



## IMPROVED PREDICTIVE SPARSE DECOMPOSITION METHOD WITH DENSENET FOR PREDICTION OF LUNG CANCER

Ibomoiye Domor Mienye <sup>1)</sup>, Yanxia Sun <sup>1)</sup>, Zenghui Wang <sup>2)</sup>

<sup>1)</sup> Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg 2006, South Africa, ibomoiye@yahoo.com, ysun@uj.ac.za

<sup>2)</sup> Department of Electrical and Mining Engineering, University of South Africa, Florida 1709, South Africa, wangz@unisa.ac.za

### Paper history:

Received 22 November 2019  
Received in revised form 24 July 2020  
Accepted 17 October 2020  
Available online 30 December 2020

### Keywords:

convolutional neural network;  
deep learning;  
denseNet;  
predictive sparse decomposition.

**Abstract:** Lung cancer is the second most common form of cancer in both men and women. It is responsible for at least 25% of all cancer-related deaths in the United States alone. Accurate and early diagnosis of this form of cancer can increase the rate of survival. Computed tomography (CT) imaging is one of the most accurate techniques for diagnosing the disease. In order to improve the classification accuracy of pulmonary lesions indicating lung cancer, this paper presents an improved method for training a densely connected convolutional network (DenseNet). The optimized setting ensures that code prediction error and reconstruction error within hidden layers are simultaneously minimized. To achieve this and improve the classification accuracy of the DenseNet, we propose an improved predictive sparse decomposition (PSD) approach for extracting sparse features from the medical images. The sparse decomposition is achieved by using a linear combination of basis functions over the  $L_2$  norm. The effect of dropout and hidden layer expansion on the classification accuracy of the DenseNet is also investigated. CT scans of human lung samples are obtained from The Cancer Imaging Archive (TCIA) hosted by the University of Arkansas for Medical Sciences (UAMS). The proposed method outperforms seven other neural network architectures and machine learning algorithms with a classification accuracy of 95%.

Copyright © Research Institute for Intelligent Computer Systems, 2020.  
All rights reserved.

## 1. INTRODUCTION

The lung is a large organ, and this means that tumors can keep growing for a long time before being detected. Therefore, many tumors are detected when they have progressed significantly [1]. Lung computed tomography (CT) scans remain one of the most effective techniques for detecting lung tumors as a marker of lung cancer due to its non-invasive nature [2]. However, with advancements in computing techniques and artificial intelligence, it is possible to minimize the problem of lung cancer misdiagnosis due to inaccurate interpretation of lung CT images. This is because differences between cancerous and non-cancerous lesions are generally not easy to detect [2].

Convolutional neural networks (CNNs) have been used extensively for image classification and

recognition over the years [3-8]. However, like other neural network structures, CNNs are also susceptible to problems of false classification due to inaccurate feature selection. An optimal sparse representation of image data matrices is a vital requirement for accurate approximation of the input matrix into the classifier. It is essential in denoising of matrices representing image data since it attempts to capture crucial details of the image matrix with the least possible number of features. Predictive sparse decomposition (PSD) is a technique that has greatly reduced the cost of calculating sparse representations [9]. The approach involves the simultaneous optimization of both a loss function (using a feedforward predictor) and the basis functions representing the image matrix (using a nonlinear regressor). The basic approach of sparse representation involves the use of sparse matrices to

represent image data. Therefore, using the feedforward predictor alone makes the representation process both time-consuming and computationally expensive [9]. Sparse representation aims to make prediction of the regressor as close as possible to the optimal set of coefficients  $R^*$  represented by Eq. (1):

$$\Gamma(\Phi, R; F) = \frac{1}{2} \|\Phi - FR\|_2^2 + \omega \|R\|_1, \quad (1)$$

where  $\Phi$  is the output signal,  $R$  is the input sparse matrix,  $F$  is the matrix of basis functions representing the input matrix,  $\omega$  is sparse representation coefficient. Equation (1) represents the idea of Basis pursuit denoising. This approach attempts to minimize reconstruction error using sparsity and linear basis function sets. However, iteratively doing this makes the minimization algorithm computationally expensive. PSD adds a nonlinear regressor to Equation (1) for the optimization of  $F$ , which makes the optimization process faster and less expensive. Therefore, we propose an optimal approach to the selection of the sparse representation coefficient in order to improve the accuracy of the sparse representation of the output signal  $\Phi$ .

