ISSN 2219-2808 (online)

World Journal of *Clinical Pediatrics*

Quarterly Volume 13 Number 4 December 9, 2024

Published by Baishideng Publishing Group Inc

W J C P

World Journal of
Clinical Pediatrics

Contents

Quarterly Volume 13 Number 4 December 9, 2024

World Journal of Clinical Pediatrics

Quarterly Volume 13 Number 4 December 9, 2024

Contents

ABOUT COVER

Editorial Board Member of *World Journal of Clinical Pediatrics*, Li-Ye Yang, MD, PhD, Chief Doctor, Precision Medical Center, People's Hospital of Yangjiang, Yangjiang 529500, Guangdong Province, China. yangleeyee@sina.com

AIMS AND SCOPE

The primary aim of the *World Journal of Clinical Pediatrics* (*WJCP*, *World J Clin Pediatr*) is to provide scholars and readers from various fields of pediatrics with a platform to publish high-quality clinical research articles and communicate their research findings online.

WJCP mainly publishes articles reporting research results and findings obtained in the field of pediatrics and covering a wide range of topics including anesthesiology, cardiology, endocrinology, gastroenterology, hematology, immunology, infections and infectious diseases, medical imaging, neonatology, nephrology, neurosurgery, nursing medicine, perinatology, pharmacology, respiratory medicine, and urology.

INDEXING/ABSTRACTING

The *WJCP* is now abstracted and indexed in PubMed, PubMed Central, Scopus, Reference Citation Analysis, China Science and Technology Journal Database, and Superstar Journals Database. The *WJCP*'s CiteScore for 2023 is 3.2 and Scopus CiteScore rank 2023: Pediatrics, perinatology and child health is 129/330.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Hua-Ge Yu*; **Production Department Director:** *Xiang Li*; **Cover Editor:** *Xu Guo*.

NAME OF JOURNAL
World Journal of Clinical Pediatrics

ISSN
ISSN 2219-2808 (online)

LAUNCH DATE
June 8, 2012

FREQUENCY
Quarterly

EDITORS-IN-CHIEF
Consolato M Sergi, Elena Daniela Serban

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2219-2808/editorialboard.htm

PUBLICATION DATE
December 9, 2024

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/GerInfo/239

ONLINE SUBMISSION
https://www.f6publishing.com

WJCP World Journal of Clinical Pediatrics

Submit a Manuscript: https://www.f6publishing.com

World J Clin Pediatr 2024 December 9; 13(4): 98472

DOI: 10.5409/wjcp.v13.i4.98472

ISSN 2219-2808 (online)

ORIGINAL ARTICLE

Retrospective Study

Prediction of cyanotic and acyanotic congenital heart disease using machine learning models

Sana Shahid, Haris Khurram, Apiradee Lim, Muhammad Farhan Shabbir, Baki Billah

Specialty type: Cardiac and cardiovascular systems

Provenance and peer review: Invited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's classification
Scientific Quality: Grade D
Novelty: Grade C
Creativity or Innovation: Grade C
Scientific Significance: Grade C

P-Reviewer: Jeyaram K

Received: June 27, 2024
Revised: August 28, 2024
Accepted: September 23, 2024
Published online: December 9, 2024
Processing time: 125 Days and 7 Hours

Sana Shahid, Department of Statistics, Bahauddin Zakariya University, Multan 60000, Punjab, Pakistan

Haris Khurram, Apiradee Lim, Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000, Thailand

Haris Khurram, Department of Science and Humanities, National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus, Chiniot 35400, Punjab, Pakistan

Muhammad Farhan Shabbir, Department of Cardiology, Chaudhary Pervaiz Elhai Institute of Cardiology, Multan 60000, Punjab, Pakistan

Baki Billah, School of Public Health and Preventive Medicine, Monash University, Melbourne 3000, Victoria, Australia

Co-first authors: Sana Shahid and Haris Khurram.

Co-corresponding authors: Haris Khurram and Apiradee Lim.

Corresponding author: Haris Khurram, PhD, Assistant Professor, Postdoctoral Fellow, Department of Mathematics and Computer Science, Prince of Songkla University, 181 Village No. 6 Charoen Pradit Road, Rusamilae, Mueang Pattani District, Pattani 94000, Thailand. haris.khurram@nu.edu.pk

Abstract

BACKGROUND

Congenital heart disease is most commonly seen in neonates and it is a major cause of pediatric illness and childhood morbidity and mortality.

AIM

To identify and build the best predictive model for predicting cyanotic and acyanotic congenital heart disease in children during pregnancy and identify their potential risk factors.

METHODS

The data were collected from the Pediatric Cardiology Department at Chaudhry Pervaiz Elahi Institute of Cardiology Multan, Pakistan from December 2017 to October 2019. A sample of 3900 mothers whose children were diagnosed with cyanotic or acyanotic congenital heart disease was taken. Multivariate outlier detection methods were used to identify the potential outliers. Different machine learning models were compared, and the best-fitted model was selected using the area under the curve, sensitivity, and specificity of the models.

RESULTS

Out of 3900 patients included, about 69.5% had acyanotic and 30.5% had cyanotic congenital heart disease. Males had more cases of acyanotic (53.6%) and cyanotic (54.5%) congenital heart disease as compared to females. The odds of having cyanotic was 1.28 times higher for children whose mothers used more fast food frequently during pregnancy. The artificial neural network model was selected as the best predictive model with an area under the curve of 0.9012, sensitivity of 65.76%, and specificity of 97.23%.

CONCLUSION

Children having a positive family history are at very high risk of having cyanotic and acyanotic congenital heart disease. Males are more at risk and their mothers need more care, good food, and physical activity during pregnancy. The best-fitted model for predicting cyanotic and acyanotic congenital heart disease is the artificial neural network. The results obtained and the best model identified will be useful for medical practitioners and public health scientists for an informed decision-making process about the earlier diagnosis and improve the health condition of children in Pakistan.

