2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Heart Failure diagnosis based on deep learning techniques*

Theofilos G. Papadopoulos, Daphni Plati, Evanthia E. Tripoliti, *Senior Member*, *IEEE*, Yorgos Goletsis, Katerina K. Naka, Aidonis Rammos, Aris Bechlioulis, Chris Watson, Kenneth McDonald, Mark Ledwidge, Rebabonye Pharithi, Joseph Gallagher and Dimitrios I. Fotiadis, *Fellow IEEE*

Abstract— **The aim of the study is to address the heart failure (HF*)* diagnosis with the application of deep learning approaches. Seven deep learning architectures are implemented, where stacked Restricted Boltzman Machines (RBMs) and stacked Autoencoders (AEs) are used to pre-train Deep Belief Networks (DBN) and Deep Neural Networks (DNN). The data is provided by the University College Dublin and the 2nd Department of Cardiology from the University Hospital of Ioannina. The features recorded are grouped into: general demographic information, physical examination, classical cardiovascular risk factors, personal history of cardiovascular disease, symptoms, medications, echocardiographic features, laboratory findings, lifestyle/habits and other diseases. The total number of subjects utilized is 422. The deep learning methods provide quite high results with the Autoencoder plus DNN approach to demonstrate accuracy 91.71%, sensitivity 90.74%, specificity 92.31% and f-score 89.36%.**

I. INTRODUCTION

Heart Failure (HF) is a chronic clinical syndrome of the cardiovascular system in which the heart cannot pump enough blood to deal with the metabolic needs of the body, that is mostly caused by reduced left ventricular myocardial function [1]. The main symptoms of HF are dyspnea and fatigue, that may cause reduced exercise tolerance, and fluid retention, and may lead to pulmonary and/or splanchnic congestion and/or peripheral edema [2]. HF is among the major causes of mortality and morbidity in western societies and it is responsible for high costs due to hospitalization [3].

Early diagnosis of HF is essential. Therefore, it is very important to establish a dynamic HF diagnostic model. Towards this direction, machine learning (ML) techniques have significantly contributed. In recent years, deep learning (DL), a branch of ML techniques based on learning representations of data with multiple levels of abstraction between the inputs and outputs of the algorithm, has gained much attention in the field of heart disease detection and prediction. DL offers a powerful alternative compared with conventional ML, due to the fact that they enable the users to perform more complex analysis [4].

Several studies have been conducted to build a model that can diagnose HF based on various DL algorithms. Choi *et al.* [5] implemented Recurrent Neural Network (RNN) models and multilayer perceptron with 1 hidden layer for early detection of HF. Both Chen *et al.* [6] and Wang *et al.* [7] utilized Heart Rate Variability (HRV) measures to predict congestive HF. Chen *et al.* [6] implemented a Deep Neural Network (DNN) whereas Wang *et al.* [7] combined the long short-term memory (LSTM) network and Convolution Network architecture. In 2019, Acharya *et al.* [8] implemented a Convolutional Neural Network (CNN) analyzing also ECG signals. Several studies [9-12] implement a DL-based analysis on ECG signals. Kim *et al.* [9] analyzed ECG and echocardiographic data from subjects with chronic HF with reduced ejection fraction (EF) (HFrEF) and HF with mid-range EF (HFmrEF). Ning *et al.* [10] also analyzed ECG signals and applied a hybrid DL algorithm that was composed of a CNN and a recursive N. Kwon *et al.* [11] developed an algorithm using a DNN and analyzing demographic information and ECG features. Le *et al.* [13] developed a multilayer perceptron neural network utilizing demographics, laboratory findings, lifestyle habits and EF features. Gjoreski *et al.* [14] recently conducted a method that combines classic ML techniques and end-to-end DL to detect chronic HF. The classic ML learns from expert features, and the DL learns from a spectro-temporal representation of the signal.

