

Prediction of Postpartum Depression Using Multilayer Perceptrons and Pruning

S. Tortajada¹; J. M. García-Gómez¹; J. Vicente¹; J. Sanjuán²; R. de Frutos²; R. Martín-Santos³; L. García-Esteve³; I. Gornemann⁴; A. Gutiérrez-Zotes⁵; F. Canellas⁶; Á. Carracedo⁷; M. Gratacos⁸; R. Guillamat⁹; E. Baca-García¹⁰; M. Robles¹

¹IBIME, Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universidad Politécnica de Valencia, Valencia, Spain;

²Faculty of Medicine, Universidad de Valencia, Valencia CIBERSAM, Spain;

³IMIM-Hospital del Mar and ICN-Hospital Clínic, Barcelona CIBERSAM, Spain;

⁴Hospital Carlos Haya, Málaga, Spain;

⁵Hospital Pere Mata, Reus, Spain;

⁶Hospital Son Dureta, Palma de Mallorca, Spain;

⁷National Genotyping Center, Hospital Clínico, Santiago de Compostela, Spain;

⁸Center for Genomic Regulation, CRG, Barcelona, Spain;

⁹Hospital Parc Tauli, Sabadell, Spain;

¹⁰Hospital Jiménez Díaz, Madrid CIBERSAM, Spain

Keywords

Multilayer perceptron, neural network, pruning, postpartum depression

Summary

Objective: The main goal of this paper is to obtain a classification model based on feed-forward multilayer perceptrons in order to improve postpartum depression prediction during the 32 weeks after childbirth with a high sensitivity and specificity and to develop a tool to be integrated in a decision support system for clinicians.

Materials and Methods: Multilayer perceptrons were trained on data from 1397 women who had just given birth, from seven Spanish

general hospitals, including clinical, environmental and genetic variables. A prospective cohort study was made just after delivery, at 8 weeks and at 32 weeks after delivery. The models were evaluated with the geometric mean of accuracies using a hold-out strategy.

Results: Multilayer perceptrons showed good performance (high sensitivity and specificity) as predictive models for postpartum depression.

Conclusions: The use of these models in a decision support system can be clinically evaluated in future work. The analysis of the models by pruning leads to a qualitative interpretation of the influence of each variable in the interest of clinical protocols.

factors have been suggested concerning its etiology. For instance, social support, partner relationships and stressful life events related to pregnancy and childbirth [4], as well as neuroticism [5] have all been pointed out as being important. With respect to biological factors, it has been shown that inducing an artificial decrease in estrogen can cause depressive symptoms in patients with PPD antecedents. Cortisol alteration, thyroid hormone changes and a low rate of prolactin are also relevant factors [6]. Treloar et al. conclude in [7], a comparative study with twin samples, that genetic factors would explain 40% of variance in PPD predisposition. In Ross et al. [8], a biopsychosocial model for anxiety and depression symptoms during pregnancy and the PPD period has been developed using structural equations. However, most of the research studies involving genetic factors are separate from those involving environmental factors. There is a remarkable exception that explains that a functional polymorphism in the promoter region of the serotonin transporter gene seems to moderate the influence of stressful life events on depression [9].

An early prediction of PPD may reduce the impact of the illness on the mother, and it can help clinicians to give appropriate treatment to the patient in order to prevent depression. The need for a prediction model rather than a description model is of paramount importance. Thus, artificial neural networks (ANN) have a remarkable ability to characterize discriminating patterns and de-

Correspondence to:

Salvador Tortajada
IBIME-Itaca
Universidad Politécnica de Valencia
Camino de Vera s/n
CP 46022 Valencia
Spain
E-mail: vesaltor@upvnet.upv.es

Methods Inf Med 2009; 48: 291–298

doi: 10.3414/ME0562

received: May 15, 2008

accepted: December 8, 2008

prepublished: March 31, 2009

1. Introduction

Postpartum depression (PPD) seems to be a universal condition with equivalent prevalence (around 13%) in different countries [1, 2] which implies an increase in medical care costs. Women suffering from PPD feel a considerable deterioration of cognitive and emotional functions that can affect mother-infant

attachment. This may have an impact on the child's future development until primary school [3]. The identification of women at risk of developing PPD would be of significant use to clinical practice and would enable preventative interventions to be targeted at vulnerable women.

Multiple studies have been carried out on PPD. Several psychosocial and biological risk

rive meaning from complex and noisy data sets. They have been widely applied in general medicine for differential diagnosis, classification and prediction of disease, and condition prognosis. In the field of psychiatric disorders, few studies have used ANNs despite their predictive power. For instance, ANNs have been applied to the diagnosis of dementia using clinical data [10] and more recently for predicting Alzheimer's disease using mixed effects neural networks [11]. EEG data from patients with schizophrenia, obsessive-compulsive disorder and controls has been used to demonstrate that an ANN was able to correctly classify over 80% of the patients with obsessive-compulsive disorder and over 60% of the patients with schizophrenia [12]. In Jefferson et al. [13], evolving neural networks overcome statistical methods in depression prediction after mania. Berdia and Metz [14] have used an ANN to provide a framework for understanding some of the pathological processes in schizophrenia. Finally, Franchini et al. [15] have applied these models to support clinical decision making for the treatment of psychopharmacological therapy.

