

---

# A DEEP LEARNING HYBRID ENSEMBLE FUSION FOR CHEST RADIOGRAPH CLASSIFICATION

Saima Sultana\*, Syed Sajjad Hussain†, Manzoor Hashmani‡, Jawwad Ahmad§,  
Muhammad Zubair¶

---

**Abstract:** Biomedical imaging, archiving, and classification is the recent challenge of computer-aided medical imaging. The popular and influential Deep Learning methods predict and congregate distinct markable features of ambiguity in radiographs precisely and accurately. This study submits a new topology of a deep learning network for chest radiograph classification. In this approach, a hybrid ensemble fusion of neural network topology can better diagnose ambiguities with high precision. The proposed topology also compares statistical findings with three optimizers and the most possible varying essential attributes of dropout probabilities and learning rates. The performance as a function of the AUCROC of this model is measured on the Chest Xpert dataset.

Key words: *chest radiographs, 2D image dataset, neoteric neural network model, SGDM, Adam, RmsProp, dropout, learning rate*

Received: August 31, 2020

DOI: 10.14311/NNW.2021.31.010

Revised and accepted: June 30, 2021

## 1. Introduction

With the start of 2020, the most dangerous and death-causing disease the whole world faces is chest disease. Some disorders are chronic, contagious, and even life-threatening. The in-time diagnosis and sufficient precautionary measures can only prevent a person from the severity of the disease [28]. Practically it is unbearable to predict the future of medications, which fetches several hopes, doubts, and confusions to the theme. Artificial Intelligence (AI) [53] provides many surprises to

---

\*Department of Computer Science; Hamdard University, Madinat al-Hikmah Hakim Mohammad Said Road, Karachi-74600, Pakistan, E-mail: [sultana.saima@hotmail.com](mailto:sultana.saima@hotmail.com),

†Syed Sajjad Hussain – Corresponding author; Faculty of Computer Science, Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, SZABIST, Block 5 Clifton, Karachi-75600, Pakistan, E-mail: [sshussainr@ieee.org](mailto:sshussainr@ieee.org), [dr.sajjad@szabist.edu.pk](mailto:dr.sajjad@szabist.edu.pk)

‡Electrical and Computer Engineering Department; Universiti Teknologi PETRONAS, Persiaran UTP, 32610 Seri Iskandar, Perak, Malaysia, E-mail: [manzoor.hashmani@utp.edu.my](mailto:manzoor.hashmani@utp.edu.my)

§Department of Electrical Engineering; Usman Institute of Technology, ST-13, Block-7, Gulshan-e-Iqbal, Abu-Hasan Isphahani Road, Karachi-75300, Pakistan, E-mail: [jawwad@uit.edu](mailto:jawwad@uit.edu)

¶Faculty of Engineering, Sciences & Technology; Faculty of Computer Science, IQRA University, North Karachi Campus, Sector 7b North Karachi Twp, Karachi-75850, Pakistan, E-mail: [zubair@iqra.edu.pk](mailto:zubair@iqra.edu.pk)

the technological world, and Deep Learning is one of the most advanced, flourished, and well-groomed arenas to unveil many wonders to humanity around the globe. A set of neural networks systemized to mimic human cognitive behaviors and infer distinctive patterns from provided imageries data to classify subjective customs. So that neural networks layers classify, absorb and aggregate a distinct set of features to bound them with high precision and accuracy [13].

The deep neural network ensures new prospects of chest disease discovery. One of the chief causes of raised mortality is a permanent slit in recognition of the diseases. Several Deep Learning networks were industrialized during recent decay to investigate digital chest radiography for abnormalities. Cope up with latent, potential shortenings, scarcity, and unavailability of human expertise [2].

## 2. Background

A crafting machine has ever fantasized pioneers that they dreamed. The times of antique Greece craving dates back, and the mythical figures Pygmalion, Daedalus, and Hephaestus may all inferred as legendary originators. Galatea, Talos, and Pandora may all regarded as artificial life. With the invention of computers having decision capabilities, a soft shock of curiosity grasps the minds of the inventors that machines could develop intelligent behavior. So Artificial intelligence (AI) is exposed to the world for the novel exciting arena of Artificial intelligence (AI) [46]. Fundamentally, Artificial intelligence affords high grades of decision-making support to the machine. Consequently, numerous progressions of human life became automated. To make artificial intelligence sufficiently advantageous many related concept visualizations adhered to novelty. Machine learning and deep learning are a sub-part of that [22]. Learning procedure leads to knowledge, reflects in absorbing novel ideas and themes [24]. Deep Learning algorithms enthused with humanoid intelligence entitled ANN (artificial neural networks) [32]. Deep Learning acquired sky-rocketed eminence in the sphere of duration. It excelled as it learned to drive a driverless car, diagnose numerous diseases, etc.; in the Deep learning network, “Deep ”represents a keen value of layers in the system [29]. The term deep learning was coined by Geoffrey Hinton, utilized his approach in the ImageNet contest won for lessen error rate from 25.7% to 15.3% [15]. CNN consent the neural network to progress parameters in many spatial spots of imagery data and enhances feature learning briskly [42].

## 3. Literature review

Chest radiography is one of the vital assessments under consideration of automation. The improved and better accuracies value for the same radiographs depicts the reliability of automated computer-aided diagnosis. So [23] utilizes and proves the validity of dataset training at DenseNet 121 in terms of AUC [41]. This study compares the state-of-the-art DL systems and populations. A comparative statement varies the feasibility among CAD4TB, Lunit INSIGHT, and qXR datasets from different regions [34]. The next study makes use of the pre-trained model on the Chest X pert dataset. It also unlocks the fact that a wrong understanding of

disease causes miss-treatment these kinds of entitlements explain about 40 % – 54 % of radiology-related medical misconduct cases. The study examines the accuracy of 4 out of 14 diseases of the Chest X Pert dataset. It proposes a network built on pre-trained VGG, developed by Oxford University. It calculates Macro averaged precision: 0.3087 Micro average precision: 0.4520. The pioneering work for chest radiographs was done in the research study [14] when the author developed a natural language processing model to extract features regarding Bacterial Pneumonia and acquired a result as Recall: 0.94 and Precision: 0.86 in comparison with the pathologist.

