Face Age Estimation Approach based on Deep Learning and Principle Component Analysis

Noor Mualla
Faculty of Computers and Informatics,
Benha University, Egypt

Essam H. Houssein
Faculty of Computers and Information,
Minia University, Egypt

Hala H. Zayed
Faculty of Computers and Informatics,
Benha University, Egypt

Abstract—**This paper presents an approach for age estimation based on faces through classifying facial images into predefined age-groups. However, a task such as the one at hand faces several difficulties because of the different aspects of every single person. Factors like exposure, weather, gender and lifestyle all come into play. While some trends are similar for faces from a similar age group, it is problematic to distinguish the aging aspects for every age group. This paper's concentration is in four chosen age groups where the estimation takes place. We employed a fast and effective machine learning method: deep learning so that it could solve the age categorization issue. Principal component analysis (PCA) was used for extracting features and reducing face image. Age estimation was applied to three different aging datasets from Morph and experimental results are reported to validate its efficiency and robustness. Eventually, it is evident from the results that the current approach has achieved high classification results compared with support vector machine (SVM) and k-nearest neighbors (K-NN).**

Keywords—***Deep learning; principal component analysis; support vector machine; K-NN; age estimation***

I. Introduction

Many applications rely on techniques for age. An example of that is found in vending machines that provide alcoholic drinks or cigarettes as they run the age prediction program to stop underage customers from illegally purchasing the products. Those processes are done through a facial recognition through the computer vision system. It worth noting that the human age estimation is a difficult task due to many a factor; primarily, identifying or labeling people within individual age classes is one hectic mission. While there have been several ways to overcome the problems facing age-estimation through facial recognition, there have been a couple of steps for the past decade that stood out the most. The algorithms designed for the task we accompanied with two components, namely, feature extraction [1]-[3] and age determination [4], [5].

In those systems, the age is recognizable by some unique face features [6]. However, it should be pointed out that age progression seen on faces cannot be accurately estimated due to the different and diverse information that can be seen in human faces. Information obtained from the human face cannot be limited for there are numerous dimensions that can be seen in the race and age. The human eye can identify known faces and the mind can estimate the age range of a human being; however, it is never always accurate despite it being "human". When it comes to the computer, there is a lot that can be read from a person's face regarding their age. Identity, gender, emotion, and age can all be deciphered in the daily communications through developing an automated method for the age estimation of an individual from the face image.

The reason why this work started was that of the need for a fast and efficient system that could provide a successful facial recognition. This work is motivated by the capacity that a system such as that could provide if it was developed successfully which could, in turn, enable numerous applications. If the outcome of any environmental setting changes and it is known, then the research work will be able to improve faster as decisions could be made up faster regardless of whether it is needed to proceed in certain directions or not. The aim of this paper is to present an automatic classification approach for age estimation. Therefore, this paper pays to focus on the classification of age estimation. The problems are analyzed on a set of records taken from Morph database. In fact, the quality of automatic identification systems performance is determined by signal representation algorithm and feature extraction schemes. In this work, a classification approach was proposed that involve PCA for extracting features as well as deep learning for the classification purpose. It is crucial for any developing approach to be tested and compared to other works in order to evaluate where it stands and whether or not it was successful. This work finishes up with a comparison to previous methods that have been used ensure the quality measures and efficacy of the proposed work such as SVM and K-NN.

This research follows the following organization: in Section II, the introduction of the work is presented; Section III follows up on the techniques and methods used within the methodology; Section IV brings forth a discussion on the deep learning approach presented; Section V exposes the results of the work from the experiments developed; and finally, Section VI contains the conclusions and the recommendations of future works proposed by the author.

II. Literature Review

In order to comprehend the steps needed to move forward with the work and bring forth advancements in the field, the work had to lay basis on previous works that relate to the topic. The works presented here vary from experiments to methodologies and even surveys developed on the matter of facial recognition and related topics, and deep learning along with topics similar to them. Several recent studies on age estimation have been presented, in the following, we will introduce some of them such as, in this work [7], a survey was done on age estimation. Another work [8] was inspired by another problem related to classification. The problem that the work focused on was make-up and plastic surgeries that divided the programs in finding suitable age groups. One more that was studied used the Active Appearance Model (AAM) [9]. A combination of shapes and models was presented in [10]. That approach was evaluated in another work [11] through employing three qualifiers.

