Original Article

Cleft prediction before birth using deep neural network

Numan Shafi
Faisal Bukhari iD
Waheed Iqbal
University of the Punjab, Pakistan

Khaled Mohamad Almustafa iD
Prince Sultan University, Saudi Arabia

Muhammad Asif iD
National Textile University, Pakistan

Zubair Nawaz
University of the Punjab, Pakistan

Abstract

In developing countries like Pakistan, cleft surgery is expensive for families, and the child also experiences much pain. In this article, we propose a machine learning–based solution to avoid cleft in the mother's womb. The possibility of cleft lip and palate in embryos can be predicted before birth by using the proposed solution. We collected 1000 pregnant female samples from three different hospitals in Lahore, Punjab. A questionnaire has been designed to obtain a variety of data, such as gender, parenting, family history of cleft, the order of birth, the number of children, midwives counseling, miscarriage history, parent smoking, and physician visits. Different cleaning, scaling, and feature selection methods have been applied to the data collected. After selecting the best features from the cleft data, various machine learning algorithms were used, including random forest, *k*-nearest neighbor, decision tree, support vector machine, and multilayer perceptron. In our implementation, multilayer perceptron is a deep neural network, which yields excellent

Corresponding author:
Faisal Bukhari, Punjab University College of Information Technology (PUCIT), University of the Punjab, P.O. Box No. 54590 Lahore, Pakistan.
Email: faisal.bukhari@pucit.edu.pk

results for the cleft dataset compared to the other methods. We achieved 92.6% accuracy on test data based on the multilayer perceptron model. Our promising results of predictions would help to fight future clefts for children who would have cleft.

Keywords
cleft prediction, cleft lip, cleft palate, machine learning, multilayer perceptron, *k*-nearest neighbor, pre-birth prediction, deep neural network

Introduction

Cleft lip and cleft palate (CLP) are conditions in which there are openings present in the upper lip and the roof of the mouth, respectively, both having a standard opening extension to the nose. These clefts of lip and palate are collectively known as orofacial clefts (OFCs). When facial tissues fail to engage correctly during baby development, such clefts occur.

The cleft in newborn babies is found around the globe. Tanaka et al.[1] gathered cleft information from 34 states over the 5 years. After the statistical analysis, it is noted that a baby with a cleft rises day by day. Internationally, every 7.94 of 10,000 children are the subject of the cleft.

In Pakistan, epidemiological studies show that the incidence of this disease is higher, particularly in northern regions. A total of 117 instances of CLP of 61,156 were recognized with a more significant proportion of affected boys compared to girls. The total percentage recorded was 1.91 of 1000 live births. 0.1% to 0.2% of newborn babies have this disease. A cleft lip happens more frequently in men relative to women. Cleft palate without cleft lip is more prevalent in women.

Figure 1 shows two clefts (https://cleftandcraniofacialinstituteofutah.com/cleft-lip/). First, it is unilateral, meaning the lip is either left or right sliced. On both ends of the lip, the bilateral cleft is the split.

Figure 2 demonstrates that CLP have distinct levels of intensity. At the ordinary level, cleft lip or cleft palate affects the outer lip in terms of the cut on lip. In the second stage, this cut becomes slightly more serious and moves to the hard palate in the mouth. This cut causes the problem in feeding, and a newborn baby is deprived of colostrum (breast milk). This cut passes on at the third stage and affects the soft palate of the throat. The cut begins with the lip and goes to the start of the trachea. It creates both feeding and breathing problems.

CLP raise mortality rates in affected children and creates a financial risk to families and imposes social stress. During the growth of the child during pregnancy, the cellular membranes and single cells from each side of the head develop toward the middle of the face and join together to create the image. In the first trimester of pregnancy, it allows unique characteristics such as ears and teeth. Tragically, the cleft lip may happen if the tissues that make up the lip fail to unite correctly, resulting in an opening in the upper lip in the form of a cut or potentially an enormous scar that can lead to the nose.

There are numerous triggers in children that trigger cleft lip and palate. The onset of the disease involves both environmental and genetic variables. Some significant risk variables include
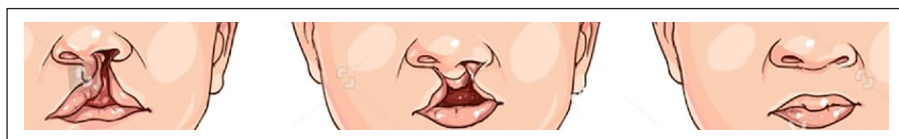

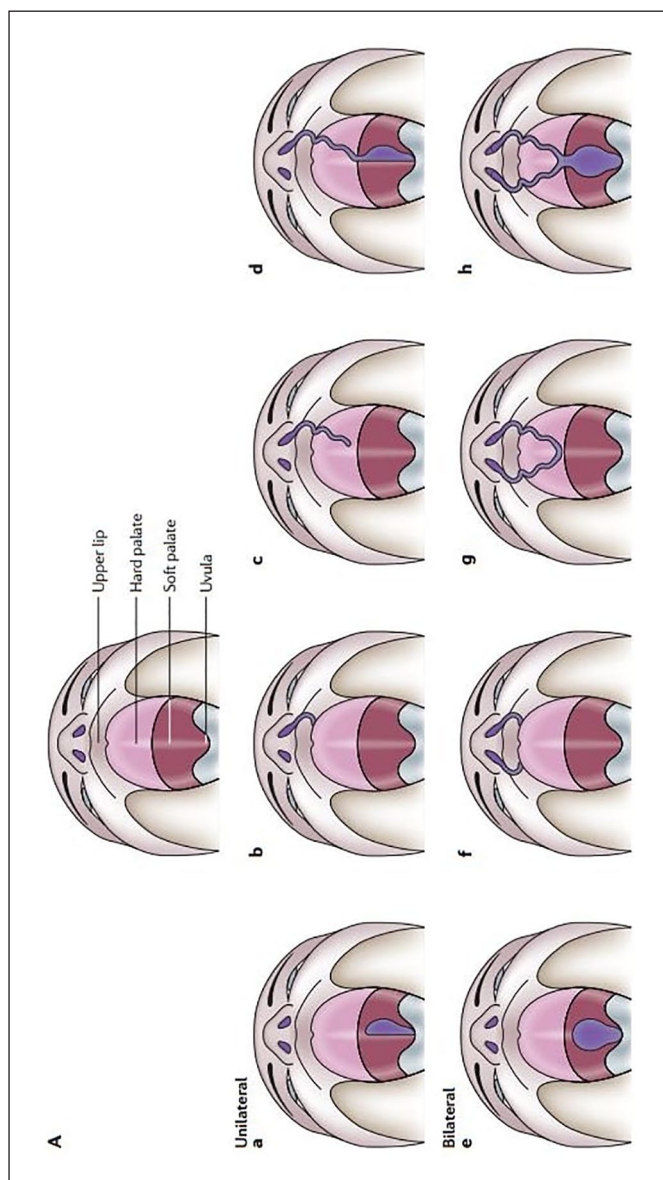**Figure 1.** Unilateral and bilateral cleft.

