Using the Breeder GA to Optimize a Multiple Regression Analysis Model used in Prediction of the Mesiodistal Width of Unerupted Teeth

F. Stoica, C.G. Boitor

Florin Stoica
Department of Mathematics and Informatics, Faculty of Sciences,
"Lucian Blaga" University of Sibiu,
Romania, 550012, Dr. Ion Ratiu 5-7, Sibiu
florin.stoica@ulbsibiu.ro

Cornel Gheorghe Boitor
Department of Preventive Dentistry, Faculty of Medicine "V. Papilian",
"Lucian Blaga" University of Sibiu,
Romania, B-dul. Victoriei, No. 10, Sibiu
boitor.cornel@ulbsibiu.ro

Abstract: For the prediction of the unerupted canine and premolars mesiodistal size, have been proposed different variants of multiple linear regression equations (MLRE). These are based on the amount of the upper and lower permanent incisors with a tooth of the lateral support. The aim of the present study was to develop a method for optimization of MLRE, using a genetic algorithm for determining a set of coefficients that minimizes the prediction error for the sum of permanent premolars and canines dimensions from a group of young people in an area Romania's central city represented by Sibiu. To test the proposed method, we used a multiple linear regression equation derived from the estimation method proposed by Mojers to which we adjusted regression coefficients using the Breeder genetic algorithm proposed by Muhlenbein and Schlierkamp. A total of 92 children were selected with complete permanent teeth which had not clinically visible dental caries, proximal restorations or orthodontic treatment that requires the decrease of the mesiodistal size of teeth. For each of these models was made a hard dental stone which was then measured with a digital calliper, the instrument having an accuracy of 0.01 mm. To improve prediction equations, we divided data into training and validation sets. The Breeder algorithm, using the training set, will provide new values for regression coefficients and error term. The validation set was used to test the accuracy of the new proposed equations.

Keywords: Genetic Algorithms (GA), mixed dentition analysis, multiple regression equations, mesiodistal teeth size.

1 Introduction

The estimation of the mesiodistal size of permanent canine and of the two premolars before their eruption is important for early evaluation of the need for space in this area and consequently to mandible and maxillary. This represents an important part of diagnosis and orthodontic treatment strategy.

Methods of estimation, performed during mixed dentition, can be grouped into three categories: those using multiple linear regression equations, those using radiographs and those using a combination of the two methods [1], [3]- [7].

Among these methods recently reported in literature, those based on MLRE have the highest predictive capacity of the mesiodistal dimension (MDD) for unerupted canines and premolars.

Prediction capacity of these methods can vary depending on the characteristics of different types constitutional areas, sometimes may vary even in the same countries [5]- [8], [9]- [17].

Our aim was to verify if optimizing through an original method the MLRE used recently in the literature [3], [14], can be predicted with sufficient accuracy the sizes of unerupted teeth from the support area in a group of children in Sibiu located in a central area of Romania.

The first objective of the study was to verify the accuracy of recently used MLRE based on known variables, namely the mesiodistal dimensions of teeth 42, 21 and 46, used in prediction of the sizes of unerupted teeth from the support area. The second objective of the study was to use an evolutive calculation method based on the genetic algorithm Breeder, to optimize the regression coefficients used in the MLRE. Thus the accuracy of the predictions can be improved [1], [9], [19], [20].

2   Subjects, materials and prediction methods

A representative public school with a population of 321 children ages 12-15 years from Sibiu (Romania) was selected for this study. From these subjects, a random simple technique was used to select 92 students (47 females and 45 males) fulfilling the selection criteria:

- To have the parents' written consent to participate in the study;

- To present the dental arches fully erupted permanent teeth (molars 3 was not considered);

- The erupted teeth show no abnormalities of shape, size or structure;

- The teeth must not have missing of substance in the mesiodistal size, due to decay, trauma or orthodontic treatments have provided striping.

Dental impressions were taken with alginate impression material and immediately poured with hard dental stone to avoid any distortion. The measure tooth size models we used a digital calliper manufactured by Mega (Germany) with an accuracy of 0.01 mm.

Measurements were performed after the procedure proposed by Seipel [12]. All models were measured 2 times by the same author and the result used was the average of two values. It was calculated the Pearson correlation coefficient between measurements and method error (ME) was calculated using the formula of Dahlberg: $ME = \sqrt{d^2/2n}$ where d is the difference between the two measurements and n is the number of patterns measured a second time.

