Neural Network Approach to Incomplete Data Applied to Assessing Cardiac Health

Joannna Grabska-Chrząstowska

Laboratory of Biocybernetics, Department of Automatics and Biomedical Engineering,
AGH University of Science and Technology, Krakow, Poland

DATABASE PARAMETERS

#1	age	in years
#2	sex	(1 = male; 0 = female)
#3	cp	chest pain type -- Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
#4	trestb	resting blood pressure (in mm Hg on admission to the hospital)
#5	chol	serum cholestoral in mg/dl
#6	fbs	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
#7	restecg	resting electrocardiographic results -- Value 0: normal --Value 1: having ST-T wave abnormality -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
#8	thalach	maximum heart rate achieved
#9	exang	exercise induced angina (1 = yes; 0 = no)
#10	oldpeak	ST depression induced by exercise relative to rest
#11	slope	slope: the slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
#12	ca	number of major vessels (0-3) colored by flourosopy
#13	thal	3 = normal; 6 = fixed defect; 7 = reversable defect
		num: diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing (in any major vessel: attributes 59 through 68 are vessels)
#14	output	the presence of heart disease in the patient 0 absence 1 presence

Table 1. Databases cardinality.

Database name	All	Healthy	Sick
Cleveland (reference database)	**297**	160	137
V.A.	**200**	51	149
Hungarian	**294**	188	106
Switzerland	**123**	8	115

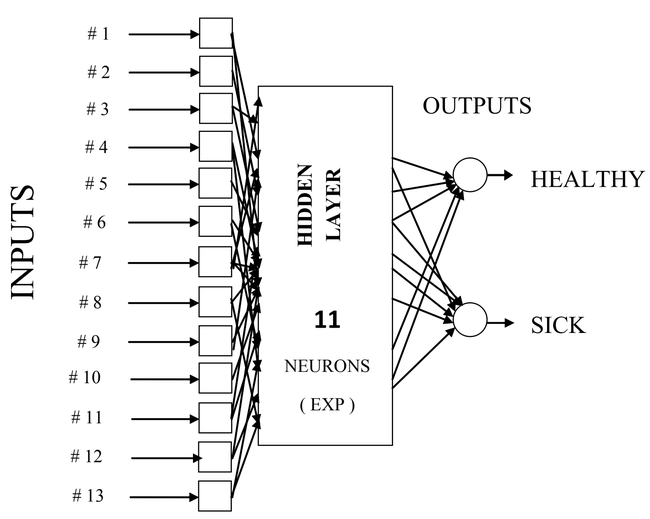
archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.switzerland.data

```
32,1,1,95,0,?,0,127,0,.7,1,?,?,1
34,1,4,115,0,?,?,154,0,.2,1,?,?,1
35,1,4,?,0,?,0,130,1,?,?,?,7,3
36,1,4,110,0,?,0,125,1,1,2,?,6,1
38,0,4,105,0,?,0,166,0,2.8,1,?,?,2
38,0,4,110,0,0,0,156,0,0,2,?,3,1
38,1,3,100,0,?,0,179,0,-1.1,1,?,?,0
38,1,3,115,0,0,0,128,1,0,2,?,7,1
38,1,4,135,0,?,0,150,0,0,?,?,3,2
38,1,4,150,0,?,0,120,1,?,?,?,3,1
40,1,4,95,0,?,1,144,0,0,1,?,?,2
41,1,4,125,0,?,0,176,0,1.6,1,?,?,2
42,1,4,105,0,?,0,128,1,-1.5,3,?,?,1
42,1,4,145,0,0,0,99,1,0,2,?,?,2
43,1,4,100,0,?,0,122,0,1.5,3,?,?,3
```