Figure 2. Types of clefts caused by genetic mutations: (a, e) show unilateral and bilateral clefts of soft palate, respectively, (b, c, d) show different degrees of unilateral cleft lip and palate, and (f, g, h) show different degrees of bilateral cleft lip and palate. Reprinted from Dixon et al.[2]

maternal smoking during pregnancy, stress, diabetes, epilepsy, obesity, elderly mother, family cleft history, and certain medicines. In children, cleft lip and palate can contribute to problems related to feeding, speech, hearing, ear infections, poorly placed teeth, nose and mouth asymmetry, oral hygiene, aesthetic issues, development, and also a psycho-social burden. Genetic factors include some transcription factors, mistakes, and multiple genes of epistasis.

In this article, we proposed a machine learning–based solution to identify cleft in the mother's womb. We collected 1000 pregnant female samples related to the cleft. A questionnaire has been designed to obtain a variety of data, such as gender, parenting, family history of cleft, the order of birth, the number of children, midwives counseling, miscarriage history, parent smoking, and physician visits. Different cleaning, scaling, and feature selection techniques were applied. After that, we applied different machine learning algorithms to predict the cleft. From different machine learning algorithms, our primary outcome is to predict the cleft with the accuracy of the 92.6%. Our proposed machine learning–based solution is capable of predicting the cleft lip and palate before the birth of the babies. The work presented in this article will also be helpful to improve the awareness for pregnant mothers to combat the future clefts.

The exact contributions of this article include the following:

1. Identify the critical factors which cause the cleft in newborn babies.
2. Dataset collection and preparation for cleft prediction.
3. Employ machine learning algorithms on the cleft dataset to predict the pre-birth cleft.
4. Use a deep neural network to improve the cleft prediction accuracy.

The rest of the article is organized as follows. Section "Related work" is about related work. The data collection method is explained in section "Data collection method." The preparation of the cleft dataset is discussed in section "Preparation of the cleft dataset." Predictive methods are described in section "Predictive methods." The result is presented in section "Results." Finally, conclusions are drawn, and future work is discussed in section "Conclusion and future work."

Related work

The creation of the cleft lip and palate is based on different factors. These factors are separately studied by various researchers. Little et al.[3] examined that smoking in females increases the chances of OFCs. Correa et al.[4] stated that diabetes mellitus is the root of many diseases regarding abnormalities and leads to the death of many infants, mostly in the United States. Werler et al. reported that seizures and epilepsy are the leading diseases to congenital disabilities. Analyzing a large number of the patients, the confidence interval (CI) was found to be 95%, and the most dangerous drug proved was valproic acid.[5] Dixon et al. studied that the genetic and environmental factors also influence the typical morphology of facial features and leads to CLP.

Different surveys and statistical experiments are performed to know the effect of the cleft on the life of the children. Ranta carried out a study in children that the formation of teeth is also affected by CLP. The experimental group showed 2.65% frequency, while the control group was with 2.83% frequency. It was concluded that the incisor's abnormality is also linked with non-cleft families, and it is not a microform of CLP.[6] Maier et al. reported that two thirds of facial defects and 80% OFCs are due to CLP. By the PLAKSS speech test, the correlation value for the automatic system was found to be 0.89, and after evaluation, it was 0.81.[7] Jocelyn et al.[8] reported that children with CLP also face communication problems (hearing and speech), cognition problems as well as social and academic issues. Asher-McDade et al.[9] assessed patients with CLP, usually by study models involving speech, hearing tests, and radiographs. Noar carried out a questionnaire survey

to know the concerns of parents and their children about their clefts. The questionnaire consisted of four different variables regarding facial appearance and speech, treatment aspects, social and environmental aspects, and the success of specialists.[10]

Various methods are used for CLP therapy. Maier et al. noted that children's speech remains unhealed even after surgery, so a reliable technique was adopted to verify phenome-level phonetic disorders. PLAKSS voice test gathered and evaluated voice information from 58 kids. During the experiment, multiple properties were assessed, including recognition, summon mapping, pronunciation, prosodic features, and energy profiles.[7] Semb outlined that craniofacial developments must be carefully handled before the child's school age, as it impacts the child's personality and makes him or her depressed by their abnormality.[11] Millard et al. compared the past therapy protocol with the latest therapy protocol for cleft lip and palate pre-surgical therapy. A new protocol called presurgical orthopedics followed by periosteoplasty and lip adhesion (POPLA) has been applied.[12] Friede and Enemark[13] researched that mid-facial development in patients with CLP is better than those who used surgical methods to delay the repair of the hard palate and push back palatal closure techniques. Geft Palaterayson and Shetor Clye showed a pre-surgical nasal-alveolar molding (NAM) method to treat patients with CLP. The NAM method was generally carried out after birth at the age of 3–4 months. Nose and alveolus were pre-surgically shaped during this method.[14] Smith conducted a dental arch survey in 88 patients with single cleft lip and palate in the Nijmegen Cleft Palate Center using Golson yardstick. The patients got pre-surgical orthopedic therapy of the warm sort. The operation was performed at the age of 6 months.[15]

To the best of our knowledge, there is no associated work on cleft prediction using machine learning algorithms and no publicly accessible dataset related to the cleft. This is an entirely new area of research in which we used different machine learning algorithms to predict the cleft before the birth of the child, in connection with variables that trigger the cleft. The correct cleft prediction will not only make baby life more comfortable but also ease the parents' difficulties. Earlier detection of clefts will decrease the likelihood of cleft surgical solution.

Data collection method

A total of 1000 mothers were sampled from three separate hospitals in Lahore, Pakistan, including Children Hospital Lahore, General Hospital Lahore, and Mayo Hospital Lahore. Two groups were created from 1000 sample points. First, the experimental group and second, the control group. The experimental group had 500 mothers whose children had clefts in their lip/palate. The control group also had 500 mothers whose children were normal and healthy. The objective of establishing two groups was to normalize data and avoid overfitting and underfitting the outcomes at the time of prediction. A questionnaire was designed, and questions were asked from mothers including parent gender, parent relationship, clefts family history, birth order, the number of children, midwife counseling, history of miscarriage, parent smoking history, doctor visits, and so on.

