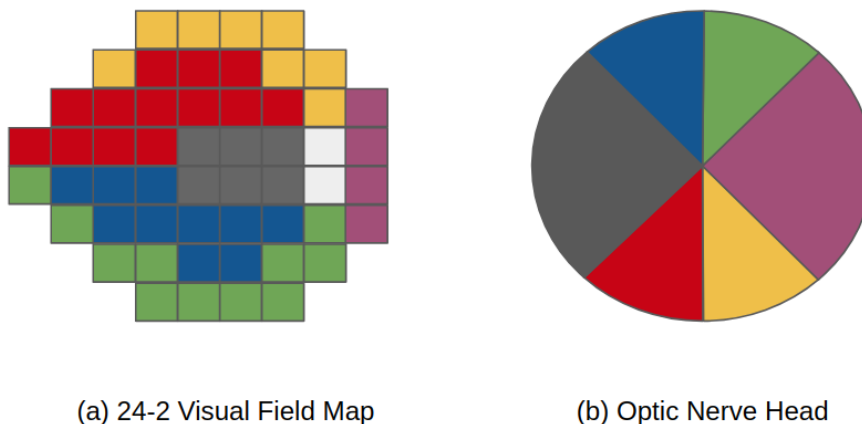


FIGURE 2. The regional relationship between VF map (a) and optic nerve head (b) in thickness map.

TABLE 1. Participant demographics and clinical characteristics.

	Training/Validation data	Testing (External) data
Subjects (n)	266	50
Left/right eyes	118/148	49/50
Age (years)	57.4 ± 12.8	50.9 ± 12.4
Follow-up period (years)	5.0 ± 1.3	3.3 ± 2.4
Number of visits (n)	5.7 ± 1.7	2.3 ± 0.6
Mean Deviation (decibel, dB)	-6.4 ± 5.1	-2.4 ± 6.0
Average RNFL thickness (μm)	71.58 ± 11.78	80.64 ± 13.42

B. MODEL

Information from previous report tests was extracted and included past VF maps, thickness maps, tomogram images, and enhanced depth imaging (EDI) OCT images. However, since the EDI data is only available for a few patients in our dataset. We decide not to use EDI data in this research. Based on the inputs, the model, shown in Fig. 3, will predict the future VF.

First, information was extracted from image data and included thickness maps, horizontal tomograms, and vertical tomograms. Before feeding to the deep learning model, all images were normalized (pixel value from -1 to 1) and resized to 224 × 224. A CNN model was applied for analyzing image data (Fig. 4). The pre-trained ResNet-50 [19] was used as the extractor and fine-tuned with the dataset. For each image as an input, the CNN extractor provides a feature vector. Notably, the thickness map provides information regarding the entire VF, while the vertical and horizontal tomograms only provide useful information regarding specific regions in the VF. We decided to focus on the thickness map features. The output vectors of the thickness maps, vertical tomograms, and horizontal tomograms were 512-d vector, 128-d vector, and 128-d vector, respectively. Because the dataset was limited, several common augmentation techniques (shift, scale, rotate) were applied to avoid overfitting. Next, the three feature vectors were concatenated into a 768-d vector. Then, a convolution layer

was applied to create a final 128-d feature vector of the image data.

The previous VFs were reshaped to 76-d vectors and normalized with a mean of 0 and a standard deviation of 1. Furthermore, the time information was added as an input in the proposed model. There are two types of time information: timeline and interval. The timeline information of a chosen image is the time between the chosen image and the first image. The interval information of a chosen image is the time between the chosen image and the previous image. Overall, for each time step, the images, VF, and time information were concatenated to create a 206-d vector. For each patient, a series of 206-d vectors were created, with each vector corresponding to a test date. This series of 206-d vectors was the input of a Long Short-Term Memory (LSTM) network [20]. The output of LSTM was a 128-d vector.

Finally, the LSTM output was used to predict the future VF. Although a structure-function relationship exists between the VF and thickness map, the relationship differs by person. In addition, though glaucoma develops in one of several common patterns, each patient has specific characteristics. To solve this customization problem, the previous VF (only the last one) was used as a reference for predicting the future VF. Based on the LSTM output and the reference VF, the prediction accuracy can be significantly improved. In this case, the LSTM output contained the direction of glaucoma development, and the reference VF was responsible

