Towards a Smart Dental Healthcare: An Automated Assessment of Orthodontic Treatment Need

Seiya Murata
Graduate School of Information Science and Technology, Osaka University
Email: murata.seiya@ais.cmc.osaka-u.ac.jp

Kobo Ishigaki
Cybermedia Center
Osaka University
Email: kobo.ishigaki@ais.cmc.osaka-u.ac.jp

Chonho Lee
Cybermedia Center
Osaka University
Email: leech@cmc.osaka-u.ac.jp

Chihiro Tanikawa
Graduate School of Dentistry
Osaka University
Email: ctanika@dent.osaka-u.ac.jp

Susumu Date
Cybermedia Center
Osaka University
Email: date@cmc.osaka-u.ac.jp

Takashi Yoshikawa
Cybermedia Center
Osaka University
Email: tyoshikawa@cmc.osaka-u.ac.jp

***Abstract*—With increasing demands for dental healthcare becoming one of the regular life health factors, this work focuses on the automation of diagnostic imaging in the field of orthodontics. The automated diagnostic imaging of oral images can evaluate the severity of malocclusion and jaw abnormality, and it is beneficial for both doctors reducing their workload and patients periodically performing self-assessment without visiting clinics. In this paper, we propose a deep learning-based model that assesses oral images and gives the severity of orthodontic treatment need. Unlike a traditional image classification model, the proposed model successfully deals with the case that one class label (e.g., the severity score) is assigned to a set of images (e.g., oral images of a patient). The experimental results show that the proposed model improves the classification accuracy by 11% (18% in the best) compared to other conventional models.**

Keywords–Orthodontic treatment; Diagnostic imaging; Deep learning.

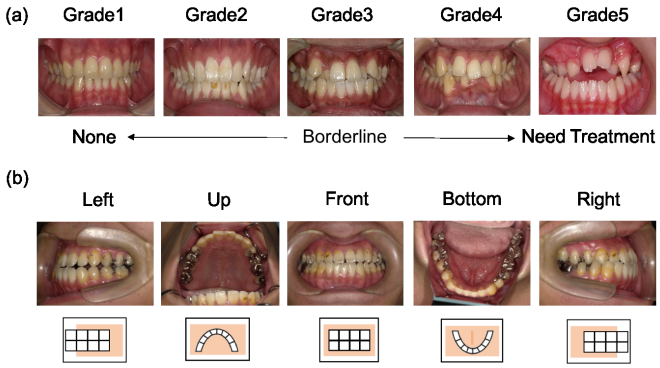

Figure 1. (a) IOTN Grades in 5 scales and the corresponding sample oral images. (b) Oral images taken from five different directions and their illustrations (used in the rest of the paper).

I. INTRODUCTION

The recent breakthrough in image recognition technology using deep convolutional neural network (CNN) model [1][2] brings further improvement in diagnostic imaging that can diagnose the presence of tuberculosis in chest x-ray images [3], detect diabetic retinopathy from retinal photographs [4], as well as locate breast cancer in pathology images [5]. The automated diagnostic imaging is eagerly desired in the field of orthodontics as well, along with the increasing demands for dental healthcare, becoming one of the regular life health factors. For example, it enables individuals to self-check the degree of malocclusion and jaw abnormality from oral and facial images, which are the causes of masticatory dysfunction, apnea syndrome and pyorrhea, etc. Moreover, it leads to providing objective diagnosis that is important for both doctors and patients because the diagnosis directly affects the treatment plan, treatment priority and insurance coverage.

Orthodontists generally use Index of Orthodontic Treatment Needs (IOTN) to determine whether individuals qualify for further orthodontic treatment. IOTN [6] is one of the severity measures for malocclusion and jaw abnormality, which determines whether orthodontic treatment is necessary. Typically, the value ranges between Grade 1 (None) and Grade 5 (Need Treatment) as shown in Figure 1(a). A primary care doctor or general dentist checks the dental healthcare of his/her patient with IOTN, and if the score is high, he/she refers the patient to the other specialist for further treatment. The IOTN assessment is a significant key process to prevent oral diseases from becoming worse. However, to provide the accurate assessment of IOTN requires special training. Many patients tend to miss the appropriate treatment timing due to an incorrect assessment by an inexperienced doctor.

