International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-06, Issue-12, MAR 2021

Pneumonia Detection Using Deep Kernel-Based Extreme Learning Machine

Dr. Afzal A.L. Assistant Professor, College of Engineering Muttathara, Thiruvananthapuram, India, afzal.a.l@gmail.com

Smitha C. S. Assistant Professor, College of Engineering Muttathara, Thiruvananthapuram, India, smithacscs@gmail.com

Abstract **Deep learning, an emerging trend in machine learning research and application, opens up a new outlook in machine learning realms. Even though the deep learning methods have been mainly investing their efforts in the neural network, recently there are many attempts to leverage these features into kernel machines. However, the gradient-based and quadratic-based parameter tuning methods enforce iterative training procedures in artificial neural networks and kernel machines respectively. Extreme Learning Machine (ELM) eliminates this hectic and highly time consumed iterative parameter tuning mechanism from its training procedure. ELM exploits a unified learning method that facilitates feature mappings like kernel functions and random feature mappings. The multilayer arc-cosine kernel exhibit deep learning features and it can be glued with ELM to build a Deep kernel-based extreme learning machine. This paper explores the possibility of detecting pneumonia from x-ray images using deep kernel-based extreme learning machines.**

***Keywords* — Arc-cosine kernel, Deep kernels, Deep learning, kernel ELM, Extreme learning Machines, pneumonia detection**

I. Introduction

Two prevalent learning approaches, neural network and kernel machines, are enduring with their iterative parameter tuning-based learning strategy. The gradient-based learning algorithms like Back Propagation (BP) and its variants are the driving force behind neural networks [1], [2], [3], [4], [5]. Because of the inter-dependencies between the parameters in different layers of neural networks, these gradient-based algorithms cannot guarantee better generalization performance or global optimum solution. Similarly, the quadratic equation with tuning parameters used in the training process of Support Vector Machines (SVM), a popular kernel machine, also enforce iterative parameter tuning mechanisms [6], [7]. In the recent past, there were some attempts to build learning models without these iterative parameters tuning [8]. The extreme learning machine (ELM) is one such prominent learning model that exploits non-iterative parameter tuning mechanism in its learning process [9]. The core of ELM is the generalized matrix inverse operation. It eases the operation to compute the output weight, random computation of input weights, and biases [10], [11]. It has already been observed that as the norm of output weights decreases the generalization performance of feed-forward neural networks improves, reaching minimal errors [12]. Unlike conventional training approaches, the ELM accomplishes both the smallest norm of output weights and minimized training error. The universal approximation and classification capabilities turn the ELM into a unified learning method. Even though ELM had been proposed as a fast learning algorithm for single-layer feed-forward networks (SLFNs), later it has been used to model different types of feature mappings, including kernel functions [13]. Kernel-based extreme learning machine (KELM) exploits kernel functions to tackle the situation where the feature mapping functions of hidden nodes are unknown to the user [14]. It proposes a unified learning model for regression and both binary and multi-class classifications. The conventional KELMs involve only a single layer kernel and hence they fit in shallow-based learning architecture. Robot execution failure prediction [15], different combinations of multiple kernels [16], the exploitation of parallel computing frameworks such as MapReduce [17] are some of the shallow architecture-based models that make KELM more popular. The deep learning models, the latest trend in machine learning, also exhibit the potential to shattered shallow-based learning models. Nowadays, there are several attempts to transform conventional shallow models into deep learning models. The eagerness over these transformations is the escalated number of layers and efficient feature extractions capabilities of deep learning architectures. Deep layered learning models have been showing their excellence in the area of neural networks. Inspired by the widespread acceptance of deep neural network models, there are some noticeable attempts to leverage deep learning capabilities in kernel machines. The kernel function, the crux in kernel machines, eliminate the explicit mapping of data in the high

 DOI : 10.35291/2454-9150.2021.0081

ISSN : 2454-9150 Vol-06, Issue-12, MAR 2021

dimensional feature space but rather computes the inner products between every pair of data in the feature space [18].

The muli-layer arc-cosine kernel provides multiple layers of feature extraction in kernel machines. Different layers in the hierarchy of arc-cosine kernel can have different activation functions and hence qualitatively different geometric properties. The noticeable characteristic of the arc-cosine kernel is that it is recursive by itself and exhibits the potential to enforce multilayer computation in learning models. This multilayer arc-cosine kernel function attributes deep learning capabilities in kernel machines like support vector machines and their variants [19], [20]. Deep core vector machine [21], scalable deep kernel machines with multiple layers of unsupervised feature extraction [22] are some of the deep kernel machines.

