Telematics and Informatics 33 (2016) 277–292

Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Disease prediction with different types of neural network classifiers

Cheng-Hsiung Weng [a], Tony Cheng-Kui Huang [b,*], Ruo-Ping Han [a]

[a] Department of Management Information Systems, Central Taiwan University of Science and Technology, Taichung, Taiwan, ROC
[b] Department of Business Administration, National Chung Cheng University, 168, University Rd., Min-Hsiung, Chia-Yi, Taiwan, ROC

A R T I C L E I N F O

Article history:
Received 24 March 2015
Received in revised form 20 July 2015
Accepted 13 August 2015
Available online 21 August 2015

Keywords:
Disease prediction
Artificial neural network
Ensemble classifier
Statistical testing

A B S T R A C T

Disease prediction has long been regarded as a critical topic. Artificial intelligence and machine learning techniques have already been developed to solve this type of medical care problem. Recently, neural network ensembles have been successfully utilized in a variety of applications including to assist in medical diagnosis. Neural network ensembles can significantly improve the generalization ability of learning systems through training a finite number of neural networks and then combining their results. However, the performance of multiple classifiers in disease prediction is not fully understood. The major purpose of this study is to investigate the performance of different classifiers, including individual classifiers involved in an ensemble classifier and solo classifiers. In addition, we use various evaluation criteria to examine the performance of these classifiers with real-life datasets. Finally, we also use statistical testing to evaluate the significance of the difference in performance among the three classifiers. The statistical testing results indicate that an ensemble classifier performs better than an individual classifier within an ensemble. However, the solo classifier does not perform worse than the ensemble classifier built with the same size training dataset.

1. Introduction

Disease prediction is very important for medical institutions in order to make the best possible medical care decisions. Incorrect decisions are likely to cause delays in medical treatment or even loss of life. A number of disease prediction models have been proposed, including the single best model and ensemble model (Bellaachia and Guven, 2006; Delen et al., 2005; Tyrer et al., 2004).

The basic concept on which ensemble method is the best has diverse perspectives on different aspects of a problem, which can be combined to produce a better quality decision. O'Leary (1998) has suggested that knowledge acquired from groups provides more correct probability solution than that acquired from individuals. Zhou and Lopresti (1997) found that a consensus vote of multiple machine learning models trained by repeated sampling always yields a net improvement in accuracy for the common distribution of interest. These findings suggest that ensembles or collections of learning models provide more accurate problem generalization than the selection of a "single best" model determined by cross validation tests (Breiman, 1996; Zhang, 1999; Zhou et al., 2002).

* Corresponding author.
E-mail address: bmahck@ccu.edu.tw (T.C.-K. Huang).

http://dx.doi.org/10.1016/j.tele.2015.08.006

Neural network ensembles can significantly improve the generalization ability of learning systems through the training of a finite number of neural networks and then combining their results. Hansen and Salamon (1990) proposed several means for improving the performance and training of neural networks for classification and used cross-validation as a tool for optimizing the network parameters and architecture. This has been shown that the remaining residual generalization errors can be reduced by invoking ensembles of similar networks. The multitude of local minima encountered in the training of individual neural networks result in errors occurring in different regions of input space. The collective decision of the ensemble is likely less to be in error than the decision made by any of the individual networks within the ensembles of similar networks. Moreover, neural network ensembles have been applied successfully to real-life applications, such as medical diagnosis (Zhou et al., 2002).

Recent related strategies using the artificial neural network (ANN) technique for disease prediction are compared in Table 1. However, these strategies discussed in the literature have the following limitations:

(1) Most studies only use one dataset for system validation which may not be reliable enough to make a conclusion. It is necessary to consider a certain number of different datasets for validation.
(2) Most studies only examine the average prediction performance of their models without considering the *F*-score and precision.
(3) Most studies do not provide statistical test results to demonstrate the level of significance of their experimental results.
(4) Most studies related to ensemble classifier do not compare the performance difference between individual classifiers and an ensemble classifier consisted of individual classifiers.

In this study, we use ANN to build three types of classifiers, the Individual Classifier (IC), Ensemble Classifier (EC), and Solo Classifier (SC), forming a research framework as shown in Fig. 1. The differences between IC, EC, and SC are described below. First, both IC and SC are single classifiers; yet, the size of the training dataset for SC is about several times, which is larger than that for IC. Second, EC consists of several ICs, making it a multiple-type classifier.

We investigate the performance of different classifiers, including an individual classifier (as part of an ensemble classifier), ensemble classifier (multiple classifiers), and solo classifier. In addition, we attempt to adopt various evaluation criteria, such as accuracy, precision, true positive rate (TPR), true negative rate (TNR), and *F*-score, to examine the performance of these classifiers with real-life datasets. Finally, statistical tests are employed to evaluate the significance of difference in performance among the three classifiers. The following research questions related to the prediction of diseases are also addressed:

(1) Is there a significant difference in *accuracy* between the ensemble classifier and individual classifier?
(2) Does the ensemble classifier outperform the individual classifier (as part of an ensemble classifier)?

Table 1
Comparison of related works.

Work	Classifier	Datasets	Evaluation methods				
			Accuracy	TPR	TNR	*F*-score	Statistics test
Temurtas et al. (2009)	ANN + LM	Pima-diabetes	Yes	No	No	No	No
Babaoglu et al. (2009)	Ensemble ANN	Coronary artery	Yes	Yes	Yes	No	No
Übeyli (2009)	Combined ANN	Psoriasis/Seboreic/...	Yes	Yes	Yes	No	No
Åström and Koker (2011)	Parallel ANN	Parkinson	Yes	Yes	Yes	No	No
Pan et al. (2012)	SVM, ANN	Parkinson	Yes	Yes	Yes	No	No
Noia et al. (2013)	Ensemble ANN	End stage kidney	Yes	Yes	Yes	Yes	No

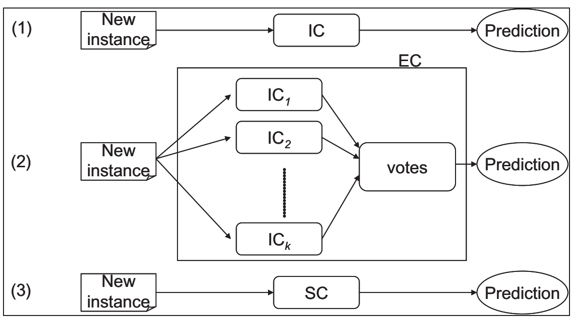

Fig. 1. The proposed framework.