Most of the above approaches focus either on ECG raw signal or HRV measures to classify subjects as HF or not. Still, clinical practice suggests that HF diagnosis is a complex procedure that should take into consideration multiple clinical manifestations. In this sense, in our study, we exploit the data mentioned in European Society of Cardiology (ESC) guidelines for HF [2]; are usually collected in clinical practice

*Research supported by the KardiaTool project (http://www.kardiatool.eu/) that has received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement No 768686. This article reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.

T. G. Papadopoulos is with the Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, Ioannina, Greece (e-mail: tpapado2011@gmail.com).

D. Plati and E. E. Tripoliti are with the Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, Ioannina, Greece (e-mail: daphni.plati@gmail.com, etripoliti@gmail.com).

Y. Goletsis is with the Department of Economics, University of Ioannina, Ioannina, Greece and with the Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, Ioannina, Greece (e-mail: goletsis@uoi.gr).

K. K. Naka, A. Rammos, A. Bechlioulis are with the 2nd Department of Cardiology, Medical School, University of Ioannina, Ioannina, Greece (e-mail: drkknaka@gmail.com, aidrammos@yahoo.gr, md02798@yahoo.gr).

C. Watson, K. McDonald, M. Ledwidge, R. Pharithi, J. Gallagher are with the University College Dublin, National University of Ireland, Belfield, Dublin, Ireland (e-mail: chris.watson@qub.ac.uk, kenneth.mcdonald@ucd.ie, mark.ledwidge@ucd.ie, rpharithi@gmail.com, jgallagher@ucd.ie).

Dimitrios I. Fotiadis is with the Department of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH, Ioannina, Greece and with the Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, Ioannina, Greece (phone: +302651009006, fax: +302651008889, e-mail: fotiadis@cc.uoi.gr).

and we apply DL in a multisource dataset emulating the medical knowledge about the disease.

II. MATERIALS AND METHODS

A. The dataset

Data is provided by the University College Dublin (UCD), Ireland, and the 2nd Department of Cardiology from the University Hospital of Ioannina. Data is collected in the framework of KardiaTool project. The dataset for HF diagnosis consists of 422 subjects (260 without HF and 162 with chronic HF belonging to NYHA class I or II, as the diagnosis of NYHA class III or IV patients is rather straightforward and does not necessarily benefits by the classification abilities of DL). The features recorded for each patient are grouped into the following categories: general demographic information, physical examination, classical cardiovascular risk factors, personal history of cardiovascular disease, symptoms, medication, echocardiogram, laboratory findings, lifestyle/habits and other diseases. In total, 28 features are recorded for each patient.

B. The proposed methodology

Data preprocessing is performed as a first step. Missing values are imputed by the most frequent value and the mean value for nominal and numeric input features respectively. Additionally, the nominal input features are transformed by one-hot encoding into dummy/indicator variables. The values of the numeric input features are normalized by scaling to the range of [0,1]. In the present study, stacked Restricted Boltzman Machines (RBMs) [15] and stacked Autoencoders (AEs) [15] are used to pre-train Deep Belief Networks (DBN) [16] and DNN [17] which operate either as feature extractors or directly as classifiers [18-22]. A combination of stacked RBMs and a Deep Autoencoder (DeepAE) [23] is implemented as feature extractor as well. The learned features from the deep feature extractors are used as input to a Random Forest (RF) classifier.

Seven DL architectures are implemented: (i) Stacked RBMs + RF classifier, (ii) DBN + RF classifier. (iii) Stacked RBMs + DeepAE + RF classifier, (iv) Stacked RBMs + DNN. (v) Stacked Autoencoders (AE) + DNN, (vi) DBNs, (vii) DNN. Three architectures were implemented as deep feature extractors and four architectures as deep classifiers. The recorded features from the data, which consist of 12 categorical and 16 numeric characteristics, feed the models.

Deep feature extractors

Stacked RBMs plus RF classifier: Two RBMs are trained to produce features layer by layer in unsupervised way [24]. The top layer features feed a standalone supervised classifier. The size of the two hidden layers is 200 and 100 respectively. Thus, 100 features enter the RF classifier (Fig. 1).

