The CSI Journal on
Computer Science and Engineering
Vol. 19, No. 2, 2025
Pages 39-44
doi.org/10.22034/jcse.2025.182045
Regular Paper

Automatic diagnosis of retinal diseases using a new method in Deep Dictionary Learning on OCT images

A. Montazeri[1], M. Shamsi[2], R. Dianat[1]

[1]Department of Information Technology, Faculty of Computer, University of Qom, Zip Code: 3716146611, Iran.

[2]Department of Computer, Faculty of Computer and Electrical, Qom University of Technology, Zip Code: 3718146645, Iran.

Abstract

The aim of this research is to modify the classification efficiency of state-of-the-art techniques using a multi-layer framework based on deep dictionary learning. This paper presents the new method in deep dictionary learning for classification tasks and applies it to diagnose diseases of the retina. The framework uses a label consistent in the K- SVD algorithm to learn a discriminative dictionary for sparse coding to learn better features in retinal OCT images. Also, we used label information with each dictionary item (columns of the dictionary matrix) to enforce discrimination in sparse codes during the multi-later dictionary learning process. This algorithm learned multiple levels of dictionaries instead of learning shallow dictionary. The performance of the proposed algorithm is evaluated on two datasets Duke and THOCT to diagnose diseases AMD and DME. The proposed method results are comparable or better. The approach proposed is based on the best algorithms and is more precise than highly tuned models and strong dictionary learning models. The proposed model of deep dictionary learning presents the idea to develop more impressive dictionary learning methods and can help to move forward the state-of-the-art.
Keywords: Sparse Representation, Deep Dictionary Learning, Optical Coherence Tomography, Classification.

1. Introduction

Mathematical models for successful data description include several types of dictionary learning (DL) and sparse representation models and deep-learning models, especially in the processing of medical images. As DL focuses on learning the basics and features through
matrix factorization methods. DL has been very interesting in the field of representational learning. The idea of DL was used in image processing and information discovery by the researchers in the late 1990's [1].
Two words, "DL" and "factoring of matrixes," and their applications have not stopped in the recent past. The objective was to learn an empirical database. To decompose into a dictionary matrix and a function matrix, it must be a data matrix. DL is currently very popular because of K-SVD. K-SVD is an algorithm for disassembly into a compact basis and sparse coefficients of the training matrix. Although before K-SVD, the concept of sparse decomposition was introduced, much work was done with the advent of K-SVD in 2006 [2]. In both areas, DL can be well used and uncontrolled. DL also applies in matters such as image restoration, video processing, image denoising and painting [3].
Inverted problems may also be solved in compressed sensing (CS). However, learning through DL was shown to provide a more customized representation than the firm foundation of Compressed Sensing [4]. Mathematical changes such as DCT, wavelet, curvelet, and Gabor were often used for the problem of image classification in the field of medical applications [5]. Many attempts have been made in medical image reconstruction and especially the retinal image processing to provide automated systems for diagnosing

various diseases. Such systems, in addition to providing the possibility of processing retina images in a large volume with a minimum of time and cost, are beyond the fatigue and other weaknesses that a detective can suffer. The new technique of OCT imaging (Optical Coherence Tomography) from the retina and providing the possibility of obtaining information and images with high accuracy and high quality from different layers of the retina has made it possible, by working on these images, the error of the algorithms provided to detect and the automatic extraction of patterns associated with retinal diseases was reduced, and new methods and algorithms for automatic analysis of these images were introduced. OCT imaging is a visualization of different layers of the eye and head of the nerve, which produces separate imaging sections using optical waves. This technology can produce cross-sectional imaging of the microscopic structure of biological tissues. It also has a resolution micrometer [6].