The possibilities of including this arc-cosine kernel in the fast non-iterative training mechanism of extreme learning machines were discussed in [23]. Nowadays, different machine learning approaches have been using in the diagnosis of common diseases. This paper explores the possibility of using a deep kernel machine such as deep kernel-based extreme learning machine to detect Pneumonia from the x-ray images.

The rest of the paper is organized as follows. The deep kernel-based extreme learning machine and its computations are described in Section II. The proposed deep kernel-based Pneumonia detection is explained in Section III. Experimental analysis and performance evaluations are given in Section IV. Section V includes conclusions and future enhancements.

II. Deep Kernel-based Extreme Learning Machines

Extreme learning machines were proposed as an alternative mechanism to train data without iterative parameter tuning. It has been achieved by randomly generated input weights and hidden layer biases. The other observed feature is that the hidden layer parameters are independent of input samples. The output weights are also calculated non iteratively as in linear systems. The basic computations in Single Layer Feedforward Network (SLFN) as depicted in Figure 1. can be described as follows[23].

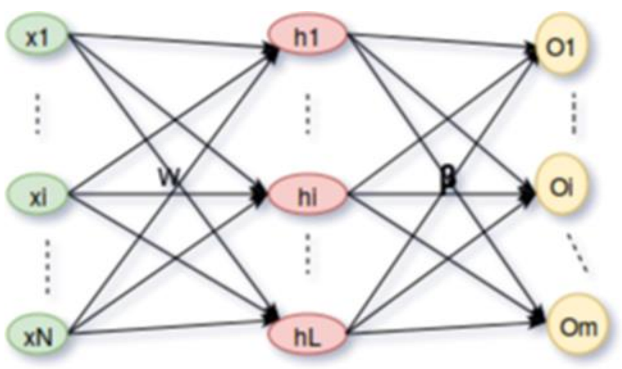

Figure 1: The basic model of single layer feed forward network

Here, the training sample can be expressed as: $\{(x_i, y_i)_{i=1}^{N} | x_i \in \Re^n, y_i \in \Re^m\}$. In Figure 1, w be the input weight matrix and it can be represented as:

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1L} \\ w_{21} & w_{22} & \cdots & w_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NL} \end{bmatrix}_{N \times L}$$

Similarly, the output weight matrix

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{L1} & \beta_{L2} & \cdots & \beta_{Lm} \end{bmatrix}_{L \times m}$$

The mathematical model of SLFN with b_i as the bias of i^{th} hidden node and F( ) as the Activation function can be expressed as [11].

$$\sum_{i=1}^{L} \beta_i F(w_i . x_j + b_i) = O_j, \quad for\ j = 1\ to\ N \tag{1}$$

The activation function F() of the hidden nodes may be the sigmoid, hyperbolic tangent, threshold, etc.

As described in the paper [10], there exist β_i, w_i and b_i such that the actual output O_j equals to the target output y_j, the SLFN approximates these N samples with zero mean square error.

$$i.e.\quad \sum_{j=1}^{N} \|O_j - y_j\| = 0$$

Now the Equation 1 becomes

$$\sum_{i=1}^{L} \beta_i F(w_i . x_j + b_i) = y_j \quad , for\ j = 1\ to\ N \tag{2}$$

In simple form, the output function of SLFN can be expressed as

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) \quad = h(x)\beta$$

and these N equations(one for each data sample) can be written in matrix form as

$$f_L(x) = H\beta = Y \tag{3}$$

where H is a $N \times L$ matrix of the form

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_i) \\ \vdots \\ h(x_n) \end{bmatrix} = \begin{bmatrix} F(w_1 . x_1 + b_1) & \cdots & F(w_L . x_1 + b_L) \\ \vdots & \cdots & \vdots \\ F(w_1 . x_N + b_1) & \cdots & F(w_l . x_N + b_L) \end{bmatrix} \tag{4}$$

β is the output weight matrix ($L \times m$) is of the form $\begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}$

and Y is the output matrix ($N \times m$). The conventional gradient-based training method of SLFN fine-tunes the parameters $w, b\ \ and\ \ \beta$ with respect to the objective function.

