Predicting child labor in Peru: A comparison of logistic regression and neural networks techniques

Christian Fernando Libaque-Saenz[1] **Juan Lazo** [1]
Karla Gabriela Lopez-Yucra[2] **Edgardo R. Bravo**[1]
[1]Universidad del Pacífico
Avenida Salaverry 2020, Jesús María, Lima 11, Peru
[2]Pontificia Universidad Católica del Perú
Av. Universitaria 1801, San Miguel, Lima 32, Peru
cf.libaques@up.edu.pe jg.lazol@up.edu.pe
karla.lopez@pucp.pe er.bravoo@up.edu.pe

Abstract

Child labor is a relevant problem in developing countries because it may have a negative impact on economic growth. Policy makers and government agencies need information to correctly allocate their scarce resources to deal with this problem. Although there is research attempting to predict the causes of child labor, previous studies have used only linear statistical models. Non-linear models may improve predictive capacity and thus optimize resource allocation. However, the use of these techniques in this field remains unexplored. Using data from Peru, our study compares the prediction capability of the traditional logit model with artificial neural networks. Our results show that neural networks could provide better predictions than the logit model. Findings suggest that geographical indicators, income levels, gender, family composition and educational levels significantly predict child labor. Moreover, the neural network suggests the relevance of each factor which could be useful to prioritize strategies. As a whole, the neural network could help government agencies to tailor their strategies and allocate resources more efficiently.

1 Introduction

Child labor is a critical problem in developing countries because it could negatively affect economic growth (Hanushek, 2013). Child labor has a negative effect on human capital, which is defined as the stock of skills that the labor force possesses (Goldin, 2016). Children who work have a high probability of becoming individuals with a low stock of skills in both quantity and quality (Becker, 1962). In fact, these children (who work) usually do not dedicate their efforts to study and sometimes they do not even attend school at all. In turn, this low level of human capital and the associated lack of skills have a negative impact on individuals earnings and income (Hanushek, 2013). Therefore, as a countrys human capital decreases, its economy decreases as well.

According to the International Labor Organization (ILO), in Latin America this phenomenon reached 12.5 million children and teenagers between 5 and 17 years old in 2014 (Lopez, 2016). Although this number has decreased from 20 million in 2010, an important fact is that the number of children working in dangerous activities has increased from 9 million in 2010 to 9.6 million in 2014 (Lopez, 2016). As for the case of Peru, the National Housing Survey (ENAHO in Spanish) shows that 21% of teenagers between 12 and 17 years old had been working in 2014 (Lopez, 2016). In other words, 1 out of 5 teenagers works in Peru.

Child labor can not only lead to gaps among countries but also within a country. In Peru, for example, the child labor rate in rural areas is twice as high as in urban areas (Sausa, 2016). By assessing child labor by region, Huancavelica presents the highest rate of child labor (58%), which is more than 10 times that for Tumbes (5%) the latter is the region with the lowest rate of child labor (Sausa, 2016). Therefore, this phenomenon could negatively impact social and economic inclusion by increasing socioeconomic differences. It is important that governments formulate adequate programs and policies to reduce child labor. It is also important that they identify those children with a high probability of becoming workers in order to allocate resources in the correct place. There are various techniques to achieve this goal. We have traditional techniques such as logit models, and modern techniques such as neural networks. The principal difference is that the former capture linear effects, while the latter can capture non-linear relationships. It is important to have a model with high predictive capability, and therefore it is necessary to compare the predictive power of the different models.

Table 1 shows a summary of the issues covered by previous research in this field. All these studies used traditional techniques; to the best of our knowledge, in this field there are few studies using modern techniques such as neural networks. For example, Rodrigues, Prata, and Silva (2015) used data from Brazil and decision trees to search for patterns in the variables explaining child labor. The objective of the present study is to compare the predictive power of traditional and modern models in regard to child labor (i.e., correctly identify those children who work). It is expected that our results will shed light on the difference between models in terms of predictive power. By identifying the antecedents to child labor and the technique with the best predictive power, we will be able to provide recommendations to the Peruvian government.