Convolutional neural network (CNN) is a deep, feedforward neural network that is used extensively in visual image analysis [10]. The idea of convolution is based on the fact that feature maps representing images are usually stacked on top of each other. Therefore, each 2D grid in the stack is convolved with the grid next to it to represent the image. Each 2D grid is represented by an  $n \times m$  matrix or feature map. The convolution between any two functions  $a$  and  $b$  measures the overlap between the functions when they are both shifted by a factor  $\lambda$  and then flipped [11].

A CNN uses cross-correlation to process input matrix and kernels, and then adds a bias in order to generate an output. Therefore, cross-correlation and the convolution process can be used interchangeably. This paper aims to present an improved method of sparse representation of the input data matrix for lung cancer prediction implemented with a recently proposed CNN architecture called densely connected convolutional network (DenseNet). The dataset is obtained from The Cancer Imaging Archive (TCIA) hosted by the University of Arkansas for Medical Sciences (UAMS). Objectives of the research include:

- Demonstration of the capability of DenseNet to accurately classify noisy and sparse datasets.
- Use of unsupervised improved PSD method to improve the performance of the DenseNet.
- Comparison of the performance of the proposed

improved PSD/DenseNet approach with other CNN architectures and machine learning algorithms.

## 2. RELATED WORKS

CNNs are particularly suited to image recognition and classification tasks. The DenseNet is an instance of a convolutional neural network which consists of densely connected layers. In this kind of CNN, features from all the preceding layers are fed into subsequent layers within the network [33]. This results in a network architecture that enables the integrity of the information fed into it to be preserved as it moves through successive layers from input to output. In recent years, there have been several applications of these neural networks in areas such as medicine, security, science, and engineering. In [12], an autoencoder CNN was used to detect stomach tumors in double contrast X-ray images. The CNN involved stages of convolution and deconvolution using 1-dimensional  $3 \times 3$  feature maps with rectified linear units (ReLU) as the activation function. The training images involved those with no tumors, as well as those with benign and malignant tumors. The performance of this proposed structure was compared to a previously proposed approach using just an autoencoder neural network. Performance indices were mean squared error (MSE) and area under the curve (AUC) based on Receiver Operating Characteristic (ROC) curves for evaluating CAD systems. The CNN outperformed the autoencoder neural network on both MSE and AUC.

In [13], a deep CNN with 16 convolutional layers was employed to identify the accurate fetal facial standard plane in order to detect malformation in the early stages of pregnancy. The input to the deep CNN was a  $224 \times 224$  RGB image. The deep CNN used five stages of max-pooling for dimensionality reduction. The proposed deep CNN was compared to 18 other CNN models and outperformed them with accuracy of up to 97%.

In [14], a CNN was used for pathology detection in chest radiographs. A 433-image dataset was employed with performance comparison to GIST, and bag-of-visual words approach. CNN was the best performer with an accuracy of up to 0.94 using AUC metric. In [15], CNN was used for the classification of breast cancer based on histology images. The deep learning approach was compared with two other coding models (namely bag of words and locality constrained linear coding). CNN outperformed the other two methods with an accuracy of 98.33% and 88.23% for binary and multi-class classification, respectively. Other

successful applications of CNNs in medical imaging can be seen in [16-25].

From these results, it is clear that CNNs are very effective tools for accurate image classification. However, the speed of prediction and classification, and a large number of training parameters are challenges that are encountered in the use of CNNs. CNNs are also prone to problems of false classification due to inaccurate feature selection. This paper proposes the use of the improved predictive sparse decomposition (PSD) approach with DenseNet to remedy the above-mentioned problems. DenseNets have been successfully applied in medical image processing and classification [26-28]. Other applications of neural networks in medical image processing can also be found in [30, 31, 34].

### 3. PROPOSED METHODOLOGY

The proposed approach involves including a nonlinear approximation of the basis function set. This approximation is added to Eq. (1) in order to improve the overall sparse representation approach. In particular, we assume a linear combination of basis functions over the L2 norm with an interval [0,1]. The linear combination is given as:

$$\mathcal{L}_p \eta = \prod_{j \in z_i} \delta_j \beta_j, \tag{2}$$

where  $\mathcal{L}_p$  represents selected basis functions,  $\eta$  represents the resulting feature map,  $\delta_j$  are basis function coefficients, and  $\beta_j$  are the orthonormal basis functions.