Key Words: Congenital heart disease; Cyanotic heart disease; Acyanotic heart disease; Logistic regression model; Artificial neural network

Core Tip: In this study, to identify and build the best model for predicting cyanotic and acyanotic congenital heart disease in children during pregnancy and identify their risk factors, we employed machine learning models and compared their performance to choose the best one. We also used multivariate outlier detection methods to determine the outlier cases. The best fit model for congenital heart disease was the artificial neural network model. Children having a positive family history are at very high risk of having cyanotic and acyanotic congenital heart disease.

Citation: Shahid S, Khurram H, Lim A, Shabbir MF, Billah B. Prediction of cyanotic and acyanotic congenital heart disease using machine learning models. *World J Clin Pediatr* 2024; 13(4): 98472

URL: https://www.wjgnet.com/2219-2808/full/v13/i4/98472.htm

DOI: https://dx.doi.org/10.5409/wjcp.v13.i4.98472

INTRODUCTION

Congenital heart disease (CHD) is most commonly seen in neonates[1] and is a major cause of pediatric illness and childhood morbidity and mortality[2]. CHD is usually the result of the abnormal embryonic development of a normal structure during the early stage of embryonic or fetal development[3]. The incidence of CHD is 8 to 10 per 1000 births in Pakistan and nearly about 50000 children are affected by CHD each year[4]. The prevalence of CHD was 4 per 1000 live births in Karachi, Pakistan and 41.7% of children had cyanotic CHD and 58.3% had acyanotic CHD[5]. Acyanotic CHD was more common than cyanotic CHD and both conditions were found to have a higher incidence in males as compared to females[6,7].

In underdeveloped countries, families of children with CHD are faced with many health care and socioeconomic problems[1]. Late diagnosis of CHD carries a high risk of avoidable morbidity, mortality, and handicap. Problem identification and modification at an early stage were crucial in avoiding complexity, improving quality of life, and reducing mortality[2]. Awareness among parents about the disease can reduce the delay in the identification of disease, which can undoubtedly prevent mortality and morbidity in the subjects[8]. In rural areas of Pakistan, the prevalence of CHD was very high as compared to urban areas[9]. There were several fetal factors associated with CHD, like premature birth, stillbirth, and low birth weight. Low birth weight, family history of CHD, maternal co-morbidities, and consanguineous marriage were associated with CHD[10]. Physical activity, nutrition, partner interaction, access to basic health care facilities, calories in food, environment, and housing conditions during pregnancy reduce the risk factors of cyanotic and acyanotic CHD[11]. The prevalence of CHD was 9.3 per 1000 Live births in Asia and 8 to 10 per 1000 Live births worldwide; 60.6% of cases were acyanotic CHD and 38.6% were cyanotic CHD[12]. The prevalence of CHD for Whites was significantly higher than for Blacks or Mexican Americans. The prevalence rate of CHD in children aged 5 to 15 years has been reported as 2 per 1000 in Sudan, 3 per 1000 in Uganda, and 3.6 per 1000 in Nigeria[13]. In India, the prevalence of CHD was reported from 8.5 to 13.6 per 1000 live births, and the 10% infant mortality was due to CHD. Acyanotic CHD was present in 79% of CHD children, 21% had cyanotic CHD, and 82.9% were diagnosed between 0 to 3 years of age. Parental age, illness during pregnancy, and advanced maternal age were found to be risk factors for CHD[14,15]. The prevalence of CHD was 5 to 10 per 1000 live births and 10.01 per 1000 in school children in Alexandria, Egypt. Parental consanguinity, positive family history, and maternal health during pregnancy were high-risk factors for CHD[16]. CHD was the most common birth defect in China and the prevalence was 7 to 8 per 1000 live births; it shows about 100000 to 150000 new cases annually. The mental stress in the mother, number of previous pregnancies, maternal infection, and education level of the mother were the risk factors for CHD[17]. CHD risk was higher among those children who had a family history of heart disease[18].

The CHD prevalence in Asia, Europe, and Africa was found to be 9.3 per 1000, 8.2 per 1000, and 1.9 per 1000 live births, respectively. The CHD prevalence was reported to be higher in the Asian region as compared with other regions[19]. Smoking status in mothers and mental stress in mothers during pregnancy were found highly associated with CHD in children[20]. The increase in the risk of CHD was associated with poor socioeconomic status, family income, occupation, and education level of mothers[21]. In Brazil, CHD was most common in newborns and reached 1% of the population of Brazil. Socioeconomic status and family income are important factors in child development, and indirectly, they affect the process and outcome of child development with home type, nutrition quality, availability of school, health care, and medical facilities[22]. CHD is associated with physical inactivity and obesity in children and adolescents. To reduce the risk of obesity and heart disease in children and adolescents, it must be necessary to adopt a healthy lifestyle[23]. The environment and lifestyle factors also influence children with CHD[24].

In recent years, machine learning models have emerged as crucial tools in revolutionizing disease prediction and diagnosis. These advanced analytical models have transformed the way that healthcare professionals approach patient care, enabling early detection, accurate diagnosis, and personalized treatment[25]. The artificial neural network (ANN) model is vital in disease prediction due to their ability to learn from vast amounts of complex data, identify suitable patterns and correlations that may elude human clinicians, adapt to new data and improve prediction accuracy over time, and provide tailored recommendations for patient care[26]. The ANN models, inspired by the human brain's structure and function, have shown remarkable promise in disease prediction due to their ability to mimic human brain function, capacity to analyze complex relationships between variables, and identification of non-linear patterns and interactions [27]. The present study aimed to identify the risk factors for cyanotic and acyanotic CHD in children, predict cyanotic and acyanotic CHD in children at the time of pregnancy, and suggest the best machine learning-based predictive model.