Further, Fu et al. [12] have proposed an age estimation method developed to analyze pictures through the employment of the state-of-the-art manifold learning methods that can aid in discovering sufficient low dimension embedding area. The results of the simulation have shown that the linear manifold learning could provide sufficient aging feature extraction for regression ways of age estimation. Also, a similar approach was used by Guo *et al.* [13] that followed an age manifold learning scheme for extracting the aging features. The method also planned a locally adjusted robust repressor for the learning process and to expect the human age.

Eventually, there were various ways that attempted to reach novel approaches for facial recognition with the emphasis of the facial expression characterization and face modeling [14]. The geometric invariants that are important aspects of strain transformation have been observed through previous studies that provided information necessary for the work at hand [15]. Another work dealt with the uncertainty through the formation of nearly definite programming problem [16] or an EM-based algorithm [17]. Other works relied on the Local Binary Pattern (LBP) [18] features to identify series of local features that were also observed by the combination of a strong classifier that enabled a successful method of age classification [19].

III. Preliminaries

A. PCA for Feature Extraction

One of the most popular methods in the field of extracting features is Principle Component Analysis (PCA). The method linearly transforms the high-dimensional input vectors into low-dimensional ones in which the components would be uncorrelated. It computes the eigenvectors of the covariance matrix of the original inputs for processing. The process has been used in developing various algorithms in the past [20]. It is also considered as one of the best means in this area of work along with a reduction in dimensionality.

Through PCA, this work aims to showcase the d-dimensional data in a low-dimensional space which would decrease the degrees of freedom and space and time complexions. The aim of using the method is to present data in a space which expresses the variations in a sum-squared error sense as well as possible. PCA is also beneficial as it divides signals and images alike from several sources. Knowing the number of independent components exist beforehand helps in getting the best out of the method, just like standard clustering methods.

The basic approach in principal components is theoretically rather simple. First, the *d*-dimensional mean vector λ and $m \times m$ covariance matrix *R* is calculated for the full data set. Then, those eigenvalues and eigenvectors are calculated and organized based on the decreasing eigenvalue. Given a group of centered vectors of input $x_t\,(t=1,...,l \text{ and } \sum_{t=1}^{l} x_t = 0)$, each of which is of m dimension $x_t = (x_t(1), x_t(2), ..., x_t(m))^T$ (usually $m < l$), PCA linearly transforms each vector x_t into a new one s_t by:

$$s_t = U^T x_t \qquad (1)$$

Where U is the $m \times m$ orthogonal matrix whose *i*th column u_i is the *i*th eigenvector of the sample covariance matrix $C = \frac{1}{l}\sum_{t=1}^{l} x_t x_t^T$. Simply put, PCA first solves the eigenvalue problem defined in (2).

$$\lambda_i u_i = C u_i, \quad i = 1, ..., m \qquad (2)$$

Where, an eigenvalue of C and the matching eigenvector is u_i. Based on the predicted u_i, the components of s_t are then computed as the orthogonal alterations of x_t:

$$s_t(i) = u_i^T x_t, \quad i = 1, ..., m \qquad (3)$$

The new components are called the principal or main components. Through employing solely the first several eigenvectors sorted in descendant order of the eigenvalues, the number of principal components in s_t can be reduced, meaning that PCA has the dimensional decrease feature. The main components of PCA also have the following properties [21].

- $s_t(i), \mathrm{i}=1,...,\mathrm{m}$ are uncorrelated.
- $s_t(i), \mathrm{i}=1,...,\mathrm{m}$ have sequentially maximum variances.
- The mean-squared approximation error in the demonstration of the original inputs by the first several principal components is minimal.

