SciencePG
Science Publishing Group

Finding the Best Performing Pre-Trained CNN Model for Image Classification: Using a Class Activation Map to Spot Abnormal Parts in Diabetic Retinopathy Image

Jihyung Kim

Walter Johnson High School, North Bethesda, Maryland, USA

Email address:
jihyungkim09@gmail.com

Abstract: Diabetic retinopathy (DR) is a common eye disease that people get from diabetes. About 33.7% of the people with diabetes have DR. With our datas, which are pictures of the eyeball with and without DR, we tried different convolutional neural network (CNN) models to get the best accuracy score. We tested our datas with a default CNN model, and 5 different pre-trained models: MobileNet, VGG16, VGG19, Inception V3, and Inception ResNet V2. The default CNN model didn't perform very well, getting only 10.4%. The pre-trained model also didn't perform as good as expected, so we decided to use GRU with the models, which increases the score. For the higher accuracy, we added bidirectional GRU to train the whole parameters in the model. The 5 different pre-trained models scored an average of 74.2% accuracy score, and Inception ResNet V2 with bidirectional GRU included scored the highest accuracy, achieving 83.57%. For additional study, we used a class activation map to spot the abnormal parts of the eyeball with DR, and we could spot abnormal veins and bleeding on the eyeball. However, our research has limitations on that we did not use the segmentation methods, which is more advanced technique compared to classification, such as U-net, Fully Convolutional Network (FCN), Deep Lab V3, and Feature Pyramid Network. Furthermore, even though our model classified 5 different classes, the fact that the highest accuracy score was lower than 90% is also a limitation. For further study, we would prepare a masking method for applying segmentation methods to our dataset.

Keywords: Diabetes, Diabetic Retinopathy, Inception-Resnet-v2, Bi-GRU, CNN

1. Introduction

1.1. Background

Diabetes is a lifelong disease that can happen from the low production of insulin. There are two different types of diabetes. The cause of Type-1 is lack of production of insulin by nature, and type-2 diabetes happens when the body resists producing insulin as growing up. The main reasons for type 2 are inactivity and being overweight. Diabetic retinopathy is a complication that patients get on their eyes when they get diabetes. Diabetic retinopathy can cause swelling in the retina, which is called diabetic macular edema (DME). DME can result in blurring the vision. Diabetic retinopathy also can cause neovascular glaucoma, which is abnormal blood vessels growing out of the retina, and blocking fluid from draining out. The most severe result of diabetic retinopathy could be blindness, with a very low chance [1]. Figure 1 shows the proportion of diabetic patients with diabetic retinopathy out of total diabetic patients. Out of 395 total patients, there were 133 patients with the diabetic retinopathy, which is 33.7%. It's not a high chance of getting retinopathy, but it is pretty common. The graph shows 34% of the patients have diabetic retinopathy. These days, a lot of scientists and doctors use deep learning techniques to diagnose diseases such as diabetic retinopathy. Deep learning performs outstanding results on these topics, too [2].
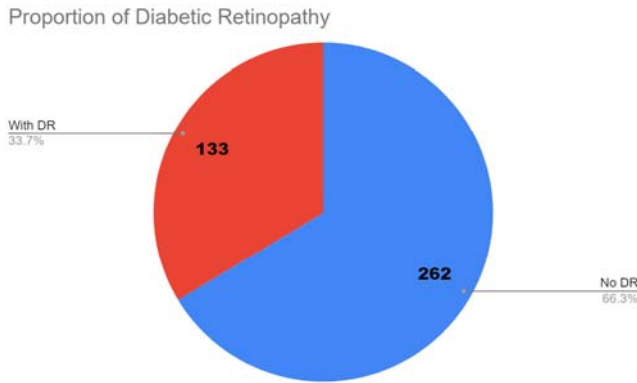
Figure 1. *Proportion of Diabetic Retinopathy.*

1.2. Objective

A lot of people with diabetes are not worried about diabetic retinopathy because only 34 percent of them get it, and people don't think that he/she will be that 34 percent. But it is really important to take care of it since the diseases can cause such severe damages. Our objective is to find a deep learning model that performs the best on the dataset to increase the accuracy and detecting abnormal parts of the eyeball to diagnose diabetic retinopathy. In order to do that, we have to try different models to find out the model that fits with the dataset the best and use the heat map feature to find out the abnormal parts of the eyeball [3].

