IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 3 July 2023, accepted 13 August 2023, date of publication 23 August 2023, date of current version 30 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3307731

RESEARCH ARTICLE

Improving the Accuracy of Adult Height Prediction With Exploiting Multiple Machine Learning Models According to the Distribution of Parental Height

JI-SUNG PARK[1] **AND DONG-HO LEE**[2]

[1]Department of Applied Artificial Intelligence (Major in Bio Artificial Intelligence), Hanyang University, Seoul 04763, South Korea
[2]Department of Artificial Intelligence, Hanyang University ERICA Campus, Ansan 15588, South Korea

Corresponding author: Dong-Ho Lee (dhlee72@hanyang.ac.kr)

This work was supported in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) funded by the Korean Government (MSIT) through the Artificial Intelligence Convergence Innovation Human Resources Development (Hanyang University ERICA Campus) under Grant RS-2022-00155885, and in part by the National Research Foundation of Korea (NRF) funded by the Korean Government (MSIT) under Grant 2022R1F1A1073208.

ABSTRACT Grade schoolers and teenagers wonder how tall they will be, as there is a tendency to prefer taller stature for many years. Child's height growth is one of the continuous interests of the parents from the past to the present for many reasons, not only their children's outer beauty but also health status of children. Pediatricians also want to make sure a child is growing as expected because the height growth of children is an important indicator for monitoring a child's nutrition and diseases. In many previous studies, adult height prediction method using growth curves is used widely. Unfortunately, growth curves are based on longitudinal cohort studies which are very challenging to conduct. That's why it is hard to find the related studies for certain ethnic group. In this study, we collected 2,687 Korean height data including parental heights and children's heights by ourselves in the same format as Galton's Height data at 1880s in the United Kingdom. Then, we focus on the influence of parental height on child's height conducting various analysis comparing Galton's and Korean height data. Especially, we find out the linearity of child's height varies depending on the combination of each parental height through visualization analysis. Finally, we propose our method of deploying the best among various machine learning techniques according to the combination of parental height. The combination is based on distribution of each parental height. And it outperforms achieving RMSE under 3.5 compared to single machine learning models which cannot achieve RMSE even under 4.0. It will be a simple and good application for many of pediatricians and parents who care a lot about their children's height growth.

INDEX TERMS Child's adult height prediction (AHP), data analysis, machine learning, healthcare.

I. INTRODUCTION

Height is one of the most crucial health indicators, determined by genetic factors that are influenced by proteins, diseases, and various environmental factors accumulated over ancestral generations. Due to its significance, the World Health Organization (WHO) has collected height data from individuals in various countries. Based on this data, WHO has established the Child Growth Standards, which are used to monitor the growth and development of children worldwide. Additionally, the Centers for Disease Control and Prevention (CDC) has also developed a similar growth chart that records the growth status of children. Both resources provide valuable data for assessing children's health and growth [1], [2], [3]. Likewise, Korean organizations such as Korean Academy of Pediatrics and the Korea Centers for Disease Control

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Magno.

and Prevention have also released Korean National Growth Chart to analyze growth status of children [4]. These growth charts and height datasets serve research purposes and play an important role in improving the health and wellbeing of children.

Numerous researchers and organizations have conducted cohort studies aimed at comprehending relationships between various factors and track, monitor the status of children's growth. Among the most notable datasets collected through these endeavors are Galton's height data [5], [6], [7], 1974/1990 Gothenburg data, and Edinburgh data. Notably, Gothenburg and Edinburgh data have been employed for longitudinal studies due to their ability to track and analyze the height growth of each child. These longitudinal studies make it possible to analyze changes in growth patten over time and pubertal growth patterns using the Quadratic Exponential Pubertal Stop (QEPS) model with height velocity [8], [9].

There are numerous reasons to study a child's height beyond simply monitoring their health status. For example, in sports, it is very important to know the adult height of young athletes in advance. Success as an athlete, hinges on the dynamic interaction of physical growth, biological maturation, and behavioral development. For this reason, numerous studies have been conducted in the field of sports to predict the height and maturity of young athletes using factors like chronological age, height, sitting height, leg length and so on [10], [11], [12], [13]. In addition, height is one of the attractive points that can appeal to people [14], [15], [16]. In recent years, traditional mass media, including TV, magazines, and social media, have fueled a growing fascination with stature. This trend is particularly evident in various countries, with South Korea being a notable example, where considerable attention is given to the height of potential partners. While preferences can vary by gender, a common thread is the preference for partners who are taller than the average.

