

## Research Article

### Comparative Features Extraction Techniques for Electrocardiogram Images Regression

<sup>1</sup>Hend A. Elsayed, <sup>2</sup>Ahmed F. Abed and <sup>2</sup>Shawkat K. Guirguis

<sup>1</sup>Department of Communication and Computer Engineering, Faculty of Engineering, Delta University for Science and Technology, Mansoura,

<sup>2</sup>Department of Information Technology, Institute of Graduate Studies and Researches, Alexandria University, Alexandria Governorate, Egypt

**Abstract:** In this study, the comparative techniques have been developed to perform features extraction for the regression of the ECG images. Two regression methods have been used that are the linear and nonlinear regression. The features extraction techniques developed in this study are the nonnegative matrix factorization used to extract the feature from the ECG images and compare the results with different techniques such as principal component analysis, kernel principal component analysis and independent principal component analysis. These features are used for image regression using two regression techniques and compare between these two regressions techniques. The performance evaluation through this comparison is the error rate that is the root mean square error between the actual data and the data predicted from the regression and the results conclude the principal component analysis technique outperforms the other techniques.

**Keywords:** Independent component analysis, kernel principal component analysis, linear regression, non linear regression, nonnegative matrix factorization, principal component analysis

## INTRODUCTION

Introductory statistics and regression texts often focus on how regression can be used to represent relationships between variables, rather than as a comparison of average outcomes. Regression can be used to predict an outcome given a function of these predictors and regression coefficients can be thought of as comparisons across predicted values or as comparisons among averages in the data development by Gelman and Hill (2007) and use of the Nonnegative Matrix Factorization (NMF) algorithm for feature extraction and identification in the fields of text mining and spectral data analysis are presented. The evolution and convergence properties of hybrid methods based on both sparsity and smoothness constraints for the resulting nonnegative matrix factors are discussed by Takeda *et al.* (2007). Nonnegative matrix factorization is a very useful tool to reduce dimension of data and is a means of extracting feature information from images. This information is used as a signature for the image and similar images should have similar signatures, these features extraction is used in machine learning and data mining by Mejía-Roa *et al.* (2015) and for face recognition based on linear regression by Fang *et al.* (2012).

Nonnegative Matrix Factorization (NMF) is a dimension-reduction technique based on a low-rank approximation of the feature space. Besides providing a reduction in the number of features. NMF technique has gained a great interest among the Bioinformatics community, since it is able to extract interpretable parts from high dimensional datasets. In addition to statistical independence proved effective in many clustering and classification tasks by Mejía-Roa *et al.* (2015), Li *et al.* (2014) and Berry *et al.* (2007). NMF only allows additive relationship. This constraint has proved to be closer to the way how humans perceive and understand the data based on this methodology, we can map the low-level image features into the additive combination of latent semantics for the clustering by Mejía-Roa *et al.* (2015) and Li *et al.* (2014). The evolution and convergence properties of hybrid methods based on both sparsity and smoothness constraints for the resulting nonnegative matrix factors are discussed by Berry *et al.* (2007). NMF is a method to obtain a representation of data using non-negativity constraints. These constraints lead to a part-based representation because they allow only additive, not subtractive, combinations of the original data.

Regression is a statistical approach to forecasting change in a dependent variable (sales revenue, for

**Corresponding Author:** Hend A. Elsayed, Department of Communication and Computer Engineering, Faculty of Engineering, Delta University for Science and Technology, Mansoura, Egypt

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

example) on the basis of change in one or more independent variables (population and income, for example). Known also as curve fitting or line fitting because a regression analysis equation can be used in fitting a curve or line to data points, in a manner such that the differences in the distances of data points from the curve or line are minimized. Relationships depicted in a regression analysis are, however, associative only and any cause effect (causal) inference is purely subjective, also called regression method or regression technique by Huang (2014). The general goal of regression analysis is to estimate the association between one or more explanatory variables and a single outcome variable. An outcome variable is presumed to depend in some way or be systematically predicted by the explanatory variables.