DBN plus RF classifier: A three-layer DBN, pre-trained by RBMs and fine-tuned by wake-sleep algorithm in unsupervised way, derives features in the last layer which feed separately a RF classifier [25]. The size of the three hidden layers is 200, 100 and 50 respectively (Fig. 2).

DeepAE plus RF classifier: A DeepAE, consisting of three encoding layers and three decoding layers initialized by RBMs pre-trained in unsupervised way, derives compact representation of features that next feed a supervised classifier [23]. The size of compression by layer is 200-100-50, accordingly (Fig. 3).

Deep classifiers

Stacked RBMs plus DNN: Two RBMs are trained sequentially in an unsupervised way using persistent contrastive divergence [26] performing one Gibbs sampling. The output of the previous RBM is used as input to the next RBM. The weights of the RBMs are used to initialize a deep neural network with three layers that is fine-tuned in supervised way by backpropagation [27]. The size of the two hidden layers is 200 and 100, respectively (Fig. 4).

Stacked AEs plus DNN: Two denoising autoencoders (d-AE) [28, 29] are trained sequentially in unsupervised way and create a corrupted copy of the input by introducing some noise. Corruption is done randomly by setting some input nodes to zero. This helps to avoid the autoencoders to copy the input to the output without learning features about the data. The learned weights are used to initialize a deep neural network with three layers, which is fine-tuned in a supervised way by backpropagation as in the previous model. (Fig. 5).

DBN: A DBN is pre-trained by RBMs and fine -tuned by wake-sleep algorithm in a supervised way. The same structure is retained as in the initial paper introducing DBNs by Hinton et al. [16]. The DBN has three layers and in the penultimate layer are added as many nodes as the output classes. The output is one-hot encoded which means the value of the correct label is set to 1 and the remainder to zero. The top two layers compose an RBM which is trained with contrastive divergence [30, 31]. The number of Gibbs sampling is scaling from 3 to 10 according to the number of epochs. The top layer operates as a feature detector which learns to model the joint distribution of the features and the labels. The size of the three hidden layers is 200, 100 and 50, respectively (Fig. 6).

DNN: A typical structure of Neural Network with three hidden layers is trained in a supervised way without pre-training and initialized by glorot-uniform initialization [32]. The size of the three hidden layers is 200, 100 and 50, respectively (Fig. 7).

Ten-fold stratified cross validation is implemented for the evaluation of the models. Cross validation error is considered to be a reliable estimate of the out of sample error. Regularization techniques such as dropout [33] and denoising [28, 29] prevent from overfitting. Unsupervised pre-training acts as a regularizer as well [19, 34]. The implementation and the evaluation of the deep models is done in Scikit-learn API combined with Keras (backended by tensorflow) upon python3. The RF classifier is provided by the WEKA software. The python-weka-wrapper3 (PyWEKA3) library is used to combine RF classifier with deep models.

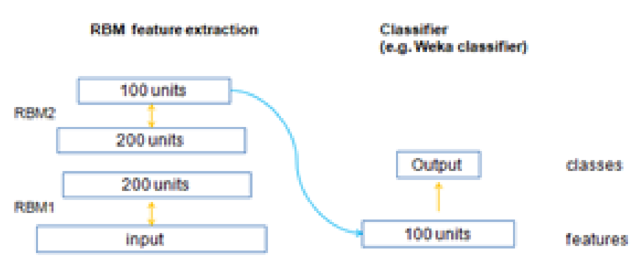

Figure 1. Structure of deep feature extractors: Restricted Boltzman Machine (RBM) with classifier.

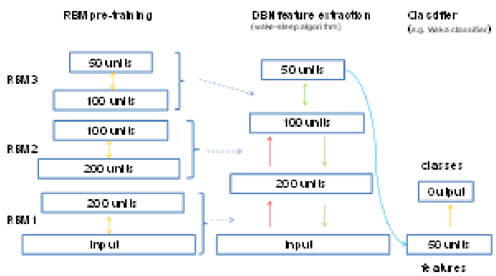

Figure 2: Structure of deep feature extractors: Deep Belief Network (DBN) with classifier.

