HAL open science

Artificial Neural Networks to Investigate the Importance and the Sensitivity to Various Parameters Used for the Prediction of Chromosomal Abnormalities

Andreas C. Neocleous, Kypros H. Nicolaides, Argyro Syngelaki, Kleanthis C. Neokleous, Gianna Loizou, Costas K. Neocleous, Christos N. Schizas

To cite this version:

Andreas C. Neocleous, Kypros H. Nicolaides, Argyro Syngelaki, Kleanthis C. Neokleous, Gianna Loizou, et al.. Artificial Neural Networks to Investigate the Importance and the Sensitivity to Various Parameters Used for the Prediction of Chromosomal Abnormalities. 8th International Conference on Artificial Intelligence Applications and Innovations (AIAI), Sep 2012, Halkidiki, Greece. pp.46-55, 10.1007/978-3-642-33412-2_5 . hal-01523076

HAL Id: hal-01523076

https://hal.science/hal-01523076

Submitted on 16 May 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Artificial Neural Networks to Investigate the Importance and the Sensitivity to Various Parameters used for the Prediction of Chromosomal Abnormalities

Andreas C. Neocleous[1], Kypros H. Nicolaides[2], Argyro Syngelaki[2], Kleanthis C. Neokleous[1], Gianna Loizou[1], Costas K. Neocleous[3], and Christos N. Schizas[1]

[1] Department of Computer Science, University of Cyprus, 1 University Avenue, P. O. Box 20537, Nicosia 1678, CYPRUS.
[2] The Harris Birthright Research Centre for Fetal Medicine, King's College Hospital Medical School, Denmark Hill, SE5 8RX, London, UNITED KINGDOM.
[3] Department of Mechanical Engineering, Cyprus University of Technology, 31 Archbishop Kyprianou, P. O. Box 50329, Lemesos 3036, CYPRUS.

Abstract. A selection of artificial neural network models were built and implemented for systematically study the contribution and the sensitivity of the main influencing parameters as important contributing factors for the non-invasive prediction of chromosomal abnormalities. The parameters that had been investigated are: the previous medical history of the pregnant mother, the nasal bone, the tricuspid flow, the ductus venosus flow, the PAPP-A value, the b-hCG value, the crown rump length (CRL), the changes in nuchal translucency (deltaNT) and the mother's age. The main conclusions drawn are: 1) The deltaNT is the most significant factor for the overall prediction, while the CRL the least significant. 2) The previous medical history of the pregnant mother is not a significant factor for the prediction of the abnormal cases. 3) The nasal bone, the tricuspid flow and the ductus venosus flow contribute significantly in the prediction of trisomy 21 but not in the prediction of the "normal" cases. 4) The PAPP-A, the b-hCG and the mother's age are of intermediate importance. Also, a sensitivity analysis of the attributes PAPP-A, b-hCG, CRL, deltaNT and of the mother's age was done. This analysis showed that the CRL and deltaNT are more sensitive when their values are decreased, the PAPP-A is more sensitive when its values are increased and the b-hCG is insensitive to variations in its values.

Keywords: Artificial neural networks, Chromosomal abnormalities, Sensitivity analysis.

1 Introduction

It is known that for an effective non-invasive first-trimester screening for fetal chromosomal abnormalities, various maternal and feto-placental parameters may be explored in an appropriate diagnostic tool ([1], [2], [7]). The most important parameters for diagnosing chromosomal abnormalities are the maternal age (MA), the previous occurrence of aneuploidies, the fetal nuchal translucency (NT) thickness, the crown rump length (CRL), the free β-human chorionic gonadotrophin (b-hCG), the

pregnancy associated plasma protein-A (PAPP-A), the nasal bone (NB), the tricuspid flow (TF) and the ductus venosus flow (DV) [2].

The traditional approach to screening for trisomy 21 (T21) (Down Syndrome), is to use statistical techniques to estimate the patient-specific risk for aneuploidy.

In a previous study [7], an accurate system for the non-invasive identification of trisomy 21 during the first three months of the pregnancy was suggested. That system was tested with 129 unknown cases of T21 and 16,898 unknown cases of normal fetuses. The correct classification yield was 99% for the unknown T21 cases and the false positive rate was 5%.