The research represented in [39] is an inter-domain comparison of three radiographic data repositories trained at DenseNet 121 for better AUC values. [21] consists of 1000 x-ray images include mutual positive and negative images for tuberculosis features identification. Make use of pre-trained networks and contributes to improving the AUC value of VGG19 [43]. It develops 121 layers CNN, named Chest X Net trained on ChestX-ray14. In that era, the primary openly obtainable chest X-ray dataset, consisting of over 100,000 frontal view X-ray images with 14 illnesses. The Performance of Chest X Net was measured in terms of F1 scores, with four Radiologists. The research study [9] applies CNN on two datasets, i.e., MNIST [10] and CIFAR-10 [30]. Calculates their accuracies and discusses the model in depth [27]. The study provides the current state of the artwork for chest radiography in deep learning. It also delivers a satisfactory investigation of various neural network architectures. Another milestone research study describes pulmonary Tuberculosis automated classification by convolutional neural network. The study is conducted at four de-identified HIPPA-compliant dataset examined at AlexNet and GoogleNet [26].

## 4. The dataset

The Deep neural network learning prerequisite is a massive dataset with dense crowd interrelated features of images. Deep Learning requires an enormous collection of images to learn the interrelated features accurately. With an adequate dense collection of images, the Deep learning model absorbs the features more precisely and much confidently to train the network [35]. These features are afterward classified into distinct classes, subjects to the grouping by the neural network model. In comparison with image datasets, medical image datasets are scarce. Relatively very few medical image datasets are available for deep learning requirements. With the obtainability of medical imaging datasets, various subjects of confidentiality and secrecy rise. Henceforth deep learning model utilizes publicly available datasets for learning features and classification.

Specifically, Chest radiograph datasets are scarce; the Chest X Pert dataset [23] is a publicly available chest radiographic dataset and nominated to prove the validity of a neoteric deep learning model. The dataset is a gigantic and labelled crowd of 224,316 grey-scaled, two-dimensional chest radiograph images; the training set consists of 65,540 patient's chest radiographs and is classified into 14 distinct classes. For validation, the current study considers five categories, named Atelectasis, Cardiomegaly, Consolidation, Edema, and No Finding. Among all, these first four are pathological findings, and the last one contains none of the symptoms

matching from any of the classes so, these are normal radiographs [41].

Fig. 1 describes the images in the quantitative manner of Normal and abnormal images for training, testing, and validation of all obsessive discovery individually. For training, the Chest X pert dataset is divide into a 70 – 30 ratio. 70% of radiograph images are dedicated to the training of the network. At the same time, the residual 30% for radiograph images in reserve for testing of the network. The collection of images for different diseases oscillates from rare to multiple thousands. So the dataset is exceptionally wonky, twisted, and shaky [34]. Fig. 2 shows positive label intensity of pathological findings selected for current study. The Chest X Pert dataset contains grey-scaled uniform, JPG format images, Which contain  $x$  and  $y$ , like height and width explanation inconsistent pixels of images.

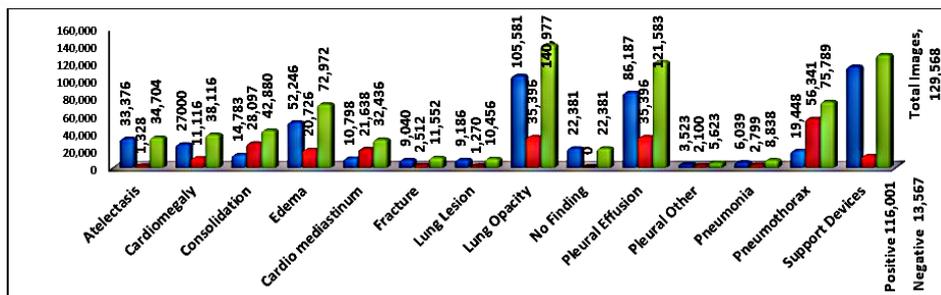


Fig. 1 Chest X pert Dataset all radiographs intensity in terms of positive and negative findings.

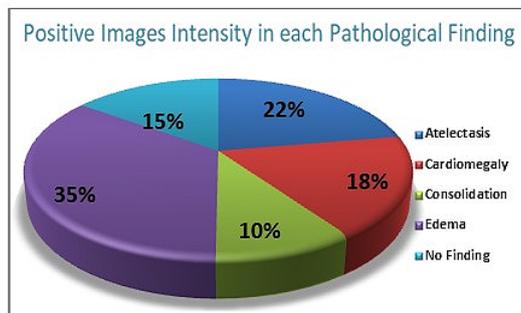


Fig. 2 Positive Label intensity in each pathological finding.

## 5. Hyper parameter optimizations

The current study considers, performs, and executes both analytical and graphical comparisons of accuracies and AUCs of all four pathological findings of chest radiographs. The systematic comparison span various learning rate values of three optimizers, SGDM, Adam, and RmsProp [3]. Each learning rate and optimizer is analyzed at three variables: Drop out probabilities, AUC and Accuracy measures.

Considering the least, moderate, and maximum probabilistic values, the current study judges the efficiency and reliability of the neoteric proposed model HNM.

## 5.1 Learning rate

The stochastic gradient descent algorithm was utilized for the training of deep learning neural networks. It is an optimization algorithm for calculating the gradient error of the recent existence of the neural network model by taking training samples from the dataset [12]. After calculation, it upgrades the bias and weights of the neural network model with the back-propagation of the error algorithm, the quantity of the weight and bias upgrade in training identified as the learning rate. Simply, the learning rate is adjustable, and the tuning parameter represented as a positive integer ranges from 0.0 to 1.0 for the training of neural networks [51]. The Learning rate is a core theme to accelerate the process of learning. A greater learning rate (probably more than 0.4) upsurges the process of learning, while during this optimization could be compromised. A comparatively lower learning rate behaves well and learns gradually. Consequently, a vibrant methodology performs better to keep the learning rate initially greater afterward decayed spontaneously. The selective compromised degrees are elected on the probing of the précised impartial and sanitizing all with experimental judgements [16][19].

## 5.2 Optimizers

For the current study, the below three optimizers, SGDM, Adam, and RmsProp, have been adopted to validate the neoteric ensemble hybrid fusion of neural networks. Based on these nominated optimizers, the functional behavior of distinct chest diseases has been assessed. Resultantly, the current study reveals the best of the best performance of classifiers after examining the network from multiple angles.