2. Related Works

Convolutional Neural Networks perform very well in detecting and predicting pictures and images. Qummar et al. used pretrained CNN models such as Resnet50, Inceptionv3, Xception, Dense121, Dense169 to predict retinopathy, and the models with fine tuning yielded 70% accuracy [3]. Kwasigroch used CNN models and worked well to achieve decent scores getting 82% accuracy score, and Kappa score equal 0.776, using the datasets provided by the EyePACKS organization [4]. Ghosh et al. also used the CNN method, and got an accuracy score of 95% for the binary classification, and an accuracy of 85% for the five classification on 3,000 validation image dataset from Kaggle [5]. Qomariah et al. used Support Vector Machine (SVM), using the CNN as the input features for classification. They got the highest accuracy values of 95.83% and 95.24% for base 12 and base 13 [6]. Mobeen-ur-Rehman et al. used 3 different pre-trained CNN models: AlexNet, VGG-16, and SqueezNet. Those pre-trained models recorded very high accuracy scores which are 93.46%, 91.82%, and 94.49%. They also used the Messidor datasets [7].

3. Materials and Methods

3.1. Data Description

In our study, we used about 30,000 different image data that are provided by Kaggle, which is available at https://www.kaggle.com/sovitrath/diabetic-retinopathy-

224x224-gaussian-filtered [8]. Kaggle collected the images from various sources. They classified the images into 5 different levels of Diabetic Retinopathy (DR): No DR, Mild DR, Moderate DR, Severe DR, and Proliferate DR. We divided the data into 2 groups, which are training data, and testing data. We divided each level of DR into those groups, and the proportion was 7:3 where training data is 70 percent, and testing data is 30 percent. The figures below show an example of each level of DR.
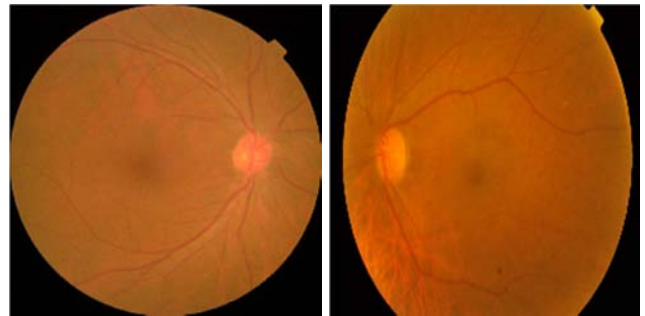

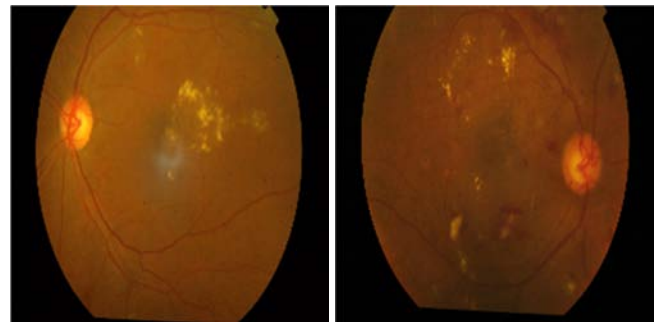**Figure 2.** *Visualization of Diabetic Retinopathy: no DR and mild DR.*

Figure 3. *Visualization of Diabetic Retinopathy: moderate DR and severe DR.*
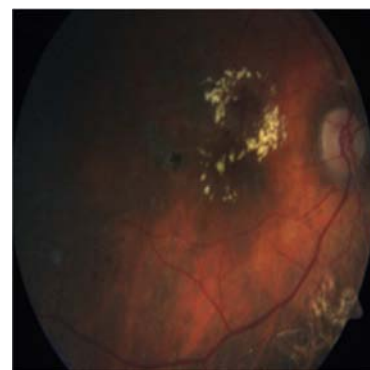