The growing interest has spurred numerous studies and approaches aimed at predicting adult height. These approaches can be categorized into two main areas. First area is to analyze hormones and genetic information that can directly influence a child's adult height. Studies on growth hormones as therapy for short stature children [17], nutrients that have a positive effect on growth hormones [18], as well as the identification of genes [19] are conducted. The focus of approaches in this area is to identify direct factors affecting adult height and to address obstacles hindering children's growth. The second area involves the analysis of human bones. Information such circumference, length, and spacing between bones, are used to predict adult height. Tanner-Whitehouse 3 (tw3) method is employed to analyze radiograph of left hand and predict bone age [20], [21]. Some studies, such as [22], extend this analysis to both hands. Furthermore, even in the field of forensic science, an approach employing metatarsal length to predict height is conducted [23]. Given the significant correlation between height and bone growth, many studies in this area also utilize height, sitting height, and leg length as parameters.

Recently, an approach involving height data and machine learning techniques has emerged [25], [26], [27]. The study [25] analyze Galton's height data and predict adult height using various machine learning techniques. The authors of this study plan to conduct further analysis by adding features and using cohort data. Another study [26], also utilizing machine learning techniques, is centered around information from children under the age of 6 in the Gothenburg and Edinburgh datasets. This study validates the use of machine learning techniques to predict adult height based on growth measurements before the age of 6. The study using Growth Curve Comparison (GCC) method [27] are based on SLOfit cohort data which consists of information of Slovenian students between 6 and 19 years of age. This method involves comparing and analyzing the similarity of growth curves among students and subsequently predicting their height at the next age by forecasting the Peak Height Velocity (PHV). In these studies, extreme gradient boosting (XGB) and XGB using PHV are employed as machine learning techniques.

In this study, we selected parental height as an indicator of genetic and environmental factors, including ethnic information, drawing inspiration from previous works [5], [6], [7], [25], [33] in the realm of Adult Height Prediction (AHP). The analysis process is conducted based on two groups: Galton's height data and Korean height data, allowing for a comparison between these two groups with variations in time, region, and race. To verify the validity of Korean height analysis and experiments, we gathered our own Korean height data, including both parental and child height information. Moreover, gender was stratified into female and male categories, considering its potential influence on the performance of AHP [24].

In section II, Data Analysis & Feature Engineering were conducted to understand data and impact of parental height on child's height with visual analysis and correlation coefficients. For accurate AHP, this section also encompassed outlier removal. In section III, we conducted the experiments of AHP based on linear regression analysis exploiting various machine learning techniques like previous studies [25], [26], [27].

As a result of section II, we observed that the linearity between parental height and a child's height was weak. Instead, a stronger relationship between parental height and a child's height was evident when specific combinations of the father's and mother's heights were considered. This implies that the entire dataset of child's heights and smaller data segments might follow different linear trends. To address this, we employed the piecewise linear regression technique [28], [29], [30] in our study. This approach takes into account both the distribution and combinations of parental heights.

The purpose of this study is to analyze the influence of parental height on child's height by conducting comparison between two different population groups. Additionally, this study aims to predict adult height of children, especially Korean children, more accurately by exploiting machine learning based segmented linear regression, taking into consideration the combination of parental height.

II. DATA ANALYSIS AND FEATURE ENGINEERING

A. DATASET

Prior to the experiments that we prepared; the most important thing is to understand the data we use. We chose Galton's height data, which consists of 898 people, which collected in the late 19th century in England. Surely, there are Gothenburg data and Edinburgh data collected in 1974 and 1990 respectively, which recorded child's heights as they grew up [16], [17], [18]. Even though, various analyses, such as analyzing the certain factors that affect growth in specific time, but it is too difficult and takes lots of time to collect these longitudinal child's height data. On the other hand, Galton's height data, including not only child's height but also parental height which has information of growth factors.

That's why we chose Galton's height data for the experiments in this paper, which is simple to collect.

Go back to Galton's height data, this data consists of 6 features.

- Family Numbers: Family ID
- Gender: F/M (Female/Male)
- Kids: The number of children in one family
- Father: Height of father (recorded in inches)
- Mother: Height of mother (recorded in inches)
- Height: Height of child (recorded in inches)

Additionally, we made some feature transformations to suit our experiments. Gender, initially recorded as characters, cannot be directly used in regression analysis with machine learning. To address this, we transformed it into 'M' represented by 1 and 'F' represented by 2. Furthermore, considering that the inch unit is not commonly used in South Korea, we converted the measurements of Father, Mother, and Height from inches to centimeters.