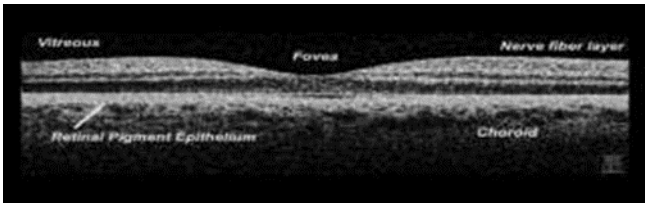

Fig 1: OCT scan in retinal tissue at 800 nm and axial resolution of 3 microns

Over the past two decades, most of the researchers involved in the processing of OCT images have been divided into two main sections, the first part of which is the segmentation of retina layers, which is not discussed in this article and section The second is the articles in this field, which we aim to address in a variety of ways the classification of OCT images, which are generally based
on image reconstruction. In the early second decade of the twentieth century, a method was proposed to identify macular area damage, including AMD and DME, using local binary patterns and using a multi-dimensional spatial pyramid and principal components analysis (PCA) for Dimension reduction, image reconstruct. Also, a model for the reconstruction of images based on the graph was proposed. In this model, the images were first divided into a four-tree set. Then, in order to analyze the four trees, the subgroup reduction technology was used and with the ability to recognize them, Common subgraphs were selected for image rebuilding. Finally, they were trained using attribute vectors, classifier, and a binary category on normal and ADM patient images. Subsequently, another model was presented for three- dimensional OCT images that initially divided the initial three-dimensional image into smaller-sized images, then reconstructed them with a tree structure, and finally, by combining extraction characteristics from the sub-graphs, the image was divided to sick and normal [7].
Deep Learning methods allows multi-layered computational models to learn multiple abstract data representation. Successful deep artificial neural networks differ a little from training to testing. Conventional learning either implies that the properties of the model family or the regularization techniques used in training are a minor generalization error.
A specific type of deep learning is the convolutional neural network (ConvNet) [8]. During the period when neural networks were not favored, it has achieved many practical achievements and recently the computer vision community has adopted it widely. Convolutional networks are strong visual models that produce feature hierarchies. For example, a color image consisting of three 2D arrays that contain pixel intensities in the three-color channels is intended for the processing of data that comes in the shape of multiple arrays [9].

When data for the training of the next layer derives from latent variables, one layer at the time is learned from Deief Belief Networks (DBN). This efficient, greedy training can either be followed by other learning methods that can adjust all weights to improve the network production or discrimination. A deep network of beliefs can also be viewed as composing simplistic learning modules, all of which are restricted-machine Boltzmann (RBM) type with a visible unit layer, representing the data and a hidden unit layer. The RBM is one of the fundamental building blocks of deep learning. RBM can find extensive applications in size reduction, feature extraction and advisory systems by modelling the probability distributions of a wide variety of input data including natural images, voice and customer ratings. It is nearly impossible to calculate the exact gradient of this loss function. However, the gradient is called a contrasting divergence gradient by stochastic

approximation. RBM is generally not supervised, but studies have been conducted in which discriminatory RBMs use class labels in training. The recent Deep Boltzmann Machines (DBM) has won significant publicity in the field of higher data representation and distribution as one of the increasing research discussions. DBM organizes a number of nonlinear latent variables into several layers, so that the single layer variables can contribute simultaneously to the probabilities or status of the following layers. The changes in one data are shown in each layer by a different factor. The DBM collected variations and structures, due to its nonlinear structure, with complex data higher than the second order, effectively organized in hidden layers [10].
In this paper using OCT images and both the feature learning popular paradigms, i.e., DL and deep learning, we propose a new method called multi-level deep dictionary learning (DDL) in order to an automatic diagnosis of retinal diseases. The proposed framework will help researchers apply more effective DL algorithms. In the proposed model, by defining the optimal dictionary, the reconstruction error and sparse coding error are minimized, and the classifying performance criterion is optimized for classifying retina OCT images into three categories: normal, Age- Related Macular Degeneration (AMD) and Diabetic Macular Edema (DME).
In the proposed multi-layered architecture, the OCT images is decomposed into a more detailed analysis, then it is encoded using the initial dictionary. The sparse coefficients obtained in each layer are used as inputs of the next layer, and the coding operations are repeated in several layers after finding the optimal dictionary. This repeatable process leads to a greater differentiation in the acquired features, and reducing reconstruction and sparse coding error in the output of the final layer. In the next phase, using the powerful softmax function and gradient descent algorithm, we complete the classification and diagnostic process. Studies show that the use of multi- layer K-SVD reduces

reconstruction and sparse coding error and, thus, obtaining an optimal dictionary during this process, leads to a significant reduction in classification error.

