EasyChair Preprint

№ 4416

A Lightweight Convolutional Neural Network Model for Child Pneumonia Classification

Khan Fashee Monowar, Md. Al Mehedi Hasan and Jungpil Shin

October 17, 2020

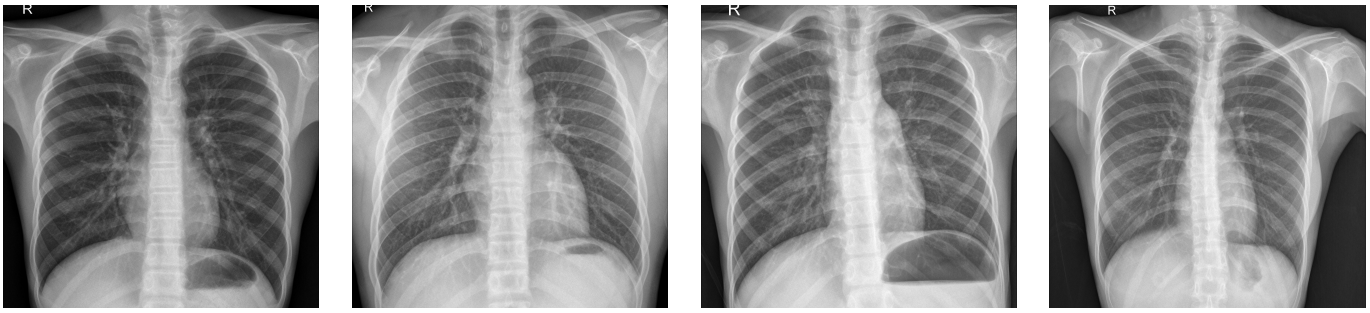
A Lightweight Convolutional Neural Network Model for Child Pneumonia Classification

***Abstract*—Pneumonia is still a serious threat for children including newborns. Each year many children died of pneumonia. Physicians diagnose pneumonia through some process including reviewing chest X-rays of patients. While reviewing, a single diagnostic mistake may cause a serious threat to patients. In recent years, Computer-aided detection system (CAD) and medical image classification are progressively turning into another research territory. CAD can reduce physician's effort to review chest X-ray. Currently, Researchers build various models to detect pneumonia. However, there is still a lack of computationally efficient models to diagnose pediatric pneumonia. Further, some off-the-shelf or pre-trained models are not always suitable for mobile and embedded vision applications since these models are not lightweight. In our research, we built a lightweight convolutional neural network model from scratch which able to learn lung texture features and detect pediatric pneumonia. We compared our proposed model performance with some off-the-shelf models. Our proposed model achieved the best AUC (98.5), test accuracy (94.0), F1 (94.1), Precision (92.3), Specificity (92.1), NPV (95.7) scores. We employed several data augmentation algorithms to increase the model's classification ability.**

***Index Terms*—Chest X-ray, Deep Learning, CNN, MobileNet, AUC, Bioinformatics, Pediatric Pneumonia, Artificial Intelligence**