One of the main goals of this paper is to obtain a classification model based on feed-forward multilayer perceptrons in order to predict PPD with high, well-balanced sensitivity and specificity during the 32 weeks after childbirth and using pruning methods to obtain simple models. This study is part of a large research project about the environment genetic interaction in postpartum depression [16]. These models can be used later in a decision support system [17] to help clinicians in the prediction and treatment of PPD. A secondary goal is to find and interpret the qualitative contribution of each independent variable in order to obtain clinical knowledge from those pruned models.

2. Materials and Methods

Data from postpartum women were collected from seven Spanish general hospitals, in the period from December 2003 to October 2004 on the second to third day after delivery. All the participants were Caucasian, none of them were under psychiatric treatment during pregnancy, and all of them were able to read and answer the clinical questionnaires.

Women whose children died after delivery were excluded. This study was approved by the Local Ethical Research Committees, and all the patients gave their informed written consent.

Depressive symptoms were assessed with the total score of the Spanish version of the Edinburgh Postnatal Depression Scale (EPDS) [18] just after delivery, at week 8 and week 32 after delivery. Major depression episodes were established using first the EPDS (cut-off point of 9 or more) at 8 or 32 weeks, and then probable cases (EPDS ≥ 9) were evaluated using the Spanish version of the Diagnostic Interview for Genetics Studies (DIGS) [19, 20] adapted to postpartum depression in order to determine if the patient was suffering a depression episode (positive class) or not (negative class). All the interviews were conducted by clinical psychologists with previous common training in the DIGS with video recordings. A high level of reliability ($K > 0.8$) was obtained among interviewers.

From the 1880 women initially included in the study, 76 were excluded because they did not correctly fill out all the scales or questionnaires. With these patients, a prospective study was made just after delivery, at 8 weeks and 32 weeks after delivery. At the 8-week follow-up, 1407 (78%) women remained in the study. At the 32-week follow-up 1397 (77.4%) women were evaluated. We compared the loss of follow-up cases with the remainder of the final sample. Only the lowest social class was significantly increased in the loss of follow-up cases ($p = 0.005$). A total of 11.5% (160) of women at baseline, 8 weeks and 32 weeks had a major depressive episode during the eight months of postpartum follow-up. Hence, from a total number of 1397 patients, we had 160 in the positive class and 1237 in the negative class.

2.1 Independent Variables

Based on the current knowledge about PPD, several variables were taken into account in order to develop predictive models. In a first step, psychiatric and genetic information was used. These predictive models are called subject models. Then, social-demographic variables were included in the subject-environment models. For each approach, we used

EPDS (just after childbirth) as an input variable in order to measure depressive symptoms. ▶ Table 1 shows the clinical variables used in this study.

All participants completed a semistructured interview that included socio-demographic data: age, education level, marital status, number of children and employment during pregnancy. Personal and family history of psychiatric illness (*psychiatric antecedents*) and *emotional alteration* during pregnancy were also recorded. Both are binary variables (yes/no).

Neuroticism can be defined as an enduring tendency to experience negative emotional states. It is measured on the Eysenck Personality Questionnaire short scale (EPQ) [21], which is the most widely used personality questionnaire, and consists of 12 items. For this study, the validated Spanish version [22] was used. Individuals who score high on neuroticism are more likely than the average to experience such feelings as anxiety, anger, guilt and depression.

The number of experiences are the number of stressful *life events* of the patient just after delivery, at an interval of 0–8 weeks and at an interval of 8–32 weeks using the St. Paul Ramsey Scale [23, 24]. This is an ordinal variable and depends on the patient's point of view.

Depressive symptoms just after delivery were evaluated by EPDS. It is a 10-item, self-report scale, and it has been validated for the Spanish population [18]. The best cut-off of the Spanish validation of the EPDS was 9 for postpartum depression. We decided to prove its initial value (i.e., at the moment of birth) as an independent variable because the goal is to prevent and predict postpartum depression within 32 weeks.

Social support is measured by means of the Spanish version of the Duke UNC social support scale [25], which originally consists of 11 items. This questionnaire is rated just after delivery, at 6–8 weeks and at week 32. For this work, the variable used was the sum of the scores obtained immediately after childbirth plus the scores obtained in week 8. Since we wanted to predict possible depression risk during the first 32 weeks after childbirth, the Duke score at week 32 was discarded for this experiment.

Genomic DNA was extracted from the peripheral blood of women. Two functional

polymorphisms of the serotonin transporter gene were analyzed^a. For the entire machine learning process, we decided to use the combination genotypes (5-HTT-GC) proposed by Hranilovic in [26]: no low-expressing genotype at either of the *loci* (HE); low-expressing genotype at one of the *loci* (ME); low-expressing genotypes at both *loci* (LE).