2. Deep Multi-Layer Label Consistent K-SVD (DMLLCK-SVD)

Our main objective is to repeat the sparse coding layer and to use an algorithm-consistent label to make the characteristics more discriminatory. The other purpose of the proposed method, which can be more efficiently processed, is to control the input patch dimensions by cutting the sparse computer intensity coding of the original image. It is like projecting a vector on a dictionary in a new space. There is no linear projection and the vector is scarce so that the vector is applied as a new entrance to the next layer. A new encoding step for the dictionary is alike to a new project for a coordination system. The process can be repeated with as many dictionaries as the number of layers.

2.1. Framework

The proposed multi-layer neural network can be characterized as follows:

1. Decompose an image in a set of patches in order to preserve its position.
2. Encoding each patch yk on a first dictionary D. The set of codes $\hat{xk}$ may be shown as 3d volume X with a profundity equivalent to the number of atoms in dictionary D.
3. 3d volume image as input of the following layer is obtained from the previous layer. For selected layers, (1) and (2) may be repeated.

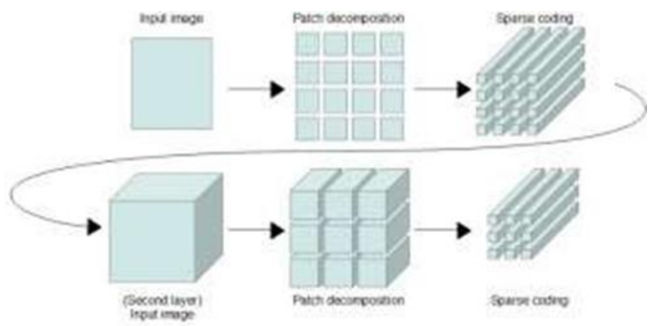

Fig 2. Schematic diagram for framework of DDL.

The optimum dictionaries at each layer need to be found in order to optimize the multi-layer classification architecture. This extension forces a constraint on label consistency in the dictionary–intuitively, class distributions, for which a dictionary element contributes, are very high in one class. In order to learn a discriminatory dictionary to make sparse coding possible, our multi-layer framework employs a label consistent with the K-SVD algorithm.

2.2. Formulation

Take a set of inputs $Y = \left\{y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)}\right\}$ associated with the label l into account. This set is generated by breaking an image into p patches. The upper index (1) is the 1-th layer and usually the k-th input signal patch in $y^{(q)}$ is a q-th layer patch. Let's take into account our algorithm with the Q layer, its formulation is:

$$\left(\widehat{x_k^{(q)}}, A, D\right) = argmin_{x^{(q)},A,D^{(q)}} ||y_k^{(q)} - D^{(q)}x^{(q)}||_2^2 + \alpha||Z^{(q)} - Ax^{(q)}||_2^2 \quad \text{s.t.} \quad \forall q \in [1, Q-1], ||x^{(q)}||_0 < T \tag{1}$$

$$min_{D^{(Q)},W} \sum_{k=1}^{n} L\left(l_k, \widehat{x_k^{(Q)}}, W\right) \tag{2}$$

Where $y_k^{(q)}$ is obtained from the output of the previous layer $\hat{x}_k^{(q-1)}$ by using transformation f that we chose max-pooling function, and α controls the relative contribution between reconstruction and label consistent regularization.

Also $Z^{(q)} = \left\{Z_1^{(q)}, Z_2^{(q)}, \dots, Z_n^{(q)}\right\} \in R^{m \times n}$ are the 'discriminative' sparse codes of input signals $y_k^{(q)}$ in q-th layer for classification. We sit that $z_i^{(q)} = \{0, \dots, 1, 1, \dots, 0\}^t \in R^m$ $\mathbb{R}^m$ is a 'discriminative' sprse code related to an input signal $y_k^{(q)}$, if the nonzero values happen at those indexes where the input signal $y_k^{(q)}$and the dictionary atom $d^{(q)}$ assign the same label. A linear transformation is defined, $g(X, A) = AX$, which transforms the main X codes into the sparse feature space $\mathbb{R}^m$ most discriminatory.

The term $\left\|Z^{(q)} - Ax^{(q)}\right\|_2^2$ represents the discriminatory sparse code error that enforces the approximation of the discriminatory Z codes in every layer by sparse codes X . It results in very sparse representations of the signals of the same class (that is, encouraging consistency of the label in the resulting sparse codes), resulting in good classification performance using a simple linear classifier.