Figure 4. *Visualization of Diabetic Retinopathy: proliferate DR.*

3.2. GRU

Gated Recurrent Units (GRU) has appeared to solve a downside of RNN (Recurrent Neural Network), which are gradient vanishing and exploding problems in backpropagation through time. GRU is composed of reset gate, update gate and candidate. The purpose of reset gate is to reset the past information by multiplying (0, 1) to the prior hidden layer through sigmoid function (1). Update gate is

similar to the forget gate and input gate of Long Short Term Memory (LSTM) and it decides the ratio of updating past and current information (2). Candidate is a stage for calculating the information candidates (3). In the hidden layer stage, the calculation is performed through adding the outcome of the update gate and candidate gate (4) [9].

3.3. CNN

CNN (Convolutional Neural Network) consists of the convolutional layer, pooling layer and fully connected layer. When the CNN gets the input image, the convolutional layer first convolutes and yields a feature map through a filter, which is also called a kernel. Then, the pooling layer reduces the size of the feature map through calculating the average or maximum values of the feature map, each process is called average pooling and max pooling. The fully connected layer is just like a deep neural network (DNN), the main purpose of this layer is to classify the target with an activation function. For multi-class classification, softmax function is used as the activation function and for the binary classification, sigmoid is mainly used [10].
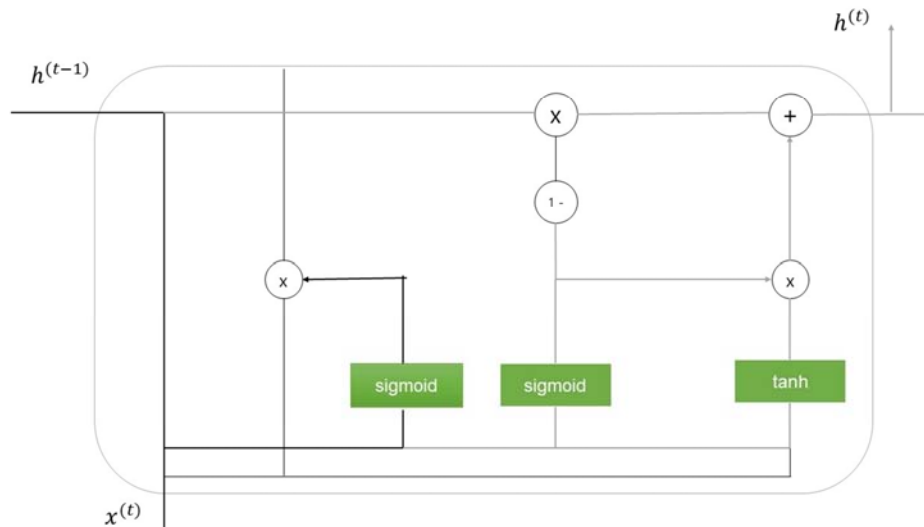

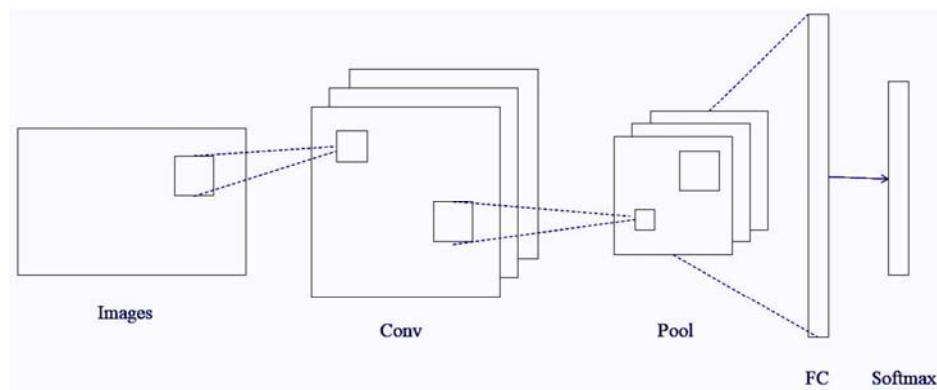*Figure 5. Overall architecture of GRU.*

Figure 6. Overview architecture of CNN.