The medical perinatal risk was measured as seven dichotomous variables: medical problems during pregnancy, use of drugs during pregnancy (including alcohol and tobacco), cesarean, use of anesthesia during delivery, mother medical problems during delivery, medical problems with more admission days in hospital, and newborn medical problems. A two-step cluster analysis was done in order to explore these seven binary variables. This analysis provides an ordinal variable with four values for every woman: no medical perinatal risk, pregnancy problems without delivery problems, pregnancy problems and delivery mother problems, and presence of both other and newborn problems.

Other psychosocial and demographic variables were considered in the subject-environment model such as age, the highest level of education achieved rated on a 3-point scale (low, medium, high), labor situation during pregnancy, household income rated on a 4-point scale (economical level), the gender of the baby, or the number of family members who live with the mother.

Every input variable was normalized in the range [0, 1]. Non-categorical variables were represented by one input unit. Missed variables were replaced by their mean, if they were continuous, or by their mode, if they were discrete. A dummy representation was used for each categorical variable, i.e., one unit represents one of the possible values of the variable, and this unit is activated only when the corresponding variable takes this value. Missed variables were simply represented by not activating any of the units.

2.2 ANNs Theoretical Model

ANNs are inspired by biological systems in which large numbers of simple units work in parallel to perform tasks that conventional

Table 1 There are 160 cases with postpartum depression (PPD) and 1237 cases without it. The second column shows the number of missing values for each independent variable, where ‘–’ indicates no missing value. The last two columns show the number of patients in each class. For categorical variables, the number of patients (percentage) is shown. For non-categorical variables, the mean \pm standard deviation is presented.

Input Variable	No. miss.	No PPD	PPD
Psychiatric antecedents	76		
No		790 (90.3%)	85 (9.7%)
Yes		374 (83.9%)	72 (16.1%)
Emotional alterations during pregnancy	–		
No		73 (81.1%)	17 (18.9%)
Yes		1164 (89.1%)	143 (10.9%)
Neuroticism (EPQN)	6	3.25 \pm 2.73	5.68 \pm 3.55
Life events after delivery	2	0.99 \pm 1.06	1.40 \pm 1.09
Life events at 8 weeks	176	0.88 \pm 1.09	1.69 \pm 1.33
Life events at 32 weeks	64	0.87 \pm 1.07	1.95 \pm 1.53
Depressive symptoms (Initial EPDS)	–	5.64 \pm 3.97	8.96 \pm 4.85
Social support (DUKE)	10	88.06 \pm 56.27	138.63 \pm 82.45
5-HTT-GC	79		
HE, No low-expressing genotype		93 (83.8%)	18 (16.2%)
ME, Low-expressing genotype at one locus		664 (87.5%)	95 (12.5%)
LE, Low-expressing genotype at both loci		408 (91.1%)	40 (8.9%)
Medical Perinatal Risk	–		
No problems		376 (88.1%)	51 (11.9%)
Pregnancy problems		426 (86.1%)	69 (13.2%)
Mother problems		117 (89.3%)	14 (10.7%)
Mother and child problems		318 (92.4%)	26 (7.6%)
Age	–	36.12 \pm 4.42	31.89 \pm 4.96
Educational level	2		
Low		324 (85.5%)	55 (14.5%)
Medium		518 (88.5%)	67 (11.5%)
High		393 (91.2%)	38 (8.8%)
Labor situation during pregnancy	4		
Employed		879 (91.1%)	86 (8.9%)
Unemployed		136 (86.1%)	22 (13.9%)
Student/Housewife		103 (85.1%)	18 (14.9%)
Leave		116 (77.9%)	33 (22.1%)
Economical level	17		
Suitable income		830 (90.9%)	83 (9.1%)
Enough income		311 (85.9%)	51 (14.1%)
Tight income		73 (79.3%)	19 (20.7%)
Economical problems		7 (53.8%)	6 (46.2%)
Gender of the baby	18		
Male		599 (89.7%)	69 (10.3%)
Female		623 (87.6%)	88 (12.4%)
Number of people living together	31	2.67 \pm 0.96	2.66 \pm 0.77

^a 5-HTTLPR in the promoter region and STin2 within intron 2

computers have not been able to tackle successfully. These networks are made of many simple processors (neurons or units) based on Rosenblatt's perceptron [27]. A perceptron gives a linear combination, y , of the values of its D inputs, x_i , plus a bias value,

$$y = \sum_{i=1}^D x_i w_i + \theta. \quad (1)$$

The output, $z = f(y)$, is calculated by applying an activation function to the input. Generally, the activation function is an identity, a logistic or a hyperbolic tangent. As these functions are monotonic, the form $f(y)$ still determines a linear discriminant function [28]. A single unit has a limited computing ability, but a group of interconnected neurons has a very powerful adaptability and the ability to learn non-linear functions that can model complex relationships between inputs and outputs.