We necessity to compute the gradients with regard to $\left\{W, D^{(Q)}\right\}$ parameters to optimize the cost of the classification function in relation to the final layer dictionary. In the layer last, we use the output $\hat{x}_k{}^{(Q)}$ as a classification feature and the feature cost over the all-training set of n images is minimized.

We also apply the cross-entropy function for classification loss because it has demonstrated fine results in subjects of multi-class classification. As an output classifier a linear classifier combined with softmax is chosen. If we regard a classification problem having C classes, the loss of cross-entropy is calculated as follows:

$$L\left(l_k, \hat{x}_k{}^{(Q)}, W\right) = -\sum_{i=1}^{C} l_{ik} \log\left(p_{ik}\right) \tag{3}$$

where p_{ik} is defined by:

$$p_{ik} = \frac{\exp\left(\hat{x}_k(Q)^T w_i\right)}{\sum_{j=1}^{C} \exp\left(\hat{x}_k(Q)^T w_j\right)} \tag{4}$$

2.3. Calculation of gradient

In this section to compute the gradient of L with regard to the last layer dictionary $D^{(Q)}$, we writing:

$$\nabla_{D^{(Q)}} L\left(l_k, \hat{x}_k{}^{(Q)}, W\right) = D^{(Q)}\beta\hat{x}_k{}^{(Q)T} + \left(y_k{}^{(Q)} - D^{(Q)}\hat{x}_k{}^{(Q)}\right)\beta^T \tag{5}$$

with β defined as:

for the indexes included in Λ set,

$$\beta_\Lambda = \left(D^{(Q)}{}_\Lambda^T D^{(Q)}{}_\Lambda\right)^{-1} \nabla_{\hat{x}_{k_\Lambda}{}^{(Q)}} L\left(l_k, \hat{x}_k{}^{(Q)}, W\right) \tag{6}$$

And $\beta_j = 0$, if
$j \notin \Lambda$
We have to calculate $\nabla_{\hat{x}_{k_\Lambda}(q)} L\left(l_k, \hat{x}_k{}^{(Q)}, W\right)$ to compute the cost function $L\left(l_k, \hat{x}_k{}^{(Q)}, W\right)$ gradient for the dictionary $D^{(q)}$ of the q-th layer. The gradient calculation can be decomposed as follows:

$$\frac{\partial L}{\partial \hat{x}^{(q)}} = \frac{\partial L}{\partial \hat{x}^{(q+1)}} \frac{\partial \hat{x}^{(q+1)}}{\partial y^{(q+1)}} \frac{\partial y^{(q+1)}}{\partial \hat{x}^{(q)}} \quad (7)$$

where:

$$\frac{\partial \hat{x}^{(q)}}{\partial y^{(q)}} = \left(D_\Lambda^{(q)T} D_\Lambda^{(q)}\right)^{-1} D_\Lambda^{(q)} \quad (8)$$

and 0 elsewhere.

The optimization procedure for this method is discussed in the following section. In the training step, D, A and X are initially calculated by Equation (1), followed by training W matrix for classification.

2.4. Optimization

Using the powerful K-SVD algorithm, the best solution for all parameters is found. It is possible to rewrite Equation (1) as:

$$\left(\widehat{x_k^{(q)}}, A, D\right) = argmin_{x^{(q)},A,D^{(q)}} \left|\left| \begin{pmatrix} \frac{y_k^{(q)}}{\sqrt{\alpha} Z^{(q)}} \end{pmatrix} - \begin{pmatrix} \frac{D^{(q)}}{\sqrt{\alpha} A} \end{pmatrix} x^{(q)} \right|\right|_2^2$$
s.t. $\forall q \in [1, Q-1], |x^{(q)}|_0 < T$ (9)

Let $y_{\text{new}} = \begin{pmatrix} y_k^{(q)} \\ \sqrt{\alpha} Z^{(q)} \end{pmatrix}, c = \begin{pmatrix} D^{(q)} \\ \sqrt{\alpha} A \end{pmatrix}$ and $X = x^{(q)}$, The optimization of Equation (1) is equivalent to solving the following problems:

$$(X, D) = \text{argmin}_{X,D} \|y_{\text{new}} - D_{\text{new}} X\|_2^2 \text{ s.t } \forall i. \|x_i\|_0 < \text{T} \quad (10)$$