To perform comparative analysis with Galton's height data, we collected Korean height data which consists of 2,687 people, who are twenties, born in the 1990s through survey. And this data is composed of 6 features, same with Galton's height data. This Korean height data consists of total 1471 families comparing to 205 families in Galton's height data. The reason why the number of families in Korean height data compared to the total data is relatively smaller rather than Galton's height data is that Koreans in the end of the 20th century tended to have fewer children than British people in the late 19th century. The results of numerical analysis on height data of each data are shown in Table 1 and Table 2.

TABLE 1. Galton's height data before outlier removal.

	Father	Mother	Son	Daughter
Count	898	898	465	433
Mean	175.85	162.77	175.844	162.84
Standard	6.27	5.85	6.68	6.02
Max	199.39	179.07	200.66	179.07
Upper Bound	191.77	177.038	193.675	177.8
Q3 (75%)	180.34	166.37	180.34	166.37
Q2 (50%)	176.53	162.56	175.768	162.56
Q1 (25%)	172.72	159.258	171.45	158.75
Lower Bound	161.29	148.59	158.11	147.32
Min	157.48	147.32	152.4	142.24

TABLE 2. Korean height data before outlier removal.

	Father	Mother	Son	Daughter
Count	2687	2687	1264	1423
Mean	173.76	160.68	175.86	162.64
Standard	6.14	5.93	6.48	5.99
Max	194	187	202.02	193
Upper Bound	190	175.57	192	179
Q3 (75%)	178	164.61	180	166.65
Q2 (50%)	173	160.3	175.925	162.78
Q1 (25%)	170	157.3	172	158.25
Lower Bound	158	146.33	160	145.65
Min	157	141.6	150	144.5

TABLE 3. Outliers of each data.

	Korean	Galton's
Father	6	3
Mother	32	5
Son	31	7
Daughter	15	4
Father / Son	2	0
Father / Daughter	2	1
Mother / Son	12	0
Mother / Daughter	6	0
Both Parents	0	0

TABLE 4. Galton's height data after outlier removal.

	Father	Mother	Son	Daughter
Count	807	807	424	383
Mean	175.72	163.1	175.77	163.01
Standard	5.77	5.16	5.33	6.18
Max	191.77	175.26	193.04	177.8
Q3 (75%)	180.34	166.37	180.34	166.37
Q2 (50%)	176.53	162.56	175.768	162.56
Q1 (25%)	172.72	159.258	171.45	158.75
Min	162.56	149.86	158.75	149.86

B. OUTLIER REMOVAL

First, for data analysis and feature engineering, removal of outlier of each data is preceded by calculating quartiles. There are a lot of methods to remove the outliers like Local Outlier Factor (LOF), calculation of residual after regression and Isolation Forest (IF). We selected two methods, quartile-based outlier removal and IF for the experiments and chose quartile-based outlier removal finally. IF method is not adequate for our overall experiments as shown in Experiment 1. More details about IF method, are explained in Experiment 1. In this section, All the analysis are conducted with quartile-based outlier removal.

The quartile-based outlier removal approach designates values above the upper bound and below the lower bound

TABLE 5. Korean height data after outlier removal.

	Father	Mother	Son	Daughter
Count	2551	2551	1178	1373
Mean	173.56	160.61	175.81	162.45
Standard	5.72	5.28	5.59	5.50
Max	190	176	192	178.66
Q3 (75%)	178	164.61	180	166.65
Q2 (50%)	173	160.3	175.925	162.78
Q1 (25%)	170	157.3	172	158.25
Min	158	147	160	146.9

as outliers, subsequently removing them from the dataset. Concerns may arise regarding the potential impact of outlier removal on regression analysis outcomes. Given that our experiments are fundamentally rooted in regression utilizing machine learning techniques, it's worth noting that regression analysis with machine learning is particularly sensitive to rare and exceptional outliers. These outliers can significantly influence the results of regression analysis. Therefore, we performed outlier removal prior to conducting experiments to ensure precise analysis. As depicted in Table 3, the number of outliers is relatively small compared to the entire height dataset of both Koreans and Galton's dataset. The premise behind the outlier removal stems from the hypothesis that exceptions might exist, such as cases involving small parents and tall children, or vice versa. Following the removal of outliers from each dataset, we present the numerical outcomes in Table 4 and Table 5. We conducted an analysis of the outliers that were eliminated during the outlier removal process. The amounts of outliers were very small. Specifically, a total of 58 outliers were identified in the Korean height data, while Galton's height data contained 17 outliers. These figures are in comparison to the total dataset sizes of 2,687 and 898, respectively. After the outlier removal, the Korean height data consists of 2,551 individuals, and Galton's height data consists of 807 individuals.