Input feature

Table 1 shows the input parameters selected to predict the cleft. The majority of input parameters are binary (0 or 1), with specific features being number values. In cleft prediction, we used these 36 input features.

Figure 3 demonstrates the combined visualization of various input parameters. Male versus female bar of Figure 3 shows that CLP occur comparatively more in males than in females. Cousin vs non-cousin bar indicates that 60.2% are cousin marriages and 39.8% are non-cousin marriages, which implies that cousin matrimony is the primary factor in the development of the CLP. Smoker

Table 1. Input feature for the prediction of cleft lip and cleft palate.

| Input features	Type	Range	Description
gender (q1)	Binary	0/1	Male or female?
parentRelation (q2)	Binary	0/1	Cousin or not cousin?
familyHistoryOfCleft (q3)	Binary	0/1	Is there any family history of cleft?
birthOrder (q4)	Number	0 to 5	Order of the affected child?
totalChildren (q5)	Number	0 to 5	How many children in the affected family?
visitsToDoctor (q6)	Number	1 to 5	The total number of visits to doctors a month?
midwifeConsulted (q7)	Binary	0/1	Is the mother having a midwife during the pregnancy?
motherGetVaccinated (q8)	Binary	0/1	During pregnancy does mother receive vaccination?
historyOfMiscarriage (q9)	Binary	0/1	How many children died before they were born?
smokingOtherDrugHistoryOfParent (q10)	Binary	0/1	Are parents taking any sort of drug like alcohol and so on?
cigarettesUsedPerDay (q11)	Number	0 to 15	How many cigarettes a mother smoked in a day?
motherAnemia (q12)	Binary	0/1	Does mother have anemia?
motherBleeding (q13)	Binary	0/1	Does mother experience an issue with bleeding during pregnancy?
motherDepression (q14)	Binary	0/1	Is mother suffering from depression during pregnancy?
motherDiabetes (q15)	Binary	0/1	During or before childbirth, does the mother experience diabetes?
motherHypertension (16)	Binary	0/1	Does mother experience tension before or during pregnancy?
motherInfection (q17)	Binary	0/1	Is there an energetic infection in the mother?
motherStress (q18)	Binary	0/1	Does mother have a problem with stress during pregnancy?
motherVomiting (q19)	Binary	0/1	Does mother face vomiting frequently during pregnancy?
fatherHasAsthma (q20)	Binary	0/1	Does father have asthma?
fatherHasCancer (q21)	Binary	0/1	Does father have cancer?
fatherHasDiabetes (q22)	Binary	0/1	Does father have diabetic?
fatherHasHcv (q23)	Binary	0/1	Does father have HCV?
fatherHasHeart (q24)	Binary	0/1	Does father have heart problem?
fatherHasHypertension (q25)	Binary	0/1	Does father have hypertension?
familyStautsOfLower (q26)	Binary	0/1	Does the family have less income?
familyStautsOfMiddle (q27)	Binary	0/1	Does the family have the status of a middle class?
familyStautsOfUpper (q28)	Binary	0/1	Does the family have an upper-class status?
medOfAnemia (q29)	Binary	0/1	Does mother get anemia medicine?
medOfBleeding (q30)	Binary	0/1	Does the mother receive bleeding medicine?
medOfDepression (q31)	Binary	0/1	Does the mother receive depression medicine?
medOfDiabetes (q32)	Binary	0/1	Does the mother receive diabetes medicine?
medOfHypertension (q33)	Binary	0/1	Does the mother receive hypertension medicine?
medOfInfection (q34)	Binary	0/1	Does the mother receive infection medicine?
medOfStress (q35)	Binary	0/1	Does the mother receive stress medicine?
medOfVomiting (q36)	Binary	0/1	Does the mother receive vomiting medicine?

HCV: hepatitis C virus.
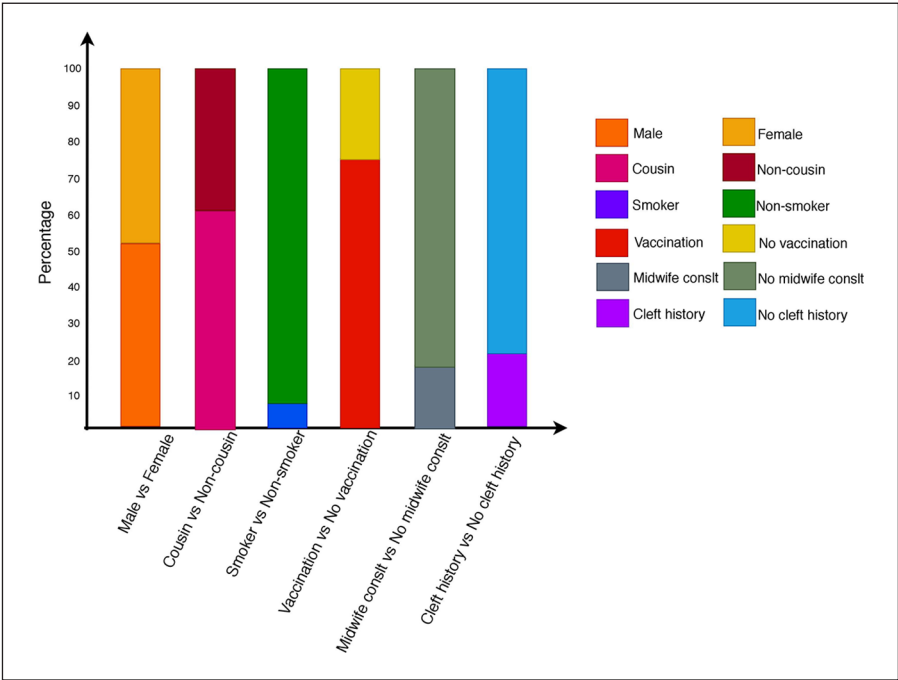
Figure 3. A statistical view of different input parameters.

vs non-smoker bar shows 7.4% of smoking mothers, but drug use in the USA has a higher level of pregnancy relative to Pakistan. While the smoking parents' proportion is small, children globally suffer from the cleft whose mothers smoked during the pregnancy. Vaccination vs no vaccination bar indicates that 74.8% of mothers were vaccinated before the baby was born. Taking no vaccination will cause many complications for the mother and newborn baby. Midwife consultation vs no midwife bar graph indicates that most (80.8%) of the mothers do not use midwife facilities during pregnancy. Cleft history vs no cleft history bar of Figure 3 shows that 79.2% of couples do not have a cleft baby family history.