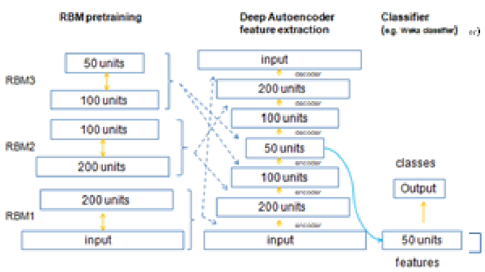

Figure 3: Structure of deep feature extractors: Deep Autoencoder with classifier.

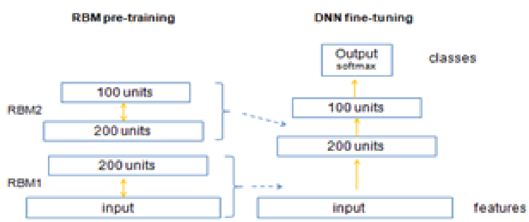

Figure 4: Deep classifier: Restricted Boltzmann Machine (RBM) and Deep Neural Network (DNN).

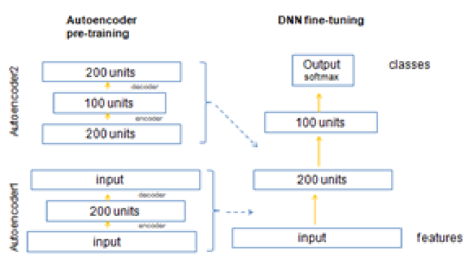

Figure 5: Deep classifier: Stacked Denoising Autoencoders with Deep Neural Network (DNN).

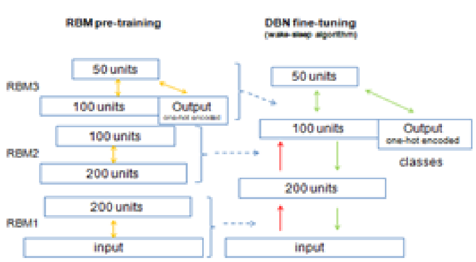

Figure 6: Deep classifier: Deep Belief Network (DBN) (embedded classifier).

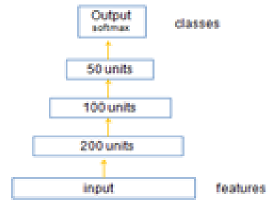

Figure 7: Deep classifier: Deep Neural Network.

III. RESULTS

The results are reported in terms of four common evaluation measures, i.e. classification Accuracy (Acc), Sensitivity (Sens), Specificity (Spec) and F-score (TABLE I). The Autoencoder plus DNN model achieves the highest results, though slightly better, in terms of accuracy (91.71%), sensitivity (90.74%), specificity (92.31%) and F-score (89.36%). Trials integrating the initial scaling of the numeric data in the cross validation loop do not show any degeneration of the results eventually due to implementation of the regularization techniques mentioned above.

TABLE I. CLASSIFICATION RESULTS FOR HF DIAGNOSIS.

Deep Feature Extractors				
	Acc %	Sens %	Spec %	F-score %
RBM plus RF	89.10	86.42	91.92	85.89
DBN plus RF	89.34	85.80	91.54	86.07
DeepAE plus RF	89.34	85.19	91.92	85.98
Deep classifiers				
	Acc %	Sens %	Spec %	F-score %
RBM plus DNN	91.47	90.12	92.31	89.02
AE plus DNN	**91.71**	**90.74**	**92.31**	**89.36**
DBN	85.78	80.86	88.85	81.37
DNN	90.28	82.18	89.14	82.66

IV. DISCUSSION

In the current study the dataset is multisource and cannot be directly compared with previous studies, since most of the existing relevant approaches (TABLE II) focus on classification based on ECG signals or HRV measures.

TABLE II. STATE OF THE ART IN DL FOR HF DIAGNOSIS.