MATERIALS AND METHODS

Study design and sample

A retrospective study design was used, and data was collected from the outpatient department, inpatient department, and ward of the Pediatric Department at the Institute of Cardiology Multan, Pakistan from December 2017 to October 2019. The data of the present study were collected from 3900 mothers whose children were diagnosed with cyanotic and acyanotic CHD by echocardiography. The sample of the current study attained a greater than 80% power of the test.

Patients' consent and ethics approval

The study was approved by the Departmental Ethics Committee and Board of Advanced Studies, Bahauddin Zakariya University, Multan, Pakistan. Also, we have taken permission from the hospital and all the families included in the study were volunteers and were well informed about the study and the confidentiality of their identity.

Operational definition of variables

The data was collected by the principal author. With the discussion of medical practitioners and based on literature survey, different factors were isolated and these factors are described as diabetes in family (diabetes in first-order relatives), smoking status in the family (smoking in first-order relatives), family history of heart disease (family history of heart disease in first-order relatives), anemia in mother during pregnancy, physically active mother during pregnancy (the mother can walk at least two and half hour in a week), use of fast food, low-calorie food, and staple food during pregnancy (mothers eating fast food more than once a week, mothers consuming less than 2000 calories per day, and mothers using cereal grain and tubers as staple food). Nutrition status during pregnancy (good nutrition: Protein more than 5 ounces, fruits up to 2 cups, vegetables up to 3 cups, grains up to 6 ounces, and dairy up to 3 cups; normal nutrition: Protein up to 5 ounces, fruits up to 1.5 cups, vegetables up to 2.5 cups, grains up to 5 ounces, and dairy up to 2.5 cups. Less than normal nutrition is considered as poor nutrition), monthly income of family, education level of parents, dwelling area (the area of the child was categorized into rural and urban areas), home environment during pregnancy, health condition of other people living in home (respiratory infections, asthma, lead poisoning, injuries, and mental health), mother interaction with their partner during pregnancy, quality of basic health care facilities during pregnancy (well-trained and motivated staff, accurate medical record, water, energy, sanitation, hand hygiene, and waste disposal facilities which are functional, reliable, and safe; adequate stocks of medicines, supplies, and equipment that is safe, effective, timely, efficient, and equitable), access to health care facilities (is there any government hospital or medical unit and government doctor available in their surroundings), housing tenure (house is rented or owned), housing condition (the good condition of house contained: Being dry, safe, and hygienic, good ventilation, good sanitation, good heating, good lighting, good facilities of cooking, availability of suitable storage for food, and good access to shop and facilities). The dependent/outcome variable was the type of CHD, which was categorized as cyanotic and acyanotic.

Data management and analysis

For data analyses, R was used. Categorical data are presented as frequencies and percentages. The data were randomly divided into two parts for modeling and validation: The first part (85%) was used for training the model, and the second part (15%) was used for validation of the model. For multivariate outlier detection in the generalized linear model, different measurements were used, *i.e.,* Cook's distance[28], modified Cook's distance[29], leverage[30], Andrew's Pregibon[31], Welsch's distance[32], and covariance ratio[33]. Those cases were considered outliers that were jointly identified by all the above methods. The prediction performance for predicting the type of CHD was evaluated using subset logistic regression (SLR)[34], subset logistic regression after deletion (SLRAD), and the machine learning model ANN[35]. The performance of the models was compared using the area under the receiver operating characteristic (ROC) curve (AUC) and its 95% confidence interval, sensitivity, and specificity. In ANN models, the best generalization is achieved by using a model whose complexity is the most appropriate to produce an adequate fit of the data. In Supplementary material, the mathematical and procedural details of the diagnostic measures of outliers and all models are described.

RESULTS

There were 53.6% of males and 46.6% of females who had acyanotic CHD, and 54.5% of males and 45.5% of females who had cyanotic CHD. The children with acyanotic CHD who had a family history of diabetes accounted for 36.0%, and 40.3% of children with cyanotic CHD had a family history of diabetes. The results of univariate analyses are presented in Table 1.

Figure 1 shows the graphs of influential diagnostic measures for CHD. In this figure, the circles show the observation of the data, the red line shows the cut point of the measure, and the points along with the observation number that are beyond the cut point were identified as influential observations for each measure. We delete those observations that were commonly identified as outliers by all the diagnostic measures.

The results of the logistic regression analysis is given in Table 2. The results of SLR showed that family history of heart disease, use of fast-food during pregnancy, use of staple food during pregnancy, poor nutrition during pregnancy, low family monthly income, uneducated parents, urban area, poor quality of health care facilities, rented house, and poor housing condition were significant risk factors for CHD. The results of SLRAD showed that family history of heart disease, use of fast-food during pregnancy, poor nutrition during pregnancy, low family monthly income, uneducated father, urban area, poor quality of health care facilities, rented house, and poor housing conditions were significant risk factors for CHD.

Figure 2 shows the weight of each input variable, and the weights were obtained by the ANN model for CHD through normalizing importance. According to importance, the most important risk factors for CHD were obtained: Father's education, family income, father's occupation, health condition, mother's education, nutrition, and number of children in family. In all the important factors, mother's education, nutrition status, and number of children in family had positive weight, while father's education, family income, father's occupation, and health condition had negative weight.

Figure 3 demonstrates the sequence of each predictor and describes the final ANN fitted model for CHD, which was generated by plotting each risk factor by normalized importance. In the ANN model for CHD, there were 20 input variables, 4 hidden variables, and 1 output variable.

Table 3 shows the comparison of all models by AUC and its 95% confidence interval, sensitivity, and specificity. The results showed that the ANN model had the highest AUC at 0.901 (95%CI: 0.892–0.910) with a sensitivity and specificity of 65.76% and 97.23%, respectively. The SLRAD model had the second highest AUC at 0.886 (95%CI: 0.876-0.896), with a sensitivity of 57.69% and specificity of 98.69%. The SLRM model had the third-highest AUC at 0.860 (95%CI: 0.849-0.871) with a sensitivity of 49.62% and specificity of 98.38%. Figure 4 also shows that the ANN model had the highest diagnostic accuracy for CHD.