In the present study, artificial neural networks (ANN) had been constructed, trained and verified with a large unknown data set in order to explore the importance and sensitivity of various contributing parameters for the appraisal of the risk for existence of chromosomal abnormalities in the fetuses.

The data were obtained from the Fetal Medicine Foundation (FMF), which is a UK registered charity that has established a process of training and quality assurance for the appropriate introduction of NT screening into clinical practice [2].

In the study, a large number of different neural network structures had been constructed and trained in a systematic manner, aiming at producing good neural classifiers/predictors for the non-invasive appraisal of the risk for the presence of chromosomal abnormalities in fetuses and hence to establish the importance and sensitivities to those parameters. The neural structures that were attempted were mainly of the feed-forward type, both of standard multi-layer, as well as of multi-slab topologies.

2 Data

The data were provided by the Fetal Medicine Foundation of London. They were obtained from the greater London area and South-East England for pregnant women attending routine clinical and ultrasound assessment for the risk of chromosomal abnormalities.

The database was made of 50,890 cases of pregnant women. The vast majority of these (50,489) were normal as far as the chromosomal abnormalities are concerned. The remaining 401 cases (0.8%) were confirmed as having trisomy 21. That is, there is a prevalence of 0.8%. This is a highly unbalanced data set that makes the prospect for building an effective neural network predictor to be a difficult task.

For each pregnant woman, a number of relevant parameters were collected, encoded/converted into appropriate numbers that made them suitable to be used for the training of neural networks. These parameters are:

1. The maternal age
2. The previous medical history
3. The fetal nuchal translucency thickness
4. The crown rump length
5. The free β-human chorionic gonadotrophin
6. The pregnancy associated plasma protein-A
7. The nasal bone
8. The tricuspid flow
9. The ductus venosus flow

The fetal CRL, the NT thickness, and the fetus heart rate were measured during a transabdominal ultrasound examination that was performed by certified sonographers

who had received the appropriate FMF certificates of competence. During such examination other possible major fetal defects were also diagnosed and recorded.

A subset of 16,807 cases (33%) were isolated and kept aside to be used as a totally unknown database to be used in order to check the predictability of each attempted neural network, and later for the evaluation of the importance of each factor as significant parameter that contributes to an accurate prediction of the risk of occurrence of the genetic abnormality of interest. In this unknown data set, there were 130 cases (0.7%) of T21 chromosomal defect. It is emphasized that these cases were never used during the learning procedures of training of the neural networks, and thus, they were a reliable way for assessing the predictability of each network. Because the number of anomalous cases is very small, and such cases cannot be artificially generated, the abnormal cases in the verification set were confined to only a small, but substantial, percentage.

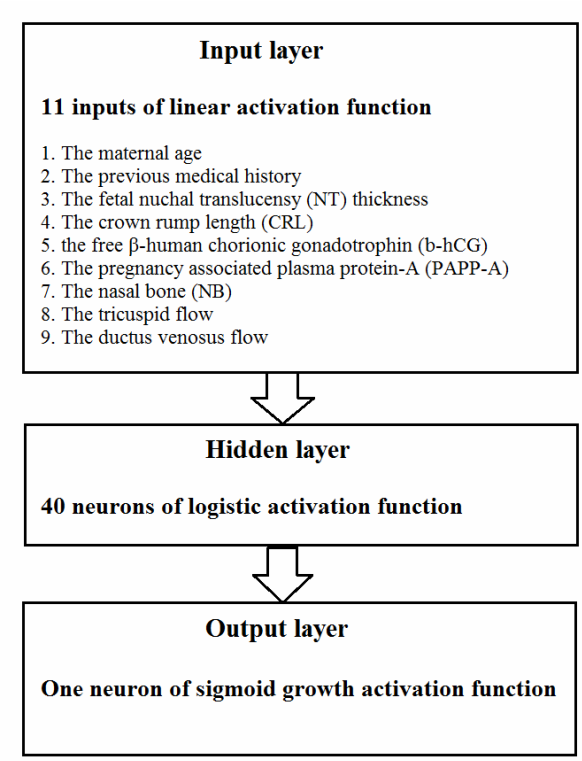

Fig. 1. The typical neural network structure that was ultimately selected and used for the diagnosis of chromosomal defects.