HF early diagnosis		
Study	***Dataset***	***Evaluation measures***
Precent Study	422 subjects 162 with chronic HF 260 controls	Accuracy 91.71% Sensitivity 90.74% Specificity 92.31% F-score 89.36%
Demographics, physical examination, classical cardiovascular risk factors, personal history of cardiovascular disease, symptoms, medications, echocardiographic features, laboratory findings, lifestyle/habits and other diseases		
Choi *et al.* [5]	3884 with HF 28903 controls	AUC 0.777
Demographics, habits, clinical and laboratory values, ICD-9 codes, CPT codes, medications		
Chen *et al.* [6]	116 subjects 44 with congestive HF 72 controls	Accuracy 72.44% Sensitivity 50.39% Specificity 84.93%
HRV measures based on the RR interval		
Wang *et al.* [7]	156 subjects 44 with congestive HF 112 controls	Accuracy 99.22%
HRV measures based on the RR interval		

HF early diagnosis		
Study	***Dataset***	***Evaluation measures***
Acharya *et al.* [8]	73 subjects 15 with congestive HF 58 controls	Accuracy 98.97% Sensitivity 98.87% Specificity 99.01%
ECG signals		
Kim *et al.* [9]	2209 subjects with chronic HF 18196 controls	ROC +4 to +6 Specificity 98% Positive predictive value 95%
ECG signals, echocardiographic features		
Ning *et al.* [10]	15 subjects with chronic HF 18 controls	Accuracy 99.93% Sensitivity 99.85% Specificity 100%
ECG signals		
Kwon *et al.* [11]	22.765 subjects 1.391 with HFrEF	AUC for identification of HFrEF was 84.30%
Demographics, ECG features, atrial fibrillation or atrial flutter, QT interval, QRS duration, R wave axis, and T wave axis, as the predictive variables		
Le *et al.* [13]	299 subjects	Accuracy 88.00%
Demographics, laboratory findings, habits, ejection fraction		
Gjoreski *et al.* [14]	947 subjects	Accuracy 92.90% sensitivity 82.30% specificity 96.20%
Heart sound characteristics		

V. CONCLUSION

In the present study, we present a method for diagnosing HF based on DL. We implemented various DL techniques on a multisource dataset, which renders our study quite innovative due to the fact that other studies are mainly based on ECG signals or HRV measures. The Autoencoder plus Deep Neural Network model achieved the highest results in terms of accuracy (91.71%), sensitivity (90.74%), specificity (92.31%) and F-score (89.36%), indicating the applicability of the proposed approach. In future work, the statistical significance of the performance can be explored comparing the models by additional runs with different random seeds.

REFERENCES

[1] S. Dassanayaka and J.P. Steven, “Recent Developments in Heart Failure,” Circulation Research, vol. 117, no 7, pp. 58-63, 2015.

[2] Ponikowski, P., et al., “2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC,” European Heart Journal, vol. 37, no. 27, pp. 2129-2200, 2016.

[3] E. Sabate, “Adherence to long-term therapies, Evidence for action,” World Health Organization, 2003.

[4] C.R. Olsen, et al., “Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure," American Heart Journal, vol. 229, pp. 1-17, 2020.

[5] E. Choi, et al., “Using recurrent neural network models for early detection of heart failure onset,” Journal of the American Medical Informatics Association, vol. 24, no. 2, pp. 361–370, 2017.

[6] W. Chen, et al., “A CHF detection method based on deep learning with RR intervals,” in 39th Annual International Conference of the IEEE Engineering in Medicine, pp. 3369-3372, 2017.

[7] L. Wang and X. Zhou, “Detection of Congestive Heart Failure Based on LSTM-Based Deep Network via Short-Term RR Intervals,” Sensors, vol. 19, no. 7, pp. 1502, 2019.

[8] U.R. Acharya, et al., “Deep convolutional neural network for the automated diagnosis of congestive heart failure using ECG signals,” Applied Intelligence, vol. 49, no. 1, pp. 16-27, 2019.

[9] K. Kim and J. Kwon, “Deep Learning for Diagnosing Heart Failure from ECG Signals,” The Journal of Heart and Lung Transplantation, vol. 38, no. 4, pp. S375, 2019.