DISCUSSION

The results of the current study show that acyanotic CHD is more common in children as compared to cyanotic CHD, which is consistent with the findings of a previous study done in Pakistan[6]. Our results show that the odds of having cyanotic CHD was 1.28 times higher for children whose mothers used fast food during pregnancy as compared to those whose mothers did not use. The odds of having cyanotic CHD was 6.22 times higher for children whose father was uneducated as compared to those whose father was educated. The odds of having cyanotic CHD was 1.49 times higher for children whose mothers had normal housing conditions as compared to those whose mothers had good housing conditions. Children who had a family history of heart disease had 0.55 times the odds of having acyanotic CHD as compared to those who had not. Children whose mothers used the staple food during pregnancy had 0.79 times the odds of having acyanotic CHD as compared to those whose mother did not. Male children were more affected by cyanotic and acyanotic CHD as compared to female children. A study in China has similar findings[17]. The result of our study shows that family history of heart disease is a risk factor for CHD, in agreement with the results of the studies in Egypt and China[16,18]. The result of the model comparison shows that the ANN model had the highest diagnostic accuracy. The result of analysis based on the ANN, the best-selected model, shows that father's education, family income, father's occupation, health condition of other people's living in home, mother's education, nutrition, and number of children in family are risk factors for cyanotic and acyanotic CHD in children. A study in China also concluded that mother's

Table 1 Descriptive analysis of categorical data of congenital heart disease, *n* (%)

Variable	Category	Acyanotic	Cyanotic	Variable	Category	Acyanotic	Cyanotic
Gender	Female	1258 (46.4)	518 (45.5)	Father's education	Uneducated	1102 (40.7)	702 (59.0)
	Male	1452 (53.6)	672 (54.5)		Primary/middle	1220 (45.0)	354 (29.7)
Diabetes	No	1734(64.0)	710 (59.7)		Secondary/higher	326 (12.0)	118 (9.9)
	Yes	976 (36.0)	480 (40.3)		Graduate	52 (1.9)	12 (1.0)
Smoking	No	1318 (48.6)	664 (55.8)		Masters or higher	10 (0.4)	4 (0.3)
	Yes	1392 (51.4)	526 (44.2)	Father's occupation	Dead/ unemployed	4 (0.1)	4 (0.3)
Family History	No	858 (31.7)	510 (42.9)		Labour/former	1866 (68.9)	826 (69.4)
	Yes	1852 (68.3)	680 (57.1)		Private job	194 (7.2)	20 (1.7)
Anemia during pregnancy	No	2598 (95.9)	1150 (96.6)		Small business	620 (22.9)	328 (27.6)
	Yes	112 (4.1)	40 (3.4)		Civil servant	26 (1.0)	12 (1.0)
Inactive	No	800 (29.5)	342 (28.7)	Area	Rural	1604 (59.2)	878 (73.8)
	Yes	1910 (70.5)	848 (71.3)		Urban	1106 (40.8)	312 (26.2)
Fast food during pregnancy	No	1356 (50.0)	486 (40.8)	Home environment	Poor	1650 (60.9)	518 (43.5)
	Yes	1354 (50.0)	704 (59.2)		Normal	590 (21.8)	400 (33.6)
Low-calorie food during pregnancy	No	1036 (38.2)	472 (39.7)		Good	470 (17.3)	272(22.9)
	Yes	1674 (61.8)	718 (60.3)	Health condition	Poor	602 (22.2)	402 (33.8)
Nutrition during pregnancy	Poor	1558 (57.5)	492 (41.3)		Normal	1660 (61.3)	522 (43.9)
	Normal	532 (19.6)	382 (32.1)		Good	448 (16.5)	266 (22.4)
	Good	620 (22.9)	316 (26.6)	Interaction with partner during pregnancy	Poor	1592 (58.7)	498 (41.8)
Staple food during pregnancy	No	1720 (63.5)	692 (58.2)		Normal	558 (20.6)	392 (32.9)
	Yes	990 (36.5)	498 (41.8)		Good	560 (20.7)	300 (25.2)
Income	< 10000	20 (0.70)	18 (1.5)	Health care quality	Poor	1540 (56.8)	506 (42.5)
	10000 to 20000	2076 (76.6)	956 (80.3)		Normal	660 (24.4)	414 (34.8)
	> 20000	614 (22.7)	216(18.2)		Good	510 (18.8)	270 (22.7)
Mother's education	Uneducated	1492 (55.1)	690 (58.0)	Health care access	No	2034 (75.1)	770 (64.7)
	Primary/middle	1002 (37.0)	404 (33.9)		Yes	676 (24.9)	420 (35.3)
	Secondary/higher	200 (7.4)	90 (7.6)	Housing tenure	Owned	2628 (97.0)	1164 (97.8)
	Graduate	8 (0.3)	6 (0.5)		Rented	82 (3.0)	26 (2.2)
	Masters or higher	8 (0.3)	0 (0.0)	Housing condition	Poor	630 (23.2)	528 (44.4)
					Normal	1742 (64.3)	548 (46.1)
					Good	338 (12.5)	114 (9.6)

education level is a risk factor for CHD[17,21]. A study in Pakistan also supports our findings, *i.e.*, health condition of other people living in home, and quality and access to basic health care facilities are risk factors of cyanotic and acyanotic CHD in children[11].

The field of machine learning has undergone significant advancements in recent years, leading to a surge in the development of innovative models that can accurately predict disease[36]. The ANN and machine learning models can analyze medical images, genetic data, and patient information to predict the risk factors of disease, detect early warning signs, and recommend preventive measures[37]. In the current study, we used different machine learning models to predict cyanotic and acyanotic CHD in children. One recent study reported that the neural network model is an accurate decision support tool in diagnosing CHD[38]. Another study shows that the ANN model yields the best accuracy while predicting CHD in children[39]. The results of another study show that the best predictive model for CHD children was machine learning models and the AUC values for those models ranged from 0.81 to 0.83[40].