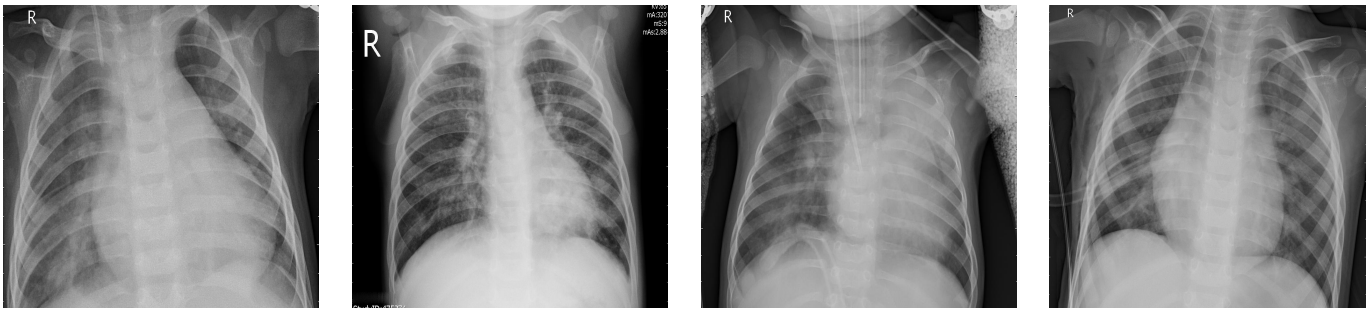
I. INTRODUCTION

According to the World Health Organization (WHO), Pneumonia still kills a lot of children aged less than 5 years [1]. Every year, more than 800,000 children died of pneumonia, most of them aged less than 2 years old, and the death rate was more than the sum of malaria, AIDS, and measles [2]. Numerically, 2200 children die of pneumonia every day, or one child dies every 39 seconds [3]. The dearth of medical resources and personnel, The death rate is even higher in underdeveloped countries. More than half of the death cases of the world due to pediatric pneumonia occur in India and Sub-Saharan countries of Africa [4]. In Bangladesh, almost 20,000 children aged less than 5 years, died of pneumonia in 2017 [5]. Scientists further warn 140,000 children may die of pneumonia in the next decade if proper steps are not taken care of [6]. Shortage of medical resources, low doctor to patient ratio, poverty lead on to a higher death rate in underdeveloped countries. Physicians of these countries need to diagnose scads of Chest X-rays daily which prone to incorrect diagnostic results. According to the care quality commission, Radiologist or clinician failed to formally review a total of 23,000 chest X-rays in 2016-2017 at Queen Alexandra Hospital in the United Kingdom [7]. Moreover, three patients with lung cancer suffered significant harm because their chest X-rays had not been properly assessed [8]. In these circumstances, deep learning based Computer-aided detection (CAD) technology can help physicians diagnosing Chest X-rays image accurately [9]. CAD decreases observational oversight and the false-negative rate [10]. In computer vision, deep learning has already confirmed its superhuman accuracy to image classification [7]. In [11], He et al. showed 3.57% top-5 error rate on ImageNet test set. In the medical image domain, Rajpurkar et al. [12] proposed an algorithm (DenseNet121, transfer learning with fine-tuning) that can detect thorax diseases from chest X-rays at a level exceeding practicing radiologists. Baltruschat et al. [7] increased AUC results up to 80.6% on the Chest X-ray 14 dataset to diagnosis thorax diseases including pneumonia. Anitha et al. [13] got over 90% test accuracy on brain tumor image classification using a two-tier classifier with an adaptive segmentation technique. Therefore, deep learning models with proper optimization, get a lot of attention to solve the various problem in the medical image domain including classification. In pediatric chest X-ray classification, several research groups have done some decent research works. Stephen et al. [14] proposed a CNN model to automatically perform pediatric chest x-ray image classification. They achieved 95.31% and 93.73% training and validation accuracy respectively. In [15], Liang et al. presented 51 layers CNN architecture by implementing transfer learning methods. Moreover, they compared their proposed model with some off-the-shelf CNN models. Off-the-shelf CNN models refers the pretrained models which are used as feature extractor and the weights of last layer are modified [7]. They achieved 90.5% training accuracy and 95.3% AUC score. Kermany et al. [16] implemented transfer learning on their CNN architecture and achieved 92% training accuracy. In [17], Gu et al. achieved 82.34% AUC score by implementing DCNN with transfer learning. In this research, we proposed a light-weight novel 6 layers CNN architecture containing fewer parameters than the various off-the-shelf architectures. To build the CNN convolution layers, we got inspiration from the study of [14]. We compared our proposed model performance with the study of Liang et al [15]. Data augmentation is applied to achieve better classification results. The remaining of this paper is organized as follows: Section II describes materials and method including the dataset we used, details about the proposed CNN model, experimental setup and the evaluation matrices we considered and data preprocessing with augmentaion. Section III portrays comparison and classification performance of proposed model. Finally, section IV concludes our research work giving some insight for further researchers.