$$\min_{w,b,\beta} \|H\beta - Y\| \tag{5}$$

The corresponding cost function is [11]

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i h_i(x_j) - y_j \right), \ where \quad h_i(x_j) = F(w_i . x_j + b_i) \tag{6}$$

ELM optimizes for the smallest error and also for the smallest norm of output weight. Taking these facts, the objective function of ELM [10] can be expressed as:

$$\min_{\beta} \|H\beta - Y\|, \ subjected\ to\ \ min\ \|\beta\| \tag{7}$$

DOI : 10.35291/2454-9150.2021.0081

The analytical model of obtaining the smallest norm least square solution of the above equation can be expressed as:

$\hat{\beta} = H^{\dagger}Y$(8)

where $H^{\dagger}$ is the Moore-Penrose generalized inverse of matrix H and it can be computed as $H^{\dagger} = H^T(\frac{I}{C} + HH^T)^{-1}$, where C is the regularization parameter. The output function of ELM can be expressed as:

$$f(x) = h(x)\beta = h(x)H^T(\frac{I}{c} + HH^T)^{-1}Y \quad (9)$$

Kernel-based ELM: Equation 9 involves a dot product HH^T and it can easily be replaced with a kernel function if the feature transformation function of ELM is unknown to the user. Now the Kernel-based ELM can be expressed as :

$$f(x) = \begin{bmatrix} k(x, x_1) \\ \vdots \\ k(x, x_N) \end{bmatrix}^T (\frac{I}{c} + K_{ELM})^{-1}Y, \quad where\ k_{ELM} = k(h(x_i), h(x_j)) \quad (10)$$

Here, the kernel matrix $K_{ELM} = HH^T$:$K_{ELM}(i,j) = \langle h(x_i), h(x_j) \rangle$ =$k(x_i, x_j)$. It can easily be migrated to the deep kernel-based extreme learning machine by inducing arc-cosine kernel in this model.

Analysis of Arc-cosine kernel: The process of attributing deep learning capabilities in kernel machines was achieved through the multi-layer arc-cosine kernel. In the paper [19], the n^{th} order kernel in the family of the arc-cosine kernel has been defined as:

$$k_n(x, y) = 2\int dw \frac{e^{-\frac{\|w\|^2}{2}}}{(2\Pi)^{\frac{d}{2}}} \Theta(w.x)\Theta(w.y)(w.x)^n(w.y)^n \quad (11)$$

where $\Theta(z) = \frac{1}{2}(1 + sign(z))$ is the Heaviside function. Two parameters on which arc-cosine kernel dependent are magnitudes of the inputs and angle between them (θ). The simple form of the arc-cosine kernel with degree n can be written as :

$$k_n(x, y) = \frac{1}{\pi}\|x\|^n\|y\|^n A(n, \theta) \quad (12)$$

where the angular dependency

$$A(n, \theta) = (-1)^n(sin\theta^{2n+1})(\frac{1}{sin\theta}\frac{\partial}{\partial\theta})^n(\frac{\pi - \theta}{sin\theta}) \quad (13)$$

The first three expressions of $A(n, \theta)$ as [20]:

$$A(0, \theta) = \pi - \theta$$
$$A(1, \theta) = sin\theta + (\pi - \theta)cos\theta$$
$$A(2, \theta) = 3sin\theta cos\theta + (\pi - \theta)(1 + 2cos^2\theta).$$

For the case n = 0, the kernel function takes the simple form:

$$k_0(x, y) = 1 - \frac{\theta}{\pi} \quad (14)$$

The angle θ between the input vectors x and y is

$$\theta = \cos^{-1}\left(\frac{(x.y)}{\|x\|\|y\|}\right) \quad (15)$$

This arc-cosine kernel can be combined to form a multi-layered arc-cosine kernel. Unlike other multiple combinations of kernel functions, each layer in multi-layer arc-cosine can have a different value of activation factor or degree. It facilitates the capability for exhibiting different geometric behavior on different layers of the arc-cosine kernel. The successive layers use the kernel matrix computed in the previous layer, with the same or different activation function in each layer [19]. l- fold arc-cosine with degree nl, $K_{nl}^l(x_i, x_j)$ can be expressed as:

$$k_{nl}^l(x_i, x_j) = \frac{1}{\pi}[k_{nl}^{(l-1)}(x_i, x_i)k_{nl}^{(l-1)}(x_j, x_j)]^{\frac{n}{2}} A(n, \theta^{(l-1)}) \quad (16)$$

where $\theta^{(l)}$ is the angle between the images of x and y in the feature space induced by the l-fold composition.