3.4. Pretrained CNN

For extracting the features from the images, we utilized pre-trained CNN models such as VGG16, VGG19, MobileNet, Inception-Resnet_v2, Inception_v3. These pretrained models were downloaded from Keras. They were pre-trained with the dataset named ImageNet. Specifically, the Inception-Resnet_v2 has an image input shape as 299 x 299 x 3. After those models extracted the features from the images, then the user-defined layers were used for classifying the targets [11].

3.5. Proposed Model

Our proposed model consists of Inception-Resnet-v1 and Bidirectional GRU (Bi-GRU). After pretrained Inception-Resnet-v1 extracts the features from the given images, the lambda layer reshapes the output of the flatten layer to transmit to Bi-GRU. Unlike GRU, Bi-GRU allows end-to-end learning, which minimizes loss from the output through training the whole parameters. Adam was used as an optimizer with learning rate 0.0001. Furthermore, all of our datasets were preprocessed through the image data generator

function because the number of given datasets were not enough to be trained efficiently. The image data generator function augmented the images through changing a width, height, brightness [12].
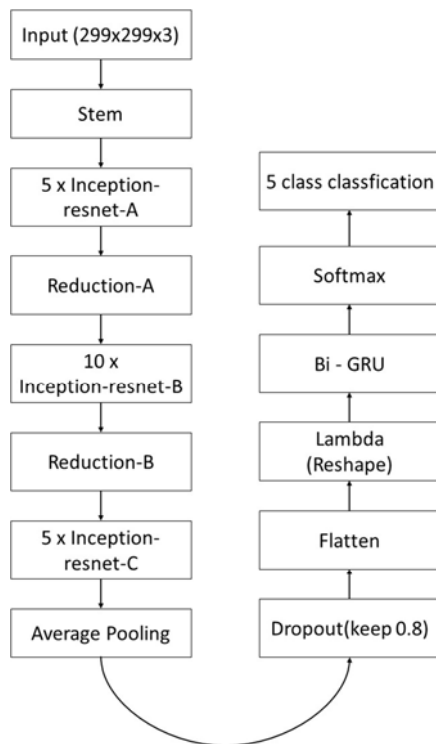

Figure 7. *Summary of proposed model.*
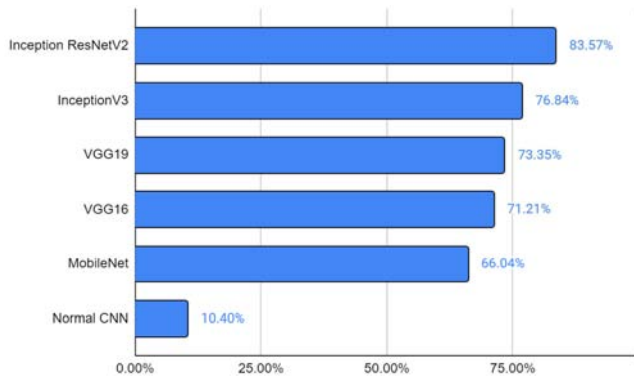
4. Results

Figure 8. *Comparison of model accuracy.*

In the beginning of our study, we tried to use a pure CNN model, but we only got a 10.4% accuracy score. So we decided to use these pretrained models from keras: MobileNet, VGG16, VGG19, Inception V3, and Inception ResNet V2. The benefit of these models is that these models are already trained very well based on countless datasets, so they usually perform better. As expected, the pre-trained model reached way higher accuracy than the normal CNN model. MobileNet reached to 66.04% accuracy score, VGG16 got 71.21% accuracy score, VGG19, a newer version of VGG16 achieved 73.35% accuracy score, inception V3

scored a very high score which is 76.84% accuracy score, and finally, Inception ResNet V2 performed the highest score, 83.57% accuracy score. Also, in order to find out the erratic part of the eyeball with the diabetic retinopathy, we used a class activation map to create a heatmap based on the picture datas we have. The figure 9 and figure 10 below are the heatmap, and the blue-purple area is showing the normal parts, and the red-yellow area is showing the abnormal area [13].
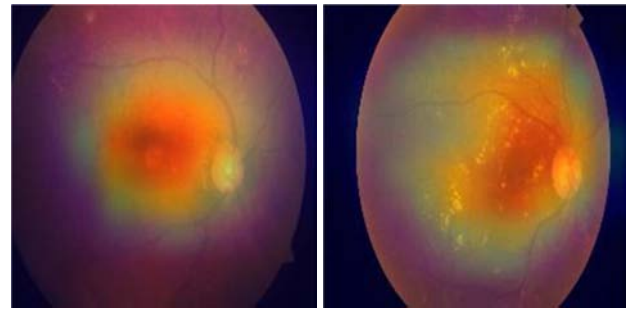