(a) (b) (c) (d)

Fig. 1: Normal Chest X-ray Images. Lung, Heart and diaphragm are clearly visible.



(a) (b) (c) (d)

Fig. 2: Pneumonia affected Chest X-ray Images. Lung, Heart and diaphragm are not clearly visible.

II. MATERIALS AND METHOD

A. Dataset Description

We evaluated our model on a pediatric Chest X-ray Image dataset. The dataset was gathered from a study of [16], [18]. This dataset (version 2) contains a library of chest X-ray images of pediatric patients from 1 to 5 years of age at the Guangzhou Women and children's Medical Center [18]. A total of 5856 anterior-posterior chest X-ray grayscale images collected and labeled carefully from retrospective pediatric patients. The original dataset contains three folders (i.e., training, testing, validation) having a total of 4273 pneumonia images and 1583 normal chest X-ray images. Kermany et al. [18] explained the collection and ground truth annotation process. Unreadable or low-quality radiographs were initially excluded and the images were labeled by three expert physicians.

B. Proposed CNN Model

The main difference between chest X-ray and other images is chest X-ray images have lung features. Hence, it is important to study the nature of the chest X-ray image to determine the architecture of the proposed model. CNN is a fabulous feature extractor however researchers can optimize it to avoid complex and expensive features [19]. Besides, Larger kernels (i.e., 9x9 or 7x7) may overlook small features and skip essential details where smaller kernels (i.e., 1x1 or 3x3) able to provide more information but it can be confusing for identifying a larger object in images. After considering the above principles, we have proposed a light-weight 6 layers convolutional neural network architecture containing 6.8M parameters. The First 4 convolutional layers extract lung texture features so they are feature extractors and the other 2 layers are fully connected layers. After trying multiple Kernels, we have achieved the best evaluation metrics result in employing 5x5 size kernel. The input image blocks are extracted by a total of 16 kernels (5x5). We choose the ReLU function in each layer for non-linear activation. Each convolution layer is followed by the Max pooling layers for extracting the highest pixel values from previous convolutional layer output. Max pooling extracts low-level features that are very effective for pneumonia image classification along with reducing the computational cost [20]. Output feature map from the 1st convolution layer then extracts by 2nd convolution layer, containing 32 kernels. Likewise, the 3rd and 4th convolution layers extract features by total of 64 and 128 kernels respectively. The resultant feature map then transforms into a 1-D vector. This method is called flattening which is fed to fully connected dense layers for final classification process [14]. Likewise, we feed the flattened vector into 2 fully connected layers. Hence, The dataset is small, therefore, We apply regularization in neural networks for preventing overfitting. The dropout layer (parameter set to 0.5) is added after flattening for regulization [21]. It stops half of the neurons randomly in each iteration to prevent each unit from being dependent on particular inputs [15]. In the end, the sigmoid activation function is applied at the last dense layer