This study, proposed a study focused on the comparison between the four techniques for features extraction and two types of regression, including linear and nonlinear.

## MATERIALS AND METHODS

This study is the work of master thesis on 2016 and in Department of Information Technology, Institute of Graduate Studies and Researches, Alexandria University, Egypt.

There are two regression methods linear and nonlinear regression in subsections A and B respectively.

**A. Linear regression:** Linear regression models the relationship between the continuous dependent variable and independent variable by fitting the data into a linear equation. The dependent variable is that we usually refer to it as outcome, or Y. The independent variable is that we refer to it as predictor, or X, which we suspect contributes to the outcome by Hoffmann and Shafer (2005).

**B. Nonlinear regression:** The nonlinear regression is the same as that of linear regression, namely to relate a response Y to a vector of predictor variables. In nonlinear regression the prediction equation depends nonlinearly on one or more unknown parameters. The linear regression is often used for building a purely empirical model but nonlinear regression usually arises when there are physical reasons for believing that the relationship between the response and the predictors follows a particular functional form by Smyth (2002).

The proposed method is image regression for ECG images by the preprocessing for the images and to obtain the features from factoring matrix of nonnegative matrix factorization and apply the linear and nonlinear regression for these features and compare the results with the features obtained from the Principal Component Analysis (PCA),

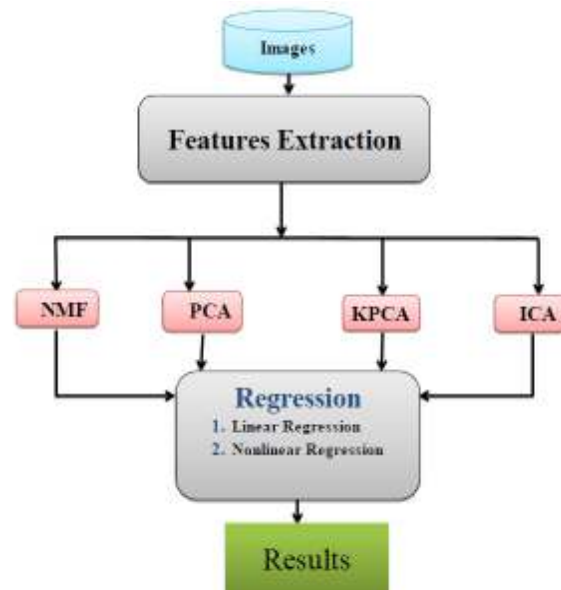


Fig. 1: The general steps for the proposed system

(ICA) Independent Component Analysis and Kernel Principal Component Analysis (KPCA) and Fig. 1 shows the steps for the proposed system.

## RESULTS AND DISCUSSION

We first, enter the images to the system. The images are twenty ECG images and apply preprocessing for those images by resizing the image to the measured rate of (256×256) pixels and then converted it from RGB to gray scale and then convert them to two-dimensional matrix.

The second is the feature extraction using a NMF algorithm from the images. The third is apply the two linear and nonlinear regression methods using different number of features and calculate the root mean square regression error for each number of feature by extracting the error that is by subtracting the actual data from the predicted data and then calculate its root mean square and then compare with the other feature extraction methods such as PCA, KPCA and ICA. An example is taken for a number of features twenty and then gets the predicted data using the two regression methods using PCA, KPCA, ICA and NMF features extraction method.