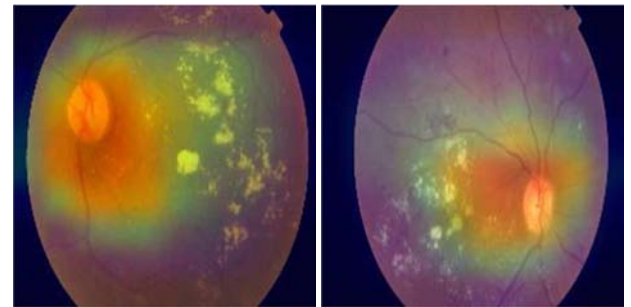***Figure 9.*** *Results of grad CAM: mild DR, moderate DR.*

Figure 10. *Results of grad CAM: severe DR, proliferate DR.*

Figure 11 and figure 12 show the accuracy and loss graph of training dataset and validation dataset. Those graphs show that accuracy from training and validation keep increasing during the epoch. Also, the loss graph from the both dataset also yields that loss kept decreasing during the epoch, and we could conclude that training was done efficiently, with a low probability of over fitting.
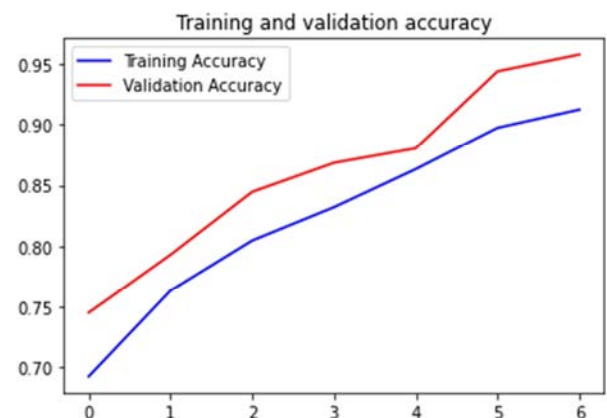

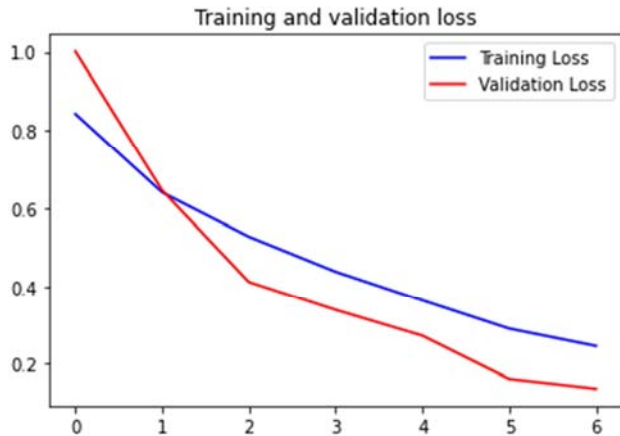***Figure 11.*** *Graph of training and validation accuracy.*

Figure 12. *Graph of training and validation loss.*

5. Discussion

5.1. Principal Finding

Our research was outstanding because we could perform very well because we tried out various kinds of models to find out the best model for the dataset. Also, we used class activation maps to figure out the abnormal part of the eye with diabetic retinopathy (DR) so it's easier to find out whether the patients have diabetic retinopathy [14]. Also, Wu et al. shows that using a CNN model with Gated Recurrent Units (GRU) helps improve the accuracy (Wu When using only Inception ResNet V2, we only got a 77.97% accuracy score, so we decided to also include GRU. Using GRU combined, our accuracy score improved by 5.6%. By combining GRU and pre-trained model, we could train it as bi-directional GRU to improve the results even better [15].