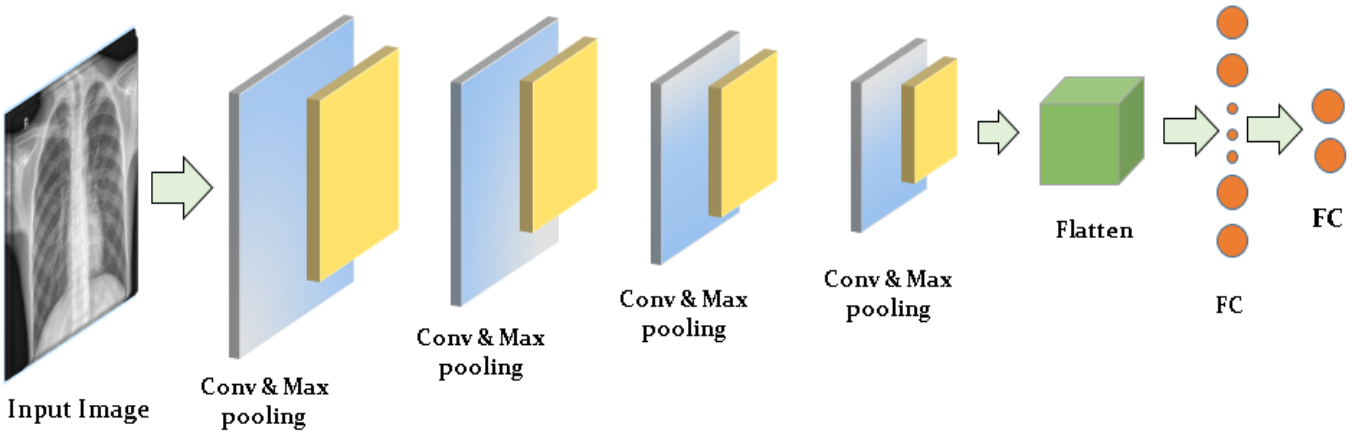


Fig. 3: Architecture of proposed convolutional neural network (CNN) model containing 4 convolutional layers and 2 fully connected layers.

TABLE I: Summary of proposed architecture containing a total of 6,824,865 trainable parameters.

Layer (type)	Kernel Quantity and Size	Output Shape	Paramaters
conv2d_1 (Conv2D)	5 x 5, 16	(None, 220, 220, 16)	1216
max_pooling2d_1 (MaxPooling2)	-	(None, 110, 110, 16)	0
conv2d_2 (Conv2D)	5 x 5, 32	(None, 106, 106, 32)	12832
max_pooling2d_2 (MaxPooling2)	-	(None, 53, 53, 32)	0
conv2d_3 (Conv2D)	5 x 5, 64	(None, 49, 49, 64)	51264
max_pooling2d_3 (MaxPooling2)	-	(None, 24, 24, 64)	0
conv2d_4 (Conv2D)	5 x 5, 128	(None, 20, 20, 128)	204928
max_pooling2d_4 (MaxPooling2)	-	(None, 10, 10, 128)	0
flatten_1 (Flatten)	-	(None, 12800)	0
dropout_1 (Dropout)	-	(None, 12800)	0
dense_1 (Dense)	-	None, 512)	6554112
visualized_layer (Dense)	-	(None, 1)	513

as we perform binary classification. Yet, Choosing the loss function and an algorithm for backpropagation optimization is crucial. An optimization algorithm called Adam is applied to update network weights iteratively which demands less memory requirement instead of traditional stochastic gradient descent procedure. The learning rate (alpha) parameter for Adam optimization is set to 0.001, epsilon is set to 1e-08. The binary cross-entropy method is applied to calculate the loss in each iteration. Further, We train this architecture from scratch without transferring weights from any pre-trained model. The model summery is shown in Table I.

C. Experimental Setup and Evaluation Matrices

The 5 classification matrices, sensitivity/recall, specificity, positive predictive value (PPV)/precision, negative predictive value (NPV) and AUC are important to determine the prediction capability of any model. For binary classification, Sensitivity/recall indicates a model's ability to predict true positive cases among all cases that have that disease. Therefore, Specificity indicates the opposite, true negatives cases among all cases having no disease in the real case. Along with, positive predictive value (PPV)/precision measures the probability of model positive prediction, If any patient really has the disease [22]. It indicates the classification performance of models over the true positive cases. Conversely, negative predictive value (NPV) measures the probability of model negative prediction, If any patient doesn't have the disease (negative). So, It indicates the classification performance of models over the normal (True negative) cases.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

$$P = \frac{TP + FN}{TP + TN + FP + FN} \tag{3}$$

$$PPV = \frac{Sn * P}{(Sn * P) + (1 - Sp) * (1 - P)} \tag{4}$$

$$NPV = \frac{Sp * (1 - P)}{(1 - Sn) * P + Sp * (1 - P)} \tag{5}$$

Here, TP, FN, FP, TN, Sn, Sp, P mean true positive, false negative, false positive, true negatives, sensitivity, specificity,

TABLE II: Comparison between classification matrices of different CNN models.