2 Theoretical background

Classification problems such as the child labor issue can be addressed by several techniques, both parametric and non-parametric. Parametric techniques (e.g., discriminant analysis, the logit model) require the prior specification of a function (or model) that relates the independent variables (X_i) with the dependent variable (Y). In practical terms, this function may be known grounded in theory or assumed. These techniques use observations of Y and X_i to estimate the parameters of the function. Once the parameters have been estimated, they can be used for prediction with new participants. One disadvantage of the parametric techniques is that they have a rigid structure (the mathematical function does not change and it only allows for estimating the parameters). Thus, these techniques may not be appropriate to represent phenomena that do not follow well-known mathematical functions.

In contrast, non-parametric techniques (e.g., artificial neural networks) do not assume a function a priori but instead approximate the function based on observation. Once the function has been approximated, it can be used to predict new cases. One relative advantage of these techniques is that they can represent complex non-linear mathematical functions. In other research arenas, this flexibility of non-parametric techniques has, under certain conditions, demonstrated the superiority of its predictive power over that of parametric techniques (e.g., Abdou et al., 2008; Altman et al., 1994).

Our research compares the logit model (parametric technique) with artificial neural networks (non-parametric technique) in the field of child labor. The application of these models for predictive purposes involves the following steps:

- The sample is randomly divided into two subsamples.
- The parameters of the model are estimated with one of the subsamples.
- The predictive capacity of the model (number of hits over total observations) is assessed.
- With these estimated parameters, prediction of the dependent variable for the other subsample is conducted.
- The predictive capacity of the model (with the test data) is assessed.

2.1 Logit model

The logit model is a method that uses independent variables to estimate the probability of occurrence of a discrete outcome in the dependent variable (Lattin et al., 2003). According to the number of discrete outcomes, this technique can be divided into binary logit or multinomial logit models (Hosmer et al., 2013; Lattin et al., 2003). The former defines a dependent variable with two discrete outcomes whereas the latter represents a logit model with more than two discrete outcomes for the dependent variable (Hosmer et al., 2013; Lattin et al., 2003). In both cases, the discrete outcomes for the dependent variable should be mutually exclusive (Lattin et al., 2003).

The logit model has a straightforward and closed functional form that is easily estimated using maximum likelihood methods (Lattin et al.,

Table 1: Literature review

Author	Topic
Rodríguez (2002)	Impact of family factors on education
Emerson and Souza (2002)	Impact of gender on child labor
Sapelli and Torche (2004)	Determinants of school desertion
Lavado and Gallegos (2005)	Characteristics of children with high probability of leaving the school
García (2006)	Relationship between home responsibilities and work
Gunnarsson, Orazem, and Snchez (2006)	Impact of child labor on education performance
Alcázar (2008)	Determinants of school desertion in rural areas
Rodríguez and Vargas (2008)	Consequences of child labor
Rodríguez and Vargas (2009)	Characteristics and nature of economic activity in child labor
Lima, Mesquita ,and Wanamaker (2015)	Effect of family wealth on the utilization of child labor
Le and Homel (2015)	Impact of child labor on education performance
He (2016)	Relationship between child labor and a child's academic achievement

2003, p. 475). The logit technique does not assume restrictions on the normality of the distribution of variables (Press and Wilson, 1978). Also, independent variables can be both continuous and categorical variables (Lattin et al., 2003). This technique is a special case of regression, which uses a transformation of the discrete dependent variable. This model assumes: 1) a categorical dependent variable with mutually exclusive outcomes, 2) independent variables can be continuous or categorical, 3) independence of observations, 4) absence of multicollinearity between independent variables, 5) a linear relationship between the continuous independent variables and the logit transformation of the dependent variable, and 6) absence of outliers.

The logit model is defined by the following function:

$$Logit(p_i) = Ln\left(\frac{p_i}{1-p_i}\right) = \alpha + X_i^T\beta + \varepsilon_i \quad (1)$$

where p_i is the probability that an observation takes a specific outcome of the dependent variable, α is the constant term; β is the corresponding vector of the coefficients; and ε_i is the error term.