The purpose of this proposed PSD approach is to provide an effective means of extracting features from images that would enable the neural network to accurately interpret key markers, hence, identifying lung cancer effectively. The method of sparse decomposition reduces the computational time required to interpret images by reducing the number of distinguishable features necessary to make an accurate classification. To get the best linear ( $L_2$ ) approximation of the basis function set, we obtain the orthonormal representation of the basis function coefficients as:

$$\|\eta - \mathcal{L}_p\|_2^2 = \prod \delta_j^2. \tag{3}$$

The best possible subset of  $n$  basis functions out of a pool size of  $N$  is obtained using the following relation:

$$b_n^2(\eta) = \prod_{n=1}^N \{\|\eta - \mathcal{L}_p\|_2^2\}. \tag{4}$$

Equation (4) also represents the summation of basis function coefficients ( $\delta_j$ ). The nonlinear representation of the selected basis functions is obtained using a nonlinear approximation operator according to:

$$\theta_j \eta = \mathcal{L}_p \eta, \tag{5}$$

where  $\theta_j$  represents the nonlinear approximation operator describing the subset of selected basis functions. Equation (5) is incorporated into the PSD minimization function as follows:

$$\min_{\sigma, \gamma} \sum_{j=1}^m (\|x^j - \sum_{k=1}^n \sigma_k^j \gamma_k\|^2 + \theta_j \eta). \tag{6}$$

From Equation (6), we replace the  $L_1$  sparsity term in the original PSD with the nonlinear  $L_2$  approximation operator. The novelty of this research lies in the improvement of the conventional PSD approach (represented by the first term of Equation (6)). This involves including a nonlinear approximation of the basis function set. The purpose of this approach is to use the nonlinear method described in Equation (6) to improve the accuracy of the resulting sparse matrix, which is the input to the DenseNet. This will consequently also enhance the accuracy of the DenseNet classification. The denoising process of the resulting nonlinear approximation model is done according to the  $L_1$  approach detailed in [29]. Thus, using the augmented Lagrangian function of the  $L_1$  norm:

$$\mathcal{L}(A, B, \Gamma) = \|D(A)\|_1 + \phi \|B\|_1 - \langle \Gamma, A + B - S \rangle + \frac{\rho}{2\|A+B-S\|^2} \tag{7}$$

where  $\Gamma$  and  $\rho$  are the local variable and positive-valued penalty parameters respectively.  $S$  is the given signal to be denoised, and decomposed into two matrices  $A$  and  $B$ .  $\phi$  is an average weighting factor which balances sparsities of  $D(A)$  and  $B$ . Based on the penalty parameter, an iterative approach is used according to:

$$A_k, B_k \in \min |\mathcal{L}(A, B, X_{k-1})|, \tag{8}$$

where  $X_k = X_{k-1} - \rho(A_k + B_k - S)$

This approach is referred to as sparse noise denoising using the alternating direction method of multipliers (ADMM).

In this paper, together with the improved PSD, we implement an instance of DenseNet [32]. The DenseNet, which is a recently proposed neural network structure, has several advantages over traditional convolutional neural networks. Some of these include enhanced propagation of learned intra-layer features throughout the neural network,

reduction of the number of network training parameters, maintenance of learning gradient throughout the learning process, and encouragement of feature reuse throughout the layers of the neural network [32]. These advantages are a result of the fact that unlike conventional CNNs, successive layers of DenseNets use the feature maps of all preceding layers as inputs to the next layer. Also, while CNNs have  $N$  connections for each of the  $N$  layers, DenseNets have  $\frac{N(N+1)}{2}$  direct connections between successive layers [32]. The DenseNet structure proposed in this paper is shown in Fig. 1.

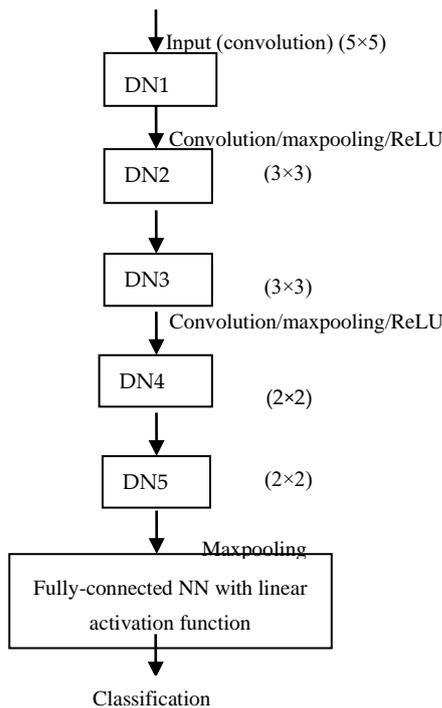


Figure 1 – Proposed 5-layer DenseNet Classifier

The feature maps for successive layers of the DenseNet are obtained according to the relation [16]:

$$\sigma_n = \Gamma_n([\sigma_0, \sigma_1, \dots, \sigma_{n-1}]), \quad (9)$$

where  $\sigma_n$  is the feature map for the  $n$ -th layer,  $\Gamma_n$  is the composite function representing normalization, pooling, convolution and rectified linear units (ReLU) for  $n$  layers of the DenseNet.