5.2. Limitation

Our limitation was that we could've used a better way to figure out the abnormal part of the eye with segmentation methods such as U-net, Fully Convolutional Network (FCN), Deep Lab V3, and Feature Pyramid Network. It is because the class activation map worked inaccurately for some of the datas. Another limitation was that the accuracy wasn't as good as expected even though we tried a lot of different ways to get higher accuracy. For the next study, we are going to aim for at least 90% using CNN models.

6. Conclusion and Recommendations

We used different kinds of CNN models to get the best accuracy score, and we used class activation map to figure out abnormal parts of the eyeball with diabetic retinopathy. As a result, we got an 83.57% accuracy score using Inception ResNet V2, which is a pre-trained model. We combined the pre-trained model with GRU to get even better results. The heatmap showed that there were more abnormal veins and bleeding on the eyeball with DR. Our study was outstanding because we did an additional study with the heatmap in order to figure out the difference between normal eyeball and eyeball with DR. But the heatmap program didn't perform well on some of the eyeball pictures, so for the next study, we will try different heatmap methods to find a program that performs better.

References

[1] *Diabetic retinopathy*. (n.d.). https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/diabetic-retinopathy.

[2] Gargeya, R., & Leng, T. (2017). Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*, *124* (7), 962–969. https://doi.org/10.1016/j.ophtha.2017.02.008.

[3] Qummar, S., Khan, F. G., Shah, S., Khan, A., Shamshirband, S., Rehman, Z. U.,… Jadoon, W. (2019). A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection. *IEEE Access*, *7*, 150530–150539. https://doi.org/10.1109/access.2019.2947484.

[4] Kwasigroch, A., Jarzembinski, B., & Grochowski, M. (2018). Deep CNN based decision support system for detection and assessing the stage of diabetic retinopathy. *2018 International Interdisciplinary PhD Workshop (IIPhDW)*. https://doi.org/10.1109/iiphdw.2018.8388337.

[5] Ghosh, R., Ghosh, K., & Maitra, S. (2017). Automatic detection and classification of diabetic retinopathy stages using CNN. *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*. https://doi.org/10.1109/spin.2017.8050011.

[6] Qomariah, D. U., Tjandrasa, H., & Fatichah, C. (2019). Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM. *2019 12th International Conference on Information & Communication Technology and System (ICTS)*. https://doi.org/10.1109/icts.2019.8850940.

[7] Mobeen-ur-Rehman, Khan, S. H., Abbas, Z., & Danish Rizvi, S. M. (2019). Classification of Diabetic Retinopathy Images Based on Customised CNN Architecture. *2019 Amity International Conference on Artificial Intelligence (AICAI)*. https://doi.org/10.1109/aicai.2019.8701231.

[8] Rath, S. R. (2020, February 18). *Diabetic Retinopathy 224x224 Gaussian Filtered*. Kaggle. https://www.kaggle.com/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered.

[9] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv: 1412.3555*.

[10] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *2017 International Conference on Engineering and Technology (ICET)*. https://doi.org/10.1109/icengtechnol.2017.8308186.

[11] Abdullah, & Hasan, M. S. (2017). An application of pre-trained CNN for image classification. *2017 20th International Conference of Computer and Information Technology (ICCIT)*. https://doi.org/10.1109/iccitechn.2017.8281779.

[12] Vukotić, V., Raymond, C., & Gravier, G. (2016). A Step Beyond Local Observations with a Dialog Aware Bidirectional GRU Network for Spoken Language Understanding. *Interspeech 2016.* https://doi.org/10.21437/interspeech.2016-1301.

[13] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

[14] Kwasniewska, A., Ruminski, J., & Rad, P. (2017). Deep features class activation map for thermal face detection and tracking. *2017 10th International Conference on Human System Interactions (HSI).* https://doi.org/10.1109/hsi.2017.8004993.

[15] Wu, L., Kong, C., Hao, X., & Chen, W. (2020). A Short-Term Load Forecasting Method Based on GRU-CNN Hybrid Neural Network Model. *Mathematical Problems in Engineering*, *2020*, 1–10. https://doi.org/10.1155/2020/1428104.