Here, we consider the automation of IOTN assessment, which brings several benefits as follows. Firstly, it helps provide an objective diagnosis that minimizes the diagnosis variation among doctors. The objective diagnosis is quite useful for Informed Consent and training inexperienced doctors. Secondly, the automated diagnostic imaging is highly expected to assist doctors reducing their workload. For example, in Osaka University Dental Hospital, a few doctors take care of over a hundred patients every day. What's more, it benefits people who are able to take their oral photo using smartphone or mobile device, and periodically perform self-assessment at remote without visiting clinics.

To achieve the automation of IOTN assessment, we employ a deep CNN model based on historical records, i.e., oral images of patients, and solve a problem as an image classification. However, there is an issue that we cannot simply apply a conventional CNN to our problem. Unlike typical image classification problems assuming that each image is paired with

 ISBN: 978-1-61208-597-5

one class or label, one class (i.e., a IOTN grade) is paired with a set of images of a patient. As shown in Figure 1(b), each patient is taken his/her oral images from five different directions, and one IOTN grade is given to each patient. There might be a case that a malocclusion at right lower molars is observed in the right image but cannot see in the left and up images, and the alignment of left teeth is clean.

In this paper, we propose a parallel CNN model that independently runs multiple CNNs, each of which deals with images taken from one direction, and then concatenates feature vectors (i.e., outputs of the multiple CNNs) to one vector, namely a patient vector. The patient vector preserves the feature information of all images of a patient. It is input to a multi-layer perceptron (MLP) whose output is one of IOTN grades. We verify that the proposed model achieves 11% (18% in the best case) improvement in its accuracy, compared to a MLP and a CNN model.

The remainder of this paper is organized as follows. Section II introduces some related work in image classification. Section III first reviews a CNN, and explains why the conventional CNN does work well to our problem, followed by the description of the proposed model. Section IV shows the evaluation and discusses some future work.

II. Related Work

A large number of researches has been done already in conventional machine learning. Image classification is typically performed in two steps, a feature extraction and a classification. It relies much more on the extraction of the features of targets in images. Even though many researches have proposed their own features, such as Histograms of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), Haar-like, etc., a great amount of time and effort has been spent.

Recently, deep learning technique has emerged as a powerful approach to solve many problems in computer vision fields [7][8]. Convolutional neural network (CNN) [1][2], especially, has been successfully applied for image classification, object recognition, segmentation, etc. It lets the model automatically learn the features of targets in images using a large scale of training data.

The deep learning technique brings further improvement in diagnostic imaging as well, which can diagnose the presence of tuberculosis in chest x-ray images [3], detect diabetic retinopathy from retinal photographs [4], as well as locate breast cancer in pathology images [5]. However, a little has been done in the filed of dentofacial orthopedics.

Typically image classification problems assume that each image is labelled with one or more classes. Unlikely, the problem we focus in this paper is that one class label (e.g., a IOTN grade) is assigned to a set of images (e.g., for a patient); thus, a conventional CNN cannot be directly applied. There is a promising technique called boosting [9][10] that creates a strong classifier/predictor by combining multiple weak or less accurate classifiers/predictors trained by sampled data. Inspired by the idea of such aggregation, we consider multiple CNNs, each of which takes care of different subset of images. However, it is difficult to find the way of how we aggregate the results of each CNN. Thus, we try to investigate a concatenated feature (i.e., a representation of a patient), implemented in the proposed model.

III. The Proposed Model

As briefly explained in Introduction, we design a deep learning-based model to classify sets of patients' oral images into the corresponding IOTN grades. This section first reviews a convolutional neural network (CNN); then explains a reason that we cannot simply apply the conventional CNN model to our problem in Section III-B, and then describes the proposed parallel CNN model in Section III-C.

A. Review of CNN

A conventional CNN for image classification consists of multiple, repeating components that are stacked in layers namely convolution, pooling, fully-connected and softmax layers.