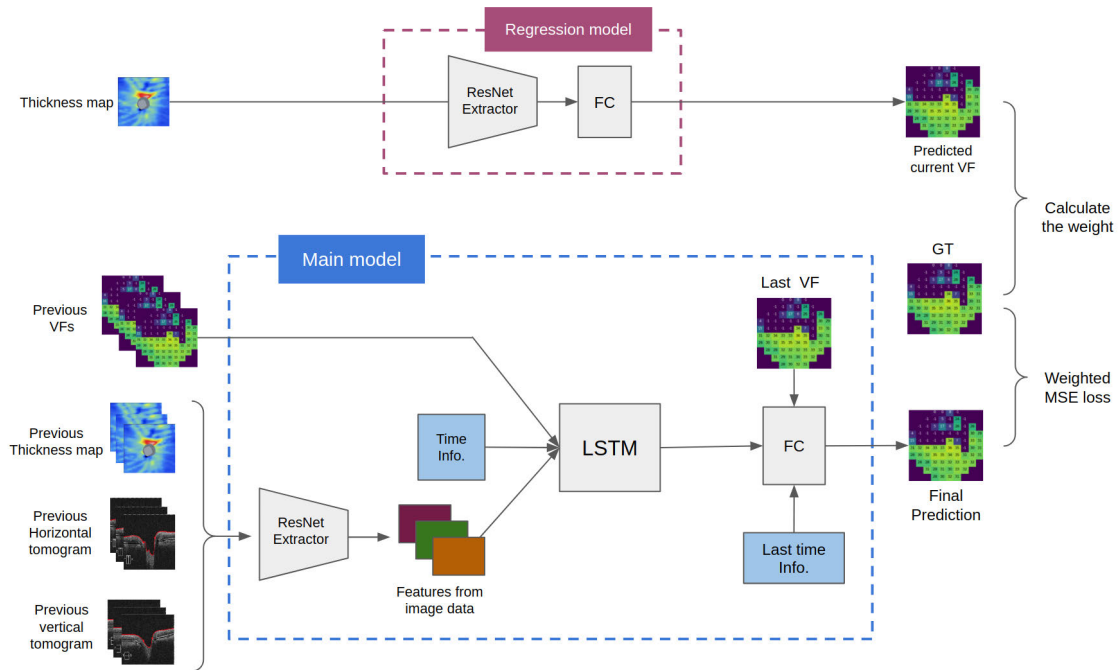


FIGURE 3. The overview of our model (FC: Fully-Connected layer, LSTM: Long Sort-Term Memory layer). The main model is used to predict future VF. The regression model is used to detect noise and calculate the weights of samples to train the main model.

for calibrating the prediction. Therefore, the LSTM output, the previous VF, and time information were combined to predict the future VF.

C. TRAINING

The VF test report is not stable because it depends on the current status of the patient, often producing a noisy VF map. To address the noisy data, a training method re-weighting the samples was introduced (Fig. 5). The OCT test was more stable than the VF test. Because the thickness maps provide consistent and reliable results, they were used for the noisy VF test, and the output was used to train the main model.

First, a regression model was built to predict VF based on the thickness map. The input was a thickness map, and the output was its corresponding VF. The backbone of the regression model was a pre-trained ResNet-50 [19]. After extracting features from the image data, fully-connected layers were added to predict 76-d vectors as VF. The loss function was mean square error (MSE loss), and data augmentation was applied simultaneously as in the main model. The regression model allows the structure-function relationship to be learned. After training, the model can predict the common pattern of glaucoma based on the thickness map.

Because the thickness map is more consistent than the VF, the noisy VF can be detected by comparing the prediction and the ground truth. A large difference indicates a noisy sample, in which the values were re-weighted. Algorithm 1 shows how the weight of VF samples was calculated based on difference. First, for each patient, the mean absolute

Algorithm 1 Calculate Weight of VFs

```

TH = 1.5
for each patient i do
  for each time step t do
     $d_i^t = \frac{\sum_{k=0}^{75} |x_{i,k,pred}^t - x_{i,k,GT}^t|}{76}$  ▷ Calculate error
    between prediction and ground truth
  end for
   $D_i = [d_i^0, \dots, d_i^T]$ 
   $D_{min} = \min(D_i)$ 
   $D'_i = D_i - D_{min}$ 
   $D'_{mean} = \text{mean}(D'_i)$ 
   $D'_i = \frac{D'_i}{D'_{mean}}$  ▷ The error array D was normalized to D'.
  for each time step t do
    if  $d_i^t \leq TH$  then
       $w_i^t = 1$ 
    else if  $d_i^t > TH$  then
       $w_i^t = e^{TH - d_i^t}$ 
    end if
  end for
end for

```

error (MAE) between the ground truth and predicted VF was calculated to obtain the list *D*. However, the mean error value varied by patient. Therefore, *D* was normalized to obtain *D'*. Next, the normalized error *D'* was used to identify the noisy samples. The threshold (TH) was set at 1.5, and samples with errors smaller than TH were considered good samples, while the others were labeled as noisy. The weight of good samples

was 1. For noisy samples, a larger error led to greater noise, and less attention should be given to those samples. Thus, the weight of a noisy sample was calculated based on an exponential function.

After obtaining all weights, the main model with weighted loss was trained to reduce the effect of noisy data. The loss function used to train the main model is weighted MSE as below:

$$\text{loss} = \sum_{i=1}^N w_i (x_{i,pred} - x_{i,GT}) \quad (1)$$

IV. RESULTS AND DISCUSSION

A. IMPLEMENTATION

The metrics we used for evaluation are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 score.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^{75} (x_{GT}^i - x_{pred}^i)^2}{76}} \quad (2)$$

$$\text{MAE} = \frac{\sum_{i=0}^{75} |x_{GT}^i - x_{pred}^i|}{76} \quad (3)$$

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} \quad (4)$$

where x_{GT}^i and x_{pred}^i are the i -th point in the ground truth and predicted VF, respectively. RSS is the sum of squares of residuals and TSS is the total sum of squares.