C. CHILD'S HEIGHT DISTRIBUTION RELATIVE TO PARENTAL HEIGHT

Figure 1 aims to elucidate the distribution of children relative to each parent and the degree of linearity between the height of children and their parents in Galton's height data.

Figure 1-(a) illustrates the distribution of child's heights relative to parental heights using Galton's height data. The left-side plot of Figure 1-(a) shows the child's height distribution relative to the father's height, while the right-side shows the child's height distribution relative to mother's height.

Figure 1-(b) shows that height distributions of sons (left) and daughters (right) relative to father's heights. Like the child's height distribution in Figure 1-(a), the height distributions of sons and daughters, stratified by gender, follow a Gaussian distribution pattern. Notably, the daughter's height distribution is narrower than the son's height distribution, as observed in Figure 1-(b). We can assume that daughter's height distribution relative to father's height has stronger linearity than son's height distribution.

Figure 1-(c) shows that the height distributions of sons (left) and daughter (right) also follow Gaussian distribution and the height data of each son and daughter relative to mother also have weak linearity.

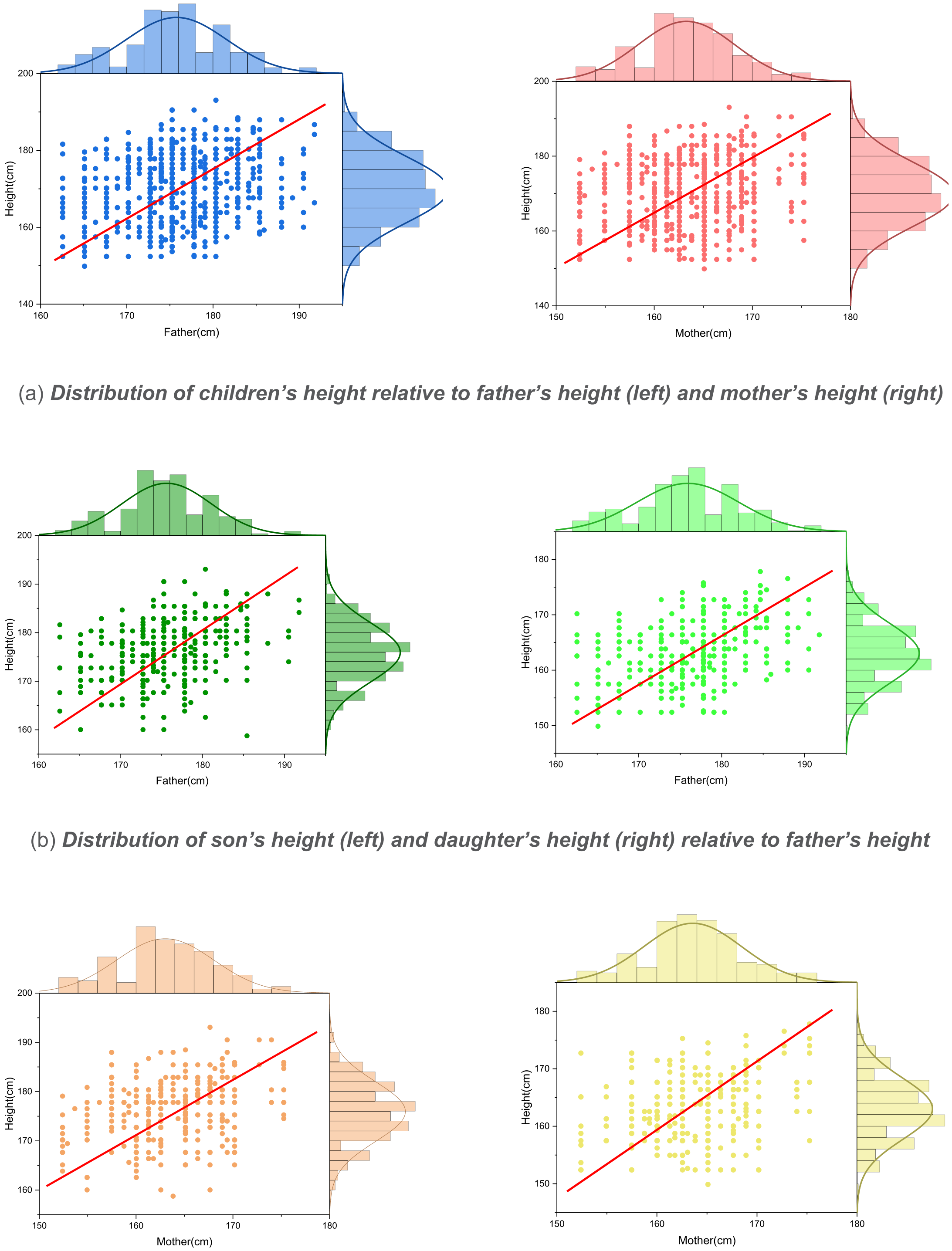
However, the analysis on child's height distribution in Galton's height data as shown in Figure 1 has weakness of the amount of data. As the amount of Galton's height data is too small and sparse, overall figures show the weak linearity between children and their parents. It means that performance of linear regression using machine learning models can also be low.