Convolution layer is to extract features from the input image. Convolution preserves the spatial relationship between pixels by learning image features using small squares called filters. The convolution operator convolves the input $\mathbf{x} = \{x_{ij}\}$ with a filter $\mathbf{w} = \{w_{pq}\}$. The output for a neuron at (i,j) in the next layer ℓ is computed by

$$\begin{aligned} z_{ij}^{(\ell)} &= f(u_{ij}) \\ &= f(\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} w_{pq} \cdot z_{i+p,j+q}^{(\ell-1)} + b_{ij}) \end{aligned}$$

where M indicates the filter size, b is a bias and $z_{ij}^{(0)} = x_{ij}$. f applies the nonlinearity to the convoluted value, namely an activation function. (Note that the layer index $^{(\ell)}$ of w and b is omitted for simplicity.) When considering the input with K channels and S filters, we need to sum up u_{ij} for all channels using each of S filters by

$$\begin{aligned} z_{ij,s}^{(\ell)} &= f(u_{ij,s}) \\ &= f(\sum_{k=0}^{K-1}\sum_{p=0}^{M-1}\sum_{q=0}^{M-1} w_{pqk,s} \cdot z_{i+p,j+q,k}^{(\ell-1)} + b_{ij,s}). \end{aligned}$$

Pooling layer is normally operated in-between successive convolution layers to reduce the spatial size of the representation and the amount of parameters. The pooling layer operates independently on every depth slice of the input using max or averaging operation. The most common max pooling operation downsamples the input using $H \times H$ filter by

$$u_{ijk} = \max_{(p,q)\in P_{i,j}} z_{pqk}$$

where P_{ij} indicates a set of pixels in any $H \times H$ subregion of input, whose center is (i, j).

Fully-connected layer is a traditional neural network layer where the features of the next layer are a linear combination of the features of the previous layer. The output value is computed by

$$y_k^{(\ell)} = f(\sum_h w_{k,h} \cdot x_h^{(\ell-1)} + b_k)$$

where y_k is the k-th neuron, and $w_{k,h}$ is the weight between x_h and y_k.

 ISBN: 978-1-61208-597-5

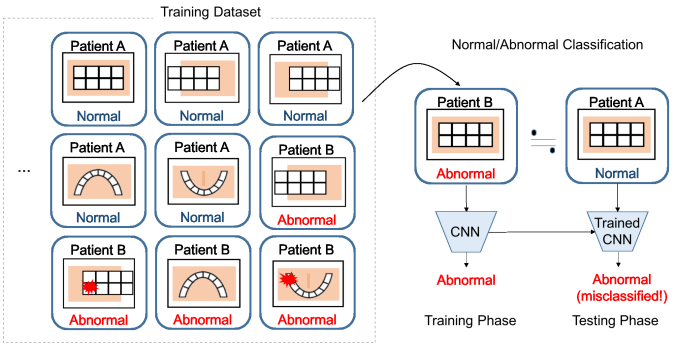

Figure 2. An illustration of misclassification due to the mislabelled training data. The blue rounded box indicates one input sample including a patient image with the corresponding label.

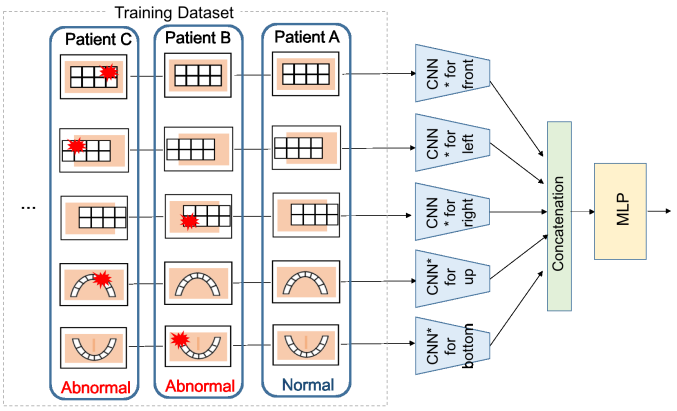

Figure 3. An illustration of the proposed parallel CNN model. The blue rounded box is one input sample including a set of five images of a patient. "CNN*" denotes a CNN without fully-connected and softmax layers

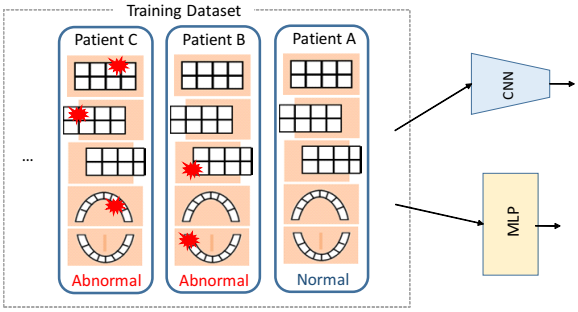

Figure 4. An illustration of alternative training dataset. The blue rounded box indicates one input sample containing one image combining five images at pixel level, and the corresponding label.