Thus, more general functions can be constructed by considering networks having successive layers of processing units, with connections running from every unit in one layer to every unit in the next layer only. A feed-forward multilayer perceptron consists of an input layer with one unit for every independent variable, one or two hidden layers of perceptrons and the output layer for the dependent variable (in the case of a regression problem), or the possible classes (in the case of a classification problem). We call a fully connected feed-forward multilayer perceptron when every unit of each layer receives an input from every unit in its precedent layer and the output of each unit is sent to every unit in the next layer.

Since PPD is considered in this work as a binary dependent variable, the activation function of the output unit was the logistic function, while the activation function of the hidden units was the hyperbolic tangent.

As a first approach, fully connected feed-forward multilayer perceptrons were used with one or two hidden layers. The learning algorithm backpropagation with momentum was used to train the networks. The connection weights of the network were updated following the descent gradient rule [29].

Although these models, and ANNs in general, exhibit a superior predictive power compared to traditional approaches, they have been labeled as "black box" methods because

they provide little explanatory insight into the relative influence of the independent variables in the prediction process. This lack of explanatory power is a major concern in achieving an interpretation of the influence of each independent variable on PPD. In order to gain some qualitative knowledge of the causal relationships about depression phenomena, we used several pruning algorithms to obtain more simple and interpretable models [30, 34].

2.2.1 Pruning Algorithms

Based on the fundamental idea in Wald statistics, pruning algorithms estimate the importance of a parameter (or weight) in the model by how much the training error increases if that parameter is eliminated. Then, it removes the least relevant one and continues iteratively until some convergence condition is reached. These algorithms were initially thought of as a way to achieve a good generalization for connectionist models, i.e., the ability to infer a correct structure from training examples and to perform well on future samples. A very complex model can lead to poor generalization or overfitting, which happens when it adjusts to specific features of the training data rather than to the general ones [31]. But pruning has also been used for feature selection with neural networks [32, 33], making their operation easier to understand since there is less opportunity for the network to spread functions over many nodes. This is important in this critical application where knowing how the system works is a major concern. The algorithms used here are based on weight pruning. The strategy consists in deleting parameters with small saliency, i.e. those whose deletion will have

the least effect on the training error. The Optimal Brain Damage (OBD) algorithm [30] and its descendant, Optimal Brain Surgeon (OBS) [34], use a second-order approximation to predict the saliency of each weight.

Pruned models were obtained from fully connected feed-forward neural networks with two hidden units, i.e., there was initially a connection between every unit from a layer and every unit of each consecutive layer. In order to select the best pruned architecture, a validation set was used to compare the networks. Then, when the best model was obtained, the interpretation of the influence of each variable was done in the following way: if an input unit is directly connected to the output unit, then a positive weight means that it is a risk factor as it increases the probability of having depression. Thus, a negative weight means that the variable is a protective factor. Let a hidden unit be connected to the output unit with a positive weight. If an input unit is connected to this hidden unit with a positive value, then the variable represented by this unit is a risk factor. If its weight is negative, then it is a protective factor. On the contrary, if the weight between the hidden unit and the output unit is negative, then a positive value in the connection between the input and the hidden unit means that the variable is a protective factor. Thus, a negative value in the weight that connects the input to the hidden unit means that it is a risk factor. ▶ Table 2 summarizes these influences. This interpretation is justified because the hidden units have a hyperbolic tangent as an activation function which delimits its output activation values between -1 and 1 .

2.2.2 Comparison with Logistic Regression

The significant variables obtained by the pruned models were compared to the ones obtained by logistic regression models. The latter models are used when the dependent variable is categorical with two possible values. Independent variables may be in numerical or categorical form. The logistic function can be transformed using the logit transformation into a linear model [35]:

$$g(x) = \sum_{i=1}^n \beta_i x_i + \beta_0. \quad (2)$$

Table 2 Summary of the nature of the variables as being a risk factor or a protective factor depending on the sign of the weights of the input-hidden connection (I-H) and the hidden-output connection (H-O).

I-H	H-O	Factor
+	+	Risk
+	-	Protective
-	+	Protective
-	-	Risk

The log-likelihood is used for estimating regression coefficients (β_i) in the model. Thus, the exponential values of the regression coefficients give the odds ratio value, which reflects the effect of the input variables as a risk or a protective factor. To assess the significance of an independent variable, we compare the value of the likelihood of the model with and without the variable. This comparison follows a chi-square distribution with one degree of freedom, so it is possible to find the associated p-value.

Thus, we have the statistical significance and the character of each factor as being a protective one or a risk one. A noteworthy fact is that the logistic regression models are limited to linear relationships between dependent and independent variables. The neural network models can overcome this restriction. Thus, the linear relationships between independent variables and the target should be found in both models. While non-linear interactions will appear only in the connectionist model.