This is precisely the problem solved by K-SVD. The k-th row in X, defined as x_R^k, is updated at a time following K-SVD, d_k and its related coefficients. Assume $E_k = \left(Y - \sum_{j \neq k} d_j x_R^j\right)$ and $\tilde{x}_R^k$ denote the result of x_R^k and E_k respectively throwing away the zero entries. By solving the following problem, d_k and $\tilde{x}_R^k$ can be calculated as follows:

$$(\tilde{x}_R^k, d_k) = \text{argmin}_{\tilde{x}_R^k, d_k} \left\|\tilde{E}_k - d_k \tilde{x}_R^k\right\|_F^2 \quad (11)$$

The SVD operation is performed for $\tilde{E}_k$, i.e., $U \sum V^t = \text{SVD}(\tilde{E}_k)$. Then d_k and $\tilde{x}_R^k$ are computed as:

$$d_k = U(:,1)\ \tilde{x}_R^k = \sum (1,1) V(:,1) \quad (12)$$

Finally, $\tilde{x}_R^k$ applyed to replace the non-zero values x_R^k.

We also need to initialize our algorithm with the parameters D_0 and A_0. For D_0, we use several K-SVD iterations in every class and then merge all K-SVD outputs. The d_k we use several K-SVD iterations in every class and then merge all K-SVD outputs. The d_k is updated during the course of the learning process. Word elements are assigned to each class uniformly with the number of items in proportion to the size of the dictionary. We use multivariable screen regression model [20] as follows for initializing A_0 in q-th layer:

$$A = \text{argmin}_A \left\|Z^{(q)} - A x^{(q)}\right\|_2^2 + \gamma \|A\|_2^2 \quad (13)$$

which provides the next solution:

$$A = \left(x^{(q)} \left(x^{(q)}\right)^t + \gamma I\right)^{-1} x^{(q)} \left(z^{(q)}\right)^t \quad (14)$$

For optimizing the classification cost function with regard to the last layer dictionary, according to approach of [11], we used the gradient descent algorithm and computed the gradients with regard to the parameters $\{W, D^{(Q)}\}$. Therefore parameters are updated as follows:

$$\text{W} \leftarrow \prod_{\text{W}} \left[\text{W} - \rho_t\left(\nabla_{\text{W}} \text{L}\left(\text{l}_k, \hat{\text{x}}_k{}^{(Q)}, \text{W}\right) + \text{vW}\right)\right] \quad (15)$$

$$D^{(Q)} \leftarrow \prod_{D(Q)} \left[D^{(Q)} - \rho_t\left(D^{(Q)} \beta \hat{x}_k{}^{(Q)^T} + \left(y_k{}^{(Q)} - D^{(Q)} \hat{x}_k{}^{(Q)}\right) \beta^T\right)\right] \quad (16)$$

where $\prod_w$ and $\prod_{D(Q)}$ are respectively orthogonal projections on the sets W and $D^{(Q)}$, also $v \in \mathbb{R}$ and ρ_t are respectively regularization and learning rate parameters and β is defined in (10). Resulted dictionary of (5) is used as initial dictionary in the algorithm. The Proposed structure with Q-layers is shown in Figure 3.

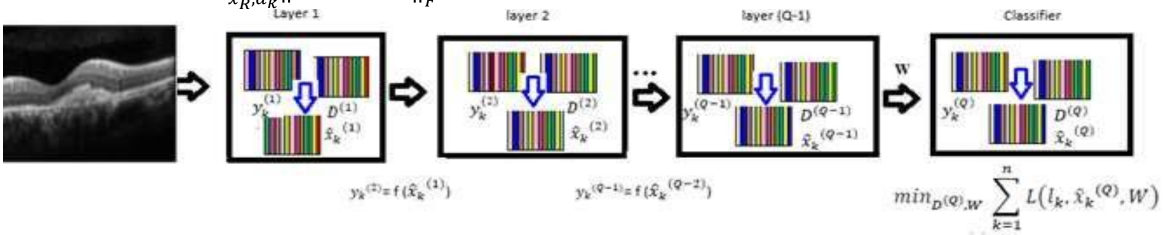

Fig 3. Steps of a deep dictionary with Q layers.