Table 2 Multivariate logistic regression models by using stepwise selection approach

Variable	Categories[4]	SLR		SLRAD	
		OR	95%CI	OR	95%CI
(Intercept)	-	2.006	0.300-13.400	0.000	-
History	Yes	0.551[1]	0.463-0.656	0.541[1]	0.454-0.646
Fast food	Yes	1.289[2]	1.027-1.618	1.331[2]	1.056-1.677
Staple food	Yes	0.794[3]	0.609-1.034	0.803	0.615-1.049
Nutrition	Normal	0.668	0.345-1.294	0.698	0.382-1.273
	Poor	0.621[2]	0.409-0.942	0.571[1]	0.382-0.853
Children	-	1.368[1]	1.267-1.476	1.365[1]	1.264-1.473
Family income	< 20000	0.276[1]	0.124-0.614	0.263[1]	0.118-0.588
	10000 to 20000	0.396[2]	0.181-0.866	0.390[2]	0.177-0.857
Mother education	Master or higher	0.000	0-2.96E+163	0.998	-
	Primary/middle	0.474	0.129-1.744	3.96E+05	-
	Secondary/higher	0.781	0.208-2.933	7.09E+05	-
	Uneducated	0.300[3]	0.082-1.106	2.58E+05	-
Father's education	Master or higher	0.391	0.053-2.907	0.000	-
	Primary/middle	1.984	0.779-5.053	2.630[3]	0.907-7.624
	Secondary/higher	1.950	0.775-4.909	2.599[3]	0.902-7.489
	Uneducated	6.221[1]	2.395-16.160	8.516[1]	2.875-25.225
Father's occupation	Dead/unemployed	0.672	0.101-4.478	8.65E+05	-
	Labour/former	0.336[2]	0.123-0.914	0.736	0.196-2.768
	Private job	0.082[1]	0.026-0.259	0.145[1]	0.035-0.611
	Small business	0.465	0.168-1.292	0.986	0.259-3.761
Dwelling area	Urban	0.582[1]	0.478-0.710	0.554[1]	0.453-0.678
Partner interaction	Normal	1.060	0.583-1.927	-	-
	Poor	0.732	0.496-1.081	-	-
Quality of health care facilities	Normal	0.468[2]	0.25-0.878	0.409[1]	0.226-0.743
	Poor	0.682	0.421-1.104	0.503[1]	0.316-0.802
Housing tenure	Rented	0.511[2]	0.276-0.946	0.437[2]	0.227-0.841
Housing condition	Normal	1.498[2]	1.053-2.131	1.554[2]	1.086-2.225
	Poor	2.852[1]	2.04-3.987	3.004[1]	2.136-4.225

[1]Significance at 1%;
[2]Significant at 5%;
[3]Significant at 10%.
[4]Reference categories are: Family history "no", use of fast food "no", use of staple food "no", nutrition "good", partner interaction "good", quality of health care facilities "good", housing condition "good", family income "< 10000", father's education "primary/middle", father's occupation "civil servant", and housing tenure "owned".

CONCLUSION

Children having a family history of heart disease are at very high risk of developing cyanotic and acyanotic CHD. The incidence of cyanotic CHD can be reduced by limiting fast food during pregnancy. Similarly, reducing the number of children can also minimize the incidence of CHD. Moreover, mothers with an uneducated partner and poor housing conditions are at high risk of birthing a child having cyanotic CHD. Similarly, the incidence of acyanotic CHD can be reduced by adopting good dietary habits (high nutrition food and rich calorie food) during pregnancy. Families with low income, uneducated mothers, and those living in urban areas are at higher risk of birthing a child having cyanotic CHD. The best fit model for our data is ANN, which can be used for earlier diagnostics. This prediction model can help medical

Table 3 Performance comparison of models

Model	AUC	95%CI	Sensitivity	Specificity
SLRM	86.01	0.849–0.871	49.62	98.38
SLRAD	88.57	0.876-0.896	57.69	98.69
ANN	90.12	0.892–0.910	65.76	97.23

AUC: Area under the curve; SLR: Logistic regression; SLRAD: Subset logistic regression after deletion; ANN: Artificial neural network.

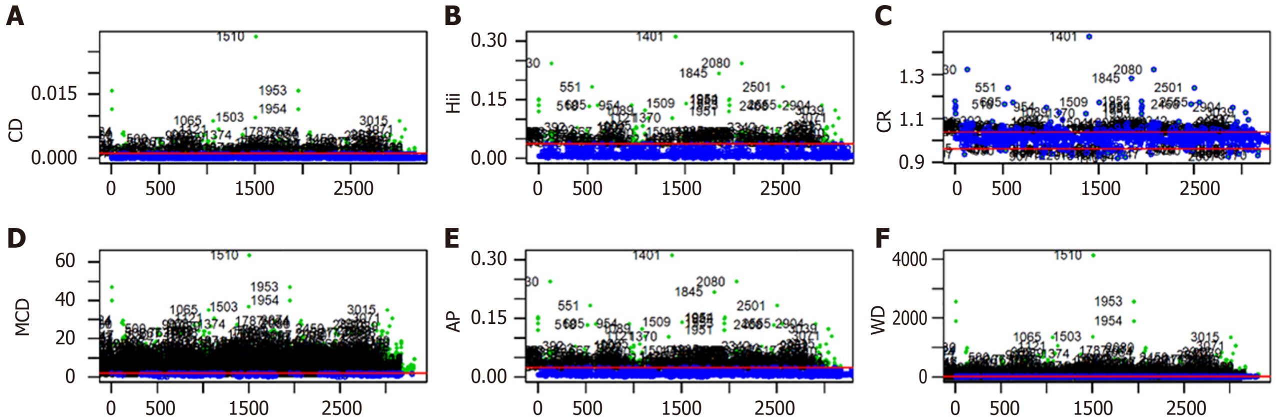

Figure 1 Graphs of influential diagnostic measures. A: Detection using Cook's distance method; B: Detection using leverage method; C: Detection using covariance ratio method; D: Detection using modified Cook's distance method; E: Detection using Andrew's Pregibon method; F: Detection using Walsh's distance method.