### 5.2.1 SGDM optimizer

To minimize the loss function at every phase, in a negative gradient loss track, the SGDM updates weights and biases in the network parameters.

$$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l). \quad (1)$$

Here,  $l$  is the iteration and learning rate is  $\alpha > 0$ , the parameter vector is  $\theta$ , and the loss function is  $E(\theta)$ . The loss function gradient  $\nabla E(\theta)$  is assessed with the whole training set. The standard gradient descent algorithm utilizes the entire data set only once [45]. The stochastic gradient descent algorithm could waver lengthways from the way of sharpest descent to the best. Including a momentum expression to the parameter updated value is a means to diminish the fluctuation. The stochastic gradient descent with momentum (SGDM) update is

$$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma(\theta_l - \theta_{l-1}), \quad (2)$$

where  $\gamma$  demonstrates the influence of the preceding gradient phase on the recent iteration, this value could be specified using the term ‘‘Momentum’’ name-value

pair argument. It is dedicated to training a neural network with the stochastic gradient descent momentum algorithm. The initial learning rate value is specified as  $\alpha$  [6]. The SGDM is an essential process; here, in this technique, momentum is accrued with the gradient of the previous step in making use of the only gradient in the current level for direction detection [4]. In this technique, the gradient is retrained from the preceding iteration then multiplied with “the coefficient of Momentum” which is the ethical part or percentage of the gradient retrained in each iteration. It adds rapidness and conjunction to the whole process.

$$v_j \leftarrow nv_j - \alpha \nabla w \sum_1^m Lm(w), \tag{3}$$

$$w_j \leftarrow v_j + w_j, \tag{4}$$

$n$  = The Coefficient of Momentum,  
 $v_j$  = The Retrained Gradient.

**Momentum** Ravines of the data create problems for SGD. These are sharp surfaces of a dimension in comparison to others and mutual among resident goals. SGD fluctuates among the grades of the gorge for such situations [20].

Momentum is the process of speeding up SGD in the pertinent track. It reduces fluctuations with the help of the addition of the value of fraction  $\gamma$  to upgrade other benefits with the current update vector.

In a practical approach, loss function  $l(\theta)$  must be minimized, and the model parameter(s) is  $\theta$ . In the modern deep learning approach, the best practice is to use SGD with step size  $\lambda$  to gear up training procedure by including momentum, applied to update rule [17].

The Gradient equation:  $gt = \nabla \theta'(\theta t - 1)$ ,

The Momentum equation:  $mt = \mu mt - 1 + gt$ ,

The Update equation:  $\theta t = \theta t - 1 - \lambda mt$  .

The tunable momentum parameter is  $\mu$ , which simultaneously gear up converge power. The optimization of neural networks proved experimentally received beneficial aspects with momentum. The  $gt$  is figured as an estimation of a mini-batch of training data. SGD could grasp the universal least value of a neural network even though none of the lines operations are inapplicable to step size for the whole process of classification [54].

### 5.2.2 ADAM optimizer

The Adam optimizer makes use of a parametric update identical to other optimizers, Root Mean Square propagation (RMSProp), with an additional momentum expression of both the parameter gradients and their squared values.

$$m_l = \beta_1 m_{l-1} + (1 - \beta_1) \nabla E(\theta_l), \tag{5}$$

$$v_l = \beta_2 v_{l-1} + (1 - \beta_2) [\nabla E(\theta_l)]^2, \tag{6}$$

$$\theta_{l+1} = \theta_l - \alpha m \frac{i}{\sqrt{v_l + \epsilon}}. \quad (7)$$

Here,  $\beta_1$  and  $\beta_2$  “*GradientDecayFactor*”, and “*SquaredGradientDecayFactor*,”, name-value pair arguments, respectively [25].

The “adam” optimizer combines the functionality of both rmsprop and SGDM as,

$$v_t = \beta_1 v_{t-1} - (1 - \beta_1) g_t, \quad (8)$$

$$s_t = \beta_2 s_{t-1} - (1 - \beta_2) g_t^2, \quad (9)$$

$$\Delta w_t = -n \frac{v_t}{\sqrt{s_t + \epsilon}} g_t, \quad (10)$$

$n$  = initial learning rate,

$g_t$  = gradient at time  $t$  laterally  $w_j$ ,

$v_t$  = Exponential standard of gradients alongside  $w_j$ ,

$s_t$  = Exponential typical of squares of the gradient along wise  $w_j$ ,

$\beta_1, \beta_2$  = Hyper constraints.

It acquaintances a standard of the gradient with the tetragonal of the gradients for each standard [11].

### 5.2.3 Rmsprop optimizer

The rmsprop stabs to reduce fluctuations, but in the entire variable manner compared to SGDM [7]. It again automates the process and takes far away from the necessity of learning rate adjustment. It equips each parameter with a variable learning rate. For every parameter, a distinct update was done through the below equation [37].

$$v_t = p v_{t-1} + (1 - p) g_t^2, \quad (11)$$

$$\Delta w_t = -\frac{n}{\sqrt{v_t + \epsilon}} g_t, \quad (12)$$

$$w_{t+1} = w_t + \Delta w_t, \quad (13)$$

$n$  = primary learning degree,

$v_t$  = Exponential mediocre of Squares of incline,

$g_t$  = gradient at time  $t$  laterally  $w_j$  [40].

### 5.3 Drop out probability

The term Dropout is a viable technique, targets the fundamental issues of neural networks. It prohibits overfitting and opens the doors of possibilities for the combination of distinct neural network architectures efficiently [38]. The procedure works by dropping out various hidden and visible units in a neural network. With the dropping unit and all its incoming and outgoing connections, the network virtually becomes shrinking in the number of nodes. Logically, every unit is sustained with a static probability ‘ $p$ ’ liberated of all other units. The value of  $p$  gets its optimal

probability grading closer to 1. The moderate at 0.5 for a wide variety of networks and tasks acquire the least grading at values closer to 0 [31].

The convolutional neural network contains numerous parametric configurations. Some help maintains and improves the quality and standard of network training, while others play an active role in keeping safe from divergence [36]. Silently, many parameters could divert the neural network training towards overfitting, which is due to the shortage of control parameters of the learning process [50].