Figure 2 shows the plots of each case like Figure 1. But unlike Figure 1 which analyzes Galton's height data, Figure 2 shows the distribution and linearity of Korean height data. Comparing to Galton's height, Korean height data which we collected by ourselves are denser. Of course, the amount of Korean height data is three times larger. Generally, an increase in the amount of the data doesn't result in making data denser. Thus, the observed result means that our collected data have common characteristic related to each parental height.

Figure 2-(a) shows the distribution of children and each parent as like Figure 1-(a). The distributions of child's height relative to each parent in Figure 1-(a) and Figure 2-(a) don't look quite different from each other except density of height data.

Figure 2-(b) shows the height distribution of children by gender relative to father's height. The data of child's height by gender relative to father's height in Korean Height data are distributed like child's height distribution in Galton's height data but more densely. It makes explain the linearity of child's height by gender and father's height more clearly. And unlike Galton's height data, the amount of Korean height data makes it possible to figure out the difference between the height distributions of son and daughter relative to father's height. It shows that daughter's height data relative to father's height are distributed narrower and denser than son's data as shown in Figure 2-(b). It means that AHP on daughter may be predicted more accurately comparing to AHP for son.

Figure 2-(c) shows distribution of son and daughter relative to mother's height. The distribution of daughter's height relative to mother's height is denser and narrower when compared to the distribution of son's height relative to mother's height, similar to Figure 2-(b). A consistent observation between Figure 2-(b) and 2-(c) is that daughter's height distributions relative to each parent exhibit a narrower and denser pattern compared to son's height distributions. It means that AHP for daughters can be more accurate than AHP for sons. To validate this observation, we will conduct correlation analysis and apply machine learning techniques. The conclusions drawn from the visual analysis using plots will be further supported by the results obtained from the correlation analysis and the various experiments presented in Section III. Experiments & Results.