2.2 Artificial neural networks

A neural network is, in a general sense, a machine designed to model the way in which the brain performs a particular task or function of interest (Haykin, 1998). The functioning of the brain is applied in this design because of its "(...) capability to organize its structural constituents, known as neurons, so as to perform certain computations (e.g., pattern recognition, perception, and motor control) many times faster than the fastest digital computer in existence today" (Haykin, 1998, p. 23). Therefore, a neural network resembles the brain mainly in two aspects: 1) the way knowledge is acquired by the network from its environment (i.e., learning process); and 2) the strength of interneuron connections (i.e., synaptic weights), which are used to store the acquired knowledge (Haykin, 1998). Accordingly, an artificial neural network is a physical cellular network that is able to acquire, store, and utilize experiential knowledge (Zurada, 1992). A fundamental unit in the operation of a neural network is the neuron. It is an information-processing unit which has three basic elements: a set of synapses or connecting links, each one with a weight or strength of its own; an adder for summing the input signals; and an activation function for limiting the amplitude of the output of a neuron (Haykin, 1998, p. 32). The neurons perform simple operations, transmitting their results to neighboring processors. Hence, the ability of a neural network to perform non-linear relationships between its inputs and outputs makes it a useful technique for pattern recognition and modeling of complex systems (Bishop, 1995).

According to their topology, neural networks can be feedforward or feedback networks. In the former, the mapping goes from an input to an output layer instantaneously since there is no delay between them. This type of network is characterized by its lack of feedback which implies that the neural network has no explicit connection between layers (Zurada, 1992). In contrast, the latter has a connection between the output and input layers (Zurada, 1992).

Another typology of neural networks is related to the learning paradigm which distinguishes between supervised learning and non-supervised learning. The first implies that the knowledge of the environment available to the teacher is transferred to the neural network through training as fully as possible (Haykin, 1998). Also, it implies an error-correction learning in which the network parameters are adjusted under the combined influence of the training vector (i.e., example) and the error signal (i.e., difference between the desired response and the actual response of the network). This adjustment is carried out step by step in order to make the neural network emulate the teacher (Haykin, 1998). On the other hand, the second does not consider a teacher to oversee the learning process. In this case, there are no labeled examples of the function to be learned by the network. The learning of an input-output mapping is performed through continued interaction with the environment or based on the optimization of its parameters in order to develop the ability to form internal representations (Haykin, 1998).

This research uses a Multilayer Perceptron neural network with a back-propagation algorithm which consists of applying a family of gradient-based optimization methods to find the optimal value of the weights based on minimizing the error norm between the desired output and the output calculated by the neural network (Rumelhart et al., 1986). In this type of network, the processing is performed by the inputs. The output obtained is compared to the expected output. From the obtained error, a process of adjustment of weights is applied, attempting to minimize the error.

2.3 Child labor in Peru

The concept of child labor varies from country to country depending on the cultural context. According to the ILO, child labor refers to a work that is dangerous and harmful to the physical, mental, or moral wellness of the child, interfering with his/her education.

In the case of Peru, the minimum age for a child to be allowed to legally work is 14 years old, as long as these activities do not harm their integrity nor negatively impact their studies (Lopez, 2016). Also, they must have the permission of their parents or legal guardians to engage in these activities. In exceptional cases, children between 12 and 14 years old could also work as long as the work meets the same requirements (Lopez, 2016). In the present research, a child was considered to be a worker if he/she helps in the family business, in domestic tasks in a house that is not his or her own, in producing products to be sold, in agriculture activities, in selling products or providing services.

According to the National Housing Survey, child labor between 6 and 13 years old in rural areas (67.5%) is twice as prevalent as child labor in urban areas (32.5%). However, in the range from 14 to 17 years old, the values are similar (49.7% and 50.3% for rural and urban areas, respectively). Another important issue is that child labor rates significantly differ between cities. For example, Huancavelica is the city with the highest rate of child labor with 79.0%, followed by Puno, Huanuco, and Amazonas with 69.0%, 65.0%, and 64.0% respectively. Trujillo has the lowest child labor rate, at about 5.0%, which is significantly lower than the others. Not surprisingly, the cities with the highest rates of child labor are also those with the lowest incomes per capita. Furthermore, according to the National Institute of Statistics and Informatics (INEI in Spanish), economic activity for females (63.3%) is considerable lower than for males (81.4%).