From the proposed structure in Fig. 1, features from 120 images representing lung cancer patients are reduced to a  $5 \times 5$  input convolution window (representing the input to the DenseNet). Eighty-five (85) images are used for training, while 35 are used for testing and validation. This feature map is passed to the next DenseNet layer (DN2) with convolution and max pooling. Max pooling reduces the feature map to a  $3 \times 3$  window, and this process continues

throughout the DenseNet. It should be noted that no convolution or max-pooling occurs between layers 2 and 3 of the DenseNet. This is because care is taken to avoid eliminating essential attributes of the feature map, which can enhance the accuracy of the final classification process. The final classification neural network is a fully connected feedforward network with a linear activation function.

The conventional mean square error (MSE) metric has a negative effect on the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) indices of the final compressed image [30]. Therefore, to improve the output of the network, the loss function is modified to become [33]:

$$\mathcal{L}_N = \sqrt{(y_n - y_g)^2 + \epsilon^2}, \quad (10)$$

where  $y_n$  is the network output,  $y_g$  is the ground truth output,  $\epsilon$  is the error constant. To avoid degradation of the learning process as information progresses through the network, we use a cross-entropy cost function.

$$\mathcal{F} = -\frac{1}{N_i} \sum_i [i_d \ln i_n + (1 - i_d) \ln(1 - i_n)], \quad (11)$$

where  $N_i$  is the total number of inputs to the network,  $i_d$  is the desired output,  $i_n$  is the input to the next layer. To prevent overfitting of the training data, we use  $L_2$  regularization, which involves adding a regularization term to the cross-entropy cost function.

$$\mathcal{F} = -\frac{1}{N_i} \sum_i [i_d \ln i_n + (1 - i_d) \ln(1 - i_n)] + \frac{\sigma}{2N_i} \sum_i w_i^2, \quad (12)$$

where  $w_i$  is the  $i$ -th weight in the network,  $\sigma$  is the regularization parameter. To address the vanishing gradient problem, we adopt the following expression for the cost/bias gradient [34]:

$$\frac{\partial \mathcal{C}}{\partial b_i} = \alpha'(\eta_1) w_2 \alpha'(\eta_2) w_3 \alpha'(\eta_3) \dots \alpha'(\eta_n) \frac{\partial \mathcal{C}}{\partial \beta_n}, \quad (13)$$

where  $\beta_n$  is the activation for neuron  $n$ ,  $\eta_i$  is the weighted input into neuron  $i$ ,  $\alpha'$  is the conversion term which converts a change in the bias of neuron  $i$  into a corresponding change in its output activation. The output neurons of the DenseNet form a softmax layer. The sum of inputs for the softmax layer forms a probability distribution that sums up to 1.

For each node  $i$  in layer  $j$ , the estimator of feature learning with respect to layer  $k$  is verified using maximum likelihood estimation [37]. This is done using a posteriori estimation with respect to the learning rate  $\eta$ . The a posteriori probability is described as:

$$P_a(\Theta|\eta) = \gamma \sum_{i=1}^m \sum_{k=1}^n \sigma_n \mathcal{F}, \quad (14)$$

where  $\mathcal{F}$  is the cross-entropy cost function,  $\sigma_n$  is a feature map for the  $n$ -th layer of the DenseNet,  $\gamma$  is the normalization constant. Concerning the estimation parameter  $\Theta$ , selected parameters are based on the size ( $\beta_s$ ) and shape ( $\beta_{sh}$ ) of cancerous cells with respect to normal cells. A ratio greater than one will likely mean that the cells are unhealthy. Therefore,

$$\Theta = (\beta_s, \beta_{sh}, I_n), \quad (15)$$

where  $I_n$  is the size of the image. The regularization of the classifier layer during the training of the DenseNet is ensured using the approach proposed in [38]. This is done by ensuring a uniform probability for each input label in the training stage by considering the probability of each basis function with regard to the input label  $y$  as:

$$p(y|x) = \frac{e^{\rho y}}{\sum_{i=1}^Y e^{\rho i}}, \quad (16)$$

where  $\rho_i$  are the unnormalized logarithmic probabilities with respect to each label  $y$ . Therefore, a single loss coefficient  $\mathcal{L}(x, y)$  based on the cross-entropy can be replaced by a loss pair  $\mathcal{L}(x, p)$ ,  $\mathcal{L}(\theta, p)$  according to:

$$\mathcal{L}(x, y) = -\sum_{y=1}^Y \log p(y)(1 - p)(y) = (1 - e)\mathcal{L}(x, p) + e\mathcal{L}(\theta, p), \quad (17)$$

where  $e$  is an error coefficient.