B. Support Vector Machines (SVM)

The classification technique helps to discriminate the unknown testing set of observations into their appropriate classes based on the training group of known annotations. A classification technique used a mathematical function named as a classifier to predict the right class of unknown observation of testing data set. SVM was a method that Vapnik and Cortes introduced [22]. SVM is a powerful classifier in biomedical science, image processing and data mining for the detection and classification purposes. SVM is an efficient classifier to classify two different sets of observations into their relevant class. It has the means to handle high-dimensional and non-linear data excellently. Base on the foundation of training data sets, it helps to guess the important characteristics of unknown testing data. SVM mechanisms are based on finding the best hyperplane that divides the data of two different classes of the category. Accordingly, the best hyperplane is figured by being the one that maximizes the margin.

The design of SVM is based on the regularization parameter, C, which is used to control the relationship between margin maximization and some misclassifications; and kernel functions of nonlinear SVMs which are used for the mapping of training data from an input space to a higher dimensional feature space. It should be noted that all the kernel roles such as linear, polynomial, radial basis function and sigmoid having some free parameters are called hyperparameters. Until the present day, the well-known kernel commonly used research was the Gaussian or radial basis function (RBF) kernel with width σ [23].

$$k(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2}) \qquad (4)$$

Where, K(x, y) is termed as the kernel function, which is built upon the dot product of two invariant x and y. suitable trade-off parameter C and the kernel parameter σ are essential to training SVM classifier and usually found by the K-fold cross-validation technique.

C. K-Nearest Neighbors Algorithm (k-NN)

Classification in K-nearest neighbor is based on a majority vote of the nearest k neighbors in the training dataset [24]. Euclidean distances between the unclassified record and the classified records are calculated and sorted. One single observation from the original dataset is selected and used accordingly as a dataset for validation whereas the remainder of observations is selected as training data. That means that each observation is only used one time as the validation data. Leave-one-out cross-validation (LOOCV) of one nearest neighbor (1-NN) is what the pre-mentioned method is known as [24].

IV. Proposed Deep Learning Approach for Classification

A wide array of missions has been accomplished through the use of deep learning using neural networks. The interest herein is in Deep Learning Architectures (DLA) that has the ability of effectively performing FE for hyperspectral data. While DLAs are newly founded developments in neural networks, they have already obtained high classification accuracies in many fields and remain to be at the peak of performance levels, over all of the other methods [25], [26]. DLAs ability is perceived to be closer to AI than other state-of-the-art learning machines [27].

SVMs were shown in the work and in literature to be outperformed through Deep Belief Network and the stacked auto-encoder; both of which are types of DLAs. They outperformed SVMs in classifying the hyperspectral data [28], [29]. The experiments were carried out to predict the effect that the depth of the networks and the principal component numbers of the data on the accuracy of the classification and the time needed to run the experiments. This work was prepared to understand the effects of a DBN on its performance through monitoring the accuracy level and impact by the width of the network and on the time required for the training purpose. Accordingly, it was found in the study that the proposed DBN structure allows for higher accuracy in classification on hyperspectral data that is remotely-sensed via an SVM and K-NN as a reference to aid in the comparison.

The process of deep learning includes a class of models that attempt to hierarchically learn deep features of the input data with very deep neural networks which are often perceived as more than three layers deep. Through unsupervised training, the network is built in the first layer. In the scheme considered for the work, high-level features can be learned through low-level features while the proper ones can be shaped for pattern classification at the end. Accordingly, deep models have the potential to contain more abstracts and complex features in higher layers. Abstract features are known to be invariant to most of the local changes of the input. Some papers [30] argue that deep models can achieve a more accurate approximation to nonlinear functions that shallower models. Orthodox deep neural network architectures contain deep belief networks (DBNs) [31], deep Boltzmann machines (DBMs) [32], restricted Boltzmann machines (RBMs) [33], pooling units, convolutional neural networks (CNNs) [34], etc. This work implements one of the deep learning models mentioned above: namely DBN, for age classification as depicted in Fig. 1.

A. Deep Belief Networks

Deep Learning (DL) can be stacked through feeding a DL the hidden victor of another as input. Layer, in this case, refers to one of the ideal parameters between two vectors of visible or two vectors of hidden units. In the layers beyond the first one, the conditional probabilities are given the following equation:

$$P(h_j^{l+1} \mid h^l) = sigmoid(c_i^{l+1} + \sum_{j \in J} W_{ij}^{l+1} h_j^l) \qquad (5)$$

Where h^l denotes the input vector of the RBM at the *l*-th layer of the network and h^0 signifies the input vector of the DBN [31].