Preparation of the cleft dataset

The cleft dataset was not publicly available for predicting the cleft in children in Lahore. A questionnaire was designed to overcome this challenge. It was based on issues such as gender, parenting, cleft family history, birth order, the number of children, counseling of midwives, history of miscarriage, smoking of parents, and visits to doctors. Each question has been read and explained to both group mothers. The responses have been registered. It was tough to build a dataset since we had at first to obtain data manually from the three different hospitals in Lahore and then transform it into a soft shape. Three primary steps were taken in the creation of the data.

Formatting

Data formatting was crucial as data were raw, and each characteristic had a distinct textual name. Gender, for instance, had two genres, male and female, which were transformed into binary format. Similarly, smoking history, birth orders, and all the other relevant attributes were converted to a

binary scale. Responses with yes were regarded as "1" and no was allocated the number "0" according to binary format. We have applied different machine learning algorithms after translating textual information into a binary standard.

Data cleaning

In this step, we mainly address the missing values and unwanted characters. Missing values were replaced by the mostly occurred value of that attribute subject to its related output, which means that we picked all the example points which have the same output to that example point, and we replace the missing value with mostly occurred values of the attribute. We deleted that example point in which more than two of the attributes were found missing. A disease in mother or father was recognized as 1, and no disease was recognized as 0. Because the disease is binary, there are only two possible chances that it will happen or not. Because the disease has a binary relevance, there are only two possible chances that the disease may or may not happen. We have, therefore, substituted all other values by 0, which show that it is not for the mother or father of the patient.

Feature extraction

We applied the Variance Threshold and SelectKBest feature selection methods. Features with small variance are eliminated in the Variance Threshold method. Only 12 of the 36 features took part in cleft predictions and achieved an accuracy of 86.95% on unseen data. In SelectKBest, the only features that contribute more to the target value are selected. We have chosen 19 features from 36 attributed and obtained 86.32% accuracy on test data, which is significantly lower than the precision of the 92.6% multilayer perceptron (MLP) model. So, we used MLP model features to predict the cleft.

Predictive methods

MLP model

An MLP is a feedforward artificial neural network classification. MLP contains more than one layer; the signal input layer, the signal output layer, makes a decision about that signal. Between these two layers, the junk of different hidden layers is an MLP computing engine. Perceptron is a linear classifier algorithm to divide a straight line input into two parts. If an MLP has a linear activation function in each perceptron, linear algebra indicates that any amount of layers can be lowered to a two-layer input-output layer. Figure 4 demonstrates the MLP model used before childbirth for cleft prediction. The input layer uses a total of 36 features and three hidden layers. The output is binary, meaning that the patient may or may not have a cleft. Since MLP is based on learning techniques for backpropagation, error on output node z with n example points is calculated from equation (1), where b is the target and y is the predicted value

$$e_z(n) = b_z(n) - y_z(n) \tag{1}$$

After calculation, equation (2) adapts the perceptron weights to minimize the output layer error where $\xi(n)$ is the total instantaneous error

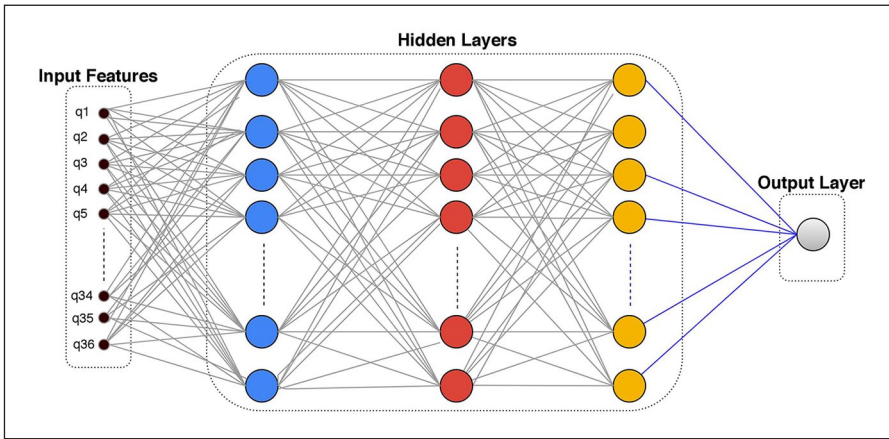$$\xi(n) = \frac{1}{2} \sum_z e_z^2(n) \tag{2}$$

Figure 4. MLP design with hidden layers used for cleft prediction.

This model uses the rectified linear unit (ReLU) as an activation function. Equation (3) depicts ReLU's mathematical model where x is the perceptron input

$$y = x^+ = max(0, x) \tag{3}$$

Hyperbolic tangent and logistic function are also used as activation functions, but ReLU converges quickly, and as with the other activation functions, there is no vanishing gradient problem with a higher value x. ReLU is acquired after all activation functions have been tuned by hyperparameter. Through hyperparameter tuning, all parameters on which this model is prepared are acquired. Epoch is a term widely used for the neural network. One epoch means that you are given example points at once. Epochs are denoted by "e," that is, a number of times data are provided. A number of epochs are used from 0 to 250. When the epoch's value is 220, the best accuracy is acquired.

K-*nearest neighbor algorithm*

K-nearest neighbor (KNN) is used in statistical estimation and pattern recognition; it is commonly used for supervised learning. Data are classified by distance function. Distance from k classes is calculated, and a point is assigned to a class with the shortest distance from that point

$$P(y = l \mid X = x) = \frac{1}{K} \sum_{i=B} I(y^{(i)} = l)$$

Here, input x gets assigned to the class with the largest probability, depending on the distance between two data points. The Euclidean distance is a popular option

$$d(x, x') = \sqrt{(x_1 - x_1')^2 + (x_2 - x_2')^2 + \cdots + (x_n - x_n')^2}$$

The above equation is used for Euclidean distance calculation where $x_1, x_2, \ldots, x_n$ shows the new points and $x_1', x_2', \ldots, x_n'$ shows the existing points up to n.

Decision tree classifier

Decision trees (DTs) used for both classification and regression are very versatile. The flow of this sort of tree is downward. It operates with conduct of "If this then that." DTs are simple to interpret, fast, and suitable for big datasets. The DT provides an optimum solution for each step without at the last stage, determining the optimum solution. DT classifier is a tree-based structure algorithm. The topmost node is root, branches are indicated by decision rules, and the leaf node is signaled by the output. The tree is recursively partitioned.