3. Experiments and Results

The performance of the proposed model is evaluated on two datasets, Duke and THOCT. Duke dataset is made available by Srinivasan et al in-Duke University, Harvard University, and the University of Michigan. This dataset consists of total 45 OCT volumes with labels AMD (15 volumes), DME (15 volumes), and normal (15 volumes). The OCT scans have tuned from 36 to 97 in every volume. We got the full dataset from [12]. Also, we applied THOCT dataset for experiments. THOCT dataset consists of 3000 retinal SD-OCT B-scans (1000 AMD, 1000 DME, and 1000 NOR), which is labeled by professional doctors [13].

We conducted experiments on Duke and THOCT datasets with different training sets and iteration times. We have replicated the experimental process 10 times with various randomly chosen images in the dataset for training and the remain for testing to achieve reliable results. The correct classification rates have always been recorded for normal, AMD, and DME subjects. Our ultimate results were reported by the mean and standard deviation of the AMD, DME, and ordinary classification rates, each. Here, we chose approximately half of the AMD and DME as well as normal training pictures and the rest as tests. We then selected AMD, DME and the normal two-thirds for training and the remainder for testing.

In this experiment, the dictionaries are trained for 5, 10, and 15 repetition times, where the number of SIFT descriptors is tuned to be 60,000. The experimental results are presented in Table 1 for Duke dataset and Table 2 for THOCT dataset.

The experimental results show that the correct classification rate approximately grows with the repetition times; for the same repetition times, the larger the training size is, the greater the correct classification rate.

Table 1: classification rate (%) compration of different proportion images for training on the duke dataset.

Table 1: classification rate (%) compration of different proportion images for training on the duke dataset.

	1/2 dataset training			**2/3 dataset training**		
Class/Iteration	**5**	**10**	**15**	**5**	**10**	**15**
Normal	96.45%	97.30%	100%	97.64%	99.50%	100%
ADM	94.20%	95.46%	97.02%	96.89%	97.61%	98.32%
DME	97.77%	100%	97.87%	97.90%	100%	97.90%

Table 2: classification rate (%) compration of different proportion images for training on the THOCT dataset.

	1/2 dataset training			**2/3 dataset training**		
Class/Iteration	**5**	**10**	**15**	**5**	**10**	**15**
Normal	94.75%	96.38%	99.89%	97.61%	99.50%	100%
ADM	93.20%	94.16%	96.02%	96.77%	97.92%	98.72%
DME	98.77%	100%	99.67%	99.79%	99.82%	100%

The performance of the proposed algorithms with popular D-KSVD [14], LC-KSVD [15], and ML-SVM [16] are compared. The proposed method also compares with the famous deep learning methods like the DBNs, which is fine tuned to a CNN (soft-max classifier). The output of DBN is [17]; the outcome of CNN is [18]. Table 3 shows the results.

Table 3. Comparing the classification accuracy of our method with available algorithms and architectures.

Method	DMLLCKSVD	LC-KSVD	D-KSVD	ML-SVM	DBNsoftmax	CNN
Accuracy	99.76	94.32	93.40	95.55	94.15	99.60

According to Table 3, it can be noted that almost always presented DDL method produces superior results than shallow DL methods such as LC-KSVD and DKSVD and multi-layer DL ML-SVM. The proposed algorithm, on the other hand, sits at the top of the existed best algorithms and provides a higher precision than very tuned models like DBN and CNN.

All of these algorithms are executed on a 3 GHz Intel (R) Core(TM) i5 machine, 8 GB RAM, Windows 10 (64 bit) on Matlab 2018a. The classification training and testing time will be almost identical for all methods until the dimensionality of the features remains identical, so that we do not include it here. As shown in Table 4, the algorithm proposed for training is almost twice as fast as a deep belief network, while testing speed is somewhat slower.

Table 4. Training/ Testing feature time (in seconds).

Table 4. Training/ Testing feature time (in seconds).

Time Dataset	Training feature DDL/DBN		Testing feature DDL/DBN	
Duke	63	127	44	21
THOCT	97	205	72	38

4. Conclusion

In this research, we proposed the idea of deep dictionary learning and applied it to diagnose some retinal diseases. We have learned several levels of dictionaries instead of learning a shallow dictionary in this algorithm. The output of each stage is used as the input to the next level and dictionaries were learned greedily one layer at a time. The 3d volume from the first level was used as the input to the next. The representation/features of one level were used to learn the next level. Thus, a simple, shallow dictionary learning algorithm, which is well known and solved, maybe the basic unit of deep dictionary learning.