3 The neural network predictor

Various feed-forward neural structures of standard multilayer type, having different number of layers and activations, as well as different neurons per layer were systematically built, trained and tested. Also, multi-slab topologies of different structures, sizes, and activation functions, were systematically built, trained and verified, in order to find the best performing structure with regards to the prediction of

the unknown verification data set. This was done in a planned and systematic manner so that the best performing architecture would be obtained and finally used. Tables 5 and 6 show a summary of the attempted structures and of the performances achieved for the unknown verification data set of the 16,807 cases.

The best performing neural structure that was found for this particular application is a feed-forward multilayer perceptron having three layers as it is typically depicted in Figure 1. All the weights were initialized to 0.1, the learning rate used for all the connections was 0.1, and the momentum rate was 0.4 for all the links. The learning scheme that was used was the momentum backpropagation.

4 Methodology

Each of the parameters was independently and individually studied. The methodology used is as follows.

For the case of the investigation of the importance of the previous medical history, 60 models were built and tested. The vector of the attributes contained three elements that are related to the history of the patient, labeled “Previous T21”, “Previous T18” and “Previous T13”, indicating a specific chromosomal abnormality that occurred in a previous pregnancy. The contribution of these three attributes was examined by creating two groups of ANN models. In the first group the entire attribute set was used, while in the other group we removed the three attributes related to the patient’s history and we compared the results.

As for the investigation of the attributes that contained information retrieved from ultrasound scans, i.e the presence or absence of the nasal bone, the tricuspid flow and the ductus venosus flow, we created 8 groups of a total of 48 models. In each group we removed one or two of the attributes that we were testing (8 combinations). The first group contained models that were built using the entire attribute set (Table 1).

In a third series of experiments we examined the contribution and the sensitivity of “Mother’s age”, “b-hCG”, “PAPP-A”, “CRL” and “deltaNT”. The contribution of these five attributes was examined by creating 6 models and comparing the results of each model with the results of the same model when the entire attribute set was used as input vector in the network. In each of the 5 models, one attribute was removed from the attribute set and in the 6th model all the five attributes were removed. The sensitivity of the attributes was tested by changing the values of each parameter in the evaluation set by a specific factor. Thus, an artificial evaluation set was created having a vector that its values are changed and is used to test the model that was initially trained with the correct values. We expect to see more misclassified cases since one vector at each time was multiplied by a factor in the range [0.001 - 10]. We observed the importance on the tolerance of a possible error for each of the attributes that were tested.

All the neural models that were built had three layers. An input layer, a hidden layer and an output layer. The number of neurons in the input layer was always the number of the attributes used as input parameters to the network. The output layer had always one neuron and the transfer function was the sigmoid. The difference in the architecture of each one of the models is in the number of neurons and the activation function used in the hidden layer. We used two activation functions and for each activation function we built models with 40, 50 and 60 neurons.

Table 1. Eight groups of models to investigate the contribution of the attributes "Nasal bone", "Tricuspid flow" and "Ductus venosus".

GROUP	**Mother's Age**	**Previous T21**	**Previous T18**	**Previous T13**	**CRL**	**PAPP-A**	**b-hCG**	**Nasal bone**	**Trcuspid flow**	**Ductus venosus**	**DeltaNT**
A	√	√	√	√	√	√	√	√	√	√	√
B	√	√	√	√	√	√	√	√	√		√
C	√	√	√	√	√	√	√		√	√	√
D	√	√	√	√	√	√	√	√		√	√
E	√	√	√	√	√	√	√	√			√
F	√	√	√	√	√	√	√		√		√
G	√	√	√	√	√	√	√			√	√
H	√	√	√	√	√	√	√				√

5 Results

From the results obtained for each group we created a decision procedure composed of two steps, aiming to assign a quality value in the comparison between the groups. That is, we needed to know if a model increases or decreases its performance by removing one attribute. If the system increases its performance after the removal of an attribute, then we concluded that the specific attribute rather confuses the predictive system instead of contributing to a better diagnostic yield.