Support vector machine

Support vector machine (SVM) is also regarded as a vector support machine, which analyzes classification data. In H dimensional space, SVM is also used for finding a hyperplane. SVM works with a subset of training points for high dimensional spaces. It does not perform well in the event of a large dataset and overlapping. Equation (4) is the equation of the hyperplane where $x = (x_1, x_2)$, $x_1 = x$, $x_2 = y$, w is the weight and b is the slope

$$w \cdot x + b = 0 \tag{4}$$

$$h(x_i) = \begin{cases} +1 & w \cdot x + b \geq 0 \\ -1 & w \cdot x + b < 0 \end{cases} \tag{5}$$

In above equation (5) $h(x_i)$ is hypothesis function. An example point is assigned to class $+1$ if it lies on or above the plane otherwise assigned to class -1. The aim of the SVM learning algorithm is basically to find a hyperplane that can correctly distinguish the data. Many such hyperplanes could be found. Moreover, we must discover the best one, often called the optimal hyperplane.

Random forest

The collection of DTs is random forests (RFs), also known as random DTs. It is used for clustering, selection of features, and statistical inference. Numerical and categorical data are used for these forests. The issue is that RF is slow and has problems of overfitting. The term forest indicates that trees are collected in some locations, and this is the same thing in the RF algorithm. When new data are obtained for classification, numerous trees are created, as each tree classifies the point according to decision-making rules. As a consequence, a new point is assigned to the class that has the highest number of tree votes.

We used Python's scikit-learn library to apply the above-mentioned machine learning algorithms as the library provides optimized implementations of machine learning algorithms.

Results

Figures 5 to 7 explain the data collected from the three different hospitals of Lahore, Pakistan. The total number of children per family is shown in Figure 5. Figure 6 demonstrates that more complications happen when a first baby is born and that more mothers come to the hospital for treatment after a second and subsequent pregnancy. Figure 6 also shows how many miscarriages in subsequent pregnancies happened. The results show that after the first pregnancy, the proportion of
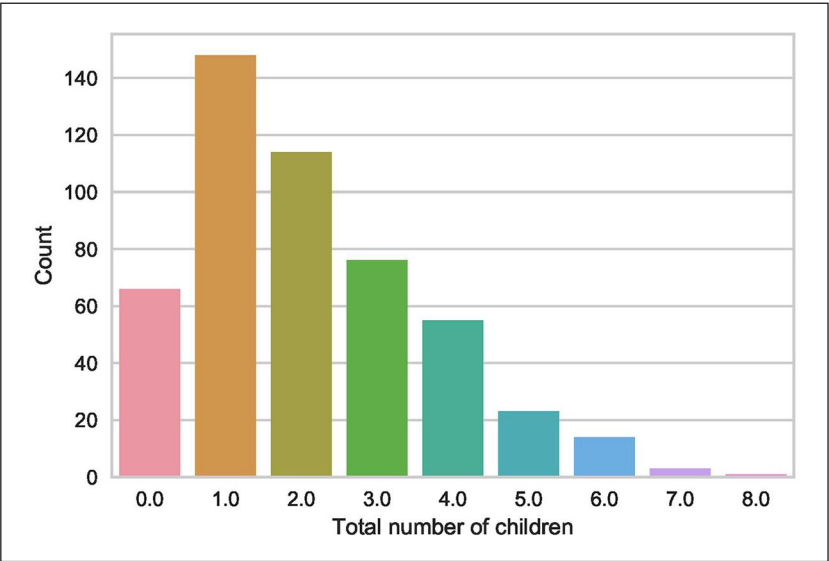
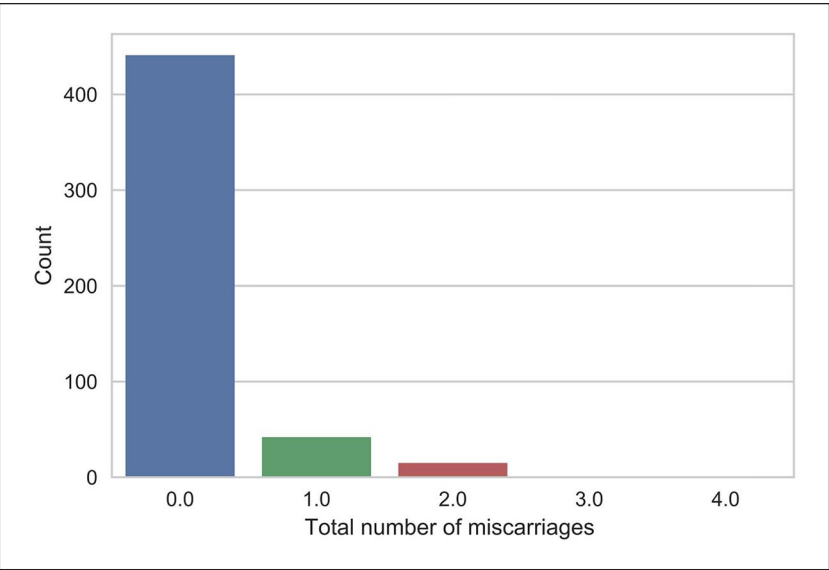**Figure 5.** Count of a different number of children.

Figure 6. Count of miscarriages to the present mother's pregnancy.

miscarriages shrinks. Figure 7 shows how many times in a month, the mother visited the doctor. The results indicate that more mothers visited the doctor only once a month.

Table 2 shows the *p* value of the different parameters of the cleft. Of 36 features, 30 are significant to predict the cleft and the remaining six are less critical in the prediction.

Table 3 shows the accuracy, precision, recall, *F*-measure, and hyperparameter settings of various algorithms for predicting the cleft.
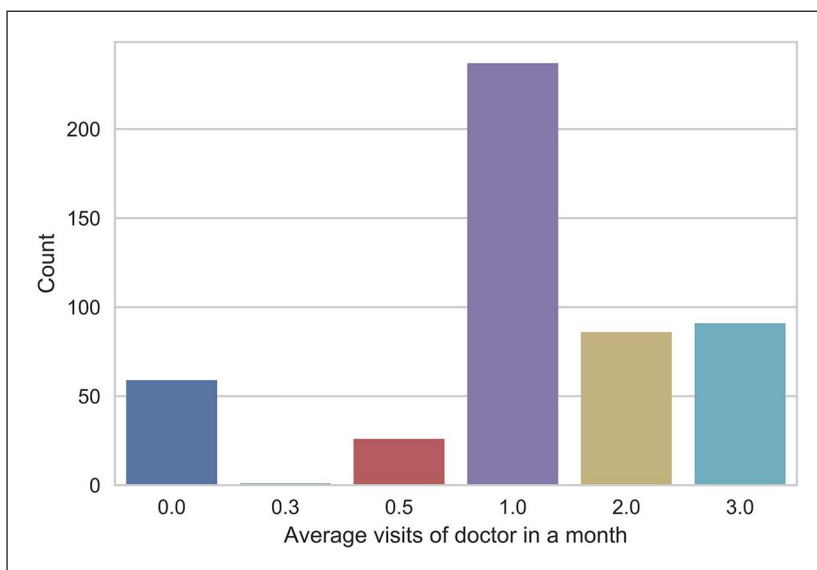
Figure 7. A number of doctor visits per month.