Finally, Softmax layer converts feature vectors into class probabilities. It normalizes the vector of scores by

$$y_k^{(L)} = \text{Prob}(\text{class} = j|\mathbf{x}, \mathbf{w}) = \frac{\exp(y_j^{(L)})}{\sum_{j=1}^{C} \exp(y_j^{(L)})}.$$

Then, the model is trained (i.e., updates their weights) in such a way that the class label with the highest probability becomes a true label. Note that multi-layer perceptron (MLP) usually indicates the neural network consisting of fully-connected layers and a softmax layer.

B. Issues to be considered

As a preliminary evaluation, we investigated the classification accuracy when using a conventional CNN model. We collected 300 patients' images (i.e., 1,500 images in total), and we intentionally assigned each image with one of two labels, *Normal* if IOTN grade is less than or equal to 3 and *Abnormal* if IOTN grade is larger than 3. In the result, we observed 60.4% of binary classification accuracy by 6-fold cross-validation.

The reason of such low accuracy comes from the mislabeling of training data because of the mixed oral images with five different directions. As explained, one IOTN grade is given to a patient based on a set of his/her oral images. There might be a case that a malocclusion at right lower molars is observed in the right images but cannot see in the left and up images. For example, as illustrated in Figure 2, Patient B has a problem in his right lower molars that can be observed in the right and bottom images, but the other front, left and up images are clean even though it was labelled as *Abnormal*. In such case, misclassification occurs when classifying Patient A's *Normal* front image, which is similar to Patient B's front image.

In fact, doctors did not set a label for each image. It is time-consuming work or nearly impossible for doctors to annotate each of all images with correct label.

C. Description of the Proposed Model

In order to solve the issue, we consider a promising technique called boosting [9][10], one of the ensemble learning approaches, which creates a strong classifier/predictor by combining multiple weak or less accurate classifiers/predictors trained by sampled data. We employ the idea of such aggregation. For doing so, we revise the format of training dataset in such that each input sample contains five images (of different directions) of a patient and one corresponding label. Then, we run multiple CNNs, each of which deals with images taken from one direction. However, the results of CNNs are still independent each other, so it is difficult to find the way of how we aggregate the results to improve the accuracy.

Thus, we design a parallel CNN model as illustrated in Figure 3. The proposed model independently runs multiple CNNs and concatenates intermediate feature vectors from the multiple CNNs to one vector preserving all information of different directions. The concatenated vector is named a patient vector, which is the input of a multi-layer perceptron (MLP) whose output is one of class labels after a softmax operation. In this model, CNNs (denoted by "CNN*" in Figure) do not have fully-connected layers and softmax layers (except the last layer of the proposed model).

As an additional investigation, we performed another experiment. In this experiment, we train a MLP or a conventional CNN using the dataset where five images are combined to one image at pixel level as illustrated in Figure 4. However, the accuracy does not improve as expected. We discuss more about it in the next section.

IV. Evaluation

This section shows the experimental results to evaluate the proposed model in terms of the classification accuracy of IOTN assessment, and compares it with a few different models.

 ISBN: 978-1-61208-597-5

TABLE I. THE PROPOSED MODEL STRUCTURE AND PARAMETER VALUES.