First, the regression model was trained to detect noisy samples. The ResNet-50-based model was trained with 25 epochs; only the last layer (regression module) in the first seven epochs was trained. Next, the last five layers were unfrozen and fine-tuned with a smaller learning rate. The optimizer was Adam. The initial learning rate was $1e-3$ and decreased to $1e-4$ and $1e-5$ after the seventh and 15th epochs, respectively. After the regression model was trained, it was used to detect noisy samples and obtain w_i with Algorithm 1. The main model was trained with weighted MSE. The number of epochs was 15. The optimizer was Adam. The initial learning rate was $1e-3$ and decreased to $1e-4$ after the seventh epoch. Because the dataset was small, cross-validation was applied (5 folds). Moreover, to guarantee that there is no overlap between partitions and series data from a patient is on the same partition, we split our data based on patient ID. To ensure fair evaluation, for each iteration, the regression and main models were trained on the same dataset. For our model and compared models, we use the same dataset for training and testing. The same pre-processing and cross-validation methods are applied for all experiments.

B. RESULTS

Table 2 shows the performance of the main model. When only using VFs, the MAE and RMSE were 3.60 ± 1.50 and 4.91 ± 1.84 , respectively. The reference module significantly increased the performance of both MAE and

RMSE. Furthermore, the use of image data improved the performance, as evidenced by MAE of 3.36 ± 1.36 and RMSE of 4.61 ± 1.75 . The final model combined VFs and image data as input and applied a weighted loss to provide the best performance of MAE of 3.31 ± 1.37 and RMSE of 4.58 ± 1.77 . The results show an adequate performance of the proposed model with the reference module. Furthermore, training with the re-weighted sample managed the noisy data and improved both metrics (MAE and RMSE). In addition, the effect of the follow-up period and age on the performance of our model was analyzed. Fig. 6 shows the MAE for each follow-up period. A shorter follow-up period resulted in a more accurate prediction. For the proposed model, the MAE for a period shorter than six months was 3.19, and for longer than 18 months was 3.48. In Fig. 7, we analyze the MAE by three age groups: 0-40, 41-65, and over 65 years old. As we can see, the older group shows a higher MAE.

The results of comparison with other methods are illustrated in Table 3 and Fig. 8. Our dataset and the external dataset are used for evaluation. Berchuck et al. applied VAE to estimate the future VF based on previous VFs [3]. Conversely, an RNN model was used to predict future VF based on previous VFs [17]. In the compared models, only previous VFs were used as input. Significance was calculated using one-sided Wilcoxon test. The present study results showed that our proposed model significantly outperforms the compared methods ($p < 0.001$). The VAE-based model [3] predicts the future VFs from the latent code (which is in a low-dimensional space), illustrating the main trend of glaucoma. However, a detailed prediction cannot be made. In particular, the 30-2 HVF test, which contains 76 test points, was used in the dataset and is more detailed than the 24-2 HVF test (which contains only 52 test points) used in the compared methods. Park's method using an RNN-based model can provide better performance than the VAE-based model. However, Park's model cannot manage a noisy dataset and can result in overfitting.

In our proposed model, without using image data, re-weighted training can provide better performance compared with other models. The proposed model showed re-weighting to be beneficial in cases with noisy data. Since the medical data contains noise, especially the VF data. As the result, re-weighting can increase the performance significantly. In addition, the OCT images can be helpful with the structure-function relationship between the VF and thickness map. The final model, which combines re-weighted training and image data, achieved the best performance in all metrics (MAE, RMSE, and R^2).

C. DISCUSSION

Overall, the best result in the proposed model was obtained at MAE of 3.31, RMSE of 4.58, and R^2 of 0.76 as shown in the point-wise MAE map in Fig. 9. As we can see, the paracentral area has a higher MAE in several points. The macular area is consisted of multiple layers of retinal ganglion cells, whereas

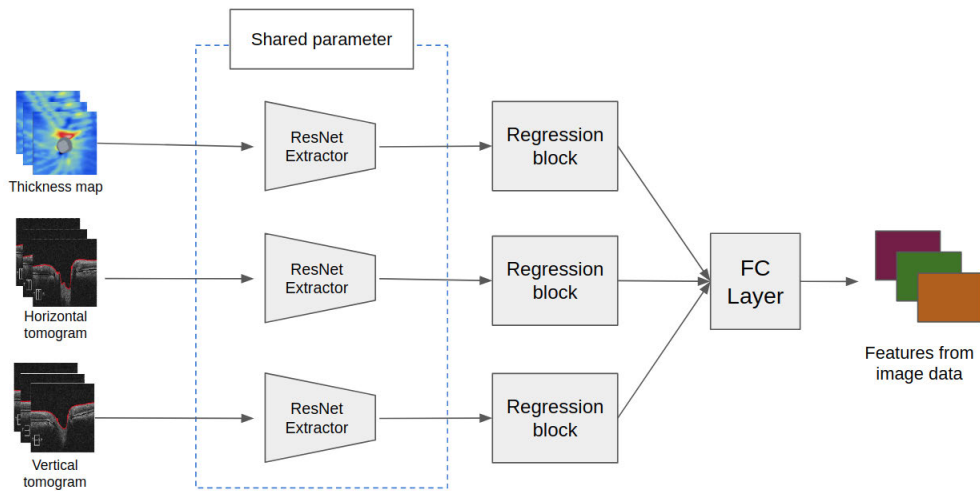


FIGURE 4. The ResNet-50 based model.