In the first step of the decision procedure, the results of each model in groups B to H are compared with the results of the models in group A. For each group 6 models were built using different architecture schemes of artificial neural networks. The results of the models in group A were kept as reference results since these were built by using the entire attribute set. Each time a comparison is done, a plus (+) or a minus (-) is added in the vector of six bin length (six models) representing the result of the comparison. In the second step of the decision procedure, we counted how many models increased and how many decreased their performance in comparison with the models in group A, by simply counting the pluses and the minuses in the vector, and divided this number by 6, which is the total number of the models in each group. In the case that the number of the models that have higher prediction compared to the models in group A is equal or higher to 4, the attribute is considered as contributing positively.

We additionally computed the average of the increment and decrement of the performance of the models. This value gives supplemental information on the power of the attribute that is tested. For example the attribute "nasal bone" proved to be very important since all of the models that were built without the "nasal bone" in the input vector had a reduced performance compared with the models that were built using the "nasal bone" in the input vector, and the average of the decrement in the performance is above 2% in the prediction of the T21 cases (Table 5).

5.1 Importance of the attributes of "previous medical history"

We examined the contribution of the three attributes "Previous T21", "Previous T18" and "Previous T13" by creating 12 models and comparing results. The results are shown in Table 2.

Table 2. Results on 12 models for the investigation of the importance of the previous medical history. The performance of the normal cases were kept the same for each pair and the performance of the prediction of the T21 cases is examined. The last row shows the difference between the performance of each model in group B and the respective model in group A.

Model	**Removed**	**Model 1**		**Model 2**	
		Normal	T21	Normal	T21
A	None	95.20%	96.92%	95.09%	96.92%
B	History	95.12%	96.15%	95.01%	96.92%
Difference B-A			-0.77%		0.00%
Model	**Removed**	**Model 3**		**Model 4**	
		Normal	T21	Normal	T21
A	None	95.30%	95.38%	97.00%	93.08%
B	History	95.54%	97.69%	97.00%	94.62%
Difference B-A			2.31%		1.54%
Model	**Removed**	**Model 5**		**Model 6**	
		Normal	T21	Normal	T21
A	None	96.67%	92.31%	96.62%	93.08%
B	History	96.62%	90.77%	96.23%	90.77%
Difference B-A			-1.54%		-2.31%

The models in group A were built using the entire attribute set, while in group B we built 6 models in which we removed the three attributes under consideration from the input vector. We present the results by keeping the performance of the normal cases to be nearly the same between the reference model and the comparison model. This was done by changing the threshold in the classification process. A higher threshold will give a higher prediction in the normal cases and lower prediction on the T21 cases. In models 1, 2 and 3 we keep the prediction of the normal cases at around 95.0%, in models 5 and 6 we keep the performance of the normal cases at around 96.6% and in model 4 at 97.0%. In this way we compare the differences of the performance on the prediction of the T21 cases. In models 1, 5 and 6 the performance in the prediction of the T21 decreased by 0.77% on average and in models 3 and 4 the performance in the prediction of the T21 increased by 0.64% on average (Table 3).

Table 3. Results on 12 models for the investigation of the importance of previous medical history. First row shows the percentage on the models that had an increment, decrement and no change in the performance after the removal of the attributes that were tested. The last row shows the average increment/decrement in the performance.

Removed		**Increased**	**Decreased**	**No diff**
History attributes	Sum of 6 models	33%	50%	17%
	Performance (avg)	0.64%	0.77%	

In model 2 the performance had no change in the prediction of the T21. From the results we conclude that the attribute "previous medical history" contributes positively, since 50% of the models built had higher prediction when the attribute was used in contrast to 33% of the models that had lower prediction when the attributes were not used.

5.2 Importance of the attributes "nasal bone", "tricuspid flow" and "ductus venosus flow"

The procedure followed in this experiment aims to investigate the importance of three attributes in the dataset. We examined the contribution of the "ductus venosus flow" (DV), "tricuspid flow" (TF) and "nasal bone" (NB), by creating 48 models and comparing the results. The results of the models are shown in Table 4.