Based on the above paragraph, we included variables capturing: 1) age and gender; 2) type of residence area such as urban/rural, region, stratum, and schooling available; and 3) socioeconomic variables such as expenses, education of the family head, type of housing, housing ownership, and housing status (adequacy, coverage of basic needs, sanitation). In addition, following (Lopez, 2016), we included family characteristics as potential antecedents to child labor. Indeed, families where both parents work are less likely to have their children working, while the number of children could increase the probability that one or more children work. In these cases, the oldest child is the one with the highest probability of engaging in economic activities. Finally, current schooling status could also be a potential factor for child labor because those children who are behind in their studies are potentially engaged in other activities.

3 Research method

3.1 Measurement model

Table 2 defines our variables and shows the measurement items used in each one.

3.2 Data collection and analysis

Data were collected from the Peruvian National Housing Survey (ENAHO) for the year 2014. We eliminated the data for the months of January, February and March to eliminate seasonality. The rationale is that those months are holidays in Peruvian schools and thus the probability of child labor is high but does not imply that children stop studying to carry it out. Data include children between 12 and 17 years old at the national level who meet the following criteria: 1) is the son/daughter of the head of the family, and 2) he/she has not yet finished school.

For analysis, we used logit and neural networks techniques to find the antecedents to child labor and to classify children according to the probability of becoming a worker. We used these two techniques to compare predictive power because a correct prediction may allow governments to correctly allocate resources to deal with this problem. The first technique is based on linear relationships, while the latter can manage non-linear effects. Thus, differences in their results are expected. In the case of the logit model, we randomly divided the full sample into 2 subsamples: 1) a training subsample consisting of 85% of the full sample, and 2) a test subsample made up of the remaining 15%. We used the training subsample to calibrate the model (i.e., estimate the parameters of the function), and the test subsample to assess the predictive power of these results. In the case of neural networks, we randomly divided the sample into 3 subsamples: 1) a training subsample (70% of the total data), 2) a validation subsample (15% of the total data), and 3) a test subsample (15% of the total data). We used the training and validation subsamples together to estimate the parameters of the model. To avoid overfitting and guarantee that the results of this stage could be generalized, we validated the predictive quality of the model with only the validation subsample every 1000 interactions. This process allows a better estimation of the weights of the network. Finally, we assessed the predictive power of the model with the test subsample.

4 Results

4.1 Logit results

We conducted a preliminary analysis including all 17 independent variables. Results show that only 9 variables were statistically significant (variables with coefficients with p-value less than 0.05) in explaining the variance of our dependent variable (WORK). The other 8 variables (p-values higher than 0.05) were not considered in the subsequent analysis given that they do not have any impact on the dependent variable. Retained variables are divided into 6 categorical variables: URBAN, AREA, STRATUM, OWN, ADEQ, and UNMET; and 3 continuous variables: EXPENSE, EDU_HEAD, and SIBLINGS. We calculated the coefficients of the model using equation (1), where p_i is the probability that child i becomes a worker.

We assessed whether assumptions of logistic regression were met. Assumptions 1, 2, and 3 were determined by the model and data collection. For assumption 4, we conducted a linear regression to obtain VIF values. All VIF values were lower than 5 (the independent variable URBAN has the highest VIF value at 2.274). Therefore, there is no evidence of multicollinearity problems in our model (Hair et al., 2011). For the fifth assumption, we used the Box and Tidwell (1962) procedure. This procedure establishes that if the interaction between an independent continuous variable and its natural logarithm transformation is found to be significant, this variable is not linearly related to the logit of the dependent variable. In addition, following Tabachnick and Fidells (2007) recommendation, we used a Bonferroni correction for the statistical significance level by dividing it by the number of independent variables running this test including the constant term. This correction provided a significance level of 0.0038 (i.e., 0.05/13, where 0.05 is the original significance level and 13 is the sum of variables including the constant term: 1 constant term, 6 categorical independent variables, 3 continuous independent variables, and 3 interaction terms). P-values for the interaction terms were 0.688 for EDU_HEAD, 0.999 for SIBLINGS, and 0.0041 for EXPENSE. Based on this assessment, all p-