When the network is built this way and trained, layer-wise, in an unsupervised style by means of (5), it is known as a Deep Belief Network (DBN). Though a DBN can be used as a generative model because of the RBM’s ability to rebuild data, the work’s interest here is in the discriminative case. The idea here focuses on classifying data rather than to reconstruct it. To use a DBN discriminatively, the output of the top-layer DL is fed as input to a classifier such as a logistic regression, at which point, the model is then trained in a supervised fashion [35], as shown in Fig. 2.

The pre-training stage is known as the unsupervised training where it initializes the model in order to improve its efficiency during the supervised training stages. Said stages are known as fine-training; where the classifier’s prediction is adjusted so that it would match the ground truth of the data. Iterations in either training stages, fine or pre, are known as epoch [36]. Accordingly, DBNs must learn to produce “good” representations.

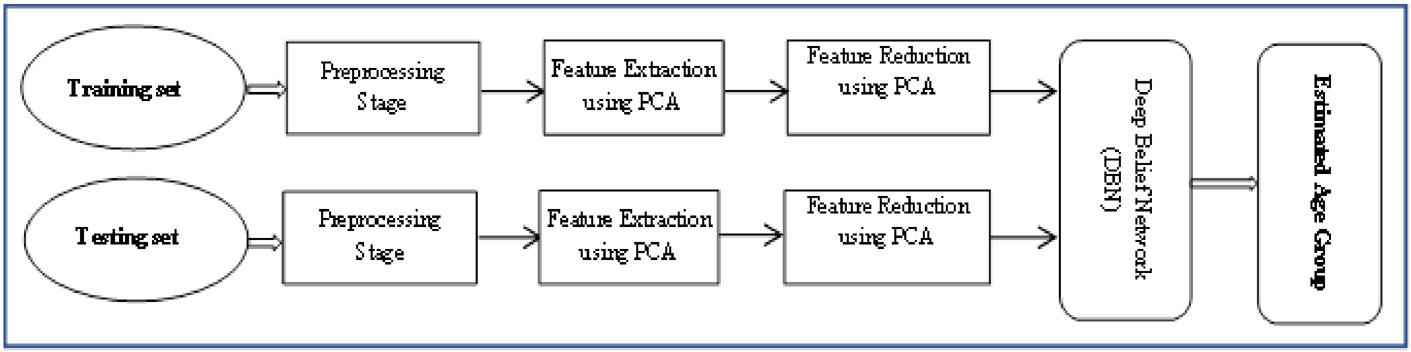
Training set
Preprocessing Stage
Feature Extraction using PCA
Feature Reduction using PCA
Testing set
Preprocessing Stage
Feature Extraction using PCA
Feature Reduction using PCA
Deep Belief Network (DBN)
Estimated Age Group

Fig. 1. Diagram of the proposed approach.

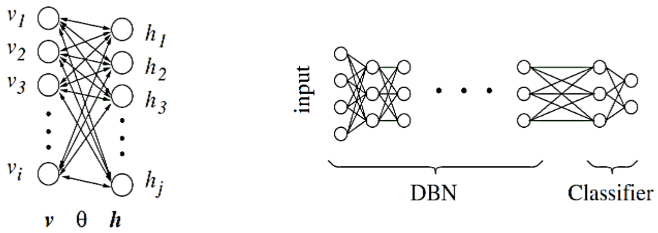

Fig. 2. Left: An RBM with i visible units, j hidden units, and model parameters θ. Right: A DBN with arbitrary width and a classifier.

A good representation is a representation that carries the most crucial underlying patterns of the data while the rest of the data is discarded. The best representation, however, is one that also catches the disentanglements of the patterns from one another, after capturing them. The best representation is similar to the human brain's ability to learn to recognize objects through disentangling them [37]. Those aforementioned definitions explain the term "representation learning" whereas the term "deep learning" refers to the implementation of this theory through a learning machine that contains many layers, such as a DBN. In this paper, the DBN is chosen over other deep constructions as the interest here is in studying one specific technique used by the DBN, pre-training, which remains partly unknown [38].