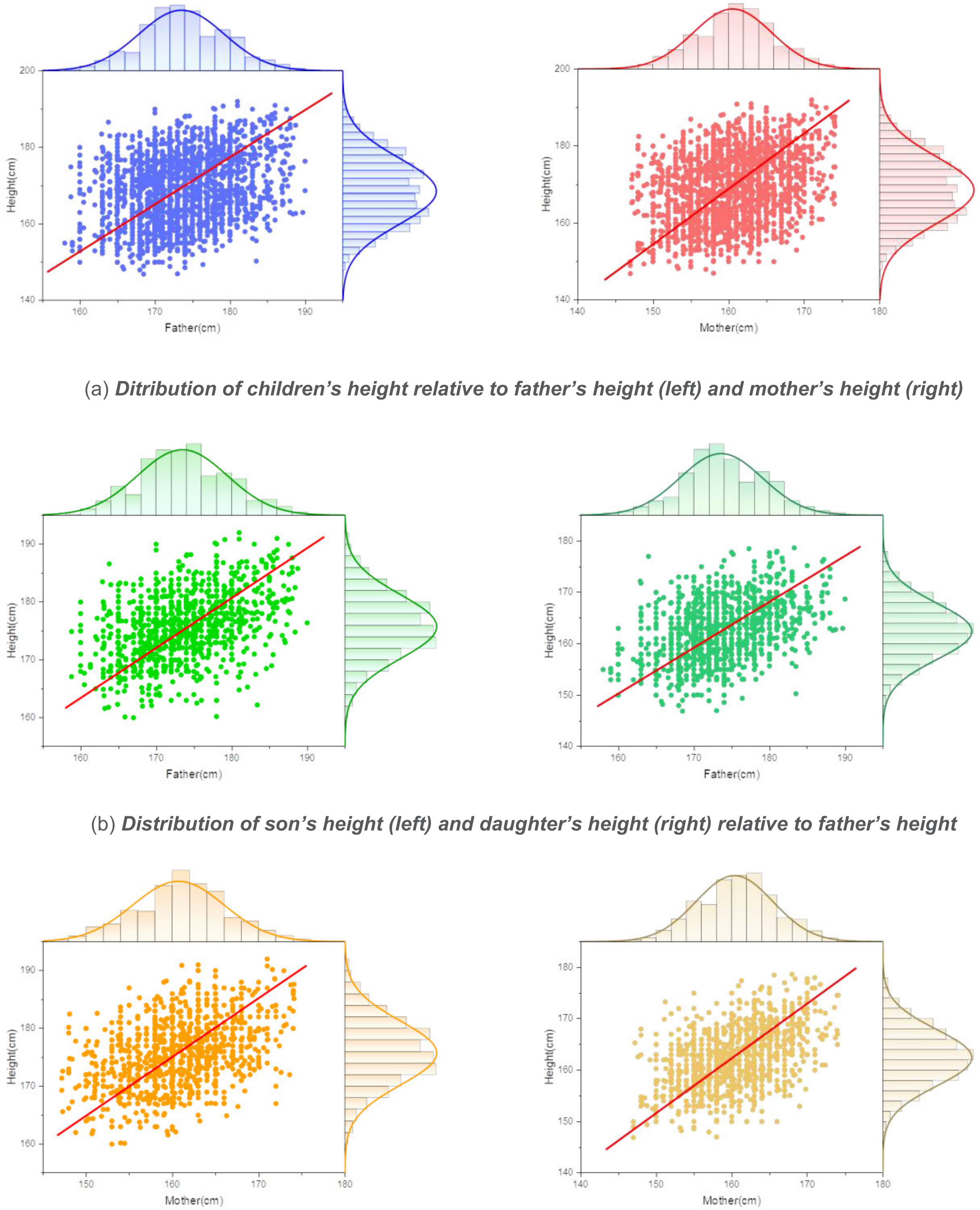
(a) ***Distribution of children's height relative to father's height (left) and mother's height (right)***

(b) ***Distribution of son's height (left) and daughter's height (right) relative to father's height***

(c) ***Distribution of son's height (left) and daughter's height (right) relative to mother's height***

FIGURE 1. Distribution of child's height depending on gender relative to each parental height in Galton's height data

Figure 3 shows the influence of both parental height on child's height visually through 3D contour plot. X-axis and Y-axis of 3D contour plot are parental height responsibly, and Z-axis is the height of children, sons, and daughters. If the color is closer to dark red, the height of the child is taller and if the color is closer to black, the child's height is smaller in 3D contour plot.



(a) ***Ditribution of children's height relative to father's height (left) and mother's height (right)***