Table 4. Results on 48 models for the investigation of the importance of the attributes "ductus venosus", "nasal bone" and "tricuspid flow".

Renoved		Increased	Decreased	No difference
DV	Sum of 6 models	33%	33%	33%
	Average performance	0.51%	0.77%	
NB	Sum of 6 models	0%	100%	0%
	Average performance	0.00%	2.56%	
TF	Sum of 6 models	83%	0%	17%
	Average performance	1.67%	0.00%	
DV + TF	Sum of 6 models	0%	100%	0%
	Average performance	0.00%	2.31%	
DV + NB	Sum of 6 models	83%	0%	17%
	Average performance	2.44%	0.00%	
NB + TF	Sum of 6 models	50%	0%	50%
	Average performance	0.77%	0.00%	
DV + TF + NB	Sum of 6 models	17%	83%	0%
	Average performance	0.13%	2.44%	

In the first row, a comparison between the models that used the entire attribute set and the models in which the "ductus venosus" was removed is presented. Similarly, in rows 2 and 3 the "Nasal bone" and the "tricuspid flow" are compared with the models that used the entire attribute set. We observe that the attribute "ductus venosus" is contributing by an average 0.77%, the attribute "nasal bone" by 2.56% while the "tricuspid flow" is confusing the system since when removed, the performance increased by 1.67%. In the case we removed the "ductus venosus flow" and the "tricuspid flow" the performance decreased by 2.31% even though we observed that the "tricuspid flow" confuses the system. In the 5th row of Table 4 we observe that even though the attributes "DV" and the "NB" are contributing to the performance, when both are removed, the performance has increased. It seems that "NB" without "DV" gives better results. In the 6th row, the "NB" and the "TF" were removed from the attribute set. We observe that the performance increased by 0.77% and this is probably because as we see at the third row where the "TF" is examined, this attribute is confusing the system. When we removed all of the three attributes, the performance decreased by 2.44%.

5.3 Importance of the attributes "Mother's age", "b-hCG", "PAPP-A", "CRL" and "deltaNT"

In this section, the importance and the sensitivity of the attributes "Mother's age", "b-hCG", "PAPP-A", "CRL" and "deltaNT" will be examined.

For an investigation of the importance of these attributes we built six models and we compared them with the model that was built using the entire attribute set. In each one of the models, an attribute was removed. The results of the models are shown in Table 5. At a first look we observe that all the attributes that were examined are

contributing to the system. More specific, when we removed all the five attributes the performance on the prediction of the T21 cases dropped from 96% to 81%. The most important attribute based on our investigations is the "deltaNT" since when it was removed from the system, the performance of the T21 dropped from 96% to 76%. The rest of the four attributes are contributing by approximately 6% each. The least significant attribute is found to be the "Mother's age".

Table 5. Importance of the attributes "Mothers age", "b-hCG", "PAPP-A", "CRL" and "deltaNT".

Removed attributes	Performance	
	Normal	**T21**
None	95.78%	96.15%
b-hCG, PAPP-A, CRL, deltaNT, Age	95.75%	80.90%
b-hCG	95.60%	90.77%
CRL	95.81%	90.77%
deltaNT	98.14%	76.15%
PAPP-A	95.74%	91.54%
Age	95.70%	93.85%

We also examined the sensitivity of each of the four attributes "CRL", "PAPP-A", "b-hCG" and "deltaNT". To test the sensitivity of each attribute we changed the values of the attribute we wanted to check with wrong values (multiply the vector by a factor in the range [0.001 – 10]) and we were re-evaluating the evaluation set. A big drop in the performance shows high sensitivity of the attribute that was tested.