DT classifier gave 88.14% accuracy. Its accuracy, recall, and the *F*-measure, precision, and recall scores are 0.88. F-measure is a harmonic mean for precision and recall. Therefore, its value is the same as precision and recall. Hyperparameter tuning indicates a maximum depth of 8, which implies a maximum number of root-to-leaf nodes is 8. Limiting the depth prevents tree overfitting. Min sample leaf is 1, which means that a minimum of 1 sample can be considered for the leaf. Min sample split is 2, which implies that a node must contain two samples for splitting. The RF classifier gave 85.77% the smallest accuracy in all classifiers, and 0.86 is the precision, recall, and *F*-measure score. The maximum depth is four with hyperparameter tuning, which implies the maximum number of root-to-leaf nodes is 4, and the *n* estimator is 4, which implies that there are four trees which contributed to forest development. *K*-neighbor classifier provided 89.72% precision when *k* value is 9. *K* value indicates how smoothly the border is defined. KNN accuracy is better than an RF classifier and DT classifier. Precision, recall, and the *F*-measure score is 0.90. SVM gave the 90.69% accuracy that is better than all the algorithms discussed above. Precision, recall, and the score for *F*-measure is 0.92. With SVM's hyperparameter tuning where *c* is 100, gamma is 0.01, and kernel is radial basis function (RBS), we obtained this accuracy. On the MLP model, we found the highest accuracy of 92.6%. Precision, recall, and *F*-measure is 0.89. This best accuracy is achieved by tuning MLP hyperparameters. We used three hidden layers and 28 perceptrons on each layer. We identified these best settings by doing an exhaustive search.

The number of layers in MLP models plays a significant role in predicting the target value, which is hard to identify. Figure 8 shows the effect of accuracy on a different number of layers in the proposed CLP-MLP model. We variate the number of layers from 1 to 10 and then computed the accuracy of the model. Three hidden layers provide a maximum of 92.4 accuracy, whereas minimum accuracy is noted for a number of layers 1 and 6 that are below 90%. Therefore, we used three hidden layers for predicting the cleft in our MLP model.

Figure 9 shows the MLP model's accuracy with a different number of perceptrons on each layer ranging from 5 to 40. Maximum accuracy is noted when there are 28 perceptrons, which is 92.4,

Table 2. Significance of cleft features based on p value using the level of significance at $\alpha = 0.05$.

Features	p value	Significant
birthOrder	0.3194	No
familyHistoryOfCleft	0.0126	Yes
gender	0.0430	Yes
historyOfMiscarriage	0.0100	Yes
parentRelation	0.0419	Yes
totalChildren	0.2274	No
visitsToTheDoctor	0.0124	Yes
cigerettesUsedPerDay	0.0059	Yes
fatherAsthma	0.0286	Yes
fatherCancer	0.0237	Yes
fatherDiabetes	0.0128	Yes
fatherHcv	0.0158	Yes
fatherHeart	0.0449	Yes
fatherHypertension	0.1998	No
familyStatusLower	0.2225	No
familyStatusMiddle	0.7121	No
familyStatusUpper	0.0375	Yes
motherAnemia	0.0172	Yes
motherBleeding	0.0178	Yes
motherDepression	0.0316	Yes
motherDiabetes	0.0059	Yes
motherHypertension	0.1058	No
motherInfection	0.0039	Yes
motherSstress	0.0178	Yes
motherVomiting	0.0196	Yes
medOfAnemia	0.0356	Yes
medOfBleeding	0.0059	Yes
medOfDepression	0.0098	Yes
medOfDiabetes	0.0029	Yes
medOfHypertension	0.0405	Yes
medOfInfection	0.0009	Yes
medOfStress	0.0049	Yes
medOfVomiting	0.0064	Yes
midwifeConsulted	0.0028	Yes
motherGetVaccinated	0.0074	Yes
otherDrugHistoryOfParent	0.0375	Yes

which implies that if we set 28 perceptrons in each layer, the maximum precision of MLP can be obtained and the minimum precision is observed at 7 below 90.5.

Figure 10 shows the area under the curve (AUC) that explains how correctly the data are classified using MLP for the cleft dataset. It implies we are checking how many patients are having cleft and classified as the cleft patient and vice versa. In Figure 10, the x-axis indicates the false positive rate and the y-axis, on a scale between 0 and 1, indicates the true positive rate. When the value is closer to 0, the data are wrongly classified, 0 is labeled with 1, and 1 is labeled with 0.

Table 3. Accuracy of different machine learning algorithms applied to cleft data.

Algorithms	Accuracy	Precision	Recall	*F*-measure	Hyperparameters
Decision tree classifier	88.14	0.88	0.88	0.88	Max depth is 8, min sample leaf is 1, and min sample split is 2.
Random forest classifier	85.77	0.86	0.86	0.86	Max depth is 4 and n_estimators are 4.
K-neighbor classifier	89.72	0.90	0.90	0.90	n_neighbors are 9.
SVM	90.69	0.92	0.92	0.92	C is 100, gamma is 0.01, and kernel is RBF.
MLPClassifier	92.6	0.89	0.89	0.89	Activation is ReLU, alpha is 0.05, the number of hidden layers is 3, hidden layer size is (28, 28, 28), learning rate is adaptive, and solver is adam.

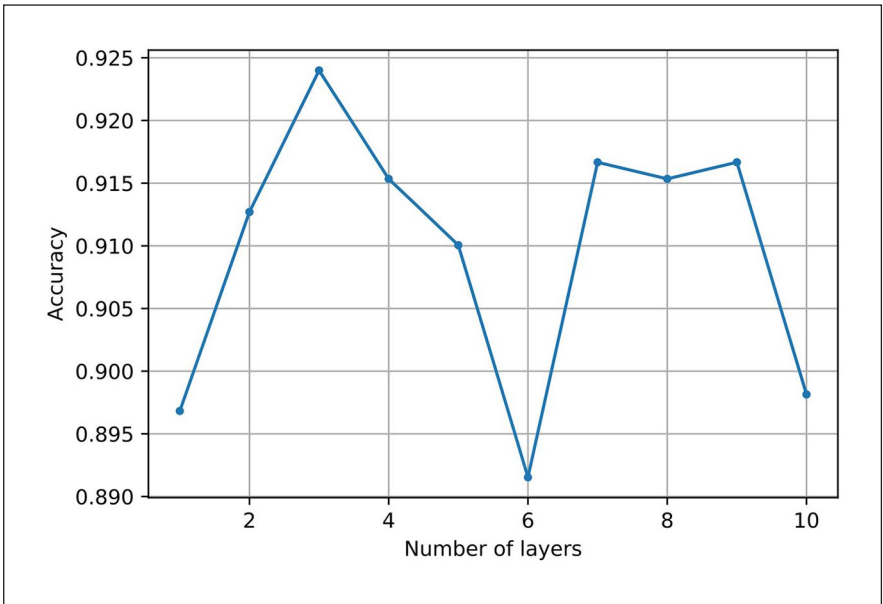SVM: support vector machine; MLP: multilayer perceptron.