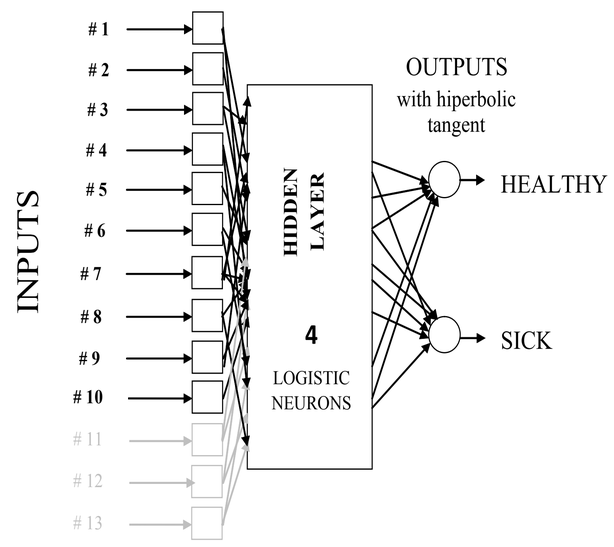
Application of genetic algorithm for all 13 input parameters - #8, #9, #12 and #13 are most important inputs (especially #8 -maximum heart rate)

Databases analisis - #11, #12, #13 are missing in most of data

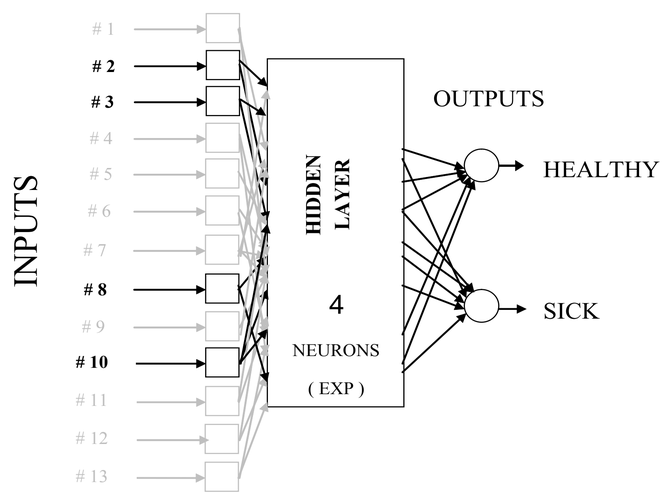
Neural network with all parameters



Neural network with 10 parameters



Neural network with only four parameters



Results for Clevland and other databases for the best classification networks

Database name	Quality [%]	Sensitivity [%]	Specificity [%]
Cleveland (reference database)	**89,6**	86,9	91,9
V.A.	**47,5**	33,6	88,2
Hungarian	**72,1**	29,2	96,3
Switzerland	**56,9**	55,7	75

Database name	Quality [%]	Sensitivity [%]	Specificity [%]
Cleveland (reference database)	**82,5**	79,6	85
V.A.	**67,5**	85,2	15,7
Hungarian	**76,9**	77,4	76,6
Switzerland	**74,8**	76,5	50

Database name	Quality [%]	Sensitivity [%]	Specificity [%]
Cleveland (reference database)	**80,8**	78,8	82,8
V.A.	**76**	96,6	15,7
Hungarian	**75,2**	83	70,7
Switzerland	**88,6**	91,3	50

Results comparison

Quality for different numbers of input parameters [%]

Numbers of param.	13	10	**4**	3
Database name	all	1- 10	**2,3,8,10**	2,3,8
Cleveland (reference database)	89,6	82,5	**80,8**	79,1
V.A.	47,5	67,5	**76**	73,5
Hungarian	72,1	76,9	**75,2**	74,1
Switzerland	56,9	74,8	**88,6**	87

CONCLUSIONS

The paper shows a general way of behavior in the case of classification in databases with a lot of parameters and a lot of missing data. It was shown that neural networks, apart from generally known benefits as classifiers, have big possibilities of classification with a limited number of inputs on condition that the user has a full reference database at his disposal (received e.g. after removing incomplete cases). In the case of cardiac data, after using the above presented method, it was possible to achieve a high ability of binary classification by a network with 4 (four) inputs. In the future one may aim at classifying heart degradation in the whole group of the sick (dividing the second class into four).