The results using linear regression are shown in subsection A and in subsection B using nonlinear regression:

**A. Linear regression:** Table 1 shows the root mean square error when using that ECG images for (1, 2, 3,....., 20) features extraction and using linear regression by using the matlab function Fitlm we note that the (PCA) is the minimum error according to the other techniques NMF, ICA and

Table 1: Linear regression error

Feature number	Linear regression			
	NMF	ICA	PCA	KPCA
1	0.4621	0.498	0.3600	0.4867
2	0.4867	0.500	0.1237	0.4867
3	0.4600	0.4858	0.1330	0.4866
4	0.4998	0.4871	0.1847	0.4938
5	0.4690	0.4867	0.2070	0.4911
6	0.4947	0.4654	0.2209	0.4891
7	0.4561	0.4967	0.2355	0.4898
8	0.4507	0.4917	0.2520	0.4904
9	0.4855	0.4944	0.2547	0.4893
10	0.4659	0.4891	0.2627	0.4923
11	0.4644	0.4999	0.2757	0.4936
12	0.4970	0.4911	0.2828	0.4947
13	0.4991	0.4976	0.2889	0.4927
14	0.4999	0.4630	0.2987	0.4913
15	0.4637	0.4914	0.3056	0.4910
16	0.4781	0.4996	0.3118	0.4902
17	0.4854	0.4865	0.3144	0.4889
18	0.4784	0.4890	0.3170	0.4904
19	0.4893	0.4981	0.3207	0.4909
20	0.4677	0.4937	0.3246	0.4918

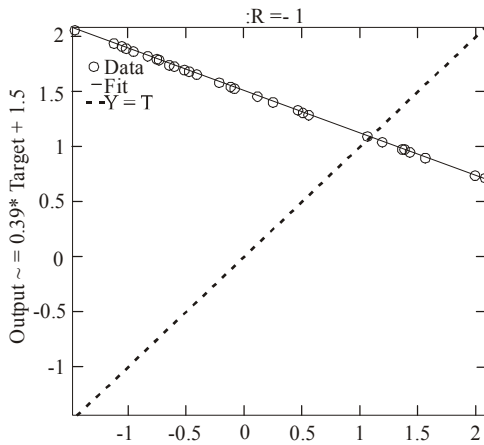


Fig. 2: Linear regression using PCA

KPCA, Fig. 2 to 5 show the actual and the predicted data for a number of features twenty using the different features extraction methods such as PCA, KPCA, ICA and NMF.

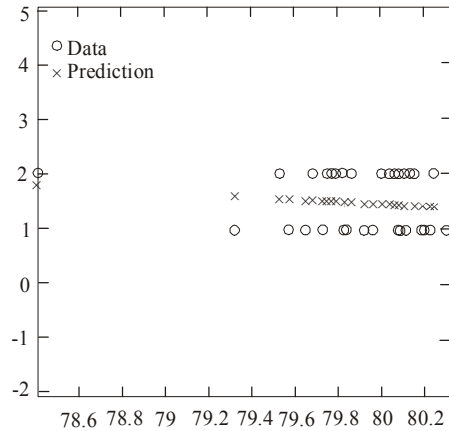


Fig. 3: Linear regression using KPCA

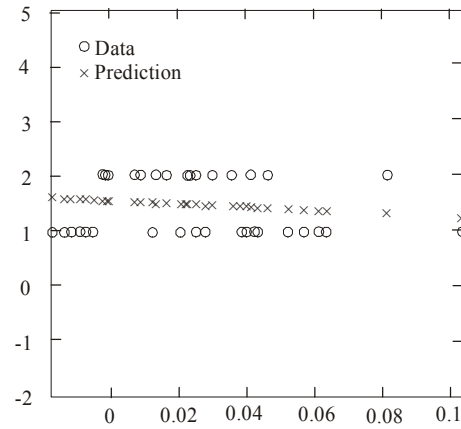


Fig. 4: Linear regression using ICA

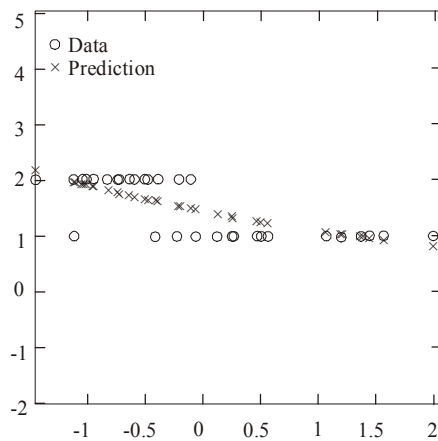


Fig. 5: Linear regression using NMF

**B. Nonlinear regression:** Table 2 shows the root mean square error when using that ECG images for (1, 2, 3,.....20) features extraction and using nonlinear regression by using the matlab function Fitnlm we note that the (PCA) is the minimum error according to the other techniques NMF, ICA