Table 2: Measurement items

Variable	Description
Dependent Variable	
Worker (WORK)	1 = If the child works 0 = If the child exclusively studies
Continuous Independent Variables	
Age (AGE)	Age of the child (in years)
Education of the family head (EDU_HEAD)	Level of schooling of the head of the family (in years)
Younger siblings (SIBLINGS)	Number of children under 5 years old in the family
Family composition (COMPO)	Ratio of the number of adults (18 years old or older) to the number of children (younger than 18 years old) in the family
Education centers (CENTER)	Ratio of the number of education centers to the number of school-age children in the province of residence of the family
Monthly expense (EXPENSE)	Natural logarithm of the total monthly expense per family member
Categorical Independent Variables	
Maleness (MALE)	1 = If the child is male 0 = If the child is female
Urban (URBAN)	1 = If the residence of the family is located in the urban area 0 = If the residence of the family is located in a non-urban area
Oldest child (OLD_CHI)	1 = If the child is the oldest in the family 0 = If the child is not the oldest in the family
School backwardness (DELAY)	1 = If the child presents school backwardness 0 = If the child does not present school backwardness
Geographic area (AREA)	1 = North Coast 2 = Center Coast 3 = South Coast 4 = North Highlands 5 = Center Highlands 6 = South Highlands 7 = Jungle 8 = Lima Metropolitan Area
Geographic stratum (STRATUM)	1 = More than 100,000 dwellings 2 = From 20,001 to 100,000 dwellings 3 = From 10,001 to 20,000 dwellings 4 = From 4,001 to 10,000 dwellings 5 = From 401 to 4,000 dwellings 6 = 400 dwellings or fewer 7 = Composite rural area 8 = Simple rural area
Type of housing (TYPE)	1 = Independent house 2 = Apartment in building 3 = Chalet 4 = Neighborhood house 5 = Shack or cottage 6 = Improvised housing 7 = Non-housing premises 8 = Other

Housing ownership (OWN)	1 = Rented 2 = Owned by the family, totally paid 3 = Owned by the family, as result of squatting 4 = Owned by the family, paying off a loan 5 = Given by the workplace of one of the members 6 = Given by other family or institution 7 = Other
Housing inadequacy (ADEQ)	1 = If the housing is inadequate 0 = If the housing is adequate
Uncovered basic needs (UNMET)	1 = If the housing has unmet basic needs 0 = If the housing has not unmet basic needs
Absence of sanitation (HYGIENIC)	1 = If the house does not have sanitation 0 = If the housing has sanitation

values were over the value of 0.0038 and thus our model satisfied the linearity assumption. For the sixth assumption, we found 4 outliers of concern which were not considered in subsequent analysis. Results of the logistic model are presented in Table 3. Our model is statistically significant ($\chi^2 = 2300.885, df = 25, p = 0.000$), and explains between 33.2% and 47.8% of the variance in child labor.

In terms of predictive value, our model correctly predicted 82.61% of cases, with 55.80% of correct positive classifications (sensitivity) and 93.02% of correct negative classifications (specificity). Accordingly, our model has an efficiency (average of sensitivity and specificity) of 74.41% and a mean absolute percentage error (MAPE) of 17.39%. Although our model has an adequate overall predictive power, the Hosmer and Lemeshow goodness of fit test was significant ($\chi^2 = 39.889, df = 8, p = 0.000$) showing that it is poor at predicting the categorical outcomes. The reason for this finding may be the difference between sensitivity and specificity. Finally, coefficients (B) were found to be significant based on the Wald test. Table 3 also shows the standard error (SE) of the coefficients and their odd ratio (OR). We then assessed the model with our test subsample. Our model correctly predicted 80.64% of all the cases in this sample, with a sensitivity of 52.78%, a specificity of 91.88%, an efficiency of 72.33%, and a MAPE of 19.36%.