We followed similar comparison procedure with the ones in previous sections by keeping the performance of the normal cases to be closely the same between the models we were comparing and we were observing differences in the performance on the T21 cases. Five erroneous vectors were created for each one of the attributes we were testing. This approach resulted in 20 evaluation sets. The five factors that were multiplied in each attribute are the numbers 10, 5, 2, 0.1, and 0.001. As shown in Table 6, the attribute CRL is not sensitive when the number is increased and is also shown that when we multiplied the vector by 10, the results are still the same with the results of the model that the values were correct. It is also shown that the attribute "CRL" is very sensitive when the number is decreased. We see that when we multiplied the vector by 0.1 the performance on the T21 dropped to 28%. On a contrary, "PAPP-A" is more sensitive when the values are being increased and less sensitive to a decrement. As shown in Table 6 in the first row esults, when the "PAPP-A" has been multiplied by a factor of ten, the performance on the T21 dropped to 22%, while even when it was multiplied by a factor of 0.001 the performance in the T21 was the same as with the cases evaluated with the correct values. The attribute "b-hCG" is the least sensitive. The performance when we multiplied the vector with a factor of 10 and a factor of 0.001 dropped from 96% to 84%. The "deltaNT" showed that it is more sensitive when the values are decreased where when the vector is multiplied by 0.001 the performance on the T21 dropped to 72%.

Table 6. Sensitivities of the attributes "CRL", "PAPP-A", "b-hCG" and "Delta NT".

Factor	CRL		PAPP-A		b-hCG		Delta NT	
	Normal	T21	Normal	T21	Normal	T21	Normal	T21
10	95.43%	95.38%	95.07%	22.31%	95.57%	83.85%	95.72%	83.85%
5	95.48%	95%	95.57%	51.54%	95.51%	95%	95.27%	87.67%
2	94.90%	98.46%	95.21%	93.85%	94.81%	96.15%	95.27%	93.08
0.1	99.69%	28.46%	95.35%	93.85%	95.54%	86%	95.91%	83.08%
0.001	99.76%	25.38%	95.42%	96.15%	95.53%	83.85%	95.32%	71.54%

6. Conclusions

The main conclusion of this research work is that the "previous medical history attributes" are not significant contributors for the diagnosis of chromosomal abnormalities. It is not clear whether eventually three previous medical history attributes are contributing positively, since some of the models had higher performance in the prediction of the T21 when these attributes were removed from the input space. We also found that the most important attribute is the "deltaNT". This attribute contributes to the prediction of the T21 by 20%. The attributes "b-hCG", "PAPP-A" and "CRL" are also contributing to the system performance, since each one was tested separately and had an increment in the performance of about 6%. The attribute "mother's age" contributed by only 3% in the prediction of the T21 cases. The attributes "nasal bone" and "ductus venosus" had contributed by 2.6% and 0.8% respectively in the prediction of the T21. Quite importantly, the parameter "tricuspid flow" confused the system and it decreased the performance by 1.7%. The attribute "CRL" is sensitive to errors lower to 10% of its value, the "PAPP-A" is sensitive to errors higher than 10% if its value and the attributes "b-hCG" and "deltaNT" are less sensitive to either lower or higher of the 10% of their values.

Acknowledgments

The FMF foundation. This is a UK registered charity (No. 1037116).
The University of Cyprus International Affairs Committee.
The Cyprus University of Technology.

References

1. Nicolaides, K.: Nuchal Translucency and Other First-trimester Sonographic Markers of Chromosomal Abnormalities. Am. J. Obstet. Gynecol, 1191:45-67 (2004a).
2. Nicolaides, K.: The 11-13+6 Weeks Scan. Fetal Medicine Foundation, London (2004b).
3. Tourassi, G., Floyd, C., Lo, J.: A Constraint Satisfaction Neural Network for Medical Diagnosis. Neural Networks, vol. 5 (1999).
4. Brause, R.: Medical Analysis and Diagnosis by Neural networks. Computer Science Department, Frankfurt am Main, Germany (2001).
5. Patel J., Goyal R.: Applications of Artificial Neural Networks in Medical Science. Curr. Clin. Pharmacol. 2(3):217-26 (2007).
6. Neocleous, C., Anastasopoulos, P., Nicolaides, K., Schizas, C., Neokleous, K., Neural Networks to Estimate the Risk for Preeclampsia Occurrence. Proceedings of the International Joint Conference on Neural Networks, Atlanta, USA (2009).
7. Neocleous, C., Nikolaides, K., Neokleous, K., Schizas, C.: Artificial Neural Networks for Non-invasive Chromosomal Abnormality Screening of Fetuses. Proceedings of the WCCI 2010. Barcelona, Spain (2010).