The procedure for the proposed improved PSD approach for image feature selection based on the  $L_2$  norm is detailed in Algorithm 1. The DenseNet architecture has been optimized to minimize the tendency for misclassification as well as decaying gradient from output to input. The procedure for achieving this is outlined in Algorithm 2.

**Algorithm 1.** Procedure for improved PSD using  $L_2$  norm

1. Start
2. Initialize  $\mathcal{L}_p, \eta, \delta_j, \beta_j$
3. Obtain  $L_2$  basis function set according to Equation (4)
4. Obtain nonlinear  $L_2$  operator according to Equation (6)
5. Include nonlinear  $L_2$  operator in original PSD relation
6. End

**Algorithm 2.** Optimization of DenseNet architecture using Probabilistic gradient descent

1. Start
2. Initialize network weights ( $w_i$ ) and biases ( $b_i$ )
3. Specify layers (number of neurons) in each level of the DenseNet (DN1 – DN5)
4. Train network using probabilistic gradient descent
5. Compute feature samples for training, validation and testing
6. Define cost function ( $c$ ) and updates
7. Define training functions for feature samples and determination of accuracy
8. Train network
9. End

#### 4. EXPERIMENTAL SETUP

The implementation and training of the proposed DenseNet with improved PSD was carried out using Matlab with a selected sample size of 120 images from the TCIA database. The dataset comprises images of non-small cell lung cancer from 211 test subjects. It consists of computed tomography (CT) and positron emission tomography (PET) images. Eighty images were from cancer patients, while 40 images were CT scans from patients without lung cancer. For the cancer images, 60 were used for training, while 20 were used for testing. For the non-cancer images, 25 were used for training, while 15 were used for testing in order to validate the network. The decision to split the dataset this way is so that approximately 70% of the whole dataset will be used for training while 30% for testing and validation of the model.

The proposed PSD approach is to improve the capability of the DenseNet to accurately classify salient features in the image that identify lung cancer. The improved capability of the  $L_2$  norm enables fast and accurate specification of the sparse matrix, which is used to train the DenseNet. To initialize the training of the DenseNet, we begin with a learning rate of 0.05 and regularization parameter  $\sigma=0.1$ . One approach that was taken to improve classification accuracy further was to artificially expand the training data by displacing the training image by one pixel (up or down, left or right). Rectified linear units (ReLU) have been used to activate neurons in various layers of the DenseNet instead of the sigmoid activation function due to their generally recognized improved performance [4, 5].

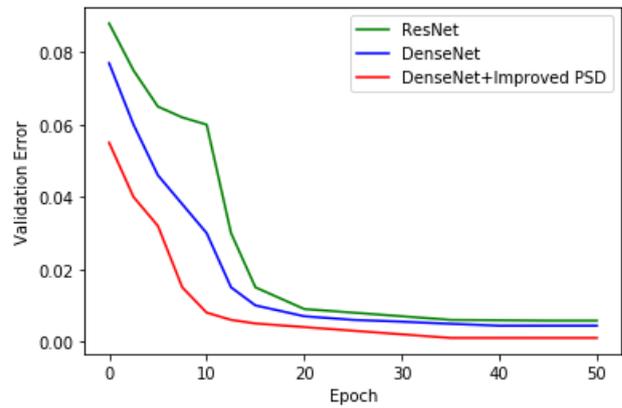
### 5. RESULTS AND DISCUSSION

The performance of the DenseNet trained using sparse images obtained from the proposed PSD method is compared to that of seven other well-known neural network architectures and machine learning algorithms, including a deeply supervised neural network proposed in [35], deep residual learning neural network (ResNet) [36], a conventional CNN, and DenseNet without proposed PSD. Performance metrics are classification accuracy, AUC, and mean of the network cost/activation gradient. The results are summarized in Table I. From the results obtained, it can be seen that the classification accuracy of the proposed method is better than the other seven methods.