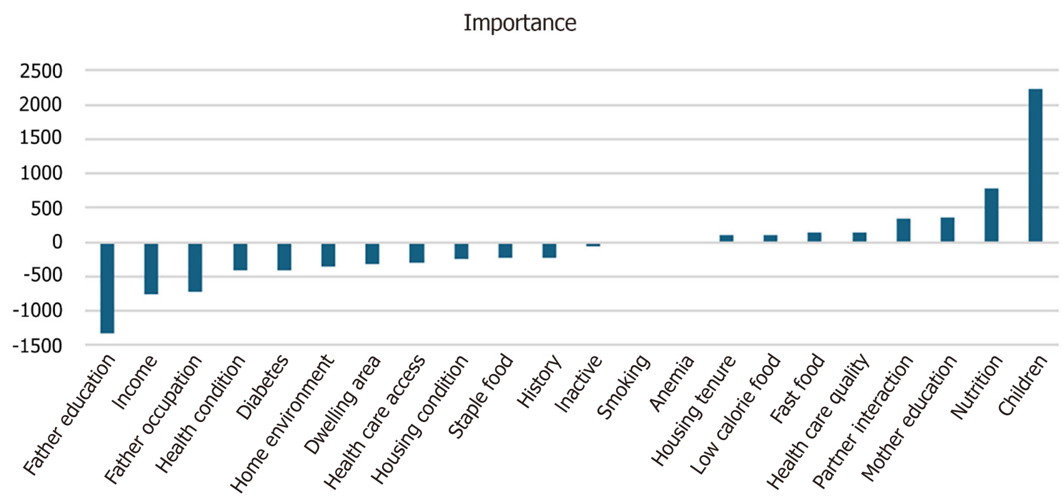

Figure 2 Weights according to importance of variables by artificial neural network.

practitioners and experts to identify the risk and make earlier diagnoses of cyanotic and acyanotic CHD during pregnancy, which will improve healthcare.

Limitations and future directions

The accuracy of the models may be limited by the quality and availability of the regional data. This can be improved by using large nationwide data. For future studies, investigating new features and feature engineering techniques can help improve model performance. Developing models that are more interpretable can help clinicians understand why certain predictions are made. Validating the models prospectively can help establish their clinical utility. Comparing the performance of different machine learning models can help identify the best approach.

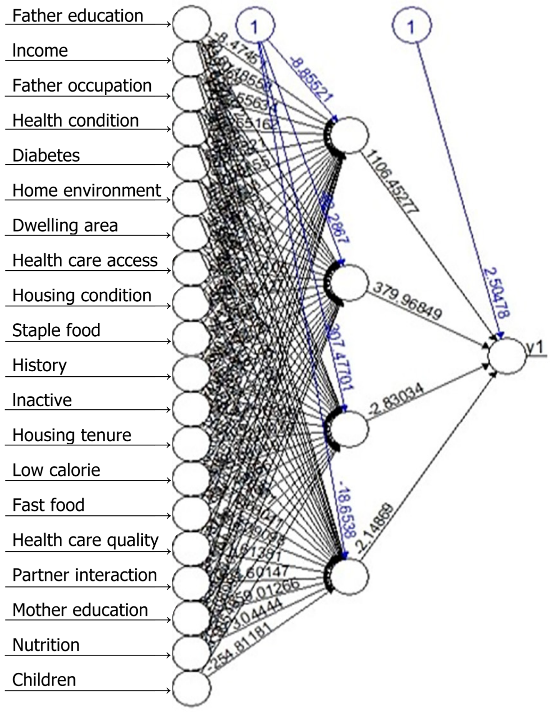

Figure 3 Modeling structure of artificial neural network with weights of each node.

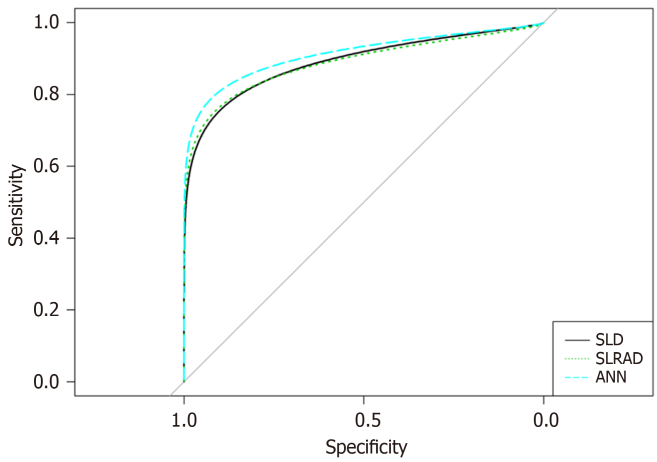

Figure 4 Receiver operating characteristic curves for comparison of subset logistic regression, subset logistic regression after deletion, and artificial neural network model. SLR: Subset logistic regression; SLRAD: Subset logistic regression after deletion; ANN: Artificial neural network.

ACKNOWLEDGEMENTS

We thank the Prince of Songkhla University for providing Haris Khurram with the post-doctoral fellowship. We also want to thank Ms. Sadia Ashfaq, Lecturer, National University of Computer Science, CFD Campus, Pakistan, for providing the honorary service for language editing, which greatly enhanced the quality of our manuscript.