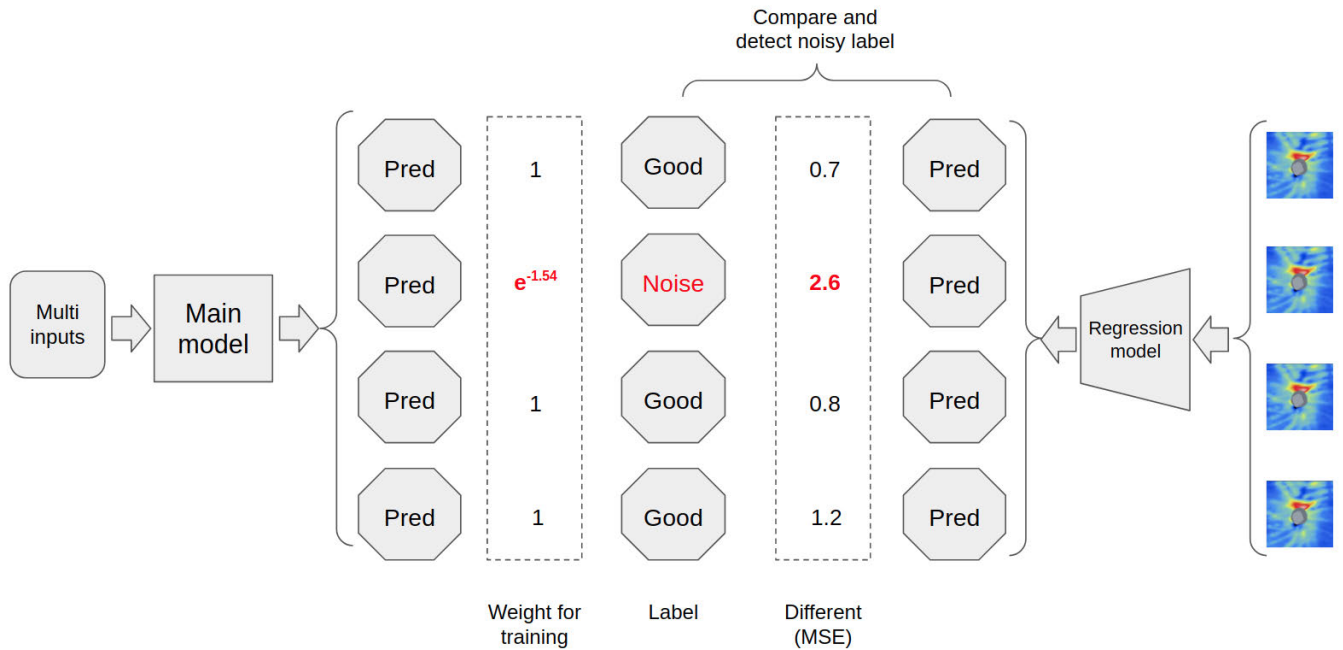


FIGURE 5. An example of detected and weight noisy samples.

TABLE 2. The ablation study with MAE and RMSE (dB).

Input	Reference Module	Weighted Loss	MAE	RMSE
VFs	✗	✗	3.60 ± 1.50	4.91 ± 1.84
VFs	✓	✗	3.34 ± 1.38	4.60 ± 1.77
VFs	✓	✓	3.33 ± 1.38	4.60 ± 1.77
VFs + images	✓	✗	3.36 ± 1.36	4.61 ± 1.75
VFs + images	✓	✓	3.31 ± 1.37	4.58 ± 1.77

other areas usually consist of one retinal ganglion cell layer [21]. Though there were retinal ganglion cell damages

in the macula, the paracentral VF might be normal. Therefore, when the structural changes in OCT occurred, the associated

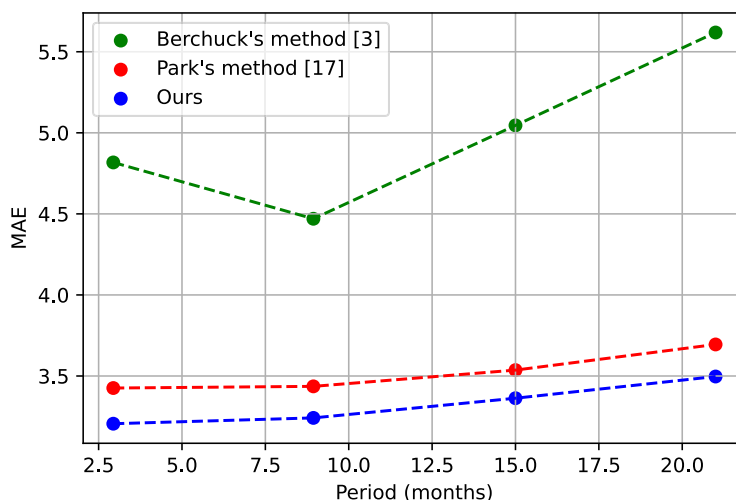


FIGURE 6. The MAE for each follow-up period.