Numerous networks usually get trained as the model combination for securing overfitting. Dropout resolves the problematic arena by dropping a part of units arbitrarily in probability  $1-p$  or  $p$ . Viably, Dropout makes the network differentiable with a distinct thinning than the standard neural networks [48]. Advantageously, it improves the model's confrontation for overfitting and is a primary cause of fastening in the training procedure. These thinned networks rituals a model blend to make all neurons involved in prediction and overfitting is prevented [49].

## 6. A neoteric topology – hybrid network model

As neural networks have ample dimensions to learn and innovate in their domain, the arena of medical imaging requires automation to chase accuracy and precision. A neoteric 'Hybrid Network Model' (HNM) is projected to obtain the subject. The network topology observes the real events in command to attain the Best accuracy. The neoteric topology is anticipated and intended primarily for chest image radiographs. It proceeded the input in two dimensions (2D) and provided to the neural network by execution in augmentation on the data. The designated Chest X pert dataset is intersected into a ratio of 70% for training resolution and residual for testing determination. For validation, a distinct 200 patient's chest radiographs were offered.

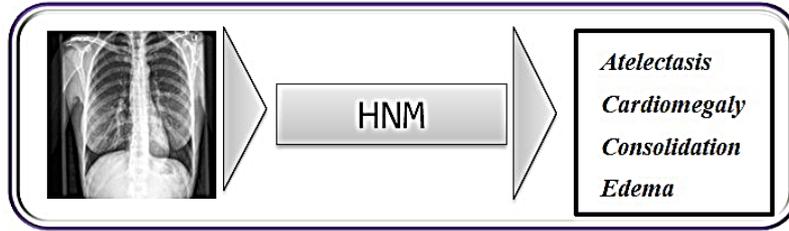
The projected neural network is trained for the radiographs to obtain the Best accuracy. Neoteric HNM is consists of many layers to recognize and classify features in a particular classification. A Hybrid Network Model (HNM) is a different, working, purposeful, and state-of-the-art influence on society. The influential target is the medical imaging arena for healthcare betterment by precise findings.

A fusion of CNN and the Bi-LSTM layer performs its convolutional and classification operation in the proposed HNM network. There are many more studies that experienced hybrid models [18]. The hybrid model comprises more than one classical network [52]. These networks complement the functionality of different neural models. The fusion of networks adheres to support and enhance the productivity of network models [1]. With the inspiration of the current state-of-the-art neural networks, a neoteric deep neural network, "Hybrid Network Model" (HNM), is developed [5]. It takes a radiograph as an input image and classifies them into four distinct classes, namely, Atelectasis, Cardiomegaly, Consolidation and Edema. It is visualized through Fig. 3.

The HNM model comprises six CNN layers combined in a parallel fashion. Fig. 4 shows it visibly. The 2D chest radiographic input reaches each CNN layer concurrently as all are joined together in parallel mode. A concatenation layer receives Its convolutional resultant. After concatenation, the resultant of parallel CNN layers are fed into the next three series CNN layers. Afterward, the only selected features

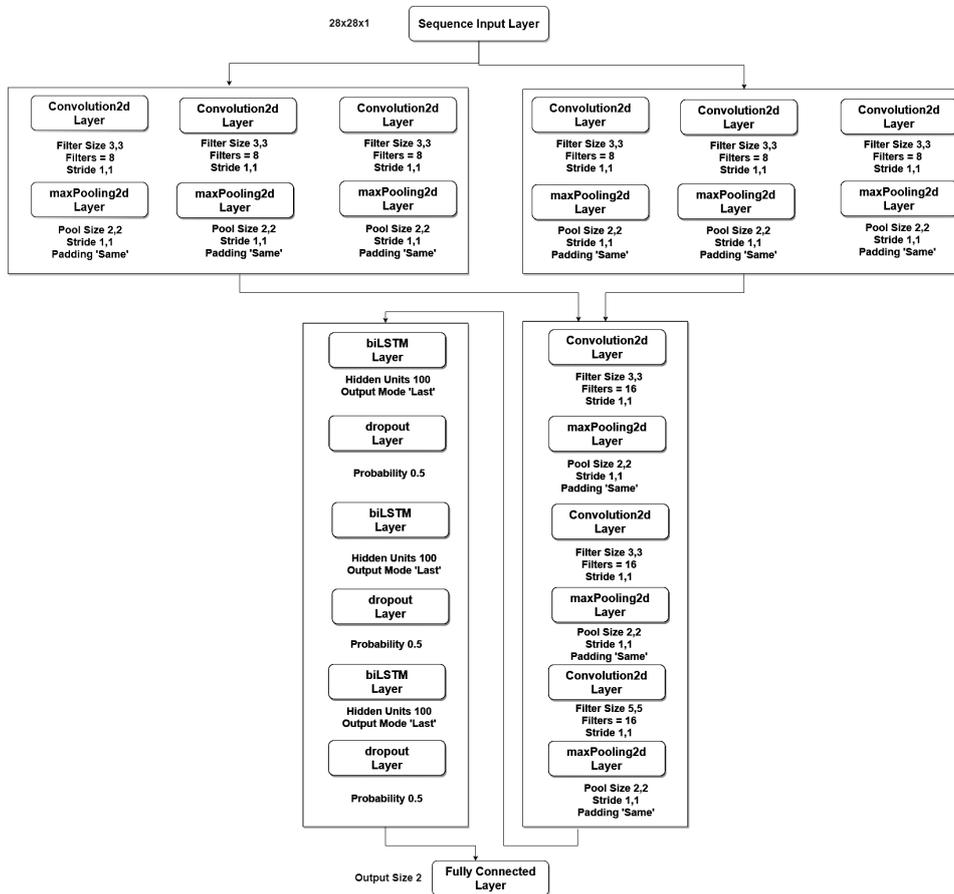
No	Layer Name	Type	Activations	Learnable
1	<b>Sequence</b> Sequence input with $28 \times 28 \times 1$ dimensions	Sequence input	$28 \times 28 \times 1$	-
2	<b>Conv_1</b> $83 \times 3 \times 1$ convolution with stride [11] and padding 'same'	Convolution	$28 \times 28 \times 8$	Weights Bias $3 \times 3 \times 1 \times 8$ $1 \times 1 \times 8$
3	<b>batchnorm_1</b> Batch normalization with 8 channels	Batch Normalization	$28 \times 28 \times 8$	Offset Scale $1 \times 1 \times 8$ $1 \times 1 \times 8$
4	<b>Relu_1</b> ReLU	ReLU	$28 \times 28 \times 8$	-
5	<b>Maxpooling_1</b> $2 \times 2$ max pooling with stride [11] and padding 'same'	Max pooling	$28 \times 28 \times 8$	-
6	<b>Concat_1</b>	Concatenation	$28 \times 28 \times 24$	-
7	<b>Flatten_1</b> Flatten	Flatten	25088	-
8	<b>Bilstm_1</b> biLSTM with 100 hidden units	BiLSTM	200	Input Weights Recurrent Weights Bias $800 \times 25088$ $800 \times 100$ $800 \times 1$
9	<b>Dropout_1</b> 50% dropout	Dropout	200	-
10	<b>Fc</b> 2 fully connected layers	Fully Connected	2	Weights Bias $2 \times 200$ $2 \times 1$
11	<b>Softmax</b> softmax	softmax	2	-
12	<b>Classoutput</b> crossentropyex	Classification Output	-	-