**Table 1. Performance comparison of the proposed method with conventional CNN, ResNet, DCNN, KNN, SVM, BPNN, and DenseNet over 100 training cycles**

NN Architecture	AUC	Classification accuracy (%)	Mean ( $\pm$ std dev) $\frac{\partial C}{\partial \beta_n}$	Comp. time (ms)
Conventional CNN	0.70	88	66.52( $\pm$ 0.21)	208
ResNet	0.84	92	74.12( $\pm$ 0.07)	176
Deeply supervised NN	0.79	83	69.19( $\pm$ 0.37)	150
KNN	0.80	90	73.72( $\pm$ 0.11)	118
SVM	0.79	92	75.16( $\pm$ 0.18)	160
BPNN	0.72	87	70.93( $\pm$ 0.26)	190
DenseNet	0.89	91	81.52( $\pm$ 0.10)	113
<b>DenseNet+Improved PSD</b>	<b>0.93</b>	<b>95</b>	<b>89.73(<math>\pm</math>0.02)</b>	<b>106</b>

Figure 2 shows a plot of validation error comparing the proposed enhanced DenseNet structure with conventional DenseNet and ResNet; 50 training cycles (epochs) are considered. With each learning step of the neural network, the proposed DenseNet performs better than the other two structures, thus confirming the results obtained in Table 1. From the plot, we can see that the proposed method has the lowest validation error. This means that the DenseNet is more efficient and accurate concerning giving the correct classification of the image. For the computational time, the proposed DenseNet also gives classification results faster than the other seven methods.



**Figure 2 – Plot of validation error for ResNet, DenseNet, and proposed method over 50 training cycles**

The effect of adding an extra layer of neurons on the performance of the DenseNet in terms of classification accuracy and mean of the cost/activation gradient was also examined. In particular, 5, 10, and 30 neurons were added to the hidden DN2, DN3, and DN4 layers of the DenseNet. The purpose of this comparison is to establish the performance of the DenseNet with  $L_2$  regularization as well as dropout. As a strategy to improve accuracy, we also considered the effect of dropout by temporarily removing half of the hidden neurons in DN2-DN4 of the DenseNet. The resulting network weights and biases were then updated, and then the dropout neurons were restored. This process is also repeated on a different set of hidden neurons within the DenseNet. The results for hidden layer expansion and dropout are summarized in Table 2 and Table 3, respectively.

**Table 2. Effect of adding a layer of 5, 10, and 30 hidden neurons to DN2, DN3, and DN4 of the proposed DenseNet**

DN2	DN3	DN4	Classification accuracy (%)	Mean ( $\pm$ std dev) $\frac{\partial C}{\partial \beta_n}$
5	10	30	92	88.68( $\pm$ 0.13)
5	30	10	96	89.22( $\pm$ 0.16)
10	5	30	90	78.51( $\pm$ 0.06)
10	30	5	93	79.25( $\pm$ 0.06)
30	5	10	78	69.85( $\pm$ 0.20)
30	10	5	81	67.35( $\pm$ 0.15)

**Table 3. Effect of dropout on hidden layers DN2, DN3, and DN4 of the proposed DenseNet**

Dropout	Classification accuracy (%)	Mean ( $\pm$ std dev) $\frac{\partial C}{\partial \beta_n}$
DN2	92	55.83( $\pm$ 0.13)
DN3	93	79.89( $\pm$ 0.10)
DN4	93	81.16( $\pm$ 0.08)

From Table 2 and Table 3, it can be seen that compared to the classification accuracy obtained using  $L_2$  regularization, the techniques of dropout and network expansion of hidden layers do not achieve superior performance. In particular, we observe that in both cases of dropout and network expansion, the performance of the proposed DenseNet degrades significantly in terms of both classification accuracy and cost/activation gradient when modifications are made to DN2. This is likely because this layer is the closest to the input layer. As a result, the effect of the vanishing gradient problem commonly associated with deep neural networks [34]. Therefore, overall performance improves when modifications (in terms of both dropout and network expansion) are done farther away from the input layer. The cost/activation gradient  $\frac{\partial c}{\partial \beta_n}$  is considered here because it provides evidence of degrading network performance as information flows from input to output. The higher the mean value, the less significant the degradation would be from input to output.

## 6. CONCLUSIONS

From the results obtained, the DenseNet trained using sparse images obtained from unsupervised improved PSD technique has outperformed seven other well-known methods. The improved  $L_2$  sparsity approach ensures that features are selected more efficiently to constitute the reduced feature map. Also, the optimized DenseNet performance using the  $L_2$  regularization approach has enabled a more efficient selection of network weights and biases in successive layers of the DenseNet.

The classification accuracy of the proposed DenseNet is up to 95% compared to the ResNet of 92% and conventional DenseNet of 91%. The reduced parameter set of the DenseNet also ensures that it classifies images faster than the other structures. The effects of both dropout and hidden layer expansion on the performance of the DenseNet have also been investigated. While there is no specific rule-of-thumb regarding how many neurons to add (or remove) and at what stage, it has been observed that making significant changes to the layer close to the input layer of the DenseNet can significantly degrade its performance. In particular, using the classification accuracy and cost/activation gradient indices, we have observed a classification accuracy decline of 3% and 20% respectively for dropout and hidden layer expansion.