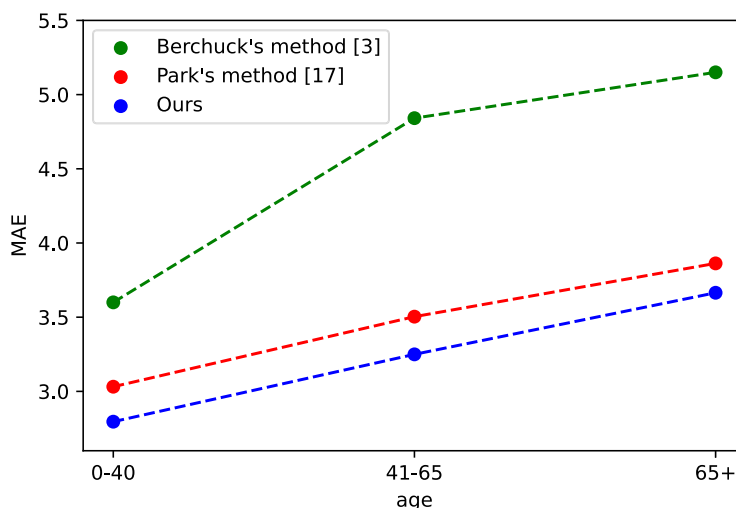


FIGURE 7. The MAE for each age group.

TABLE 3. The comparison of MAE and RMSE results (dB).

Model	Validation data			Test (External) data		
	MAE	RMSE	R^2	MAE	RMSE	R^2
Berchuck et al. [3]	4.76 ± 2.09	6.34 ± 2.44	0.53	4.79 ± 2.74	6.57 ± 3.33	0.38
Park et al. [17]	3.51 ± 1.46	4.81 ± 1.83	0.73	3.78 ± 1.52	5.11 ± 2.02	0.60
Our model	3.31 ± 1.37	4.58 ± 1.77	0.76	3.63 ± 1.52	5.02 ± 2.08	0.61

VF changes might be less predictable in macula than in other areas.

Glaucoma is a progressive optic neuropathy with accompanying glaucomatous VF defects [22]. Glaucoma has several types of VF defect patterns such as accurate defects, nasal steps, and other patterns associated with RNFL defects. Although glaucoma patients can have good visual acuity, some might have difficulties in activities of daily life due to VF defect pattern [23], [24], [25]. Thus, predicting the final VF pattern is important [26], [27].

In several previous studies, only previous VFs were used to predict the future VF because they show the development of glaucoma over time. Yousefi et al. [4] used a Gaussian mixture model to predict glaucoma progression using previous VFs. However, our proposed model showed that the addition of OCT images can help improve the performance.

In previous studies, the common technique for analyzing VF was to transform it into a typically low-dimensional latent space and study glaucoma progression. Since glaucoma

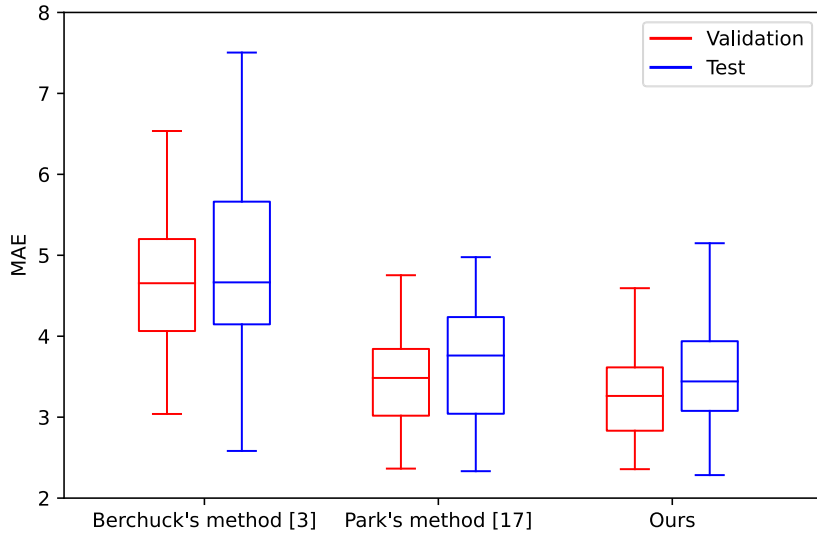


FIGURE 8. The MAE of the proposed and compared methods.