V. Experimental Results and Discussion

In order to estimate the performance of each model, input facial image dataset is split into training and test sets, then k-fold cross-validation (leave one out) was subsequently run [39]. In this paper, the value of k is set to 3; hence, the facial dataset was divided into 3 distinct parts. Two parts are used to train the classifier, while the third part is used to test classification except for K-NN because this is a non-parametric lazy learning algorithm.

The most popular kernel functions RBF, polynomial, linear and sigmoid kernel functions in this paper for SVM, Gaussian radial basis function (RBF) is applied. Further, in this paper, PCA is used to extract the significant feature and those features are used as input to the different classifier used in this paper. DBN, SVM, and K-NN classifiers are employed to classify dataset. With PCA applied to extract 750 features. Then DBN and SVM classifier was repeatedly trained in order to produce a model that is more precise for age estimation classification. In this section, a number of experiments have been done on a standard data set to measure the accuracy of the facial age approximation method suggested.

A. *Dataset Description*

The used dataset is built using images from Morph database II [40]. Morph database II the largest publicly available aging face database. It contains facial images of males and females with ages ranging from 16 to 77 years. It consists of 55,134 images of 13,000 persons and it took four years to be collected. Sample images are shown in Fig. 3.

The used dataset consists of 2494 images of males and females that are classified as into four age groups. The details of each age group including the age ranges and the number of images in each group are shown in Table I.

Fig. 3. Sample images from the used Morph dataset.

TABLE I. The Description of the used Dataset

Age Group	Age Range	Number of Images
Group I	31-40	624
Group II	41-50	1424
Group III	51-60	450
Group IV	61-70	46

A. *Experimental Results*

The comparison of the performance was done over 3-k fold cross validation as every group of images was divided into a 3-fold where each fold is always saved for testing and the remaining nine were used for training.

Fig. 4 shows the graphical representation of the performance measures for SVM classifier.

Fig. 5 shows the graphical representation of the performance measures for K-NN classifier.

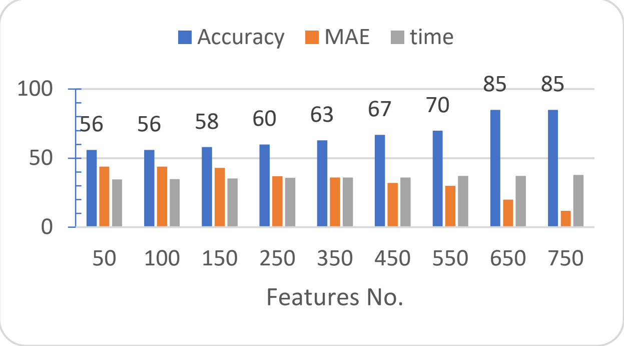

Fig. 4. Performance measurements for SVM.

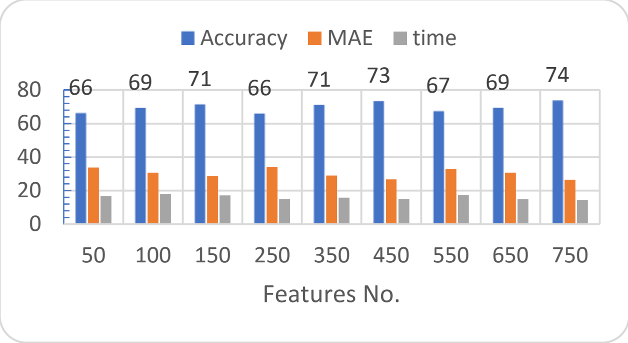

Fig. 5. Performance measurements for K-NN.

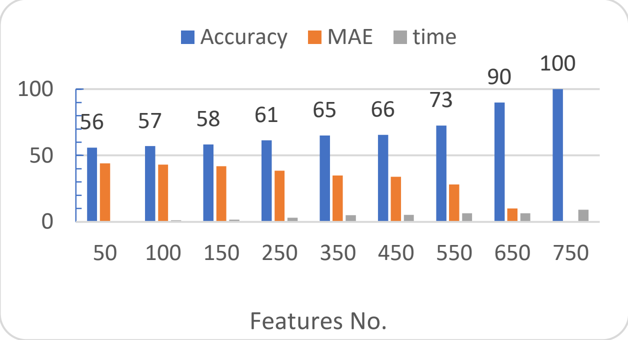

Fig. 6. Performance measurements for DBN.