Tab. 1 The HNM Network Scheme Detail applicable to all parameters.



**Fig. 3** The Schematic diagram of HNM classifies the radiographs into four distinct classes.

are passed into three Bi-LSTM layers consecutively, attached serially. Eventually, the resultant is submitted by Fully Connected Layer to classify them accordingly in respective categories [52].



**Fig. 4** The Schematic network model of Ensemble Hybrid Model – HNM.

## 7. Discussion

The performance evaluation takes place in consideration of three varying Drop out probabilities as 0.1, 0.5 and 0.9, in contrast with two varying Learning Rates as 0.01 and 0.02 for three considered optimizers as ‘SGDM’, ‘adam’ and ‘rmsprop’ [8]. Each pathological finding is gone through every possible combination of assessment. Each pathological finding is assessed based on the AUC-ROC values found [33]. A higher AUC value advocates better performance [44]. The current study considers the pathological finding radiographs for assessing image size  $28 \times 28$  with a single channel, as shown in Tab. I. The network model HNM takes input 2D data in  $28 \times 28 \times 1$  format. 28 is the height, 28 is the width of the image in pixels, and 1 is the number of channels. The CNN layers consist of the Convolutional 2D layer, Batch Normalization Layer, Relu Layer, and MaxPooling2DLayer. At the same time, Bi-LSTM layers are accompanied by dropout layers. These Dropout layers play a vital role and affect the whole selection and classification process of the radiograph [47]. The configuration is described in the Tab. I. All assessments took place at a single CPU Intel processor. The time for processing any pathological finding is not measured by any means.

The existing state-of-the-art scores [39] of DenseNet 121 in Fig. 5 show that each pathological finding is assessed at Adam optimizer. At the same time, all pathologies are compared with a neoteric fusion of hybrid ensemble using varying dropout and learning rate values at different selected optimizers.

For the Atelectasis AUC values in Fig. 6, comparison at numerous parameters, the current state-of-the-art AUC DenseNet 121 value is **0.693**. While at HNM, the Atelectasis class achieves the highest AUC at **0.7899**. For the Cardiomegaly AUC values comparison, in Fig. 7, DenseNet 121 is compared with AUC value **0.8687**, while HNM compares the same class at numerous parameters and achieves the highest AUC at **0.8841**. The Consolidation current state-of-the-art AUC value **0.732** is compared at HNM with multiple parametric states and achieves the highest AUC as **0.8399** shown in Fig. 8. While, the current state-of-the-art comparison of the Edema AUC values in Fig. 9, is **0.834**, and the same class achieves the high-

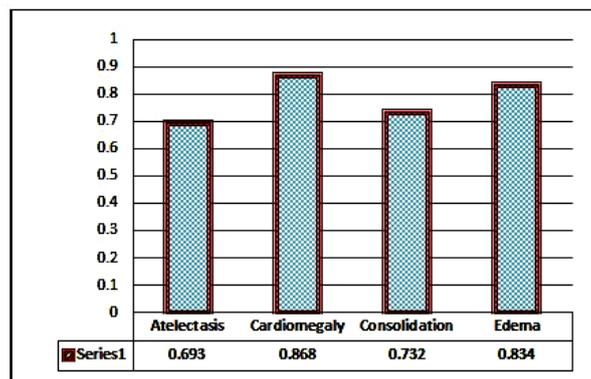


Fig. 5 Dense Net 121 AUC Values from state of the art for selected findings.

est AUC as **0.8691** in Fig. 5, at HNM by considering various parametric variables and Tab. II shows all analytical facts about current research. Fig. 10 advocates the research training process while Fig. 11 summarizes the research in terms of confusion matrix for complete research findings.

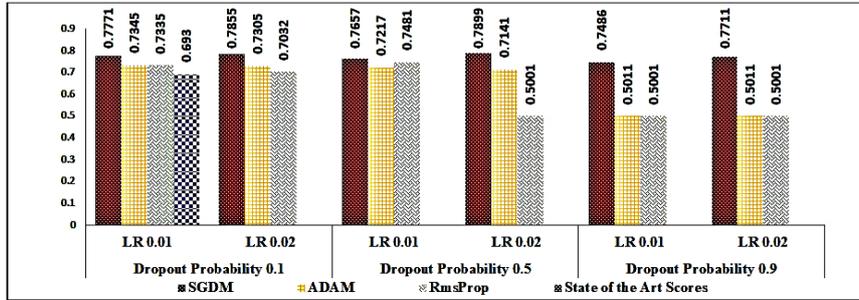


Fig. 6 Atelectasis AUC is compared along with state of the art and many essential parameters.

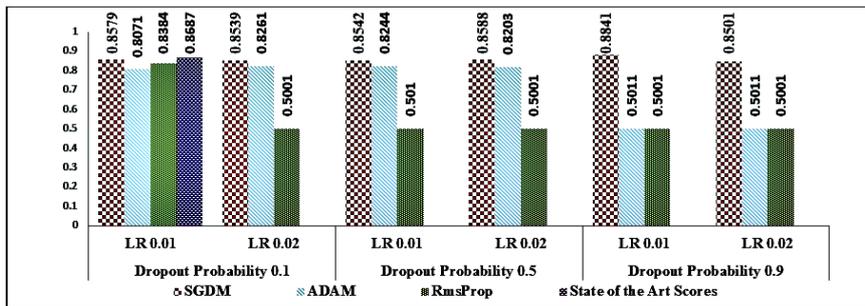


Fig. 7 Cardiomegaly AUC comparison of DenseNet 121 and HNM parameters.