Name	Filter size	Stride	Output size	Activation
input	-	-	90 × 120 × 3	-
conv1	3 × 3	1	90 × 120 × 32	ReLU
pool1	2 × 2	2	45 × 60 × 32	-
conv2	3 × 3	1	45 × 60 × 64	ReLU
pool2	2 × 2	2	23 × 30 × 64	-
conv3	3 × 3	1	23 × 30 × 128	ReLU
pool3	2 × 2	2	12 × 15 × 128	-
conv4	3 × 3	1	12 × 15 × 256	ReLU
pool4	2 × 2	2	6 × 8 × 256	-
conv5	3 × 3	1	6 × 8 × 512	ReLU
pool5	2 × 2	2	3 × 4 × 512	-
concat	-	-	3 × 4 × (512 × 5)	-
flat	-	-	1 × 30720	-
fc1	-	-	1 × 4096	ReLU
fc2	-	-	1 × 512	ReLU
fc3	-	-	1 ×# of classes	Softmax

TABLE II. THE CLASSIFICATION ACCURACY OF DIFFERENT MODELS (%).

Model	2-class		5-class		2-class - G3	
	Mean	Best	Mean	Best	Mean	Best
MLP	58.2	59.5	20.9	21.2	-	-
CNN	60.4	61.4	21.3	21.6	-	-
Parallel CNN	71.8	79.0	40.2	45.0	73.1	81.3

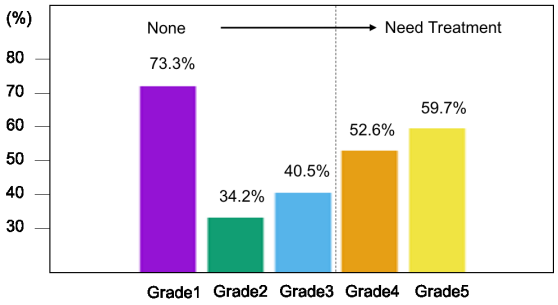

Figure 5. The classification accuracy of 5-class case for IOTN grades.

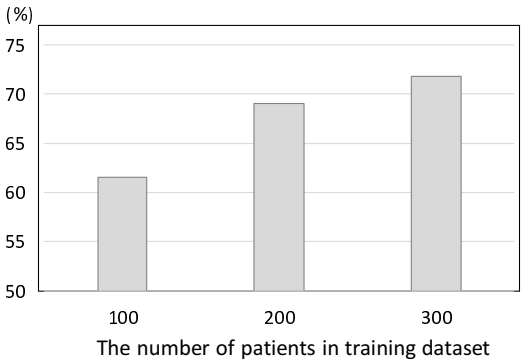

Figure 6. The classification accuracy of the model trained over the different size of training samples.

A. Environment Settings

For this experiment, we use a machine of Windows10 with Intel(R) Xeon(R) CPU E5-1620 v3 @ 3.50GHz and 16GB memory. The proposed model is trained with GeForce GTX TITAN x 12GB. We implement the model using TensorFlow [11].

Department of Orthodontics and Dentofacial Orthopedics in Osaka University Dental Hospital provides a dataset containing the oral images (taken from five directions) of 300 patients for this experiment. Each patient is assessed by one IOTN grade, and there are 60 patients per each grade.

B. Experimental Results

Table I shows a list of model parameters such as the filter size of convolution and pooling layers, and their stride size. The model consists of five pairs of convolution and pooling layers followed by a concatenation layer and three layers of MLP. We use Retified Linear Unit (ReLU) function [12] for an activation function. During a training phase, we set a learning rate to 1e-4, and use Adam [13] for the optimizer. Although there are various choices of hyper-parameters for layers, activation functions, learning rates, and optimizers, we showed the best case among several trial and error. For the accuracy evaluation, we perform 6-fold cross-validation by 90%/10% of training/validation data.

Table II shows the classification accuracy of different models. ”2-class” indicates a binary classification of Normal (Grades 1,2,3) and Abnormal (Grades 4,5). ”5-class” indicates the classification of five IOTN grades. ”2-class-G3” indicates a binary classification of Normal except Grade 3 and Abnormal. The proposed model successfully improves the mean and best of cross-validation accuracy by 11% (18% in the best case) for 2-class (*Normal/Abnormal* classification) case and 19% (24% in the best case) for 5-class (IOTN grade classification) case, compared with the accuracy using a MLP and a conventional CNN (as described in Figure 4).