$$\theta^{(l)} = cos^{-1}\left(\frac{k_{nl}^{(l)}(x_i, x_j)}{\sqrt{k_{nl}^{(l)}(x_i, x_i)k_{nl}^{(l)}(x_j, x_j)}}\right) \quad (17)$$

It can be computed iteratively and the hierarchical method starts with initial kernels: $k^0(x_i, x_i) =< x_i, x_i >$, $k^0(x_j, x_j) =< x_j, x_j >$ and $k^0(x_i, x_j) =< x_i, x_j >$.

Deep Kernel-based extreme learning machine: Deep kernel-based extreme learning machine can be modeled by inducing the multi-layer arc-cosine kernel in kernel-based ELM. It can be expressed mathematically as :

$$f(x) = \begin{bmatrix} k_{nl}^l(x, x_1) \\ \vdots \\ k_{nl}^l(x, x_N) \end{bmatrix}^T (\frac{I}{c} + K_{ELM})^{-1}Y, \quad where\ k_{ELM} = k_{nl}^l(h(x_i), h(x_j)) \quad (18)$$

Here $K_{nl}^l(x_i, x_j)$ representing l- fold arc-cosine kernel with degree nl. The iterative computation of arc-cosine kernel gives the deep kernel matrix Ω_{ELM}. The output weight can be computed as $Outputwt = (\frac{I}{C} + \Omega_{ELM})^{-1}.Y$ This output weight vector is used to predict the class label for training and testing dataset as $Predicted\_Targets =$ dot($\Omega_{ELM}, Outputwt$). In the case of testing, Ω_{ELM} represent the kernel matrix corresponding to the test dataset.

III. Pneumonia Detection Using Deep Kernel-Based Extreme Learning Machine

Disease diagnosis and management involve an immense amount of images and their classification. It seems to be very hectic for human experts to classify those images and converge to a proper decision. There were many attempts to leverage machine learning algorithms in the realm of disease diagnosis and management. These conventional frameworks mainly relied on handcrafted object segmentation and shallow-based classification models to classify them [24]. The recent trend in machine learning, the deep learning architecture exhibit the potential to extract pattern automatically from input dataset through multiple layers of computations. The disease diagnosis and management process are progressively rolled out to these Artificial Neural networks (ANN) based deep learning methods [25]. Besides these deep ANN and kernel-based models, ELM also showing its capabilities in medical object detection [26] and image classifications[27].

It has been reported that the mortality rate of pneumonia among children is comparatively high [28]. Pneumonia may be caused by either Bacterial or viral pathogens and it needs to be handled separately. Antibiotic treatment has been usually practiced against the bacterial pathogen. Supportive caring mechanisms are taken for viral Pneumonia. Doctors are usually exerting chest X-rays for diagnosing and

 DOI : 10.35291/2454-9150.2021.0081

differentiate bacterial and viral Pneumonia. Figure 2 shows the chest X-ray images of normal, bacterial, and viral pneumonia patients.

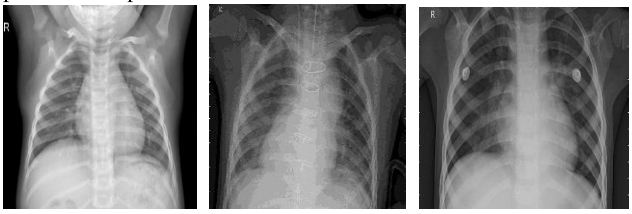

Figure 2: (a) Normal (b) Viral (C) Bacterial X-rays

This paper explores the possibility of applying the deep kernel-based extreme learning machine, the Kernel-based ELM with multi-layer arc-cosine kernel, on X-ray images to identify pneumonia patients.

IV. Experimental Result and Evaluation

Experiments were performed on the dataset taken from [29]. Out of 5840 images, 5216 images were taken under the consideration of training, and 624 were taken for testing purposes. The composition of normal, viral, and bacterial x-rays images in the training and test data sets are depicted in Figure 3.