For estimation the size of the unerupted canines and premolars we chose a recently proposed equation [3] based on known variables 21, 42 and 46. The form of this equation is: $Y = X_1*A_1+X_2*A_2+X_3*A_3+A$, where Y is the outcome expected, X_1, X_2, X_3 are independent variables determined by the size of teeth 42, 46 and 21, A_1, A_2 and A_3 are regression coefficients for used teeth and A is a specific constant. The values of constant A and regression coefficients of the equation are presented in Table 1:

Canines premolars group	Constant A	A_1 (42)	A_2 (46)	A_3 (21)
Maxillary	6,563	0,822	0,595	0,411
Mandible	3,350	0,872	0,710	0,538

Table 1: Parameters of multiple linear regression equation used [3]

In the following is presented our approach in optimization of the regression coefficients presented above, to provide a more accurate method for prediction of the mesiodistal width of unerupted permanent canines and premolars.

Because parameters of the multiple linear regression equation are real values, we are using a Breeder genetic algorithm, in order to avoid a weak point of classical GAs, represented by their discrete representation of solutions, which implies a limitation of the power of the optimization process [20].

The Breeder genetic algorithm, proposed by Mühlenbein and Schlierkamp-Voosen [2] represents solutions (chromosomes) as vectors of real numbers, much closer to the reality than normal GA's.

The selection is achieved randomly from the $T\%$ best elements of current population, where T is a constant of the algorithm (usually, $T = 40$ provide best results). Thus, within each generation, from the $T\%$ best chromosomes are selected two elements, and the crossover operator is applied over them. On the new child obtained from the mate of the parents is applied the mutation operator. The process is repeated until are obtained $N - 1$ new individuals, where N represents the size of the initial population. The best chromosome (evaluated through the fitness function) is inserted in the new population (1-elitism). Thus, the new population will have also N elements.

2.1 The Breeder genetic operators

Let be $x = \{x_1, x_2, ..., x_n\}$ and $y = \{y_1, y_2, ..., y_n\}$ two chromosomes, where $x_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$, $i = \overline{1, n}$. The crossover operator has a result a new chromosome, whose genes are represented by values $z_i = x_i + \alpha_i(y_i - x_i)$, $i = \overline{1, n}$, where α_i is a random variable uniformly distributed between $[-\delta, 1 + \delta]$, and δ depends on the problem to be solved, typically in the interval $[0, 0.5]$.

The probability of mutation is typically chosen as $1/n$. The mutation scheme is given by $x_i = x_i + s_i \cdot r_i \cdot a_i$, $i = \overline{1, n}$ where: $s_i \in \{-1, +1\}$ uniform at random, r_i is the range of variation for x_i, defined as $r_i = r \cdot domain_{x_i}$, where r is a value in the range between 0.1 and 0.5 (typically 0.1) and $domain_{x_i}$ is the domain of the variable x_i and $a_i = 2^{-k \cdot \alpha}$ where $\alpha \in [0, 1]$ uniform at random and k is the number of bytes used to represent a number in the machine within is executed the Breeder algorithm (mutation precision).

2.2 The Breeder genetic algorithm

The skeleton of the Breeder genetic algorithm may be defined as follows [19]:

Procedure Breeder
begin
$t = 0$

Randomly generate an initial population $P(t)$ of N individuals
Evaluate $P(t)$ using the fitness function
while (termination criterion not fulfilled) **do**
    for $i = 1$ to $N - 1$ **do**
        Randomly choose two elements from the $T\%$ best elements of $P(t)$
        Apply the crossover operator
        Apply the mutation operator on the child
        Insert the result in the new population P'(t)
    end for
    Choose the best element from $P(t)$ and insert it into $P\prime(t)$
    $P(t + 1) = P\prime(t)$
    $t = t + 1$
end while
end

2.3   The optimization process

The aim of the Breeder genetic algorithm is to find new values for the parameters of multiple linear regression equation presented in table 1, in order to reach a better prediction.

Each chromosome contains four genes, representing the real values $A_i, i = \overline{1,3}$ and A. The fitness function for chromosomes evaluation is represented by the number of cases from the training set having an approximation error obtained with the new equation (in absolute value) bigger than prediction error provided by original equation. In our tests, parameters of Breeder algorithm are assigned with following values: $\delta = 0$, $r = 0.1$ and $k = 8$. The initial population has 1500 chromosomes and algorithm is stopped after 30000 generations.

Data provided by our study models was randomly divided in two sets: the training set, containing 50 cases and the validation set, composed by 42 study models.

Implementation of our new optimization method was accomplished in Java language, using Net Beans 7.01.