For 5-class classification, it seems that less than 50% accuracy is a quite low, but there are a few clear reasons as follows. As seen in Figure 5, the accuracy for Grade 2 is a very low compared to the others. Most of Grade 2 samples are classified as Grade 1 because the images of “Perfect” and “Slight” are very similar. In addition, the accuracy for Grades 3 and 4 is also relatively low. Correctly classifying Grades 3 and 4 is really significant to determine whether if a patient needs treatment. However, we observe that most of Grade 3 samples are classified as Grade 4. In practice, doctors also tend to diagnosis a patient of Grade 3 as Grade 4. This difficulty can be seen in the performance improvement when the model is trained on all training samples except Grade 3 samples. In this case, the accuracy improves to 73.1% (81.3% in the best case) as shown in Table II.

Another reason of the low accuracy might be the size of training data. To learn clear features of five grade samples, we need more data with correct label. Figure 6 shows the results when training the parallel CNN model on 100, 200 and 300 patients’ samples. We observe the trend of increasing performance, so we believe that the accuracy will increase as the number of samples increases. We are now collecting more samples for training, and also perform data cleansing to correctly label the samples.

 ISBN: 978-1-61208-597-5

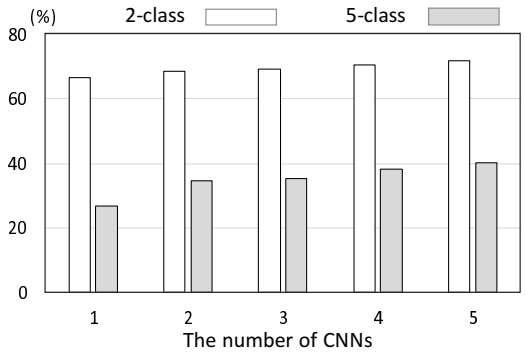

Figure 7. The classification accuracy of the model with the different number of CNNs.

Finally, we evaluate the effectiveness of running multiple CNNs in parallel. For doing it, we train the model on only front images; and then, we perform the experiment on additional set of images such as left images (i.e., front and left images). We repeat the same experiment by adding the other sets of images one by one. The result of accuracy comparison is shown in Figure 7. As the number of CNNs for the different set of images increases, the accuracy also increases.

V. Conclusion

In this paper, we proposed a parallel CNN based image classification model that assesses IOTN grades. Technically, it deals with a training dataset including pairs of a set of images and its corresponding class label. We verify that the proposed model outperforms the other conventional models in terms of classification accuracy.

In future work, we will increase the number of accurate data to retrieve features that clearly separate IOTN Grades 3 and 4 samples. Eventually, we plan to build a dental healthcare application that fully or semi-fully automates the process of IOTN assessment and treatment plan generation using smartphone or mobile devices. The successful remote or automated diagnostic imaging will also be expanded to other fields, such as otolaryngology (ear and nose) and ophthalmology (eye).

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17KT0083, and partly supported by JSPS KAKENHI Grant Number JP16H02802 and JP17K00168. The authors would like to thank Prof. Takashi Yamashiro and Assistant Prof. Kazunori Nozaki, Osaka University Dental Hospital, for setting up environment and medical dataset for the experiments.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems, vol 1, pp. 1097-1105, 2012.

[2] J. Gu, et al., "Recent advances in convolutional neural networks," arXiv preprint arXiv:1512.07108, Jan. 2017.

[3] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks," Radiology, vol. 284, no. 2, pp. 574-582 2017.

[4] V. Gulshan, et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, 316(22), pp. 2402-2410, 2016.

[5] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," Journal of Pathology Informatics, 7(292), 2016.

[6] Jarvinen S., "Indexes of orthodontic treatment need," the American Journal of Orthodontics and Dentofacial Orthopedics, 120(3), 2001.

[7] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Transactions on Pattern Analysis. Mach. Intell., vol. 35, no. 8, pp. 1798-1828, Aug. 2013.

[8] Y. Lecun, Y. Bengio, and G Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 5 2015.

[9] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," In Proc. of the 13th International Conference on Machine Learning, vol. 96, pp. 148-156, Jul. 1996.

[10] T. Kanamori, K. Hatano, and O. Watanabe, "Boosting - Design method of learning algorithms," , Mori Publishing, Sep. 2006.

[11] TensorFlow, https://www.tensorflow.org, retrieval: Aug. 2017.

[12] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," Aistats, vol. 15, no. 106, pp. 275, Apr. 2011.

[13] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, Dec. 2014.

 ISBN: 978-1-61208-597-5