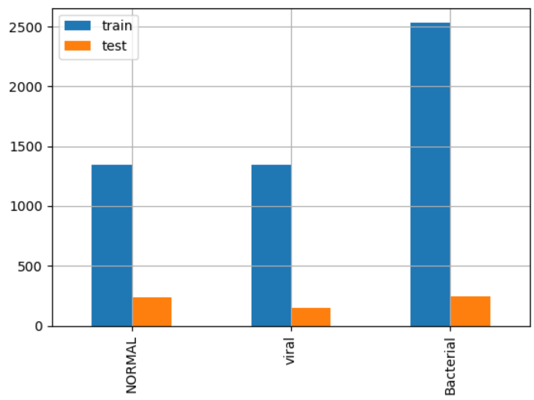

Figure 3: Dataset composition

In our experiment first, we implement a deep kernel-based extreme learning machine by incorporating the arc-cosine kernel [1] as an external kernel in a classical ELM package [2]. A three-layer arc-cosine kernel is evaluated in our experiment. Even though any values can be used as the activation value (degree) for each layer in the arc-cosine kernel, we have tried the combinations of activation values 0, 1, and 2. Experiments have been tried in binary class classification mode (Normal vs Pneumonia) and it achieved an accuracy of 93.01 % for the combination {2,0,1}. The average precision, recall, F1-sore, and accuracy of our deep kernel-based ELM against various combinations of activation values are shown in Table 1. Bold values indicate maximum accuracy and it has been more elaborated in the confusion matrix as shown in Figure 4. In this confusion matrix, it is clear that out of 234 normal test images 199 (86 %) were correctly identified as the normal case itself. In affected cases out of 390 images, only less than 3% were wrongly diagnosed as normal all others are detected as affected patients. It shows the efficiency of this model.

Table 1: precision, recall, F1-sore, and accuracy against various activation list

Activation List	Precision	Recall	F1-Score	Accuracy
[0, 1, 2]	.920	.916	.918	.914
[0, 2, 1]	.920	.919	.922	.919
[1, 0, 2]	.910	.878	.894	.882
[1, 2, 0]	.914	.912	.913	.908
[2, 0, 1]	**.931**	**0.929**	**0.930**	**.931**
[2, 1, 0]	0.891	0.910	.900	.902

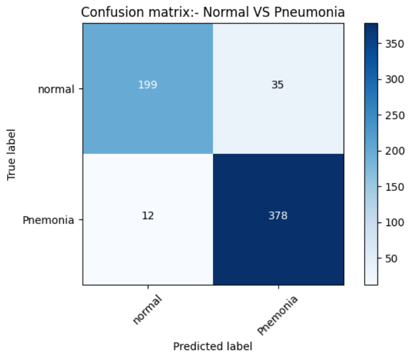

Figure 4: Confusion matrix against maximum accuracy.

V. Conclusion and Future Works

Neural networks and kernel machines have been revolving around the traditional iterative parameter tuning mechanism for their training purpose. The exploration towards the elimination of the hindrance of iterative parameter turning reached the formation of Extreme Learning Machines (ELMs). In ELM, the output weight is computed by analytical methods which are the source behind the elimination of iterative parameter tuning. It is facilitated by assigning random weight in the hidden nodes. Even though ELM has been modeled as a fast learning algorithm for single-layer feed-forward neural networks (SLFNs), it gradually rolled out different models like kernel-based ELM with its universal approximation and classification capabilities. The conventional kernel-based ELMs include only shallow-based (single-layered) kernels. The multi-layer arc-cosine kernel exhibits the potential for being a deep kernel. The inclusion of this multi-layer arc-cosine kernel in ELM explored the possibility of the deep kernel-based extreme learning machine. This paper discussed the applicability of the deep kernel-based extreme learning machine with the multi-layer arc-cosine kernel on X-ray images for the rapid radiologic interpretation of bacterial and viral Pneumonia to provide urgent referral interventions. The experimental results show the potential of deep kernel-based extreme learning machines in the domain of disease diagnosing and management. This model can be enhanced for multi-class classification such as normal vs bacterial vs viral. This work can be further explored with more layers in the arc-cosine kernel. However, it may raise memory exceptions as the size of the kernel matrix increases.