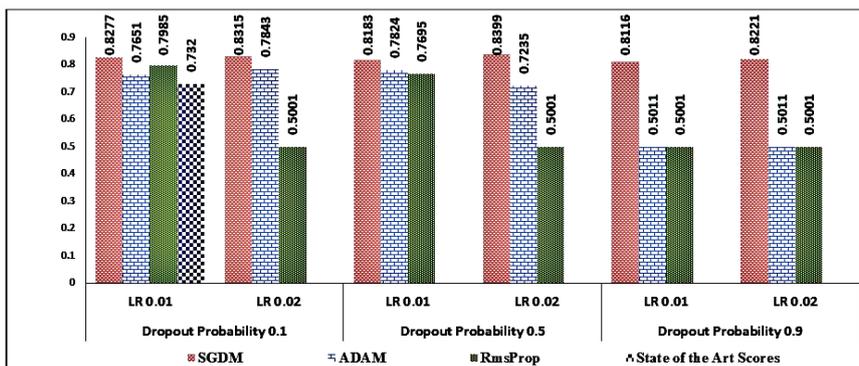


Fig. 8 Consolidation AUC, in comparison of parametric values among HNM and DenseNet 121.

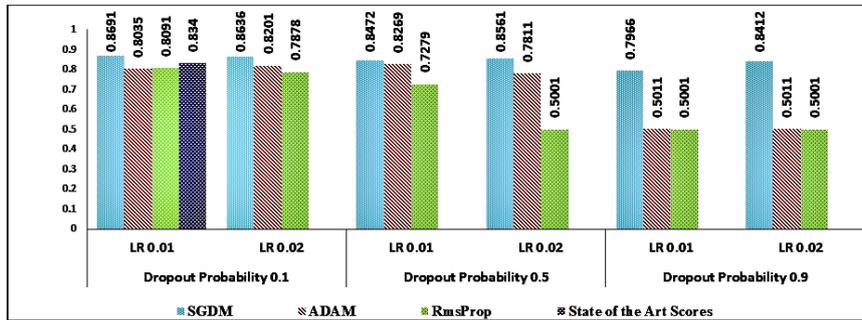


Fig. 9 Edema class HNM AUC values parametric comparison with DenseNet 121.

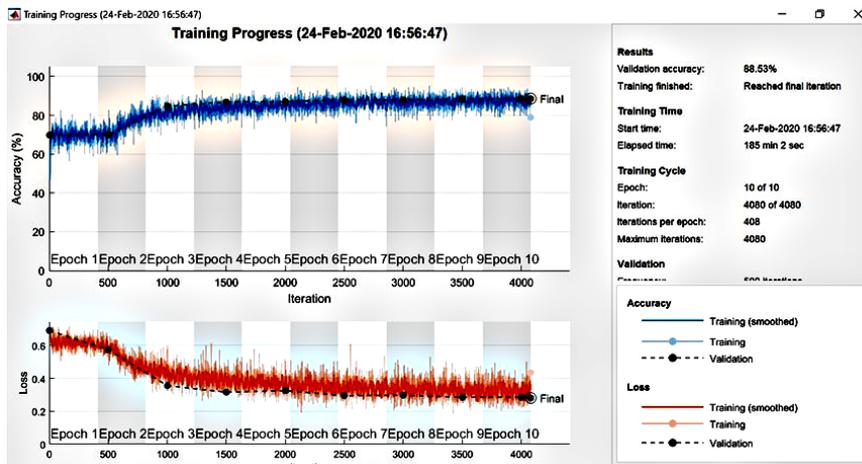


Fig. 10 The Training process.

	SGDM		ADAM		RMS Prop	
Atelectasis	8070	1518	8124	2299	3423	1365
	1943	5196	1889	4415	1012	5349
Cardiomegaly	6989	1173	6970	1421	6976	1238
	1111	5541	1130	5293	1124	5476
Consolidation	3665	984	3423	1365	3456	1224
	770	5730	1012	5349	979	5490
Edema	1414	1103	423	326	1459	2104
	1531	5611	190	321	1077	4610

Confusion Matrix

Fig. 11 The Confusion Matrix for the Best of each class finding.

Optimizer	SGDM						ADAM						RMS Prop						
	0.1		0.5		0.9		0.1		0.5		0.9		0.1		0.5		0.9		
LR	Drop Out Probability	AUC	Accuracy																
0.01	Comparative Measure	0.7771	80.18	0.7657	79.41	0.7486	77.17	0.7345	74.96	0.7217	74.95	0.5011	59.86	0.7335	75.65	0.7481	76.31	0.5001	59.86
	Atelectasis	0.8579	85.85	0.8542	85.95	0.8841	84.58	0.8071	81.01	0.8244	82.78	0.5011	54.67	0.8384	84.05	0.501	54.67	0.5001	54.67
	Cardiomegaly	0.8277	84.02	0.8183	83.61	0.8116	82.57	0.7651	78.11	0.7824	77.35	0.5011	60.22	0.7985	80.24	0.7695	77.71	0.5001	60.22
	Consolidation	0.8691	88.23	0.8472	88.88	0.7966	86.63	0.8035	85.13	0.8269	85.01	0.5011	70.01	0.8091	85.79	0.7279	82.99	0.5001	70.01
0.02	Edema	0.7855	79.96	0.7899	79.31	0.7711	79.06	0.7305	76.01	0.7141	73.91	0.5011	59.86	0.7082	73.14	0.5001	59.86	0.5001	59.86
	Atelectasis	0.8539	85.67	0.8588	86.11	0.8501	84.93	0.8261	83.02	0.8203	81.69	0.5011	54.67	0.5001	54.67	0.5001	54.67	0.5001	54.67
	Cardiomegaly	0.8315	84.32	0.8399	84.27	0.8221	82.99	0.7843	78.67	0.7235	75.63	0.5011	60.22	0.5001	60.22	0.5001	60.22	0.5001	60.22
	Consolidation	0.8636	88.87	0.8561	88.88	0.8412	88.53	0.8201	84.64	0.7811	84.01	0.5011	70.01	0.7878	84.63	0.5001	70.01	0.5001	70.01
	Edema																		

Tab. II The experimental analytics is captured to compared against numerous valuable parameters.