FOOTNOTES

Author contributions: Shahid S, Khurram H, and Lim A conceptualized and designed the research; Shahid S and Khurram H organized the dataset, performed statistical analysis and data interpretation, and wrote the first draft of the manuscript with the help of Lim A; Khurram H and Lim A played important and indispensable roles in the experimental design, data interpretation, and manuscript preparation as the co-corresponding authors; Shahid S and Khurram H made crucial and indispensable contributions towards the completion of the project and were thus qualified as the co-first authors of the paper; Lim A proofread the draft and gave valuable suggestions to improve the manuscript; Shabbir MF reviewed the final draft from a medical perspective. Billah B reviewed the final draft from a statistical perspective. All authors contributed to the manuscript revision, and read and approved the final draft.

Institutional review board statement: The study was reviewed and approved by the Advance Studies & Research Board, Bahauddin Zakariya University, Multan, Pakistan (No. 8973).

Informed consent statement: All study participants or their legal guardians gave informed verbal consent prior to study inclusion.

Conflict-of-interest statement: All authors have no conflicts of interest to disclose.

Data sharing statement: The data and code of R language are available from the corresponding author [Email: Hariskhurram2@gmail.com; haris.khurram@nu.edu.pk] upon reasonable request.

Country of origin: Thailand

ORCID number: Haris Khurram 0000-0003-1814-4742.

S-Editor: Liu JH
L-Editor: Wang TQ
P-Editor: Yu HG

REFERENCES

1 **Shabana NA**, Shahid SU, Irfan U. Genetic Contribution to Congenital Heart Disease (CHD). *Pediatr Cardiol* 2020; **41**: 12-23 [PMID: 31872283 DOI: 10.1007/s00246-019-02271-4]

2 **Mohammad N**, Shaikh S, Memon S, Das H. Spectrum of heart disease in children under 5 years of age at Liaquat University Hospital, Hyderabad, Pakistan. *Indian Heart J* 2014; **66**: 145-149 [PMID: 24581115 DOI: 10.1016/j.ihj.2013.12.041]

3 **Balat MS**, Sahu SK. Congenital heart disease: factor affecting it and role of RBSK in dealing with situation. *Int J Community Med Public Health* 2018; **5**: 4437 [DOI: 10.18203/2394-6040.ijcmph20183990]

4 **Humayun KN**, Atiq M. Clinical profile and outcome of cyanotic congenital heart disease in neonates. *J Coll Physicians Surg Pak* 2008; **18**: 290-293 [PMID: 18541084]

5 **Masood N**, Sharif M, Asghar R, Qamar M, Hussain I. Frequency of congenital heart diseases at Benazir Bhutto Hospital Rawalpindi. *Ann Pak Inst Med Sci* 2010; **6**: 120-123

6 **Farooqui R**, Haroon UF, Niazi A, Rehan N, Butt T, Niazi M. Congenital heart diseases in neonates. *JRMC* 2010; **14**: 31-33

7 **Pathan IH**, Bangash SK, Khawaja AM. Spectrum of heart defects in children presenting for paediaric cardiac surgery. *Pak Heart J* 2016; 49

8 **Hussain M**, Hussain S, Krishin J, Abbasi S. Presentation of congestive cardiac failure in children with ventricular septal defect. *J Ayub Med Coll Abbottabad* 2010; **22**: 135-138 [PMID: 22455281]

9 **Rizvi SF**, Mustafa G, Kundi A, Khan MA. Prevalence of congenital heart disease in rural communities of Pakistan. *J Ayub Med Coll Abbottabad* 2015; **27**: 124-127 [PMID: 26182756]

10 **Ul Haq F**, Jalil F, Hashmi S, Jumani MI, Imdad A, Jabeen M, Hashmi JT, Irfan FB, Imran M, Atiq M. Risk factors predisposing to congenital heart defects. *Ann Pediatr Cardiol* 2011; **4**: 117-121 [PMID: 21976868 DOI: 10.4103/0974-2069.84641]

11 **Shahid S**, Akbar A. Conventional and non-conventional risk factors of cyanotic and acyanotic congenital heart diseases in children of southern Punjab, Pakistan. *Pak Heart J* 2020; **53** [DOI: 10.47144/phj.v53i2.1698]

12 **Pate N**, Jawed S, Nigar N, Junaid F, Wadood AA, Abdullah F. Frequency and pattern of congenital heart defects in a tertiary care cardiac hospital of Karachi. *Pak J Med Sci* 2016; **32**: 79-84 [PMID: 27022350 DOI: 10.12669/pjms.321.9029]

13 **Aman W**, Sherin A, Hafizullah M. Frequency of congenital heart diseases in patients under the age of twelve years at Lady Reading Hospital Peshawar. *JPMI* 2006; **20**

14 **Kapoor R**, Gupta S. Prevalence of congenital heart disease, Kanpur, India. *Indian Pediatr* 2008; **45**: 309-311 [PMID: 18451451]

15 **Abqari S**, Gupta A, Shahab T, Rabbani MU, Ali SM, Firdaus U. Profile and risk factors for congenital heart defects: A study in a tertiary care hospital. *Ann Pediatr Cardiol* 2016; **9**: 216-221 [PMID: 27625518 DOI: 10.4103/0974-2069.189119]

16 **Settin A**, Almarsafawy H, Alhussieny A, Dowaidar M. Dysmorphic Features, Consanguinity and Cytogenetic Pattern of Congenital Heart Diseases: a pilot study from Mansoura Locality, Egypt. *Int J Health Sci (Qassim)* 2008; **2**: 101-111 [PMID: 21475491]

17 **Liu S**, Liu J, Tang J, Ji J, Chen J, Liu C. Environmental risk factors for congenital heart disease in the Shandong Peninsula, China: a hospital-based case-control study. *J Epidemiol* 2009; **19**: 122-130 [PMID: 19398851 DOI: 10.2188/jea.je20080039]