[1] libSVM-2.91:Available at https://www.csie.ntu.edu.tw/cjlin/libsvm/oldfiles/
[2] The ELM package elm-0.1.1 is available at https://pypi.python.org/pypi/elm

 DOI : 10.35291/2454-9150.2021.0081

ISSN : 2454-9150 Vol-06, Issue-12, MAR 2021

References

[1] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.

[2] G. E. Hinton, "Learning multiple layers of representation," *Trends in cognitive sciences*, vol. 11, no. 10, pp. 428–434, 2007.

[3] R. Salakhutdinov and G. E. Hinton, "Deep boltzmann machines." in *AISTATS*, vol. 1, 2009, p. 3.

[4] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle *et al.*, "Greedy layer-wise training of deep networks," *Advances in neural information processing systems*, vol. 19, p. 153, 2007.

[5] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proceedings of the 26th annual international conference on machine learning*. 1em plus 0.5em minus 0.4em ACM, 2009, pp. 609–616.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[7] V. Vapnik, S. E. Golowich, A. Smola *et al.*, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in neural information processing systems*, pp. 281–287, 1997.

[8] G.-B. Huang, Q.-Y. Zhu, and C. K. Siew, "Real-time learning capability of neural networks," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 863–878, 2006.

[9] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, no. 3, pp. 376–390, 2014.

[10] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2. 1em plus 0.5em minus 0.4em IEEE, 2004, pp. 985–990.

[11] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[12] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE transactions on Information Theory*, vol. 44, no. 2, pp. 525–536, 1998.

[13] G.-B. Huang, L. Chen, C. K. Siew *et al.*, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.

[14] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 513–529, 2012.

[15] B. Li, X. Rong, and Y. Li, "An improved kernel based extreme learning machine for robot execution failures," *The Scientific World Journal*, vol. 2014, 2014.

[16] X. Liu, L. Wang, G.-B. Huang, J. Zhang, and J. Yin, "Multiple kernel extreme learning machine," *Neurocomputing*, vol. 149, pp. 253–264, 2015.

[17] X. Bi, X. Zhao, G. Wang, P. Zhang, and C. Wang, "Distributed extreme learning machine with kernels based on mapreduce," *Neurocomputing*, vol. 149, pp. 456–463, 2015.

[18] Wikipedia, "Kernel method — Wikipedia, the free encyclopedia"https://en.wikipedia.org/wiki/Kernel_method , 2021, [Online; accessed 16-February-2021].

[19] Y. Cho and L. K. Saul, "Kernel methods for deep learning," in *Advances in neural information processing systems*, 2009, pp. 342–350.

[20] Y. Cho and L. K. Saul, "Analysis and extension of arc-cosine kernels for large margin classification," *arXiv preprint arXiv:1112.3712*, 2011.

[21] A. L. Afzal and S. Asharaf, "Deep kernel learning in core vector machines," *Pattern Analysis and Applications*, pp. 1–9, 2017.

[22] A. Afzal and S. Asharaf, "Deep multiple multilayer kernel learning in core vector machines," *Expert Systems with Applications*, vol. 96, pp. 149–156, 2018.

[23] A. Afzal, N. K. Nair, and S. Asharaf, "Deep kernel learning in extreme learning machines," *Pattern Analysis and Applications*, pp. 1–9, 2020.

[24] M. Goldbaum, S. Moezzi, A. Taylor, S. Chatterjee, J. Boyd, E. Hunter, and R. Jain, "Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images," in *Proceedings of 3rd IEEE international conference on image processing*, vol. 3. 1em plus 0.5em minus 0.4em IEEE, 1996, pp. 695–698.

[25] C. S. Lee, D. M. Baughman, and A. Y. Lee, "Deep learning is effective for classifying normal versus age-related macular degeneration oct images," *Ophthalmology Retina*, vol. 1, no. 4, pp. 322–327, 2017.

[26] W. Zhu, W. Huang, Z. Lin, Y. Yang, S. Huang, and J. Zhou, "Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation," *Multimedia Tools and Applications*, vol. 75, no. 5, pp. 2815–2837, 2016.

[27] Q. Li, Q. Peng, J. Chen, and C. Yan, "Improving image classification accuracy with elm and csift," *Computing in Science & Engineering*, vol. 21, no. 5, pp. 26–34, 2018.

[28] 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.

[29] "Kaggle dataset, x-ray images," https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia , accessed: 2010-09-30.

 DOI : 10.35291/2454-9150.2021.0081