Table 2: Nonlinear regression error

Feature number	Nonlinear regression			
	NMF	ICA	PCA	KPCA
1	0.5008	0.4980	0.3601	0.5
2	0.5008	0.5007	0.0798	0.5
3	0.5008	0.4859	0.0882	0.5
4	0.5008	0.4875	0.1341	0.5
5	0.5008	0.4868	0.1556	0.5
6	0.5008	0.4655	0.1631	0.5
7	0.5008	0.4975	0.2237	0.5
8	0.5008	0.4918	0.2335	0.5
9	0.5008	0.4945	0.2441	0.5
10	0.5008	0.4891	0.2581	0.5
11	0.5008	0.5000	0.2743	0.5
12	0.5008	0.4912	0.2786	0.5
13	0.5008	0.4977	0.2733	0.5
14	0.5008	0.4630	0.2919	0.5
15	0.5008	0.4914	0.2988	0.5
16	0.5008	0.4996	0.3108	0.5
17	0.5008	0.4866	0.3131	0.5
18	0.5008	0.4891	0.3151	0.5
19	0.5015	0.5015	0.3146	0.5
20	0.5008	0.4938	0.3199	0.5

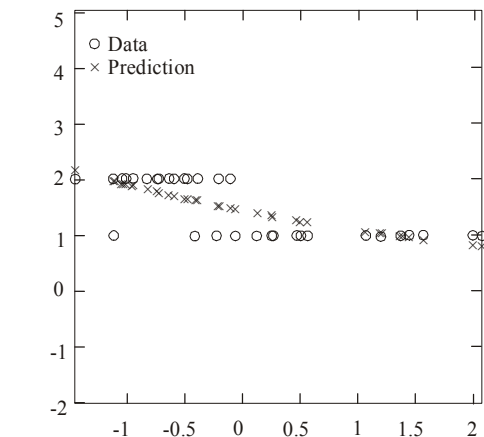
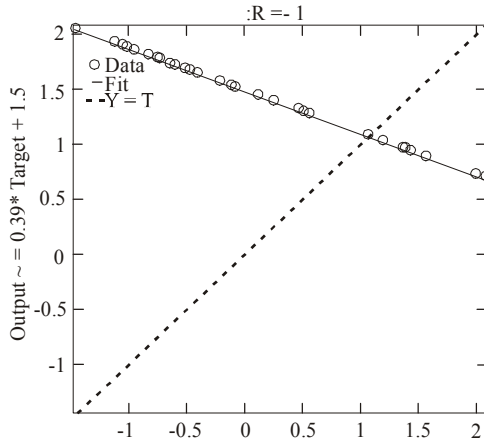


Fig. 6: Nonlinear regression using PCA

and KPCA, Fig. 6 to 9 show the actual and the predicted data for a number of features twenty using the different features extraction methods such as PCA, KPCA, ICA and NMF.