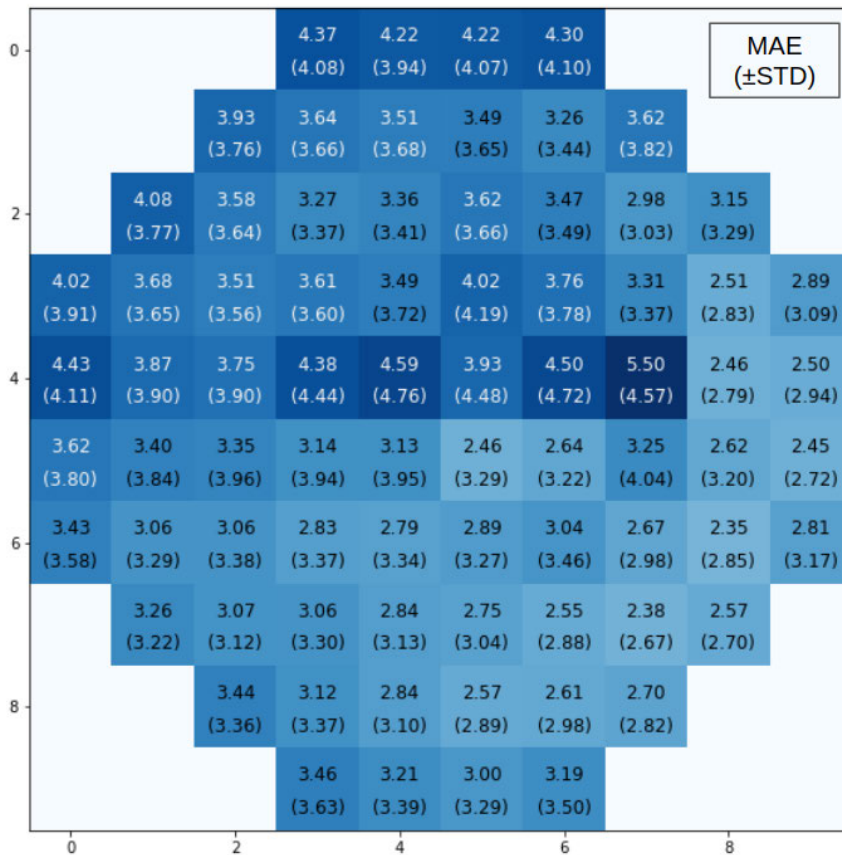


FIGURE 9. The MAE (± standard deviation) in a point-wise map.

progression can be divided into several common patterns, the goal of transformation to latent space is to group patients based on patterns. Then, progression can be determined in

the latent space. Consequently, the complexity of data is reduced, the linear model is strengthened, and overfitting is reduced [13]. This method was used with deep learning

models in several studies [5], [28]. However, patients are unique even if they have the same pattern, leading to incorrect predictions. Although reducing dimensions with latent space can manage overfitting, it fails to provide detailed results.

In previous studies, HFA 10-2 [28] or 24-2 VF maps [3] (with 68 or 54 test points, respectively) was mainly used. In contrast, the dataset of HFA 30-2 VF maps with 76 test points was used in the present study. With additional test points, reconstructing a detailed HFA 30-2 VF map from latent space is difficult. In the present study, the VF was directly analyzed only using a deep learning model. Furthermore, a reference module was applied to refine the predicted results.

In the latent space, the common method is to use a linear model for exploring the data. Yuhui et al. [14] used a linear regression method in latent space for predicting glaucoma progression. Furthermore, a deeply regularized latent-space linear regression model was applied in another study [5]. However, due to the noisy data, the linear models failed to capture the correct glaucoma progression, especially when only a short series of previous VF was provided.

Recently, the deep learning models were applied for VF analysis. For estimating future VF, Berchuck et al. [3] used a VAE model, and an RNN model was applied in another study [17]. The RNN model is good at handling sequential data. Although linear models often perform poorly in cases with noisy data, the RNN model can handle a non-linear trend. Furthermore, because Garway-Heath [8] showed a relationship between OCT image and VF map, the use of retinal thickness images for predicting VF was investigated in numerous studies. The 3D CNN model was used in other studies [5], [14].

Although deep learning can be promising, the overfitting problem should be considered because medical data are usually limited and noisy. To handle noisy data, Yuhui et al. [14] used a regularization method with matrix factorization. However, the method only showed the difference between the ground truth and the common patterns, and detailed results cannot be provided because the model attempted to predict VFs that follow common patterns. In the present study, the simple ResNet-50 model was used instead of other complicated models to reduce the number of parameters. Since our dataset is relatively small, the simple ResNet-50 can reduce the overfitting problem. Moreover, the weighted loss was used to handle the noisy data. Because stable data (OCT images) were used to detect noisy samples, the proposed model can train on noisy data without losing individuality. In addition, data augmentation was applied.

V. CONCLUSION

In this paper, a deep learning model to predict future VF based on multiple inputs using previous VF and OCT images (thickness maps, vertical and horizontal tomograms) was introduced. We developed a CNN-RNN model to analyze both series data (previous VF) and image data (OCT images). To handle noisy data, a mechanism to detect and re-weight the

samples was introduced. The quantitative results showed that a re-weighting method can improve the performance in cases with or without OCT images.