3   Results

The MLRE method is using two equations, one for mandible and other for maxillary.

Because there are differences in measurements of teeth between left and right quadrants for mandible and respectively for maxillary, in order to improve the prediction, we are using four equations, one for each quadrant.

Using the data from training set, the Breeder algorithm finds new values for the parameters of the initial multiple linear regression equation (Table 2). The accuracy of prediction made by optimized equations was verified using the validation set.

The order of reliability of both compared prediction methods is the same. As we can see from the table 3, the correlation coefficient r calculated for the all four linear regression equations is almost the same for the original MLRE equations as for the Breeder optimized equations.

In the following figures are evaluated the original and respectively the optimized multiple linear regression equations, using as criteria the number of cases better evaluated.

A comparison of prediction error in estimating the mesiodistal widths of the canines and premolars in the mandible and maxilla using multiple linear regression equations in original form and respectively in optimized form is presented in figures 3-6:

Quadrant	A	A_1	A_2	A_3
1	5.1917	0.7571	0.85332	0.28341
2	5.16292	0.90463	0.68192	0.41011
3	3.31241	0.89357	0.72022	0.51352
4	3.28732	0.70242	0.84793	0.47736

Table 2: Optimal values of parameters for multiple linear regression equations provided by the Breeder genetic algorithm

Quadrant	Linear regression equations	
	Original MLRE	Optimized with Breeder
1	0.546	0.572
2	0.509	0.510
3	0.671	0.671
4	0.625	0.664

Table 3: The correlation coefficients r for multiple linear regression equations
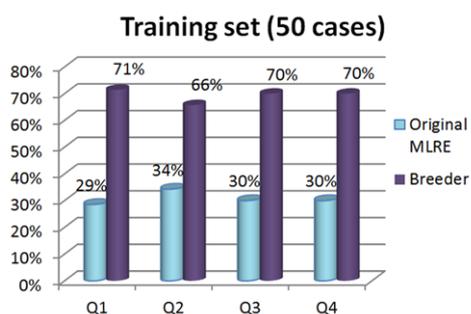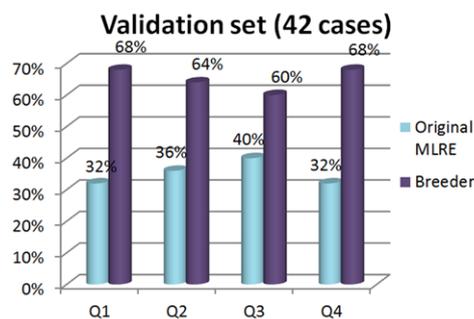

Figure 1: Predictions on the training set

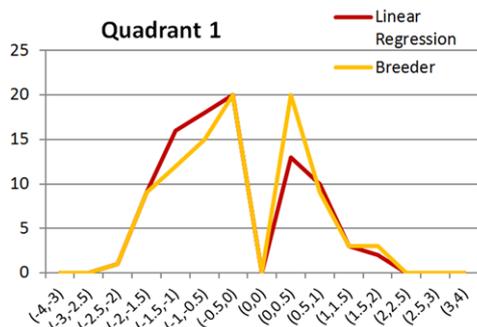Figure 2: Predictions on the validation set

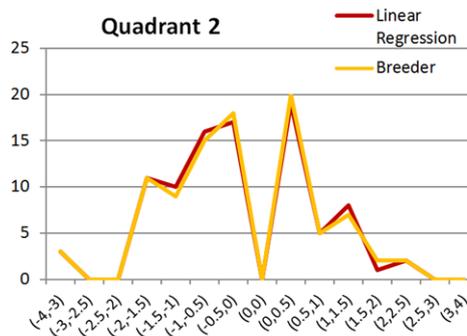Figure 3: The comparison of prediction error in quadrant 1

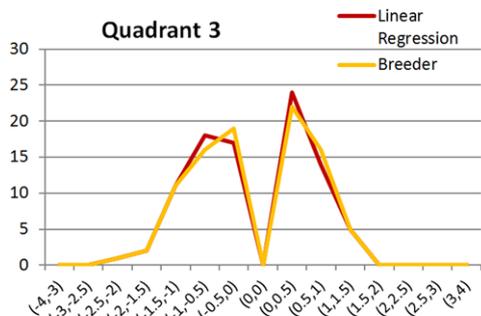Figure 4: The comparison of prediction error in quadrant 2

