Table2
Characteristics of the participants in the dyslipidemia and normal groups.

Predictors	Categories	Dyslipidemia*		OR [CI 95%]	P-value
		No	Yes		
Age (years)		14.28 ± 2.26	14.64 ± 2.39	–	0.058
Sex	Male	49.28	46.61	0.90 [0.67,1.21]	0.477
	Female			–	
Region	Urban	64.80	71.71	1.38 [1.01,1.89]	0.049
	Rural			–	
Family history of diabetes	No	70.54	66.14	–	0.207
	Yes			1.23 [0.89,1.68]	
Family history of obesity	No	68.32	70.12	–	0.604
	Yes			0.92 [0.67,1.27]	
Family history of cancer	No	83.23	78.88	–	0.137
	Yes			1.33[0.91,1.93]	
Family history of CVD	No	87.16	92.43	–	0.023
	Yes			0.55 [0.33,0.93]	
Abdominal obesity	No	88.41	61.59	–	<0.001
	Yes			4.76 [3.26,6.94]	
BMI category (WHO criteria)	Under weight	25.85	19.52	0.76 [0.52,1.09]	0.007
	Normal	58.22	58.17	-	
	Over weight	8.36	10.76	1.29 [0.77,2.15]	
	Obese	7.57	11.55	1.53 [0.91,2.56]	
Physical activity	Mild	25.47	45.82	2.03 [1.43,2.87]	<0.001
	Moderate	40.37	35.86	–	
	High	34.16	18.32	0.60 [0.40,0.89]	
Birth weight	Low	11.67	16.73	1.54 [1.0,2.34]	0.249
	Normal	79.58	74.10	–	
	High	8.75	9.17	1.13 [0.67,1.89]	
Systolic blood pressure (mm Hg)		101.87 ± 13.16	104.16 ± 13.09	–	0.025
Diastolic blood pressure (mm Hg)		65.89 ± 10.74	66.69 ± 10.61	–	0.338
Fast blood sugar (mg/dL)		87.6 ± 11.85	84.32 ± 11.85	–	0.002
HDL-C (mg/dL)		59.95 ± 18.22	29.40 ± 12.37	–	<0.001
LDL-C (mg/dL)		75.43 ± 28.35	92.55 ± 38.09	–	<0.001
Total cholesterol (mg/dL)		149.66 ± 29.50	154.46 ± 30.20	–	0.061
Triglyceride (mg/dL)		86.06 ± 33.08	93.35 ± 34.35	–	<0.001

*: Results are reported as mean ± standard deviation (for interval variables) and percentage (for categorical variables). CVD: cardio-vascular disease; BMI: body mass index; WHO: world health organization; HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; OR: Odds ratio (a categorical level was set to reference for each categorical variable); CI: confidence interval. In each dyslipidemia group, the frequency percentage of one of the categories in binary variables was shown.

that none of them performed properly on the third subset. The proposed classifier on the set 2 significantly outperformed than sets 1 and 3 (P-value < 0.05). Also, the results of Set 1 was significantly better than those of Set 3 (P-value < 0.05).

The selected features of the proposed classifier on the Set 1 were CETP TaqIB [rs708272], CETP A373P [rs5880], LPL D9N [rs1801177], ApoE, ABCAI R1587K [rs2230808], APOA5 C-1131T [rs662799], LPL HindIII [rs320], APOC3 SstI [rs5128], family history of obesity, and diabetes, and APOA1 MspI [rs2893157]. Such features for Set 2 were CETP TaqIB [rs708272], ApoE, LPL D9N [rs1801177], ABCAI R1587K [rs2230808], age, birth weight, family history of obesity and for Set 3 were abdominal obesity, birth weight, physical activity, family history of diabetes, and BMI category. The performance of the best classifiers in each subset (i.e. the proposed classifier) was further assessed using 4-fold cross validation (Table5).

The proposed prediction system showed limited discriminant power (DP = 1.3), excellent diagnosis accuracy (AUC ROC = 0.94), excellent agreement with the gold standard (Kappa = 0.87) and high correlation with the gold standard (MCC=0.87) on the second subset (Table4). The average statistical power and Type I error (α) were 93 % and 0.07, respectively based on the cross-validation on the second subset (Table5). The training time of the proposed system was 26.1 ± 2.2 (s), 33.6 ± 3.0 (s) and 20.5 ± 3.1 (s)in the first, second, and third subsets, respectively. The average running time was the average of 3 runs over 363 subjects in the training set (hold-out 50%) on an Intel Core i7-6500uCPU with 8 GB of RAM.

4. Discussion

Identifying high-risk children based on gene polymorphisms (sets 1, and 2), at the first place, is useful for further dietary, and life-styletreatments and screening. Using life-style, anthropometric indicators and family history of diseases (set 3), on the other hand, could identify the high-risk population in low-income countries.

4.1. The risk factors of dyslipidemia

Although the environment is very important in the development of dyslipidemia, genetic components are also critical [81]. CETP TaqIB [rs708272] was selected by the proposed dyslipidemia prediction system in both sets 1 and 2. In the literature, Genome wide association studies (GWAS) in adults showed a high correlation between CETP and plasma lipid concentrations [82]. However, such an association is less distinct in children [33,83]. It was shown in the literature that such a mutation has the protective effect on dyslipidemia [33] and Myocardial Infarction (MI) [84]. This was in agreement with our findings, where the OR of CT/TT vs. CC was 0.15 (P-value<0.001) (Table3).

ApoE was also selected in both sets. ApoE, playing an important function in lipid metabolism, has three isoforms, Apo-e2, Apo-e3, and Apo-e4. They are in fact translated into three alleles of the gene. It was shown in the literature that ApoE , and particularly, its e4 isoform, is associated with plasma lipid parameters and CVD risks [85,86]. Similarly, in our study, the prevalence of dyslipidemia was 85% in subjects with ApoE-e4 isoforms. Moreover, the OR of e2/e4 vs. e3 was 1.73 (P-value < 0.001) (Table3).

ABCAI R1587K [rs2230808] was the other selected feature in both sets 1 and 2. Several ABCA1 gene polymorphisms including R1587K [rs2230808], were identified. Dean et al. showed that this SNP is associated with the HDL-C concentration [87], thus affecting dyslipidemia. In our study, the OR of AG/GG vs. AA was 2.21 (P-value < 0.001) (Table3). Thus, such polymorphisms increased the risk of dyslipidemia.