2.3 Evaluation Criteria

The evaluation of the models was made using a hold-out validation where the observations were chosen randomly to form the validation and the evaluation sets. In order to obtain a good error estimation of the predictive model, this database had to be split into three different datasets: the training set with 1006 patients (72%), the validation set with 112 patients (8%), and the test set with 279 patients (20%). Each partition followed the prevalence of the original database (►see Table 3). The best network architecture and parameters were selected empirically using the validation set and then evaluated with the test set. Overfitting was avoided using the validation set to stop the learning procedure when the validation medium square error function reached its minimum. Section 3 shows that using a single hidden layer was enough to obtain a good predictive model.

There is an intrinsic difficulty in the nature of the problem: the dataset is imbalanced [36, 37], in the sense that the positive class is underrepresented compared to the negative class. Thus, with this prevalence on the negative examples (89%), a trivial classifier consisting in assigning the most prevalent

Table 3 Number of samples per class of each partition of the original database. The prevalence of the original dataset is observed in each one: 11% for the positive class (major postpartum depression) and 89% for the negative class (no depression).

Dataset	No depression	Major depression	Total
Training	891	115	1006
Validation	99	13	112
Evaluation	247	32	279
Total	1237	160	1397

class to a new sample would achieve an accuracy of around 89%, but its sensitivity would be null.

The main goal is to obtain a predictive model with a good sensitivity and specificity. Both measures depend on the accuracy on positive examples, $a+$, and the accuracy on negative examples, $a-$. Increasing $a+$ will be done at the cost of decreasing $a-$. The relation between these quantities can be captured by the ROC (Receiver Operating Characteristic) curve [38]. The larger the area under the ROC curve (AUC), the higher the classification potential of the model. This relation can also be estimated by the geometric mean of the two accuracies, $G = \sqrt{a+ \cdot a-}$, reaching high values only if both values are high and in equilibrium. Thus, if now we use the geometric mean to evaluate our trivial model (which always assigns the class with the maximum a priori probability) we see that $G = 0$, which

Table 4 Results for the best models with the subject feature models (SUBJ) and the subject-environment feature models (SUBENV). We show the G-mean, the accuracy of the model with its confidence intervals at 5% of significance and its sensitivity and specificity. Varying the threshold of the classifier we obtain a continuous classifier for which the AUC value is shown. The architecture points out the number of input units, hidden units and the output unit. When pruning a network, we see that some input variables were discarded because their connections towards any hidden unit were eliminated. Thus, these pruned models are simpler than the original ones and may be more interpretable, although they might lose some sensitivity.

Model	Pruning	Architecture	G	Acc (95% CI)	Sen	Spe	AUC
SUBJ	No	16–14–1	0.82	0.81 (0.76, 0.86)	0.84	0.81	0.82
SUBENV	No	31–3–1	0.81	0.84 (0.80, 0.88)	0.78	0.85	0.84
SUBJ	Yes	9–1–1	0.77	0.78 (0.73, 0.83)	0.75	0.78	0.80
SUBENV	Yes	13–2–1	0.80	0.84 (0.80, 0.88)	0.75	0.84	0.84

means that the model is the worst we can obtain.

3. Results

►Table 4 shows the results of the best connectionist models obtained from the first approach. Two models were trained based on differential input variables: the subject model (SUBJ) and the subject-environment model (SUBENV). Both models included psychiatric antecedents, emotional alterations, neuroticism, life events, depressive symptoms, genetic factors, social support and medical perinatal risk. The SUBENV also included social and demographic features, such as age, economical and educational level, family members and labor situation.

The best model (SUBJ with no pruning) achieved 0.82 of G and 0.81 of accuracy (95% CI: 0.76–0.86) with 0.84 of sensitivity and 0.81 of specificity. In general, SUBENV and non-pruned models tend to have a better behavior than SUBJ and pruned ones, but a χ^2 test with Bonferroni correction shows that there is no statistical significance. Also, notice that the accuracy confidence intervals are overlapped (►see Table 4). On the other hand, the use of pruning methods leads to a more understandable model at the expense of a small loss of sensitivity.

A logistic regression has been done for the SUBJ and SUBENV sets of variables to compare and confirm the significant influence of the pruned selected features. It is expected that the linear relationships between independent variables and the target should be found

in the logistic regression as well as in the neural network models. In the best pruned SUBJ model the most relevant features appear as statistically significant ($\alpha = 0.05$) in the logistic regression model. Neuroticism, life events from week 8 to week 32, social support and depressive symptoms are considered risk factors. Moreover, the influence of the 5-HTT-GC combination of low-expressing genotypes, LE, is also significant and appears as a protective factor. The rest of the input variables in the logistic regression model: emotional alterations, psychiatric antecedents, pregnancy problems and the 5-HTT-GC combination of no low-expressing genotype, HE, are not significant, but in the pruned model, these four variables are seen as risk factors. The difference between significant factors of the pruned models and logistic regression may be explained by non-linear interactions of a higher order between variables because the indepen-

dent variables interact with each other as explained in Section 2.2.