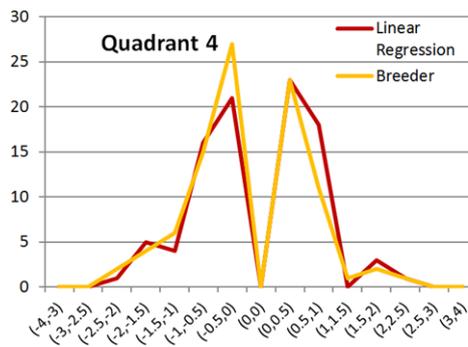Figure 5: The comparison of prediction error in quadrant 3

Figure 6: The comparison of prediction error in quadrant 4

4 Comparative analysis

The optimization using the Breeder genetic algorithm was made on all four quadrants, providing the following equations:

$$Y_{Q1} = 5.1917 + 0.7571 * X_{42} + 0.85332 * X_{46} + 0.28341 * X_{21}$$
$$Y_{Q2} = 5.16292 + 0.90463 * X_{42} + 0.68192 * X_{46} + 0.41011 * X_{21}$$
$$Y_{Q3} = 3.31241 + 0.89357 * X_{42} + 0.72022 * X_{46} + 0.51352 * X_{21}$$
$$Y_{Q4} = 3.28732 + 0.70242 * X_{42} + 0.84793 * X_{46} + 0.47736 * X_{21}$$

Figure 7: The optimized equations using genetic algorithm

where Y_{Qi} denote the outcome expected for the quadrant $i \in \{1, 2, 3, 4\}$ and X_{mn} represents the mesiodistal width of the tooth specified by index mn.

In our study, if the difference in millimetres between the measured and predicted value of the sum of the mesiodistal sizes of unerupted canines and premolars is situated in interval $[-0.75, 0.75]$, the prediction is considered as a correct estimation, if the difference is $< -0.75\ mm$ we have an overestimation, and a prediction error $> 0.75\ mm$ is considered an underestimation.

A comparison of correct estimations, overestimations and underestimations, provided by the original MLRE equations [3] and respectively by the optimized equations from figure 7, is presented in the table 4:

Canines premolars group	Method	over-estimations %	correct estimations %	under-estimations %
Maxillary	Original MLRE	35	51	14
	Breeder	32	54	14
Mandible	Original MLRE	24	63	13
	Breeder	21	66	13

Table 4: Correct estimations, overestimations and underestimations in percents

Maximum errors in predicting the sizes of the canines and premolars in the mandible and maxilla are presented in Table 5:

Quadrant	Original MLRE		Breeder	
	over-estimating	under-estimating	over-estimating	under estimating
1	-2.32	1.50	-2.11	1.23
2	-3.97	2.21	-3.56	1.84
3	-2.13	1.14	-1.86	1.01
4	-2.15	2.25	-2.01	1.93

Table 5: Maximum errors in estimating the sum of the mesiodistal
sizes of unerupted canines and premolars

Comparing predictions provided by the new and respectively old method, we can conclude that Breeder genetic algorithm is capable to provide the best values for parameters of multiple linear regression equations, and thus our equations are optimized for best performance. The results obtained by the new multiple linear regression equations are significant better than those provided by some classical statistical approaches [3], [15], [17].

The proposed technique is an adaptive tool for predicting the sizes of unerupted canines and premolars with greater accuracy than standard linear regression analyses, the fitness function ensuring optimization of predictions for data collected from different groups selected from different countries.

5   Conclusions

Using a Breeder genetic algorithm, we can automatically find the optimal values for the parameters of multiple linear regression equations used in prediction of the mesiodistal width of unerupted permanent canines and premolars.

After evaluation, we found that our new parameters, used in the regression equations, are providing a better prediction than original MLRE method.

Thus, the prediction error rates of the optimized equations using the Breeder genetic algorithm are smaller than those provided by the multiple linear regression equations proposed in [3].

Using a fitness function related to the prediction error provided by original linear regression equations, the evolution process guided by our implementation of Breeder genetic algorithm was capable to find new MLRE equations which outperform the original equations in terms of qualitative results of the prediction process.

Bibliography

[1] Moyers, R. E.; Handbook of orthodontics Chicago: Year Book Medical Publishers, 1988.

[2] Mühlenbein H., Schlierkamp-Voosen; D-The science of breeding and its application to the breeder genetic algorithm, *Evolutionary Computation*, 1:335-360, 1994.

[3] Boboc, A.; Dibbets, J.; Prediction of the mesiodistal width of unerupted canines and premolars: a statistical approach, *American J. of Orthodontics and Dentofacial Ortopedics*, 137(4):503-507, 2010.