Figure 8. Accuracy with a different number of layers for cleft prediction.

If the value is 0.5, then half is correctly labeled and half is incorrectly labeled. If the value is close to 1, then the data are accurately classified. In Figure 10, AUC is 0.98, which is closer to 1 so that data are more accurately and correctly classified.

Figure 11 indicates epoch tuning ranging from 0 to 250. Epoch is the neural network term. One epoch is to pass all the dataset to the network once and then update the weights. In the beginning, the graph more fluctuates from 0 to 10 epochs, the graph quickly changes, and the accuracy increases rapidly. Subsequently, when epochs are 220, the graph becomes smooth, and maximum precision is attained, which implies that if we pass the dataset 220 times over to the neural network, we are more accurately known than weight and can achieve the best accuracy.
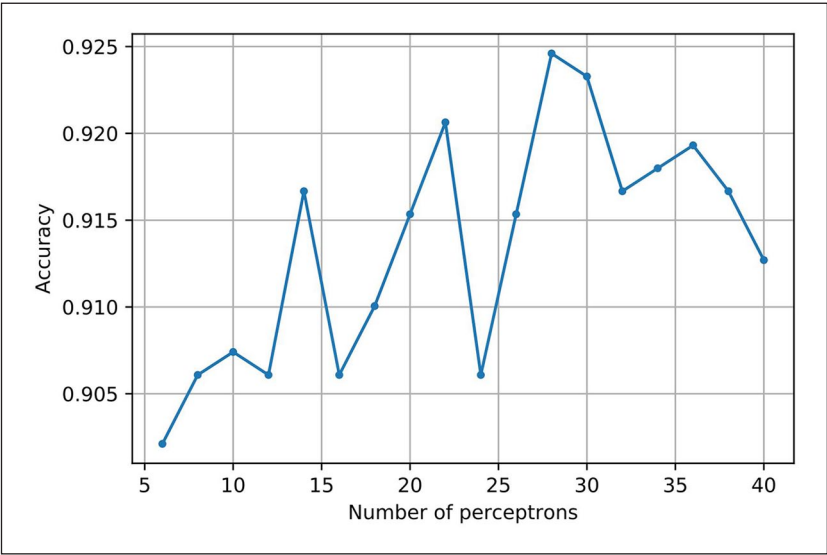
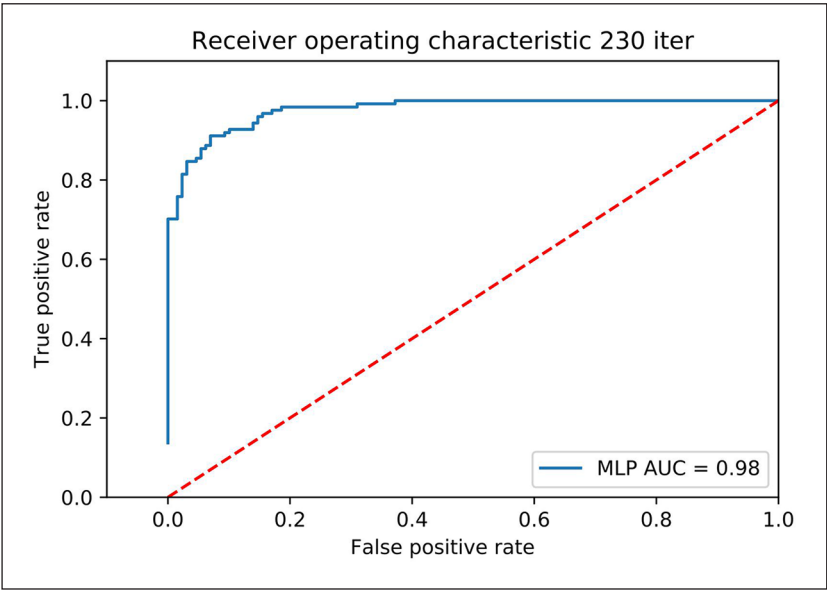**Figure 9.** Accuracy for cleft prediction with a different number of perceptrons.

Figure 10. AUC to check how much accurately cleft data are classified.

Discussion

Cleft prediction prior to birth in babies is a challenging task. There are various parameters involved in building clefts. In this work, we have collected a dataset containing important factors contributing to building clefts. Factors like gender, parent relation, family history of clefts, birth order, the number of children, midwife consultancy, medicine used by mother, medicine
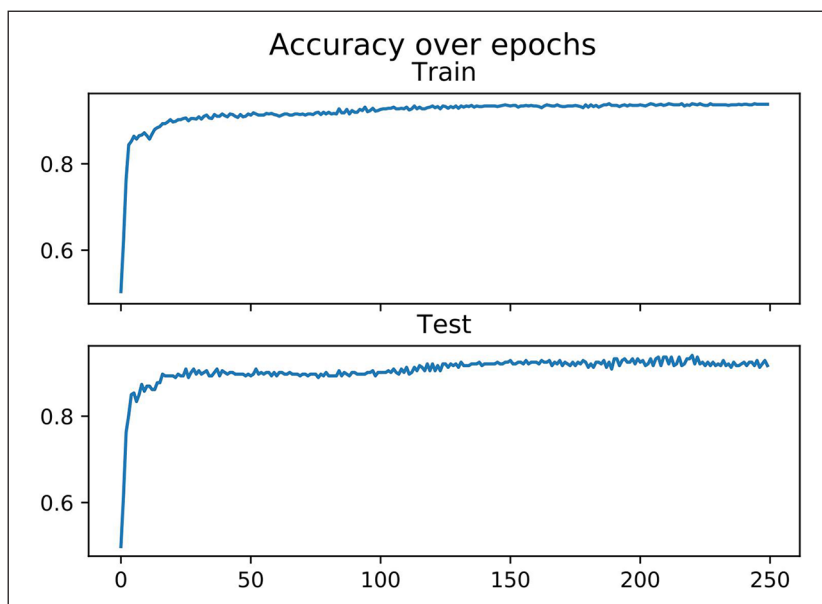
Figure 11. Accuracy with different numbers of epochs in MLP model.

used by father, miscarriage history, smoking history of parents, and visits to doctor participated in cleft prediction. After performing the *p* value test, we discovered that usage of medicine during pregnancy, smoking, parent relation, family history, and gender plays a more critical role in predicting the cleft.