This paper has presented a novel approach to implementing the PSD for feature selection in digital images. A nonlinear operator was also proposed for selecting the basis function set for the  $L_2$  linear operator. The  $L_2$  regularization approach for

DenseNet as optimization also yielded better performance compared to both dropout and hidden layer expansion approaches. From the results obtained, it can be seen that the combination of the improved PSD approach for sparse images with the enhanced capability of the DenseNet gave better performance compared to other methods. To further improve classification accuracy, future research will focus on using a larger size of training data to mitigate the vanishing gradient problem in deep neural networks.

## ACKNOWLEDGMENTS

This research is supported partially by South African National Research Foundation Grants (No. 112108 and 112142), and South African National Research Foundation Incentive Grant (No. 95687), Eskom Tertiary Education Support Programme (Y. Sun, Z. Wang), Research grant from URC of University of Johannesburg

## 7. REFERENCES

- [1] N. Emaminejad *et al.*, "Fusion of quantitative image and genomic biomarkers to improve prognosis assessment of early stage lung cancer patients," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 5, pp. 1034–1043, May 2016.
- [2] A. P. Tafti, F. S. Bashiri, E. LaRose, and P. Peissig, "Diagnostic classification of lung CT images using deep 3D multi-scale convolutional neural network," *Proceedings of the 2018 IEEE International Conference on Healthcare Informatics (ICHI)*, 2018, pp. 412–414.
- [3] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," *Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 2017, pp. 721–724.
- [4] J. Ran, Y. Chen, and S. Li, "Three-dimensional convolutional neural network based traffic classification for wireless communications," *Proceedings of the 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2018, pp. 624–627.
- [5] G. Liang, H. Hong, W. Xie, and L. Zheng, "Combining convolutional neural network with recursive neural network for blood cell image classification," *IEEE Access*, vol. 6, pp. 36188–36197, 2018.
- [6] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Transactions on Semiconductor*

- Manufacturing*, vol. 31, no. 2, pp. 309–314, May 2018.
- [7] P. Zhang, X. Wang, W. Zhang, and J. Chen, “Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 1, pp. 31–42, Jan. 2019.
- [8] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon, “Using deep convolutional neural network architectures for object classification and detection within X-ray baggage security imagery,” *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2203–2215, Sep. 2018.
- [9] K. Kavukcuoglu, M. Ranzato and Y. LeCun, *Fast Inference in Sparse Coding Algorithms with Applications to Object Recognition*, New York University, New York, 2008.
- [10] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook*, Berlin: Springer, 2018.
- [11] A. Zhang, Z. C. Lipton, M. Li and A. J. Smola, *Dive into deep learning*, 2019. [Online]. Available at: <https://d2l.ai/>
- [12] M. Tanaka, T. Isokawa, N. Matsui, T. Yumoto, and N. Kamiura, “A convolutional autoencoder for detecting tumors in double contrast X-ray images,” *Proceedings of the 2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, 2018, pp. 384–387.
- [13] Z. Yu *et al.*, “A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 874–885, May 2018.
- [14] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, and H. Greenspan, “Chest pathology detection using deep learning with non-medical training,” *Proceedings of the 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 294–297.
- [15] D. Bardou, K. Zhang, and S. M. Ahmad, “Classification of breast cancer based on histology images using convolutional neural networks,” *IEEE Access*, vol. 6, pp. 24680–24693, 2018.
- [16] G. Yang *et al.*, “Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module,” *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3790–3795.
- [17] S. Wang, Y. Shen, D. Zeng, and Y. Hu, “Bone age assessment using convolutional neural networks,” *Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2018, pp. 175–178.
- [18] J. Rathod, V. Waghmode, A. Sodha and P. Bhavathankar, “Diagnosis of skin diseases using convolutional neural networks,” *Proceedings of the 2nd Int. Conf. on Electronics, Communication, and Aerospace Technology (ICECA 2018)*, 2018, pp. 1048–1051.
- [19] C. Chin, M. Chin, T. Tsai, and W. Chen, “Facial skin image classification system using convolutional neural networks deep learning algorithm,” *Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST)*, 2018, pp. 51–55.
- [20] H. S. Baweja and T. Parhar, “Leprosy lesion recognition using convolutional neural networks,” *Proceedings of the 2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2016, vol. 1, pp. 141–145.
- [21] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum, “Automatic segmentation of MR brain images with a convolutional neural network,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1252–1261, May 2016.
- [22] N. Tajbakhsh *et al.*, “Convolutional neural networks for medical image analysis: Full training or fine tuning?” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.
- [23] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez, “Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1273–1284, May 2016.
- [24] H. Salehinejad, E. Colak, T. Dowdell, J. Barfett, and S. Valaee, “Synthesizing chest X-ray pathology for training deep convolutional neural networks,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 5, pp. 1197–1206, May 2019.
- [25] Z. Liu, C. Cao, S. Ding, Z. Liu, T. Han, and S. Liu, “Towards clinical diagnosis: Automated stroke lesion segmentation on multi-spectral MR image using convolutional neural network,” *IEEE Access*, vol. 6, pp. 57006–57016, 2018.