Considering the SUBENV model, most of the relevant features appear as significant input variables in the logistic regression: social support, neuroticism, life events from week 8 to week 32, depressive symptoms, leave labor situation and female baby are risk factors in both models. Pregnancy problems for the mother and the baby appears as a protective factor, which is explained by the proportion of mothers with postpartum depression in the observations (see Table 1). On the other hand, the age and the number of people that the patient lives with appear as protective factors in both models, but they have no statistical significance in the regression model, whereas psychiatric antecedents is a risk factor without statistical significance. Again, we find these differences due to the interactions between variables as explained before.

In ►Table 5, the SUBJ model shows that neuroticism, social support, life events and depressive symptoms are the most outstanding features and that they are risk factors in the prediction of PPD. In the SUBENV model these variables are also main risk factors, but the variable age and the number of people that the patient lives with are both protective factors although in the regression model they have no statistical significance.

4. Discussion

The main objective of this study was to fit a feed-forward ANN classification model to predict PPD with a high sensitivity and specificity during the first 32 weeks after the delivery. The predictive model showing the best G was selected ensuring a balanced sensitivity and specificity, as Table 4 shows. With this model, we achieved around 81% of accuracy. From our results, SUBENV models did not significantly improve SUBJ models for prediction.

The major concern for the medical staff is how the PPD is influenced by the variables. These independent variables have different influences on the output of the classification model and they depend on the connections between nodes. While logistic regression models detect only linear relationships between the independent variables and the dependent variable, the neural network models can also detect non-linear relationships. Thus, the comparison with logistic regression aims to confirm that the neural network model is not inferring wrong linear influences between independent variables and the dependent variable. We expect that if a linear relationship is found to be significant in the logistic regression model then it should be also considered by the neural network pruned model. But non-linear relationships are only going to be detected by the neural network model since logistic regression cannot detect these relations. In the case where the logistic regression finds an independent variable as significant but the neural network fails to detect it, then it would be an evidence of a wrong trained model. But this situation was not found in this work as Table 5 shows. In future work, some quantitative techniques will be used in order to achieve a numeric measure of the influence of each input feature

Table 5 Independent variables selected for the SUBJ pruned model and the SUBENV pruned model for PPD. risk: risk factor; protect: protective factor; pruned: pruned variable. The table shows which variables were significant for the pruned models and for the logistic regression. If a variable is pruned in the neural network, then it is not considered significant. In the case of a logistic regression, a variable is significant if and only if the p-value < 0.05 . As expected, every significant variable in logistic regression was also significant in the neural network model.

Variable	SUBJ		SUBENV	
	Pruned Net	Log Reg (p-value)	Pruned Net	Log Reg (p-value)
Neuroticism (EPQN)	risk	risk(= 0.004)	risk	risk(= 0.004)
Social support (DUKE)	risk	risk(< 0.001)	risk	risk(< 0.001)
Depressive symptoms (Initial EPDS)	risk	risk(< 0.001)	risk	risk(< 0.018)
5-HTT-GC, HE	risk	not significant	pruned	not significant
5-HTT-GC, LE	protect	protect (= 0.041)	pruned	not significant
Emotional alteration	risk	not significant	pruned	not significant
Psychiatric antecedents	risk	not significant	risk	not significant
Pregnancy problems	risk	not significant	pruned	not significant
Life events at 32 week	risk	risk(< 0.001)	risk	risk(< 0.001)
Life events at 8 week	risk	not significant	risk	not significant
Gender girl	–	–	risk	risk(< 0.007)
Labor leave	–	–	risk	risk(< 0.008)
Labor active	–	–	protect	protect (< 0.026)
Mother-child problems	–	–	protect	protect (= 0.008)
Age	–	–	protect	not significant
No. of people living together	–	–	protect	not significant

and its interactions following rule extraction methods [39] or numeric methods [40] for ANNs. Therefore, these prevention models would give the clinicians a tool to gain knowledge on the PPD.

A classification model with this good performance, i.e., high accuracy, sensitivity and specificity, may be very useful in clinical settings. In fact, the ability of neural networks to tolerate missing information could be relevant when part of the variables are missing thus giving a high reliability in the clinical field. Since no comparison was established with other machine-learning techniques, it could be interesting to try Bayesian network models as they can also deal with missing information, find probabilistic dependencies and show good performance [41].

However, our models provided better results than the work done by Camdeviren et al. in [42] on the Turkish population. Although the number of patients was comparable, our study included more independent variables than Camdeviren's study, where a logistic regression model and a classification tree were compared to predict PPD. Based on logistic regression, they reached an accuracy of 65.4% with a sensitivity of 16% and a specificity of 95%, which means a G of 0.39. With the optimal decision tree, they obtained an accuracy of 71%, a sensitivity of 22% and a specificity of 94%, which gives a G of 0.45. As they explained, there is also a maximal tree that is very complex and overfitted, thus the generalization of this tree is very limited.