The main limitation was the size of our dataset. The deep learning technique can benefit from a larger dataset. In the present study, only 266 patients with 1,527 samples were available for training and validation, 50 patients with 225 samples were available for testing. Although, many regularization methods are used such as data augmentation, cross-validation, and re-weighting, increasing the size of the dataset is more advantageous. In addition, information other than that of the VF and OCT tests was ignored, although non-image factors (gender, age, or treatment information) can be helpful in such analysis. These factors should be further explored in future studies.

REFERENCES

- [1] R. N. Weinreb, T. Aung, and F. A. Medeiros, "The pathophysiology and treatment of glaucoma: A review," *J. Amer. Med. Assoc.*, vol. 311, no. 18, pp. 1901–1911, 2014.
- [2] L. Quaranta, I. Riva, C. Gerardi, F. Oddone, I. Floriano, and A. G. P. Konstas, "Quality of life in glaucoma: A review of the literature," *Adv. Therapy*, vol. 33, no. 6, pp. 959–981, Apr. 2016.
- [3] S. I. Berchuck, S. Mukherjee, and F. A. Medeiros, "Estimating rates of progression and predicting future visual fields in glaucoma using a deep variational autoencoder," *Sci. Rep.*, vol. 9, no. 1, p. 18113, Dec. 2019.
- [4] S. Yousefi, T. Kiwaki, Y. Zheng, H. Sugiura, R. Asaoka, H. Murata, H. Lemij, and K. Yamanishi, "Detection of longitudinal visual field progression in glaucoma using machine learning," *Amer. J. Ophthalmol.*, vol. 193, pp. 71–79, Sep. 2018.
- [5] L. Xu, R. Asaoka, H. Murata, T. Kiwaki, Y. Zheng, M. Matsuura, Y. Fujino, M. Tanito, K. Mori, Y. Ikeda, T. Kanamoto, and K. Yamanishi, "Improving visual field trend analysis with OCT and deeply regularized latent-space linear regression," *Ophthalmol. Glaucoma*, vol. 4, no. 1, pp. 78–88, Jan. 2021.
- [6] K. Nouri-Mahdavi, V. Mohammadzadeh, A. Rabiolo, K. Edalati, J. Caprioli, and S. Yousefi, "Prediction of visual field progression from OCT structural measures in moderate to advanced glaucoma," *Amer. J. Ophthalmol.*, vol. 226, pp. 172–181, Jun. 2021.
- [7] Y. H. Hwang, Y. Y. Kim, H. K. Kim, and Y. H. Sohn, "Ability of cirrus high-definition spectral-domain optical coherence tomography clock-hour, deviation, and thickness maps in detecting photographic retinal nerve fiber layer abnormalities," *Ophthalmology*, vol. 120, no. 7, pp. 1380–1387, Jul. 2013.
- [8] A. Ferreras, E. Luis Pablo, F. David Garway-Heath, P. Fogagnolo, and J. Garcia-Feijoo, "Mapping standard automated perimetry to the peripapillary retinal nerve fiber layer in glaucoma," *Investigative Ophthalmol. Vis. Sci.*, vol. 49, no. 7, pp. 3018–3025, Jul. 2008.
- [9] J. Denniss, A. Turpin, and A. M. Mckendrick, "Relating optical coherence tomography to visual fields in glaucoma: Structure–function mapping, limitations and future applications," *Clin. Exp. Optometry*, vol. 102, no. 3, pp. 291–299, May 2019.
- [10] N. M. Jansonius, J. Schiefer, J. Nevalainen, J. Paetzold, and U. Schiefer, "A mathematical model for describing the retinal nerve fiber bundle trajectories in the human eye: Average course, variability, and influence of refraction, optic disc size and optic disc position," *Exp. Eye Res.*, vol. 105, pp. 70–78, Dec. 2012.
- [11] A. Tsai, J. Caprioli, and L. Q. Shen, "Coupled parametric model for estimation of visual field tests based on OCT macular thickness maps, and vice versa, in glaucoma care," *Med. Image Anal.*, vol. 16, no. 1, pp. 101–113, Jan. 2012.
- [12] T. Uesaka, K. Morino, H. Sugiura, T. Kiwaki, H. Murata, R. Asaoka, and K. Yamanishi, "Multi-view learning over retinal thickness and visual sensitivity on glaucomatous eyes," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association for Computing Machinery, 2017, Aug. 2017, pp. 2041–2050.