[10] W. Ning, et al., “Automatic Detection of Congestive Heart Failure Based on a Hybrid Deep Learning Algorithm in the Internet of Medical Things,” IEEE Internet of Things Journal, 2020.

[11] J. Kwon, et al., “Development and Validation of Deep-Learning Algorithm for Electrocardiography-Based Heart Failure Identification,” Korean circulation journal, vol. 49, no. 7, pp. 629-639, 2019.

[12] D. Li, et al., “Automatic staging model of heart failure based on deep learning,” Biomedical Signal Processing and Control, vol. 52, pp. 77-83, 2019.

[13] M.T. Le, et al., “Predicting heart failure using deep neural network,” in 2020 International Conference on Advanced Technologies for Communications (ATC). pp. 221-225, 2020.

[14] M. Gjoreski, A. Gradisek, B. Budna, M. Gams, G. Poglajen, “Machine Learning and End-to-End Deep Learning for the Detection of Chronic Heart Failure From Heart Sounds,” IEEE Access, vol. 8, pp. 20313-20324, 2020.

[15] Y. Bengio, et al., “Greedy layer-wise training of deep networks,” Advances in Neural Information Processing Systems, pp. 153-160, 2007.

[16] G.E. Hinton, et al., “A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18, no. 7, pp. 1527-1554, 2006.

[17] Y. Lecun, et al., “Deep learning,” Nature, vol. 521, no. 7553, pp. 436-444, 2015.

[18] Y. Bengio, “Practical recommendations for gradient-based training of deep architectures,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 437-478, 2012.

[19] D. Erhan, et al., “Why does unsupervised pre-training help deep learning?,” Journal of Machine Learning Research, vol. 11, pp. 625-660, 2010.

[20] D. Erhan, et al., “Undertanding representations learned in deep architectures,” Technical Report 1355 Universite de Montreal/DIRO, 2010.

[21] G.E. Hinton, “A practical guide to training restricted boltzmann machines,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 599-619. 2012.

[22] H. Larochelle, et al., “Exploring strategies for training deep neural networks,” Journal of Machine Learning Research, vol. 10, pp. 1-40, 2009.

[23] G.E. Hinton, R.R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504-507, 2006.

[24] H. Larochelle, M. Mandel, R. Pascanu, Y. Bengio, “Learning algorithms for the classification restricted Boltzmann machine,” Journal of Machine Learning Research, vol. 13, pp. 643-669, 2012.

[25] G.E. Hinton, “To recognize shapes, first learn to generate images,” Progress in Brain Research, pp. 535-547, 2007.

[26] T. Tieleman, “Training restricted boltzmann machines using approximations to the likelihood gradient,” Proceedings of the 25th International Conference on Machine Learning, pp. 1064-1071, 2008.

[27] D.E. Rumelhart, et al., “Learning representations by back-propagating errors,” Nature, vol. 323, no. 6088, pp. 533-536, 1986.

[28] P. Vincent, et al., “Extracting and composing robust features with denoising autoencoders,” Proceedings of the 25th International Conference on Machine Learning, pp. 1096-1103. 2008.

[29] P. Vincent, et al., “Stacked denoising autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion,” Journal of Machine Learning Research, vol.11, pp. 3371-3408, 2010.

[30] G.E. Hinton, “Training products of experts by minimizing contrastive divergence”, Neural Computation, vol. 14, no. 8, pp. 1771-1800, 2002.

[31] Y. Bengio, O. Delalleau, “Justifying and generalizing contrastive divergence,” Neural Computation, vol. 21, no. 6, pp. 1601-1621, 2009.

[32] X. Glorot, Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” Journal of Machine Learning Research, vol. 9, pp. 249-256, 2010.

[33] N. Srivastava, et al., “Dropout: A simple way to prevent neural networks from overfitting,” Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014.

[34] Y. Bengio, “Learning deep architectures for AI,” Foundations and Trends in Machine Learning, vol. 2, no 1, pp. 1-27. 2009.