4.2 Neural network results

For purposes of comparison, we chose a simple neural network. Accordingly, we used a hidden layer with activation functions Hyperbolic tangentsigmoidy, and an output layer with activation functions Log-sigmoid. The value of weights and bias are updated according to gradient descent momentum and an adaptive learning rate. The training parameters of the neural network were: Maximum number of epochs to train: 40000, learning rate: 0.01, momentum constant: 0.7, performance goal: 10-5. These values were set following current literature (Haykin, 1998; Zurada, 1992). They were also adjusted during the training process using an adaptive algorithm to find better parameters.

The first neural network used the 17 proposed independent variables (inputs). With the training and validation subsamples we obtained the best neural network made up of 38 neurons in the hidden layer and 1 neuron in the output layer. This model predicted 88.26% of all the cases, with a sensitivity of 90.97% and specificity of 87.21%. The efficiency of the model was thus 89.09%, and the MAPE was 11.74%. When applying this model to the test subsample, it predicted 85.11% of all the cases, with a sensitivity of 90.02%, a specificity of 79.42%, an efficiency of 84.72%, and a MAPE of 14.89%.

In addition, by analyzing the weight of the inputs of the neural network, we ranked the independent variables from the highest to the lowest effect: AREA (7.7), EXPENSE (7.4), HYGIENIC (7.4), STRATUM (7.0), MALE (6.2), OWN (6.0), SIBLINGS (5.9), TYPE (5.9), OLD_CHI (5.9), AGE (5.7), EDU_HEAD (5.5), COMPO (5.5), ADEQ (5.1), DELAY (5.0), CENTER (4.9), URBAN (4.6), and UNMET (4.4).

The second neural network used only the 9 variables that were statistically significant in the logit

Table 3: Logistic regression with training sample

Variables	Model 1 (N=700)			
	B	**SE**	**Wald**	**OR**
URBAN	3.649	0.441	68.563***	38.455
EDU_HEAD	-0.056	0.01	28.969***	0.946
SIBLINGS	0.181	0.081	4.978*	1.198
AREA	SS	SS	360.946***	SS
STRATUM	SS	SS	71.179***	SS
OWN	SS	SS	13.571*	SS
ADEQ	0.644	0.146	19.408***	1.905
UNMET	-0.366	0.117	9.721**	0.693
EXPENSE	-0.262	0.084	9.707**	0.769
Constant	-1.423	0.663	4.604	
-2log likelihood	4459.518			
Chi-square (Model)	2300.885*** (df=25, p-value=0.000)			
Hosmer & Lemeshow	39.889*** (df=8, p-value=0.000)			
Cox & Snell R2	33.20%			
Nagelkerke R2	47.80%			
Overall predicted %	82.61%			
Sensitivity	55.80%			
Specificity	93.02%			

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$
B=Coefficients; SE=Standard error; OR=Odds ratio
SS=Skipped for simplicity. (For categorical variables with more than 2 categories, there is a coefficient for each category. We are choosing not to report them all because our focus is the predictive power of the model.)

model for a straight comparison. In this model, with the training and validation subsamples we obtained the best neural network made up of 30 neurons in the hidden layer and 1 neuron in the output layer. Our model achieved 84.45% of correct total predictions, with a sensitivity of 79.61%, a specificity of 86.34%, an efficiency of 82.97%, and a MAPE of 15.55%. When using our models parameters on the test subsample, it predicted 81.69% of all the cases, with a sensitivity of 78.86%, specificity of 84.23%, efficiency of 81.55%, and a MAPE of 18.31%.

For this model, the ranking of the inputs according to their weights is: AREA (12.9), STRATUM (12.6), EDU_HEAD (12.4), SIBLINGS (12.3), URBAN (10.9), ADEQ (10.6), UNMET (9.7), OWN (9.5), and EXPENSE (9.1).

4.3 Technique comparison

The results of the previous section are summarized in Table 4. Considering that the logit model used 9 variables (8 were not considered because they have no significant impact on the dependent variable), the neural network used these same 9 variables and the same instances to ensure a fair comparison. In addition, Table 4 shows the results of the neural network technique with the complete 17 variables to assess if this non-linear model could extract important information from those 8 variables without a linear impact on the dependent variable. This table shows that overall neural network technique performed better than the logit model. In fact, the neural network obtained the highest values of accuracy (correct total - positive and negative - predictions). Also, considering that it is more important to predict when a child has high probabilities of becoming a worker than to predict that a child will be non-worker, sensitivity stands as our most important metric when comparing models. By an inspection of Table 4, sensitivity of the neural network technique was superior to the values obtained from the logit model. In spite of these results, the logit model was superior in terms of specificity. However, specificity is a metric for correct predictions of non-workers, which is not relevant in our case. In addition, Figure 1 shows the ROC curve of prediction for these techniques.