We evaluated the performance of the proposed algorithm on the Duke OCT Images dataset and also THOCT dataset and the detection of two key eye diseases named the Age-Related Macular Degeneration (AMD) and Diabetic Macular Edema (DME). Many image classification techniques are based on the retinal layer segmentation process, while the method proposed does not rely on the retinal layer segmentation in our datasets. In the case of retinal diseases, this is a significant feature that alters the layer of the retina and complicates the layer segmentation process.

Also, AMD and DME diseases are detected using a multiclass linear classifier based on K-SVD in preprocessing OCT images. A system is proposed here, where retinal diseases from OCT images can be automatically detected. Then all the features extracted are fused into a vector that passes to the K-SVD classification system for automatic diagnosis of retinal pathology. The system proposed is quite robust when detecting models of small diseases in OCT images.

Furthermore, the proposed algorithm is compared with existing methods such as the deep belief net and CNN. The proposed method results on our datasets are comparable or better. The approach proposed is based on the best algorithms and is more precise than highly tuned models such as DBN and CNN and strong dictionary learning models such as LC-KSVD, D-KSVD, and MLSVM. Similar to the progress made in deep learning, the proposed formulation of deep dictionary learning forms the basis for developing more efficient dictionary algorithms and can help in promoting state-of-the-art technology.

References

[1] J. Ren, Z. Zhang, S. Li, Y. Wang, and G. Liu, "Learning hybrid representation by robust dictionary learning in factorized compressed space," *IEEE Trans. Image Process.*, vol. 19, 2020.

[2] S. Mukherjee, R. Basu, and C. S. Seelamantula, "ℓ1-K-SVD: A robust dictionary learning algorithm with simultaneous update," *Signal Process.*, Elsevier, 2016.

[3] J. Mitro, D. Bridge, and S. Prestwich, "Denoising dictionary learning against adversarial perturbations," *arXiv preprint arXiv:1801.02257*, 2018.

[4] M. Pan, F. Guan, H. Hu, and Y. Yang, "Compressed sensing based on dictionary learning for reconstructing blade tip timing signals," in *Proc. Prognostics and System Health Management Conf.*, 2017.

[5] M. Saha, M. K. Naskar, and B. N. Chatterji, "Wavelet and curvelet transforms for biomedical image processing," *Biomed. Signal Process.*, 2018.

[6] K. Alsaih *et al.*, "Machine learning techniques for diabetic macular edema (DME) classification on SD-OCT images," *Biomed. Eng.*, 2017.

[7] Y. Huang, V. De Bortoli, F. Zhou, and J. Gilles, "Review of wavelet-based unsupervised texture segmentation: Advantage of adaptive wavelets," *IET Image Process.*, 2018.

[8] H. Agh Atabay, "Deep learning for horse breed recognition," *CSI J. Comput. Sci. Eng. (JCSE)*, vol. 15, no. 1, Summer 2017.

[9] R. Entezari, M. M. Arzani, M. Fathy, and H. Bayat, "A deep learning method to estimate 3D point of regard by joint head and eye information," *CSI J. Comput. Sci. Eng. (JCSE)*, vol. 13, no. 2, Winter 2016.

[10] A. Khatami, A. Khosravi, T. Nguyen, and C. P. Lim, "Medical image analysis using wavelet transform and deep belief networks," *Expert Syst. Appl.*, vol. 86, pp. 190–198, Nov. 2017.

[11] Z. Sun and Y. Sun, "Automatic detection of retinal regions using fully convolutional networks for diagnosis of abnormal maculae in optical coherence tomography images," *J. Biomed. Opt.*, 2019.

[12] P. P. Srinivasan *et al.*, "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images," *Biomed. Opt. Express*, vol. 5, no. 10, pp. 3568–3577, 2014.

[13] Y. Sun, S. Li, and Z. Sun, "Fully automated macular pathology detection in retina optical coherence tomography images using sparse coding and dictionary learning," *J. Biomed. Opt.*, vol. 22, no. 1, 2017, Art. no. 016012.

[14] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010.

[15] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013.

[16] H. Asil and J. Bagherzadeh, "Proposing a new method of image classification based on the AdaBoost deep belief network hybrid method," *Telkomnika*, 2019.