[4] Pancherez, H.; Schaffer, C.; Individual-based prediction of the supporting zones in the permanent dentition. A comparison of the Moyers method with a unitary prediction value *J. Orofacial Orthopedic*, 60(4):227-2355, 1999.

[5] Legovic, M.; Novosel, A.; Legovic, A.; Regression equation for determining mesiodistal crown diameters of canines and premolars, *Angle Orthodontic*, 73(3):314-318, 2003.

[6] Legovic, M.; Novosel, A.; Skrinjaric, T.; Legovic, A.; Madi, B.; Ivancic, N.; A comparison of methods for predicting the size of unerupted permanent canines and premolars, *European Journal Orthodontic*, 28(5):485-490, 2006.

[7] Memon, S.; Fida, M.; Development of a prediction equation for the estimation of mandibilar canine and premolar widths from mandibular first permanent molar and incisor widths, *European Journal Orthodontic*, May 9 [Epub ahead to print], 2011.

[8] Alhaija Abu, E.S.; Qudeimat, M.A.; Mixed dentition space analysis in a Jordanian population: comparison of two methods, *International Journal Paediatric Dentistry*, 16(2):104-110, 2006.

[9] Aquino Melgaco, C.; Araujo de Sousa, M.T.; Roules de Oliveira, A.C. (2007); Mandibular permanent first molar and incisor width as predictor of mandibular canine and premolar width, *American J. of Orthodontics and Dentofacial Orthopedics*, 132(3):340-345, 2007.

[10] Moghimi, S.; Talebi, M.; Parisay, I.; Desing and implementation of a hybrid genetic algorithm and artificial neutral network system for predicting the sizes of unerupted canines and premolars, *European Journal Orthodontic*, 34(4):480-486, 2011.

[11] Van der Merwe, W.S.; Rossouw, P.; Van Wyk Kotze, T.J.; Trutero, H.; An adaptation of the Moyers mixed dentition space analysis for a Western Cape Caucasian population, *The Journal of Dental Association South Africa*, 46(9):475-479, 1999.

[12] Melgaco, C.A.; Araujo, M.T.; Ruellas, A.C.O.; Applicability of three tooth size prediction methods for white Brazilians, *Angle Orthodontic*, 76(4), 644-649, 2006.

[13] Barnabe, E.; Flores-Mir, C.; Apparaising number and clinical significance of regression equations to predict unerupted canines and premolars, *American Journal of Orthopedics and Dentofacial Orthopedics*, 126(2):228-230, 2004.

[14] Barnabe, E.; Flores-Mir, C.; Are the lower incisors the best predictors for the unerupted canine and premolars sums? *An analisis of a Peruvian sample, Angle Orthodontist*, 75(2):202-207, 2005.

[15] Martinelli, F.L.; De Lima, E.M.; Rocha, R.; De Sousa Araujo, M.T.; Prediction of lower permanent canine and premolars width by correlation methods, *Angle Orthodontist*, 75(3):805-808, 2005.

[16] Nourallah, A.W.; Gesch, D.; Khordaji, M.N.; Splieth, C. (2002); New ecuations for predicting the size of unerupted canines and premolars in contemporary population, *Angle Orthodontist*, 72(3):216-221, 2002.

[17] Bonetti, G.A.; Verganti, S.; Zamarini, M.; Bonetti, S.; Mixed dentition space analisis for a northern Italian population: new regression equations for unerupted teeth, *Progress in Ortodontics*, 12(2):94-96, 2011.

[18] Philip, N.I.; Prabhakar, M.; Arora, D.; Chopa, S.; Applicability of the Mojers mixed dentition probability tables and new prediction aids for a contemporary population in India, *American J. of Orthopedics and Dentofacial Orthopedics*, 138(3):339-345, 2011.

[19] Stoica, F.; Simian D.; Optimizing a New Nonlinear Reinforcement Scheme with Breeder genetic algorithm, *Proc. of the 11'th International Conference on EVOLUTIONARY COMPUTING (EC'10)*, Iaşi, Romania, ISSN: 1790-2769, ISBN: 978-960-474-194-6, 273-278, 2010.

[20] Stoica, F.; Cacovean, L.F.; Using genetic algorithms and simulation as decision support in marketing strategies and long-term production planning, *Proceedings of the 9'th International Conference on SIMULATION, MODELLING AND OPTIMIZATION (SMO'09)*, Budapest Tech, Hungary, ISSN: 1790-2769 ISBN: 978-960-474-113-7, 435-439, 2009.