Table 4: Comparison of results of predictive capacity in the test subsample

Predictive measures	Logit	Neural network	
	9 Variables	9 Variables	17 Variables
Accuracy	80.64%	81.69%	85.11%
Sensitivity	52.78%	78.86%	90.02%
Specificity	91.88%	84.23%	79.42%
Efficiency	72.33%	81.55%	84.72%
MAPE	19.36%	18.31%	14.89%

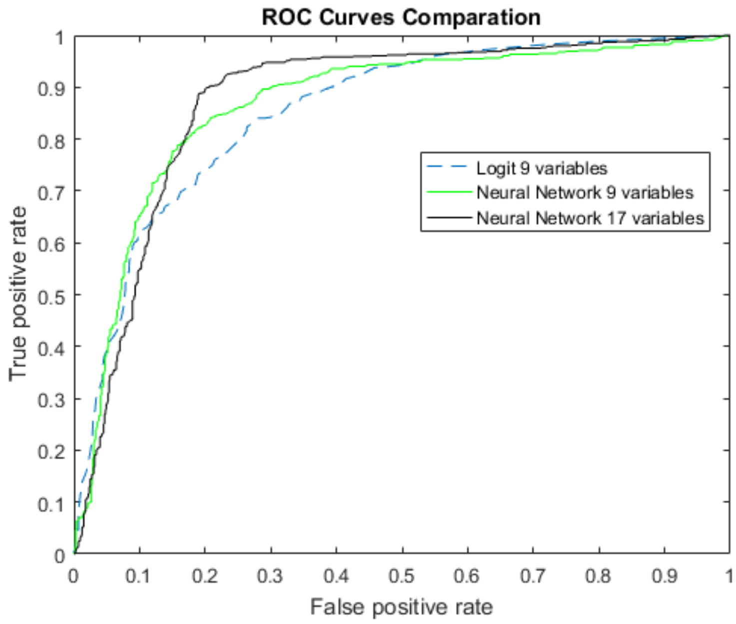

Figure 1: ROC Curves Comparation

5 Discussion

Overall, the results show that the neural network technique surpasses the logit model in predictive capacity of child labor (sensitivity). Indeed, this phenomenon may have a more complex structure than is assumed by the logit model. In consequence, the neural network (which adopts non-linear relationships) could capture sources of variation that are not identified by the logit technique. An accurate prediction of this phenomenon could be used by policy makers and government agencies to design adequate strategies or to invest scarce resources efficiently to deal with this problem.

Also, our findings show that the neural network model with 17 variables performed better than the 9-variables models (logit or neural network). This result suggests that this additional set of variables capture an important variability in explaining child labor. In other words, the neural network model with 17 variables does not ignore information that is relevant to the prediction. This result could be used by decision makers to avoid discarding relevant factors when dealing with this phenomenon.

Another important result is that the neural network model shows that geographical indicators, income levels, gender, family composition and educational levels significantly predict child labor. These results are aligned with those of the logit model showing that stratum, geographic area, and housing conditions have a significant impact on our dependent variable. These results can be used to determine the relevance of each factor. In turn, this relevance-based ranking of factors could further help government agencies to better allocate their resources and implement their strategies to reduce child labor.

Finally, previous studies in this field have used linear statistical models to predict child labor. Our study shows that the use of computational intelligence techniques, such as the neural network, could provide better predictions, which leads to better decision making.

References

Hussein Abdou, John Pointon, and Ahmed El-Masry. 2008. Neural nets versus conventional techniques in credit scoring in Egyptian banking. *Expert Systems with Applications* 35(3):1275–1292.

Lorena Alcázar. 2008. *Asistencia y deserción en escuelas secundarias rurales del Perú*, pages 41–82.