- [13] H. Sugiura, T. Kiwaki, S. Yousefi, H. Murata, R. Asaoka, and K. Yamanishi, "Estimating glaucomatous visual sensitivity from retinal thickness with pattern-based regularization and visualization," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association for Computing Machinery, 2017, Jul. 2018, pp. 783–792.
- [14] Y. Zheng, L. Xu, T. Kiwaki, J. Wang, H. Murata, R. Asaoka, and K. Yamanishi, "Glaucoma progression prediction using retinal thickness via latent space linear regression," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association for Computing Machinery, 2017, Jul. 2019, pp. 2278–2286.
- [15] K. Shon, K. R. Sung, and J. W. Shin, "Can artificial intelligence predict glaucomatous visual field progression? A spatial–ordinal convolutional neural network model," *Amer. J. Ophthalmol.*, vol. 233, pp. 124–134, Jan. 2022.
- [16] J. C. Wen, C. S. Lee, P. A. Keane, S. Xiao, A. S. Rokem, P. P. Chen, Y. Wu, and A. Y. Lee, "Forecasting future humphrey visual fields using deep learning," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0214875.
- [17] K. Park, J. Kim, and J. Lee, "Visual field prediction using recurrent neural network," *Sci. Rep.*, vol. 9, no. 1, p. 8385, Jun. 2019.
- [18] R. Asaoka, L. Xu, H. Murata, T. Kiwaki, M. Matsuura, Y. Fujino, M. Tanito, K. Mori, Y. Ikeda, T. Kanamoto, K. Inoue, J. Yamagami, and K. Yamanishi, "A joint multitask learning model for cross-sectional and longitudinal predictions of visual field using OCT," *Ophthalmol. Sci.*, vol. 1, no. 4, Dec. 2021, Art. no. 100055.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [21] C. A. Curcio and K. A. Allen, "Topography of ganglion cells in human retina," *J. Comparative Neurol.*, vol. 300, no. 1, pp. 5–25, Oct. 1990.
- [22] Y. H. Kwon, J. H. Fingert, M. H. Kuehn, and W. L. M. Alward, "Primary open-angle glaucoma," *New England J. Med.*, vol. 360, no. 11, pp. 1113–1124, Mar. 2009.
- [23] P. Y. Ramulu, S. K. West, B. Munoz, H. D. Jampel, and D. S. Friedman, "Driving cessation and driving limitation in glaucoma: The salisbury eye evaluation project," *Ophthalmology*, vol. 116, no. 10, pp. 1846–1853, 2009.
- [24] P. Y. Ramulu, S. K. West, B. Munoz, H. D. Jampel, and D. S. Friedman, "Glaucoma and reading speed: The Salisbury eye evaluation project," *Arch. Ophthalmol.*, vol. 127, no. 1, pp. 82–87, Jan. 2009.
- [25] S. Hu, N. D. Smith, L. J. Saunders, and D. P. Crabb, "Patterns of binocular visual field loss derived from large-scale patient data from glaucoma clinics," *Ophthalmology*, vol. 122, no. 12, pp. 2399–2406, Dec. 2015.
- [26] H.-C. Cheng, C.-Y. Guo, M.-J. Chen, Y.-C. Ko, N. Huang, and C. J.-L. Liu, "Patient-reported vision-related quality of life differences between superior and inferior hemifield visual field defects in primary open-angle glaucoma," *J. Amer. Med. Assoc. Ophthalmol.*, vol. 133, no. 3, pp. 269–275, Mar. 2015.
- [27] H. Sawada, T. Yoshino, T. Fukuchi, and H. Abe, "Assessment of the vision-specific quality of life using clustered visual field in glaucoma patients," *J. Glaucoma*, vol. 23, no. 2, pp. 81–87, Feb. 2014.
- [28] Y. Hashimoto, R. Asaoka, T. Kiwaki, H. Sugiura, S. Asano, H. Murata, Y. Fujino, M. Matsuura, A. Miki, K. Mori, Y. Ikeda, T. Kanamoto, J. Yamagami, K. Inoue, M. Tanito, and K. Yamanishi, "Deep learning model to predict visual field in central 10° from optical coherence tomography measurement in glaucoma," *Brit. J. Ophthalmol.*, vol. 105, no. 4, pp. 507–513, 2021.



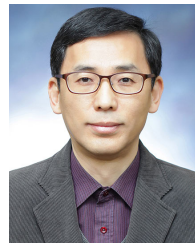
QUANG T. M. PHAM received the B.S. degree from the Ho Chi Minh City University of Technology, Vietnam, in 2017. He is currently pursuing the Ph.D. degree in computer engineering with the Media System Laboratory, Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea. His research interests include GAN and medical image analysis.



JONG CHUL HAN received the M.D. degree from Sungkyunkwan University, in 2005, and the Ph.D. degree from the School of Medicine, Sungkyunkwan University, in 2019. Since 2016, he has been an Assistant Professor with the Department of Ophthalmology, Samsung Medical Center, Seoul, South Korea. His current research interest includes translational ophthalmology research, especially in the field of glaucoma.



DO YOUNG PARK received the M.D. and M.S. degrees from Sungkyunkwan University, in 2008 and 2013, respectively, and the Ph.D. degree from the Korea Advanced Institute of Science and Technology, in 2017. Since 2020, he has been a Professor with the Department of Ophthalmology, Yeungnam University College of Medicine, Daegu, South Korea. His current research interest includes translational ophthalmology research, especially in the field of glaucoma.



JITAE SHIN (Member, IEEE) received the B.S. degree from Seoul National University, in 1986, the M.S. degree from the Korea Advanced Institute of Science and Technology, in 1988, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, USA, in 1998 and 2001, respectively.

For former industrial experiences, he worked with Korea Electric Power Corporation and the Korea Atomic Energy Research Institute, from 1988 to 1996. He has been a Professor with the School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, South Korea, since 2002. His current research interests include image/video signal processing using deep learning, medical image processing, and machine learning for wireless/mobile communication systems.

• • •