In the best model achieved, neuroticism, life events, social support and depressive symptoms just after delivery were the most important risk factors for PPD. Therefore, women with high levels of neuroticism, depressive symptoms during pregnancy and high HTT genotype are the most likely to suffer from PPD. In this subgroup, a careful postpartum follow-up should be considered in order to improve the social support and help to cope with the life events [43]. In a long term, the final goal is the improvement of clinical management of patients with possible PPD. In this sense, ANN models have been shown to be valuable tools by providing decision support, thus reducing the workload on clinicians. The practical solution to integrate the pattern recognition developments in the clinical routine workflows is the design of clinical decision support systems (CDSSs)

taking into account also clinical guidelines and user preferences [44]. There are relatively few published clinical trials and they need more rigorous methodologies of evaluation, but the general conclusion is that CDSSs can improve practitioners performances [45, 46].

In conclusion, four models for predicting PPD have been developed using multilayer perceptrons. These models have the ability to predict PPD during the first 32 weeks after delivery with high accuracy. The use of G as a measure for selecting and evaluating the models yields to a high, well-balanced sensitivity and specificity. Moreover, pruning methods can lead to simpler models, which can be easier to analyze in order to interpret the influence of each input variable on PPD. Finally, the models achieved should be incorporated, integrated and clinically evaluated in a CDSS [17] to give this knowledge to clinicians and improve the prevention and premature detection of PPD.

Acknowledgments

This work was partially funded by the Spanish Ministerio de Sanidad (PIO41635, Vulnerabilidad genético-ambiental a la depresión posparto, 2006–2008) and the Instituto de Salud Carlos III (RETICS Combiomed, RD07/0067/2001). The authors acknowledge to Programa Torres Quevedo from Ministerio de Educación y Ciencia, co-founded by the European Social Fund (PTQ05–02–03386 and PTQ–08–01–06802).

References

- Oates MR, Cox JL, Neema S, Asten P, Glangeaud-Freudenthal N, Figueiredo B, et al. Postnatal depression across countries and cultures: a qualitative study. *British Journal of Psychiatry* 2004; 46 (Suppl): s10–s16.
- O'Hara MW, Swain AM. Rates and risk of postnatal depression – a meta analysis. *International Review of Psychiatry* 1996; 8: 37–54.
- Cooper PJ, Murray L. Prediction, detection and treatment of postnatal depression. *Archives of Disease in Childhood* 1997; 77: 97–99.
- Beck CT. Predictors of postpartum depression: an update. *Nursing Research* 2001; 50: 275–285.
- Kendler KS, Kuhn J, Prescott CA. The interrelationship of neuroticism, sex and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry* 2004; 161: 631–636.
- Bloch M, Daly RC, Rubinow DR. Endocrine factors in the etiology of postpartum depression. *Comprehensive Psychiatry* 2003; 44: 234–246.

- Treloar SA, Martin NG, Bucholz KK, Madden PAF, Heath AC. Genetic influences on post-natal depressive symptoms: findings from an Australian twin sample. *Psychological Medicine* 1999; 29: 645–654.
- Ross LE, Gilbert EM, Evans SE, Romach MK. Mood changes during pregnancy and the postpartum period: development of a biopsychosocial model. *Acta Psychiatrica Scandinavica* 2004; 109: 457–466.
- Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 2003; 301: 386–389.
- Mulsant BH, Servan-Schreiber E. A connectionist approach to the diagnosis of dementia. In: *Proc. 12th Annual Symposium on Computer Applications in Medical Care*; 1988. pp 245–249.
- Tandon R, Adak S, Kaye JA. Neural networks for longitudinal studies in Alzheimers disease. *Artificial Intelligence in Medicine* 2006; 36: 245–255.
- Zhu J, Hazarika N, Chung-Tsoi A, Sergejew A. Classification of EEG signals using wavelet coefficients and an ANN. In: *Pan Pacific Conference on Brain Electric Topography*. Sydney, Australia; 1994. p 27.
- Jefferson ME, Pendleton N, Lucas CP, Lucas SB, Horan MA. Evolution of artificial neural network architecture: prediction of depression after mania. *Methods Inf Med* 1998; 37: 220–225.
- Berdia S, Metz JT. An artificial neural network stimulating performance of normal subjects and schizophrenics on the Wisconsin card sorting test. *Artificial Intelligence in Medicine* 1998; 13: 123–138.
- Franchini L, Spagnolo C, Rossini D, Smeraldi E, Bellodi L, Politi E. A neural network approach to the outcome definition on first treatment with sertraline in a psychiatric population. *Artificial Intelligence in Medicine* 2001; 23: 239–248.
- Sanjuán J, Martín-Santos R, García-Esteve L, Carot JM, Guillamat R, Gutiérrez-Zotes A, et al. Mood changes after delivery: role of the serotonin transporter gene. *British Journal of Psychiatry* 2008; 193: 383–388.
- Vicente J, García-Gómez JM, Vidal C, Martí-Bonmati L, del Arco A, Robles M. SOC: A distributed decision support architecture for clinical diagnosis. *Biological and Medical Data Analysis*; 2004. pp 96–104.
- García-Esteve L, Ascaso L, Ojuel J, Navarro P. Validation of the Edinburgh Postnatal Depression Scale (EPDS) in Spanish mothers. *Journal of Affective Disorders* 2003; 75: 71–76.
- Nurnberger JI, Blehar MC, Kaufmann C, York-Cooler C, Simpson S, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies and training. *Archives of Genetic Psychiatry* 1994; 51: 849–859.
- Roca M, Martín-Santos R, Saiz J, Obiols J, Serrano MJ, Torrens M, et al. Diagnostic Interview for Genetic Studies (DIGS): Inter-rater and test-retest reliability and validity in a Spanish population. *European Psychiatry* 2007; 22: 44–48.
- Eysenck HJ, Eysenck SBG. *The Eysenck Personality Inventory*. London: University of London Press; 1964.
- Aluja A, García O, García LF. A psychometric analysis of the revised Eysenck Personality Questionnaire short scale. *Personality and Individual Differences* 2003; 35: 449–460.