## 8. Conclusion

The innovation and prior finding play an utmost vital role in saving a life. The current study concludes via an important yet vital contribution in Biomedical imaging in hybrid ensemble fusion for Chest radiographs classification. It complements one more stride towards a better verdict of classification and the proposed neural network HNM is proposed and consequences are validated at Chest X pert dataset. Judgmentally compares and proves the enhanced and purified upgraded credentials of classification for distinct features efficiently. The proposed ensemble of CNN and Bi-LSTM model HNM evidence the enriched AUC values at three pathological abnormalities Atelectasis, Consolidation, and Edema in contrast with state of the art Dense Net 121. In inspiration for the current research study, further novel research could be conducted in the arena of Deep Learning. More study could lead towards the significant proof of HNM by consideration of other datasets of the same subject or could move towards innovations of Deep Learning neural networks. In the same domain, some of the classes' performance is poor for the proposed network, so an improvement of the same study is predictable.

## References

- [1] ALAYBA A.M., PALADE V., ENGLAND M., IQBAL R. A combined CNN and LSTM model for arabic sentiment analysis. In: *International cross-domain conference for machine learning and knowledge extraction*, 2018, pp. 179–191,
- [2] ALLEN JR B., SELTZER S.E., LANGLOTZ C.P., DREYER K.P., SUMMERS R.M., PETRICK N., MARINAC-DABIC D., CRUZ M., ALKASAB T.K., HANISCH R.J., et al. A road map for translational research on artificial intelligence in medical imaging: from the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. *Journal of the American College of Radiology*. 2019, 16(9), pp. 1179–1189.
- [3] ARDAKANI A.A., KANAFI A.R., ACHARYA U.R., KHADEM N., MOHAMMADI A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. *Computers in biology and medicine*. 2020, 121, pp. 103795.
- [4] ARJMAND A., ANGELIS C.T., CHRISTOU V., TZALLAS A.T., TSIPOURAS M.G., GLAVAS E., FORLANO R., MANOUSOU P., GIANNAKEAS N. Training of deep convolutional neural networks to identify critical liver alterations in histopathology image samples. *Applied Sciences*. 2020, 10(1), pp. 42.
- [5] BAE S.H., CHOI I., KIM N.S. Acoustic scene classification using parallel combination of LSTM and CNN. In: *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE2016)*, 2016, pp. 11–15,
- [6] BAŞARAN E., CÖMERT Z., ÇELİK Y. Convolutional neural network approach for automatic tympanic membrane detection and classification. *Biomedical Signal Processing and Control*. 2020, 56, pp. 101734.
- [7] BERA S., SHRIVASTAVA V.K. Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification. *International Journal of Remote Sensing*. 2020, 41(7), pp. 2664–2683.

- [8] BOLUKI S., ARDYWIBOWO R., DADANEH S.Z., ZHOU M., QIAN X. Learnable Bernoulli dropout for Bayesian deep learning. In: *International Conference on Artificial Intelligence and Statistics*, 2020, pp. 3905–3916,
- [9] CHAUHAN R., GHANSHALA K.K., JOSHI R. Convolutional neural network (CNN) for image detection and recognition. In: *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, 2018, pp. 278–282,
- [10] CHOI K.S., SHIN J.-S., LEE J.-J., KIM Y.S., KIM S.-B., KIM C.-W. In vitro trans-differentiation of rat mesenchymal cells into insulin-producing cells by rat pancreatic extract. *Biochemical and biophysical research communications*. 2005, 330(4), pp. 1299–1305.
- [11] DÉFOSSEZ A., BOTTOU L., BACH F., USUNIER N. On the convergence of Adam and Adagrad. *arXiv e-prints*. 2020, pp. arXiv–2003.
- [12] EDE J.M., BEANLAND R. Adaptive learning rate clipping stabilizes learning. *Machine Learning: Science and Technology*. 2020, 1(1), pp. 015011.
- [13] ESTEVA A., ROBICQUET A., RAMSUNDAR B., KULESHOV V., DEPRISTO M., CHOU K., CUI C., CORRADO G., THRUN S., DEAN J. A guide to deep learning in healthcare. *Nature medicine*. 2019, 25(1), pp. 24–29.
- [14] FISZMAN M., CHAPMAN W.W., ARONSKY D., EVANS R.S., HAUG P.J. Automatic Detection of Acute Bacterial Pneumonia from Chest X-ray reports. *Journal of the American Medical Informatics Association*. 2000, 7(6), pp. 593–604.
- [15] FUJITA H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiological physics and technology*. 2020, 13(1), pp. 6–19.
- [16] GUPTA C., NANGIA C., KUMAR C. Self Learning Robot using Real-Time Neural Networks. *International Journal of Research in Engineering and Technology*. 2020, 07(10), pp. 52–59.
- [17] HAAS T., SCHUBERT C., EICKHOFF M., PFEIFER H. BubCNN: Bubble detection using Faster RCNN and shape regression network. *Chemical Engineering Science*. 2020, 216, pp. 115467.
- [18] HAO J., LUO S., PAN L. A Novel Sleep Staging Algorithm Based on Hybrid Neural Network. In: *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, 2019, pp. 158–161,
- [19] HASHEMI S.M.R., HASSANPOUR H., KOZEGAR E., TAN T. Cystoscopic Image Classification Based on Combining MLP and GA. *International Journal of Nonlinear Analysis and Applications*. 2020, 11(1), pp. 93–105.
- [20] HE K., FAN H., WU Y., XIE S., GIRSHICK R. Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 9729–9738,
- [21] HEO S.-J., KIM Y., YUN S., LIM S.-S., KIM J., NAM C.-M., PARK E.-C., JUNG I., YOON J.-H. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers’ health examination data. *International journal of environmental research and public health*. 2019, 16(2), pp. 250.
- [22] HUA Y., MOU L., ZHU X.X. Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image classification. *ISPRS journal of photogrammetry and remote sensing*. 2019, 149, pp. 188–199.