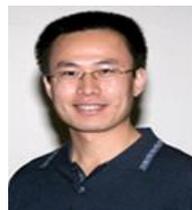
- [26] X. Xu, J. Lin, Y. Tao, and X. Wang, "An improved DenseNet method based on transfer learning for fundus medical images," *Proceedings of the 2018 7th International Conference on Digital Home (ICDH)*, 2018, pp. 137–140.
- [27] G. Yang, U. B. Gewali, E. Ientilucci, M. Gartley, and S. T. Monteiro, "Dual-channel DenseNet for hyperspectral image classification," *Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium IGARSS'2018*, 2018, pp. 2595–2598.
- [28] K. Zhang, Y. Guo, X. Wang, J. Yuan, and Q. Ding, "Multiple feature reweight DenseNet for image classification," *IEEE Access*, vol. 7, pp. 9872–9880, 2019.
- [29] S. Suzuki *et al.*, "Mass detection using deep convolutional neural network for mammographic computer-aided diagnosis," *Proceedings of the 2016 55th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, 2016, pp. 1382–1386.
- [30] Y. F. Riti, H. A. Nugroho, S. Wibirama, B. Windarta, and L. Choridah, "Feature extraction for lesion margin characteristic classification from CT Scan lungs image," *Proceedings of the 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2016, pp. 54–58.
- [31] R. Kumar, R. Srivastava, and S. Srivastava, "Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features," *Journal of Medical Engineering*, vol. 2015, 2015, pp. 1–14.
- [32] R. Zahedinasab and H. Mohseni, "Using deep convolutional neural networks with adaptive activation functions for medical CT brain image classification," *Proceedings of the 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)*, 2018, pp. 1–6.
- [33] J. Xu, Y. Chae, B. Stenger, and A. Datta, "Dense ByNet: Residual dense network for image super resolution," *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 71–75.
- [34] M. Nielsen, *Neural Networks and Deep Learning*. [Online]. Available at: <http://neuralnetworksanddeeplearning.com>.
- [35] C.-Y. Lee, S. Xie, P. W. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," *Journal of Machine Learning Research*, vol. 38, pp. 562–570, 2015.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9908, LNCS, pp. 630–645, 2016.
- [37] Z. Luo, L. Liu, J. Yin, Y. Li, and Z. Wu, "Latent ability model: A generative probabilistic learning framework for workforce analytics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 923–937, May 2019.
- [38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.



**IBOMOIYE D. MIENYE** received the B.Eng degree in Electrical and Electronic Engineering, in 2012, and M.Sc degree in Computer Systems Engineering, in 2014, both from the University of East London. He is currently pursuing the Ph.D. degree in Electrical and Electronic Engineering with the University of Johannesburg, South Africa. His research interests include machine learning, deep learning, bioinformatics, and object recognition.



**Yanxia Sun** received her joint qualification: DTech in Electrical Engineering, Tshwane University of Technology, South Africa and PhD in Computer Science, University Paris-EST, France in 2012. She has therefore an approach that brings together computing and electrical engineering. She has more than 10 years teaching and research experience. Currently she is serving as an Associate Professor University of Johannesburg, South Africa. She has published 82 journal or conference papers. She is/was the principal or co-principle investigator of 11 projects including national research grants and industrial projects.



**Zenghui Wang** received the B.E degree in Automation, Naval Aviation Engineering Academy, China, in 2002 and Ph.D. degree in Control Theory and Control Engineering, Nankai University, China, in 2007. Currently he is a Professor in the Department of Electrical and Mining Engineering, University of South Africa (UNISA), South Africa. He has published more than 120 articles including more than 50 ISI indexed journal papers. His research interests are model predictive control, nonlinear control, engineering optimization, image/video processing, artificial intelligence and chaos.