	Model	Accuracy	Precision	F1	Sensitivity	Specificity	NPV	AUC
This Paper	Proposed Model	**94.0**	**92.3**	**94.1**	95.9	**92.1**	**95.7**	**98.5**
Other Paper [15]	VGG16	74.2	72.3	82.1	95.1	39.3	82.9	84.0
	DenseNet121	81.9	79.1	87.0	96.4	57.7	90.6	76.9
	InceptionV3	85.2	91.6	87.7	84.1	87.1	76.7	65.5
	Xception	87.8	85.7	90.8	**96.7**	73.1	93.0	93.0
	Their Proposed Method	90.5	89.1	92.7	**96.7**	80.3	93.5	95.3

and prevalence respectively. In this research, We compared our model performance by test accuracy, Precision/PPV, F1, Sensitivity, Specificity, NPV and AUC socres.

D. Data Preprocessing and Augmentation

The raw dataset was split into training, testing, validation set. We gathered all images in one folder than applied an 84%/10%,6% split. A total of 4906 images (using 3798 pneumonia images, 1108 normal images) were fed to our proposed CNN model for training. Besides, 631 (316 pneumonia and 315 normal) and 315 (158 pneumonia and 157 normal) images combined into the test and validation set respectively. Class imbalance could cause overfitting so it was taken care of by class balancing technique. Furthermore, all images of this dataset were resized to 224 x 224 pixels initially. Deep learning requires tons of data for training. However, we had a small dataset, we performed data augmentation on each iteration so that model could be fed with the augmented counterpart of each image which helped to reduce overfitting and enhanced the model's generalization ability [15]. Each image was converted to 3 channels (RGB). Histogram equalization was performed to each image for improving contrast. Moreover, Rotation (maximum 7 degrees), width shift range (0.05), height shift range (0.05), zooming (0.45), horizontal flip, these operations were performed to expand the train data. Though the chest is not symmetric, we employed horizontal flip operation to evaluate its impact on prediction. Each image was normalized to reduce computational cost. No data augmentation operation was applied to the validation and test set.

III. RESULTS AND DISCUSSION

In our research, We trained the proposed model continuously through 32 epoch. Accordingly, we ran the training process several times to get the best result. Our proposed model achieved 93.35% training accuracy and 94.01% validation accuracy, obtaining 0.1570 training and 0.1078 validation loss. We observed that being a lightweight model, our model performed very well. Liang et al. measured the classification performance of some off-the-shelf models (VGG 19, DenseNet 121, Inception V3, Xception) along with their 51 layers proposed model [15]. However, We witnessed a large diversity between the classification performance between pretrained, off-the-shelf, and proposed models. The ROC curve reflects the model's predictive ability between classes [23]. In the medical image classification domain, the Higher AUC (area under the ROC curve) score indicates the best diagnostic model. In this research, Our proposed model achieved better accuracy, PPV/precision, F1, specificity, NPV, and AUC scores. Best sensitivity/recall score is achieved by Xception and the proposed method of Liang et al. Our proposed model achieved 98.5% AUC score, which is far better than off-the-shelf models (highest AUC score is 93.0%, which is achieved by Xception among off-the-shelf models.

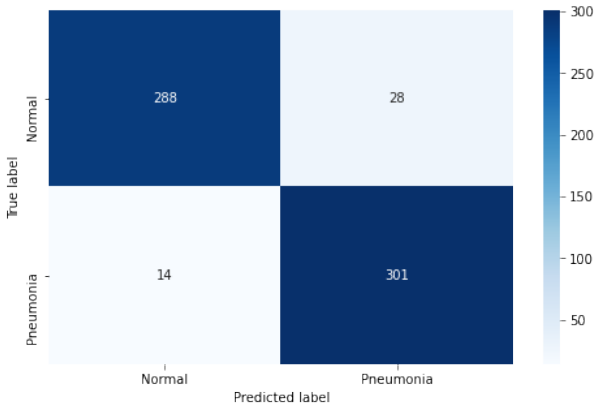


Fig. 4: Confusion matrix based on the proposed model predicting test set.

TABLE III: Comparison between the number of parameters of the proposed model and off-the-shelf models.