- [23] IRVIN J., RAJPURKAR P., KO M., YU Y., CIUREA-ILCUS S., CHUTE C., MARKLUND H., HAGHGOO B., BALL R., SHPANSKAYA K., et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In: *Proceedings of the AAAI conference on artificial intelligence*, 01, 2019, pp. 590–597,
- [24] IWAN T., KAVI O., YILDIRIM E. Learning Machines from Simulation to Real World. *arXiv preprint arXiv: 2002.10853*. 2020.
- [25] KINGMA D., LEI BA J. Adam: A method for Stochastic Optimization. In: *3rd International Conference on Learning Representations*, San Diego, CA, USA: ICLR 2015, 2015, pp. 1–15,
- [26] LAKHANI P., SUNDARAM B. Deep learning at chest radiography: Automated Classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017, 284(2), pp. 574–582.
- [27] LEE S.M., SEO J.B., YUN J., CHO Y.-H., VOGEL-CLAUSSEN J., SCHIEBLER M.L., GEFTER W.B., VAN BEEK E.J., GOO J.M., LEE K.S., et al. Deep learning applications in chest radiography and computed tomography. *Journal of thoracic imaging*. 2019, 34(2), pp. 75–85.
- [28] LENG M., SCHULZ H., SAALBACH A. Continual learning for domain adaptation in chest x-ray classification. In: *Medical Imaging with Deep Learning*, 2020, pp. 413–423,
- [29] MAHALAKSHMI S.D., GEETHANJALI V. Plant Classification using Deep Learning. *Journal of International Pharmaceutical Research*. 2019, 46(3), pp. 745–749.
- [30] MCCRARY M.B. Urban multicultural trauma patients. *ASHA*. 1992, 34(4), pp. 37–40.
- [31] MELE B., ALTARELLI G. Lepton spectra as a measure of b quark polarization at LEP. *Physics Letters B*. 1993, 299(3-4), pp. 345–350.
- [32] MIRJALILI S., DONG J.S., LEWIS A. *Nature-inspired optimizers*. 2020.
- [33] MU Y., XIE F., SUN T. Clinical value of seven autoantibodies combined detection in the diagnosis of lung cancer. *Journal of Clinical laboratory analysis*. 2020, 34(8), pp. e23349.
- [34] NAIR A., NAIR U., NEGI P., SHAHANE P., MAHAJAN P. Detection of Diseases on Chest X-ray Using Deep Learning. *Cikitsa Journal for Multidisciplinary Research*. 2019, 6(5), pp. 298–305.
- [35] OZTURK T., TALO M., YILDIRIM E.A., BALOGLU U.B., YILDIRIM O., ACHARYA U.R. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine*. 2020, 121, pp. 103792.
- [36] PAL A., LANE C., VIDAL R., HAEFFELE B.D. On the Regularization Properties of Structured Dropout. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7671–7679,
- [37] PENG W., DAI Y.-H., ZHANG H., CHENG L. Training GANs with centripetal acceleration. *Optimization Methods and Software*. 2020, 35(5), pp. 955–973.
- [38] PIOTROWSKI A.P., NAPIORKOWSKI J.J., PIOTROWSKA A.E. Impact of deep learning-based dropout on shallow neural networks applied to stream temperature modelling. *Earth-Science Reviews*. 2020, 201, pp. 103076.
- [39] POOCH E.H., BALLESTER P.L., BARROS R.C. Can we trust deep learning models diagnosis? The impact of domain shift in chest radiograph classification. *arXiv preprint arXiv: 1909.01940*. 2019.

- [40] POSTALCIOĞLU S. Performance analysis of different optimizers for deep learning-based image recognition. *International Journal of Pattern Recognition and Artificial Intelligence*. 2020, 34(02), pp. 2051003.
- [41] QIN Z.Z., SANDER M.S., RAI B., TITAHONG C.N., SUDRUNGROT S., LAAH S.N., ADHIKARI L.M., CARTER E.J., PURI L., CODLIN A.J., et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Scientific reports*. 2019, 9(1), pp. 1–10.
- [42] RAGHU M., SCHMIDT E. DRAFT: A Survey of Deep Learning for Scientific Discovery. 2020, pp. 1–37.
- [43] RAJPURKAR P., IRVIN J., ZHU K., YANG B., MEHTA H., DUAN T., DING D., BAGUL A., LANGLOTZ C., SHPANSKAYA K., et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*. 2017, pp. 3–9.
- [44] RICCIARDI C., EDMUNDS K.J., RECENTI M., SIGURDSSON S., GUDNASON V., CARRARO U., GARGIULO P. Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions. *Scientific reports*. 2020, 10(1), pp. 1–13.
- [45] SRINIVAS B., DHATRI R., DEDEEPIYA C., AKHIL C., TEJESH C. Performance Analysis of Brain Tumor Classification using Performance Analysis of Brain Tumor Classification using Supervised Learning Based Model. *Alochana Chakra Journal*. 2020, 9(4), pp. 3721.
- [46] TOPOL E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*. 2019, 25(1), pp. 44–56.
- [47] VERBAKEL J.Y., STEYERBERG E.W., UNO H., DE COCK B., WYNANTS L., COLLINS G.S., VAN CALSTER B. ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. *Journal of Clinical Epidemiology*. 2020, 126, pp. 207–216.
- [48] WANG S.-H., HONG J., YANG M. Sensorineural hearing loss identification via nine-layer convolutional neural network with batch normalization and dropout. *Multimedia Tools and Applications*. 2020, 79(21), pp. 15135–15150.
- [49] WANG S.-H., MUHAMMAD K., HONG J., SANGAIAH A.K., ZHANG Y.-D. Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization. *Neural Computing and Applications*. 2020, 32(3), pp. 665–680.
- [50] WU H., GU X. Towards Dropout training for Convolutional Neural Networks. *Neural Networks*. 2015, 71, pp. 1–10.
- [51] WU Y., GAO R., YANG J. Prediction of coal and gas outburst: A method based on the BP neural network optimized by GASA. *Process Safety and Environmental Protection*. 2020, 133, pp. 64–72.
- [52] ZHU F., YE F., FU Y., LIU Q., SHEN B. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Scientific reports*. 2019, 9(1), pp. 1–11.
- [53] ZHU H. Big data and artificial intelligence modeling for drug discovery. *Annual review of pharmacology and toxicology*. 2020, 60, pp. 573–589.

Sultana S. et al.: A deep learning hybrid ensemble fusion for chest radiograph. . .

- [54] ZIYIN L., WANG Z.T., UEDA M. Laprop: a better way to combine momentum with adaptive gradient. *arXiv preprint arXiv: 2002.04839*. 2020.