We have applied various machine learning techniques on the collected dataset and identified that deep neural network (MLP model) performs best for cleft prediction. We identified the best parameters for the proposed MLP model using the exhaustive searching technique. The number of layers in MLP models plays a significant role in predicting the target value. After testing results on different numbers of layers, three hidden layers provide a maximum of 92.4 accuracy. MLP model's accuracy with a different number of perceptrons on each layer gave maximum accuracy when there are 28 perceptrons, which is 92.4. In the MLP model, accuracy is also tested with different numbers of epochs. When epochs were 220, the graph became smooth, and maximum precision was attained. After hyperparameter tuning, we combined the best parameters in the MLP model and tested the accuracy of the cleft prediction, which is 92.6%. AUC is also created to check how correctly the data are classified using MLP for the cleft dataset and 98% data correctly classified.

This is the first work to use machine learning for cleft prediction in babies prior to birth. Our effort, presented in this article, will significantly help the individuals as well as healthcare providers to predict the cleft timely and take appropriate measures to minimize the cleft in newborn babies. Our contribution toward the medical field will also help the new researcher to discover more hidden patterns in predicting similar types of disease, which still required attention from the machine learning community.

Conclusion and future work

Cleft prediction before birth is a challenging task. In this article, we address this challenge by collecting a dataset consist of 1000 samples and identify critical features that can be used to build a

cleft prediction model with good accuracy. We have evaluated various machine learning methods to evaluate the cleft prediction. Our experimental evaluations show that the CLP-MLP is a better model for cleft data classification, which yields 92.6% accuracy on unseen test data. Our research reported in this article is to predict cleft before birth. Once a cleft is produced, it cannot be treated but can be minimized by avoiding the use of medicine and drugs during pregnancy. Moreover, if the mother is suffering from stress, anxiety, epilepsy, and anemia, then she should not take any medicine related to these diseases without the suggestion of the doctor to avoid the cleft. For families having cleft previously, they should pay a regular visit to the doctor to reduce the chances of cleft in newborn babies. Our research can help identify the chances of the cleft and then take the necessary medical attention from the doctor to avoid it.

In the future, we intend to increase the dataset size to improve the model. We also aim to build a mobile application for pregnant women and healthcare providers to use cleft predictions.

Author contributors

F.B. gave the paper's main idea and undertook the paper. He also contributed to the writing of the paper and the statistical analysis of it. N.S. contributed in terms of data collection, cleaning, scaling, and analyzing results and in writing the paper. W.I. helped in applying different machine learning approaches to predict the cleft. M.A. and Z.N. reviewed the paper and helped in data collection, normalizing, and scaling the data. K.M.A. helped in addressing the comments of the reviewers in the revision of the paper. He also helped in improving English of the paper and doing the statistical analysis.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Faisal Bukhari [iD] https://orcid.org/0000-0002-7703-9742

Khaled Mohamad Almustafa [iD] https://orcid.org/0000-0003-2129-7686

Muhammad Asif [iD] https://orcid.org/0000-0003-1839-2527

References

1. Tanaka SA, Mahabir RC, Jupiter DC, et al. Updating the epidemiology of cleft lip with or without cleft palate. *Plast Reconstr Surg* 2012; 129(3): 511e–518e.
2. Dixon MJ, Marazita ML, Beaty TH, et al. Cleft lip and palate: understanding genetic and environmental influences. *Nat Rev Genet* 2011; 12(3): 167–178.
3. Little J, Cardy A and Munger RG. Tobacco smoking and oral clefts: a meta-analysis. *Bull World Health Organ* 2004; 82(3): 213–218.
4. Correa A, Gilboa SM, Besser LM, et al. Diabetes mellitus and birth defects. *Am J Obstet Gynecol* 2008; 199(3): 237.e1–237.e9.

5. Werler MM, Ahrens KA, Bosco JL, et al. Use of antiepileptic medications in pregnancy in relation to risks of birth defects. *Ann Epidemiol* 2011; 21(11): 842–850.
6. Ranta R. A review of tooth formation in children with cleft lip/palate. *Am J Orthod Dentofacial Orthop* 1986; 90(1): 11–18.
7. Maier A, Hönig F, Bocklet T, et al. Automatic detection of articulation disorders in children with cleft lip and palate. *J Acoust Soc Am* 2009; 126(5): 2589–2602.
8. Jocelyn LJ, Penko MA and Rode HL. Cognition, communication, and hearing in young children with cleft lip and palate and in control children: a longitudinal study. *Pediatrics* 1996; 97(4): 529–534.
9. Asher-McDade C, Brattström V, Dahl E, et al. A six-center international study of treatment outcome in patients with clefts of the lip and palate: part 4. Assessment of nasolabial appearance. *Cleft Palate Craniofac J* 1992; 29(5): 409–412.
10. Noar JH. Questionnaire survey of attitudes and concerns of patients with cleft lip and palate and their parents. *Cleft Palate Craniofac J* 1991; 28(3): 279–284.
11. Semb G. A study of facial growth in patients with unilateral cleft lip and palate treated by the Oslo CLP team. *Cleft Palate Craniofac J* 1991; 28(1): 1–21; discussion 46.
12. Millard DR, Latham R, Huifen X, et al. Cleft lip and palate treated by presurgical orthopedics, gingivo-periosteoplasty, and lip adhesion (POPLA) compared with previous lip adhesion method: a preliminary study of serial dental casts. *Plast Reconstr Surg* 1999; 103(6): 1630–1644.
13. Friede H and Enemark H. Long-term evidence for favorable midfacial growth after delayed hard palate repair in UCLP patients. *Cleft Palate Craniofac J* 2001; 38(4): 323–329.
14. Grayson BH and Shetye PR. Presurgical nasoalveolar moulding treatment in cleft lip and palate patients. *Indian J Plast Surg* 2009; 42(Suppl.): S56–S61.
15. Enemark H, Friede H, Paulin G, et al. Lip and nose morphology in patients with unilateral cleft lip and palate from four Scandinavian centres. *Scand J Plast Reconstr Surg Hand Surg* 1993; 27: 41–47.