Fig. 6 shows the graphical representation of the performance measures for DBN classifier.

In summary, Fig. 4, 5 and 6 shows a graphical comparison among all output accuracy measures for DBN, SVM, and K-NN classifiers. The results provided evidence that the DBN outperforms the comparison classifier. Comparing results, it is shown that best results are obtained by DBN classifier was 100% compared to best accuracies by SVM and K-NN classifier.

VI. Conclusion and Future Works

The research area of automated facial age estimation has gained an increasing attention from the researchers due to its relevance to several daily life applications. In this paper, we have proposed an automated facial age estimation method based on several well-known classifiers famous in the machine learning domain such as deep belief network, SVM and K-NN which able to estimate the human age based on the face image. The proposed approach consists of four steps: image preprocessing, feature extraction using PCA, feature reduction using PCA, and DBN based classification process. The proposed approach has been evaluated via a dataset that includes Morph II database images. The experimental results have shown that the proposed approach has a promising performance with achieved classification accuracy up to 100% compared with SVM and K-NN. For future works, different modifications are to be added to the proposed approach such as employing wavelet transform and linear discriminant analysis for feature extraction and nature-inspired algorithms for feature selection and classifier's parameter optimization.

References

[1] Y. Fu and T.S. Huang, "Human age estimation with regression on a discriminative aging manifold", IEEE Transactions on Multimedia (TMM) 10 (4) (2008), pp. 578–584.

[2] X. Geng, Z.H. Zhou and K. Smith-Miles, "Automatic age estimation based on facial aging patterns", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 29 (12) (2007), pp. 2234–2240.

[3] Noor Mualla, Essam H. Houssein, Hala H. Zayed, "Automated Facial Age Estimation Using Deep Belief Network", International Journal of Advancements in Computing Technology(IJACT), Vol. 9, No.3, 2017.

[4] K.Y. Chang, C.S. Chen and Y.P. Hung, "A ranking approach for human age estimation based on face images", in Proceedings of the International Conference on Pattern Recognition (ICPR), 2010, pp. 3396–3399.

[5] S. Yan, H. Wang, X. Tang and T.S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels", in Proceedings of the International Conference on Computer Vision (ICCV), 2007, pp. 1–8.

[6] X. Geng, Z.H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation", in Proceedings of the Fourteenth Annual ACM International Conference on Multimedia, 2006, pp. 307–316.

[7] N. Ramanathan, R. Chellapa and S. Biswas, "Age progression in human faces: a survey", J. Vis. Lang. Comput. 15 (2009) 3349–3361.

[8] [8] Y.H. Kwon and N.D.V. Lobo, "Age classification from facial images", in Proceedings of the 1994 IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 762–767.

[9] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models", IEEE Trans. Pattern Anal. Mach. Intell. (1998), pp. 484–498.

[10] A. Lanitis, C. Taylor and T. Cootes, "Toward automatic simulation of aging effects on face images", IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002), pp. 442–455.

[11] A. Lanitis, C. Draganova and C. Christodoulou, "Comparing different classifiers for automatic age estimation", IEEE Trans. Syst. Man Cybern. 34 (1) (2004), pp. 621–628.

[12] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in IEEE Conf. ICME'07, 2007, pp. 1383–1386

[13] G. Guo, Y. Fu, C. Dyer, and T. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression", IEEE Trans. Image Process. 17 (7) (2008), pp. 1178–1188.

[14] S. Z. Li and A. K. Jain. "Handbook of Face Recognition". Springer, New York.

[15] L.S. Mark, J.T. Todd, and R.E. Shaw. "Perception of growth: A geometric analysis of how different styles of change are distinguished", Journal of Experimental Psychology: Human Perception and Performance, pp.855–868, 1981.

[16] S. Yan, H. Wang, X. Tang and T. Huang, "Learning auto-structured regressor from uncertain nonnegative labels", in Proceedings of the Eleventh IEEE International Conference on Computer Vision, 2007, pp. 1–8.