18 **Pei L**, Kang Y, Zhao Y, Yan H. Prevalence and risk factors of congenital heart defects among live births: a population-based cross-sectional survey in Shaanxi province, Northwestern China. *BMC Pediatr* 2017; **17**: 18 [PMID: 28086762 DOI: 10.1186/s12887-017-0784-1]

19 **van der Linde D**, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, Roos-Hesselink JW. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J Am Coll Cardiol* 2011; **58**: 2241-2247 [PMID: 22078432 DOI: 10.1016/j.jacc.2011.08.025]

20 **Feng Y**, Yu D, Yang L, Da M, Wang Z, Lin Y, Ni B, Wang S, Mo X. Maternal lifestyle factors in pregnancy and congenital heart defects in offspring: review of the current evidence. *Ital J Pediatr* 2014; **40**: 85 [PMID: 25385357 DOI: 10.1186/s13052-014-0085-3]

21 **Yu D**, Feng Y, Yang L, Da M, Fan C, Wang S, Mo X. Maternal socioeconomic status and the risk of congenital heart defects in offspring: a meta-analysis of 33 studies. *PLoS One* 2014; **9**: e111056 [PMID: 25347676 DOI: 10.1371/journal.pone.0111056]

22 **Mari MA**, Cascudo MM, Alchieri JC. Congenital Heart Disease and Impacts on Child Development. *Braz J Cardiovasc Surg* 2016; **31**: 31-37 [PMID: 27074272 DOI: 10.5935/1678-9741.20160001]

23 **Barbiero SM**, D'Azevedo Sica C, Schuh DS, Cesa CC, de Oliveira Petkowicz R, Pellanda LC. Overweight and obesity in children with congenital heart disease: combination of risks for the future? *BMC Pediatr* 2014; **14**: 271 [PMID: 25323400 DOI: 10.1186/1471-2431-14-271]

24 **Wacker-Gussmann A**, Oberhoffer-Fritz R. Cardiovascular Risk Factors in Childhood and Adolescence. *J Clin Med* 2022; **11** [PMID: 35207409 DOI: 10.3390/jcm11041136]

25 **Alowais SA**, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ* 2023; **23**: 689 [PMID: 37740191 DOI: 10.1186/s12909-023-04698-z]

26 **Byeon H**, Gc P, Hannan SA, Alghayadh FY, Soomar AM, Soni M, Bhatt MW. Deep neural network model for enhancing disease prediction using auto encoder based broad learning. *SLAS Technol* 2024; **29**: 100145 [PMID: 38750819 DOI: 10.1016/j.slast.2024.100145]

27 **Taherdoost H**. Deep Learning and Neural Networks: Decision-Making Implications. *Symmetry* 2023; **15**: 1723 [DOI: 10.3390/sym15091723]

28 **Belle V**, Papantonis I. Principles and Practice of Explainable Machine Learning. *Front Big Data* 2021; **4**: 688969 [PMID: 34278297 DOI: 10.3389/fdata.2021.688969]

29 **Shahid S**. Statistical Modeling of Epidemiology of Cardiovascular Diseases in Children (Doctoral dissertation, Bahauddin Zakariya University Multan, Pakistan). 2022

30 **Belsley DA**, Kuh E, Welsch RE. Regression diagnostics: Identifying influential data and sources of collinearity. New York and Chichester: John Wiley & Sons, 1980 [DOI: 10.1002/0471725153]

31 **Bagheri A**, Midi H, Imon A. The Effect of Collinearity-influential Observations on Collinear Data Set: A Monte Carlo Simulation Study. *J App Sci* 2010; **10**: 2086-2093 [DOI: 10.3923/jas.2010.2086.2093]

32 **Van den Broeck J**, Cunningham SA, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med* 2005; **2**: e267 [PMID: 16138788 DOI: 10.1371/journal.pmed.0020267]

33 **Ullah MA**, Pasha GR. The origin and developments of influence measures in regression. *PJS* 2009; **25**

34 **Hosmer DW**, Lemeshow S, Sturdivant RX. Applied logistic regression. Canada: Jhon Wiley & Sons, 2013 [DOI: 10.1002/9781118548387]

35 **Shahid S**, Khurram H, Billah B, Akbar A, Shehzad MA, Shabbir MF. Machine learning methods for predicting major types of rheumatic heart diseases in children of Southern Punjab, Pakistan. *Front Cardiovasc Med* 2022; **9**: 996225 [PMID: 36312229 DOI: 10.3389/fcvm.2022.996225]

36 **Shivahare BD**, Singh J, Ravi V, Chandan RR, Alahmadi TJ, Singh P, Diwakar M. Delving into Machine Learning's Influence on Disease Diagnosis and Prediction. *TOPHJ* 2024; **17** [DOI: 10.2174/0118749445297804240401061128]

37 **Kumar Y**, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput* 2023; **14**: 8459-8486 [PMID: 35039756 DOI: 10.1007/s12652-021-03612-z]

38 **Hoodbhoy Z**, Jiwani U, Sattar S, Salam R, Hasan B, Das JK. Diagnostic Accuracy of Machine Learning Models to Identify Congenital Heart Disease: A Meta-Analysis. *Front Artif Intell* 2021; **4**: 708365 [PMID: 34308341 DOI: 10.3389/frai.2021.708365]

39 **Rani S**, Masood S. Predicting congenital heart disease using machine learning techniques. *JDMSC* 2020; **23**: 293-303 [DOI: 10.1080/09720529.2020.1721862]

40 **Guo K**, Fu X, Zhang H, Wang M, Hong S, Ma S. Predicting the postoperative blood coagulation state of children with congenital heart disease by machine learning based on real-world data. *Transl Pediatr* 2021; **10**: 33-43 [PMID: 33633935 DOI: 10.21037/tp-20-238]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: office@baishideng.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com