Model	Number of Parameters (Per 1 Million)
Our Proposed Model	**6.8**
DenseNet121	8.1
Xception	22.9
InceptionV3	23.8
VGG19	138.35

Comparison of classification matrices and number of parameters are shown in Table II and Table III respectively.

IV. Conclusion

We built a lightweight CNN model from scratch to detect pediatric pneumonia from Chest X-ray. Different data augmentation algorithms are applied to enhance the model's classification ability. Being a lightweight architecture, we compared its classification performance with the study of Liang et al. [15]. Properly comparing with VGG 16, Inception net, Densenet 121, Xception net, our proposed model achieved better test accuracy, precision, F1, specificity, NPV, and AUC score. The specificity score achieved by our proposed model is 92.1% which is incomparably better than any model in the study of [15]. Though, In the medical image classification domain, Sensitivity, specificity, and AUC should be 100%. Therefore, Future work will include investigation of other model architecture and new architectures to increase these scores.

References

[1] I. Rudan, C. Boschi-Pinto, Z. Biloglav, K. Mulholland, and H. Campbell, "Epidemiology and etiology of childhood pneumonia," *Bulletin of the world health organization*, vol. 86, pp. 408–416B, 2008.

[2] R. A. Adegbola, "Childhood pneumonia as a global health priority and the strategic interest of the bill & melinda gates foundation," *Clinical infectious diseases*, vol. 54, no. suppl_2, pp. S89–S92, 2012.

[3] UNICEF, "UNICEF Data," https://data.unicef.org/topic/child-health/pneuhttps://www.overleaf.com/project/5e557cf95b881f0001b78de9monia/, 2019, [Online; accessed 27 February 2020].

[4] D. Onyango, G. Kikuvi, E. Amukoye, and J. Omolo, "Risk factors of severe pneumonia among children aged 2-59 months in western kenya: a case control study," *Pan African Medical Journal*, vol. 13, no. 1, 2012.

[5] B. Dadonaite, "Pneumonia," *Our World in Data*, 2018, https://ourworldindata.org/pneumonia.

[6] UNICEF, " Press release 140,000 children in Bangladesh could die in the next decade unless more is done to fight pneumonia ," https://www.unicef.org/bangladesh/en/press-releases/140000-children-bangladesh-could-die-next-decade-unless-more-done/fight-pneumonia/, 2020 note = "[Online; accessed 27 February 2020]".

[7] "Naturecomparison of deep learning approaches for multi-label chest x-ray classification," https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6476887/, accessed: 2020-02-26.

[8] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest x-ray classification," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[9] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017, pMID: 28301734. [Online]. Available: https://doi.org/10.1146/annurev-bioeng-071516-044442

[10] R. A. Castellino, "Computer aided detection (cad): an overview," *Cancer Imaging*, vol. 5, no. 1, p. 17, 2005.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[12] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[13] V. Anitha and S. Murugavalli, "Brain tumour classification using two-tier classifier with adaptive segmentation technique," *IET computer vision*, vol. 10, no. 1, pp. 9–17, 2016.

[14] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, "An efficient deep learning approach to pneumonia classification in healthcare," *Journal of healthcare engineering*, vol. 2019, 2019.

[15] G. Liang and L. Zheng, "A transfer learning method with deep residual network for pediatric pneumonia diagnosis," *Computer methods and programs in biomedicine*, p. 104964, 2019.

[16] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled optical coherence tomography (oct) and chest x-ray images for classification," *Mendeley data*, vol. 2, 2018.

[17] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography," in *Proceedings of the 3rd International Conference on Multimedia and Image Processing*, 2018, pp. 88–93.

[18] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.

[19] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, no. 1, p. 113, 2019.

[20] A. Giusti, D. C. Cireşan, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Fast image scanning with deep max-pooling convolutional neural networks," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 4034–4038.

[21] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] A. K. Akobeng, "Understanding diagnostic tests 1: sensitivity, specificity and predictive values," *Acta paediatrica*, vol. 96, no. 3, pp. 338–341, 2007.

[23] J. Huang and C. X. Ling, "Using auc and accuracy in evaluating learning algorithms," *IEEE Transactions on knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.