Edward I Altman, Giancarlo Marco, and Franco Varetto. 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance* 18(3):505–529.

Gary Becker. 1962. Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy* 70.

Christopher M Bishop. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA.

George EP Box and Paul W Tidwell. 1962. Transformation of the Independent Variables. *Technometrics* 4(4):531–550. https://doi.org/10.1080/00401706.1962.10490038.

Patrick M. Emerson and André Portela Souza. 2002. Bargaining over sons and daughters: Child labor, school attendance and intra-household gender bias in brazil.

Luis García. 2006. The supply of child labor and household work. MPRA Paper 31402, University Library of Munich, Germany.

Claudia Goldin. 2016. *Human Capital*, Springer Verlag, Heidelberg, Germany.

Victoría Orazem Peter F & Sánchez Mario A Gunnarsson. 2006. Child labor and school achievement in latin america .

Joe F Hair, Christian M Ringle, and Marko Sarstedt. 2011. PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice* 19(2):139–152. https://doi.org/10.2753/MTP1069-6679190202.

Eric A Hanushek. 2013. Economic growth in developing countries: The role of human capital. *Economics of Education Review* 37:204–212.

Simon Haykin. 1998. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition.

Huajing He. 2016. Child labour and academic achievement: Evidence from gansu province in china. *China Economic Review* 38(C):130–150.

D W Hosmer, S Lemeshow, and R X Sturdivant. 2013. *Applied Logistic Regression*. Wiley Series in Probability and Statistics. Wiley.

James M Lattin, J D Carroll, and P E Green. 2003. *Analyzing Multivariate Data*. Number v. 1 in Analyzing Multivariate Data. Thomson Brooks/Cole.

Pablo Lavado and José Gallegos. 2005. La dinámica de la deserción escolar en el Perú: un enfoque usando modelos de duración. Working Papers 05-08, Departamento de Economía, Universidad del Pacífico. https://ideas.repec.org/p/pai/wpaper/05-08.html.

Huong Thu Le and Ross Homel. 2015. The impact of child labor on children's educational performance: Evidence from rural Vietnam. *Journal of Asian Economics* 36:1–13.

Luiz Renato Lima, Shirley Mesquita, and Marianne Wanamaker. 2015. Child labor and the wealth paradox: The role of altruistic parents. *Economics Letters* 130:80–82.

Karla Lopez. 2016. Determinantes del trabajo infantil y la deserción escolar en menores de 12 a 17 años en el Perú para los años 2006 y 2014. Bachelor's degree, PUCP, Perú.

S James Press and Sandra Wilson. 1978. Choosing between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association* 73(364):699–705.

Diego C Rodrigues, David N Prata, and Michel A Silva. 2015. Exploring social data to understand child labor. *International Journal of Social Science and Humanity* 5(1):29.

Jose Rodriguez. 2002. *Adquisición de educación escolar básica en el Perú: uso del tiempo de los menores en edad escolar*. Departamento de Economía - Pontificia Universidad Católica del Perú. http://EconPapers.repec.org/RePEc:pcp:pucotr:otr-2002-02.

Jose Rodríguez and Silvana Vargas. 2009. *Trabajo infantil en el Perú. Magnitud y perfiles vulnerables. Informe Nacional 2007-2008*. Departamento de Economía - Pontificia Universidad Católica del Perú.

José Rodríguez and Silvia Vargas. 2008. Escolaridad y trabajo infantil: patrones y determinantes de la asignación del tiempo de niños y adolescentes en Lima Metropolitana. Technical report, PUCP, Peru.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning internal representations by error propagation. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, MIT Press, pages 318–362.

Claudio Sapelli and Arístides Torche. 2004. Deserción Escolar y Trabajo Juvenil: ¿Dos Caras de Una Misma Decisión? *Cuadernos de economía* pages 173–198.

Mariela Sausa. 2016. *El trabajo infantil es más alto y más penoso en las zonas rurales*. Perú 21.

BG Tabachnick and LS Fidell. 2007. Multivariate analysis of variance and covariance. In *Using Multivariate Statistics*, Allyn and Bacon Boston.

Jacek M Zurada. 1992. *Introduction to artificial neural systems*. West St. Paul.