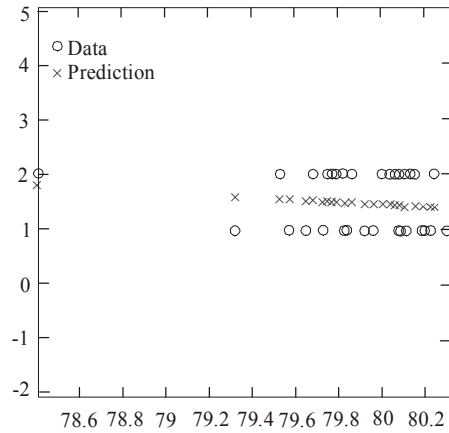


Fig. 7: Nonlinear regression using KPCA

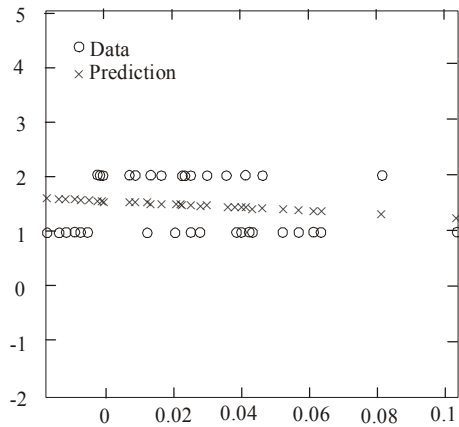


Fig. 8: Nonlinear regression using ICA

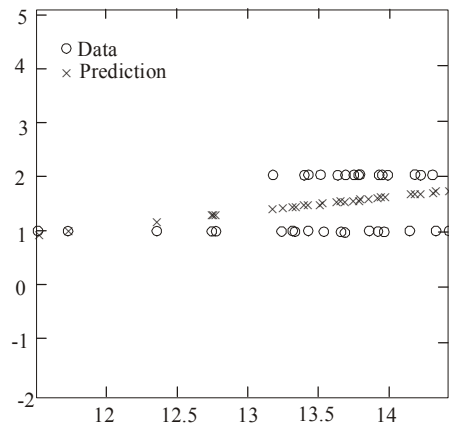


Fig. 9: Nonlinear regression using NMF

### CONCLUSION

This study, proposed a study focused on performance evaluation through the comparison between these four techniques for features extraction and two types of regression, including linear and nonlinear. The disparity between the results depends on the used data and the techniques used for features extraction and for the regression. From this study, we

concluded that the PCA is the best technique of the four techniques used PCA, NMF, KPCA and ICA and it was providing lowest error and then the highest percentage of performance among these technologies in the used ECG images.

#### REFERENCES

- Berry, M.W., M. Browne, A.N. Langville, V. Paul Pauca and R.J. Plemmons, 2007. Algorithms and applications for approximate nonnegative matrix factorization. *Comput. Stat. Data An.*, 52(1): 155-173.
- Fang, W., Z. Cheng and D. Yang, 2012. Face recognition based on non-negative matrix factorization and linear regression. *Int. J. Adv. Comput. Technol.*, 4(16): 93-100.
- Gelman, A. and J. Hill, 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York.
- Hoffmann, J.P. and K. Shafer, 2005. *Linear Regression Analysis: Assumptions and Applications*. Department of Sociology, Brigham Young University, USA.
- Huang, N., 2014. Scenarios Where Utilizing a Spline Model in Developing a Regression Model is Appropriate. Paper No. 1760-2014. Retrieved from: <http://support.sas.com/resources/papers/proceedings14/1760-2014.pdf>.
- Li, L., J. Yang, K. Zhao, Y. Xu, H. Zhang and Z. Fan, 2014. Graph regularized non-negative matrix factorization by maximizing correntropy. arXiv Preprint arXiv: 1405-2246.
- Mejía-Roa, E., D. Tabas-Madrid, J. Setoain, C. García, F. Tirado and A. Pascual-Montano, 2015. NMF-mGPU: Non-negative matrix factorization on multi-GPU systems. *BMC Bioinformatics*, 16: 43.
- Smyth, G.K., 2002. Nonlinear Regression. In: El-Shaarawi, A.H. and W.W. Piegorsch (Eds.), *Encyclopedia of Environmetrics*. John Wiley and Sons Ltd., Chichester, 3: 1405-1411.
- Takeda, H., S. Farsiu and P. Milanfar, 2007. Kernel regression for image processing and reconstruction. *IEEE T. Image Process.*, 16(2): 349-366.