[17] S. Yan, H. Wang, T. Huang, Q. Yang and X. Tang, "Ranking with uncertain labels", in Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, 2007, pp. 96–99.

[18] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002), pp. 971–987.

[19] Z. Yang and H. Ai, “Demographic classification with local binary patterns”, in S.W. Lee, S. Li (Eds.), Advances in Biometrics, Lecture Notes in Computer Science, vol. 4642, 2007, pp. 464–473.

[20] Diamantaras, Konstantinos I., and Sun-Yuan Kung. “Principal component neural networks: theory and applications”. John Wiley & Sons, Inc., 1996.

[21] Cao, L. J., et al. “A comparison of PCA, KPCA, and ICA for dimensionality reduction in support vector machine”. Neurocomputing 55.1 (2003), pp. 321-336.

[22] Cortes, C., and Vapnik, V., “Support-vector networks”, Machine learning, (1995), 20(3), pp. 273-297.

[23] Andrew, A.M., “An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods” by Nello Christianini and John Shawe-Taylor, Cambridge University Press, Cambridge, 2000.

[24] Cover, T., and Hart, P, “Nearest neighbor pattern classification”, IEEE transactions on information theory, (1967) 13(1), pp. 21-27.

[25] Le, J.H., Yazdanpanah, A.P., Regentova, E.E., and Muthukumar, V. “A Deep Belief Network for Classifying Remotely-Sensed Hyperspectral Data. In International Symposium on Visual Computing, (2015), pp. 682-692. Springer, Cham.

[26] Makantasis, K., Karantzalos, K., Doulamis, A. and Doulamis, N., “Deep supervised learning for hyperspectral data classification through convolutional neural networks”, In: IGARSS, pp. 1771–1800 (2015).

[27] Makantasis, K., Karantzalos, K., Doulamis, A., and Loupos, K., “Deep learning-based man-made object detection from hyperspectral data”, In International Symposium on Visual Computing (2015), pp. 717-727, Springer, Cham.

[28] Chen, Y., Zhao, X. and Jia, X. “Spectral-spatial classification of hyperspectral data based on deep belief network”. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, pp. 2381–2392 (2015)

[29] Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y. “Deep learning-based classification of hyperspectral data”. IEEE Journal of Selected topics in applied earth observations and remote sensing, (2014), 7(6), pp. 2094-2107.

[30] N. LeRoux and Y. Bengio, “Deep belief networks are compact universal approximators,” Neural Comput., vol. 22, no. 8, pp. 2192–2207, Aug. 2010.

[31] G.E. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets,” Neural Comput., vol. 18, no. 7, pp. 1527–1554, Jul. 2006.

[32] R. Salakhutdinov and G.E. Hinton, “Deep Boltzmann machines,” in Proc. Int. Conf. Artif. Intell. Statist. Clearwater Beach, FL, USA, 2009, pp.448–455.

[33] G. E. Hinton, “Apractical guide to training restricted Boltzmann machines,” Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. UTML TR2010-003, 2010.

[34] Y. LeCun et al., “Backpropagation applied to handwritten zip code recognition,” Neural Comput., vol. 1, no. 4, pp. 541–551, Apr. 1989.

[35] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biol. Cybern., vol. 36, no. 4, pp. 193–202, Apr. 1980.

[36] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., Montral, U.D. and Qubec, M., “Greedy layer-wise training of deep networks”. In: Sch¨olkopf, B., Platt, J., Hoffman, T. (eds.) NIPS. MIT Press, Cambridge (2007)

[37] Hinton, G.E., Osindero, S. and Teh, Y.W., “A fast learning algorithm for deep belief nets”, Neural Comput. 18, pp. 1527–1554 (2006)

[38] Bengio, Y., Courville, A.C. and Vincent, P., “Unsupervised feature learning and deep learning: a review and new perspectives”, CoRR abs, pp. 1206.5538 (2012)

[39] Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J., “Data Mining: Practical machine learning tools and techniques”. Morgan Kaufmann.

[40] K. Ricanek and T. Tesafaye, “Morph: a longitudinal image database of normal adult age-progression”, in: Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition, 2006, pp. 341–345.