23. Paykel ES. Methodological aspects of life events research. *Journal of Psychosomatic Research* 1983; 27: 341–352.
24. Zalsman G, Huang YY, Oquendo MA, Burke AK, Hu XZ, Brent DA, et al. Association of a triallelic serotonin transporter gene promoter region (5-HTTLPR) polymorphism with stressful life events and severity of depression. *American Journal of Psychiatry* 2006; 163: 1588–1593.
25. Bellón JA, Delgado A, Luna JD, Lardelli P. Validity and reliability of the Duke-UNC-11 questionnaire of functional social support. *Atención Primaria* 1996; 18: 158–163.
26. Hranilovic D, Stefulj J, Schwab S, Borrmann-Hassenbach M, Albus M, Jernej B, et al. Serotonin transporter promoter and intron 2 polymorphisms: relationship between allelic variants and gene expression. *Biological Psychiatry* 2004; 55: 1090–1094.
27. Rosenblatt F. The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review* 1958; 65 (6): 386–408.
28. Bishop CM. *Neural Networks for Pattern Recognition*. Oxford, UK: Clarendon Press; 1995.
29. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. *The MIT Press* 1986. pp 318–362.
30. Le Cun Y, Denker JS, Solla A. Optimal brain damage. *Advances in Neural Information Processing Systems* 1990; 2: 598–605.
31. Duda RO, Hart PE, Stork DG. *Pattern Classification*. New York, NY: Wiley-Interscience; 2001.
32. Mao J, Jain AK. Artificial neural networks for feature extraction and multivariate data projection. *IEEE Transactions on Neural Networks*. 1995; 6 (2): 296–317.
33. Leray P, Gallinari P. Feature selection with neural networks. *Behaviormetrika* 1999; 26: 145–166.
34. Hassibi B, Stork DG, Wolf G. Optimal brain surgeon and general network pruning. In: *Proceedings of the 1993 IEEE International Conference on Neural Networks*. San Francisco, CA; 1993. pp 293–300.
35. Hosmer DW, Lemeshow S. *Applied logistic regression*. Wiley-Interscience; 2000.
36. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. In: *Proc. 14th International Conference on Machine Learning*. Morgan Kaufmann; 1997. pp 179–186.
37. Japkowicz N, Stephen S. The class imbalance problem: a systematic study. *Intelligent Data Analysis Journal* 2002; 6 (5): 429–449.
38. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters* 2006; 27 (8): 861–874.
39. Saad EW, Wunsch DC. Neural network explanation using inversion. *Neural Networks*. 2007; 20 (1): 78–93.
40. Heckerling PS, Gerber BS, Tape TG, Wigton RS. Entering the black box of neural networks. A descriptive study of clinical variables predicting community-acquired pneumonia. *Methods Inf Med* 2003; 42: 287–296.
41. Sakai S, Kobayashi K, Nakamura J, Toyabe S, Akazawa K. Accuracy in the diagnostic prediction of acute appendicitis based on the bayesian network model. *Methods Inf Med* 2007; 46: 723–726.
42. Camdeviren HA, Yazici AC, Akkus Z, Bugdayci R, Sungur MA. Comparison of logistic regression model and classification tree: an application to postpartum depression data. *Expert Systems with Applications* 2007; 32: 987–994.
43. Dennis CL. Psychosocial and psychological interventions for prevention of postnatal depression: systematic review. *BMJ* 2005; 331 (7507): 15.
44. Fieschi M, Dufour JC, Staccini P, Gouvert J, Bouhaddou O. Medical Decision Support Systems: Old dilemmas and new paradigms? *Methods Inf Med* 2003; 42: 190–198.
45. Lisboa PJ, Taktak AFG. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Networks* 2006; 19(4): 408–415.
46. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005. [bmj.38398.500764.8F](https://doi.org/10.1136/bmj.38398.500764.8F).