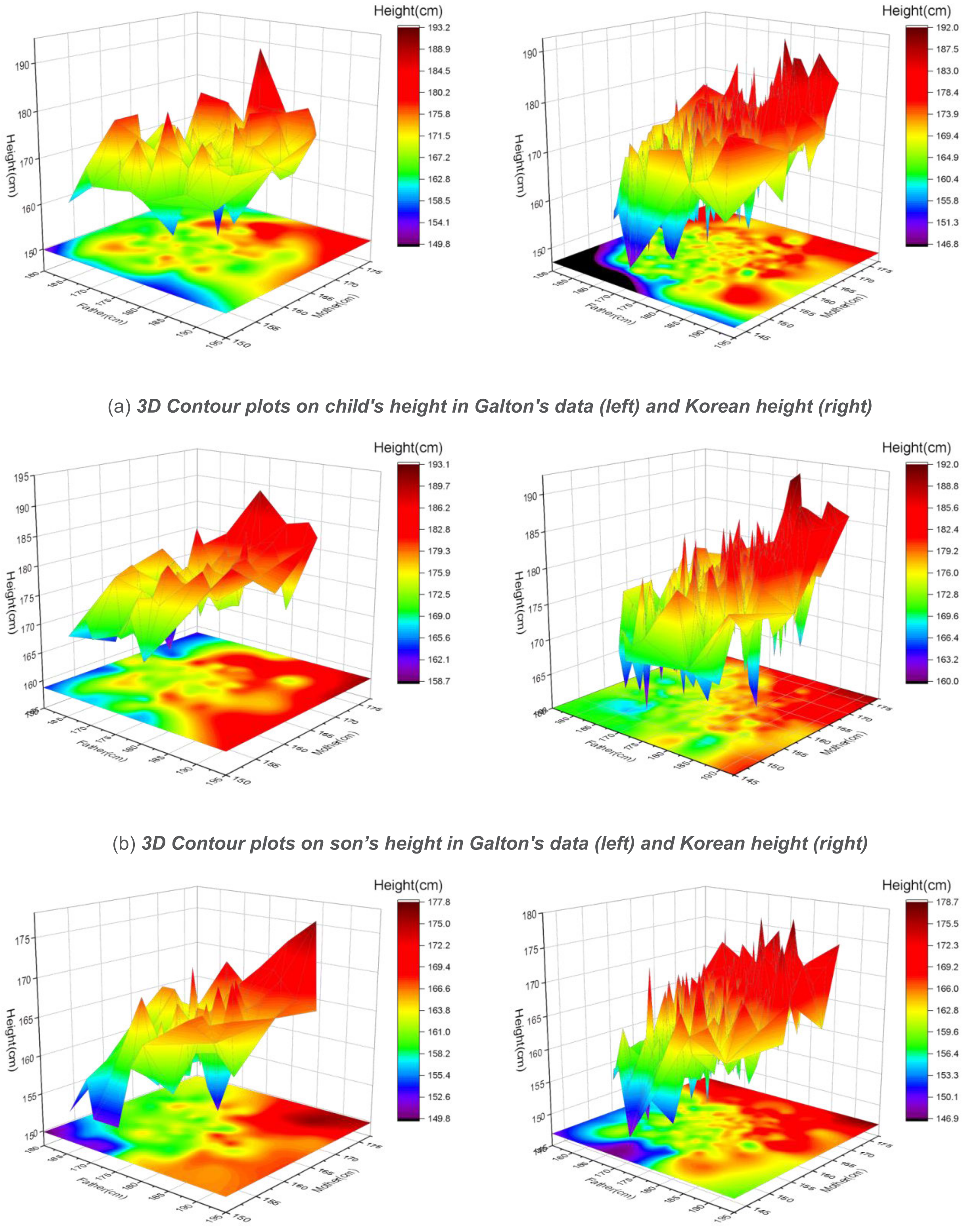
(b) ***Distribution of son's height (left) and daughter's height (right) relative to father's height***

(c) ***Distribution of son's height (left) and daughter's height (right) relative to mother's height***

FIGURE 2. Distribution of child's height depending on gender relative to each parental height in Korean height data

The left-side of Figure 3 from (a) to (c) shows the influence of both parental height on child's height in Galton's height data.

Both left and right side of Figure 3-(a) are the 3D contour plots of child's height relative to father's height and mother's height simultaneously in Galton's height data and Korean height data. Child's height in this 3D contour plot is not divided into gender. The son's height and daughter's height can introduce noise to each other to analyze the influence of parental height on child's height visually. To ensure accurate



(a) ***3D Contour plots on child's height in Galton's data (left) and Korean height (right)***

(b) ***3D Contour plots on son's height in Galton's data (left) and Korean height (right)***

(c) ***3D Contour plots on daughter's height in Galton's data (left) and Korean height (right)***

FIGURE 3. 3D Contour plots on child's height depending on gender in Galton's data (left) and Korean data (right)

visual analysis, child's height is subsequently divided into the heights of sons and daughters.

In Galton's height data, both the height of sons and daughters tends to be taller than the median if the father's height surpasses 185cm, regardless of the mother's height. Conversely, the child's height, regardless of gender, seems to be smaller than the median when the father's height is below 170cm, even if the mother's height exceeds 175cm. This implies that if the distribution of a child's height changes frequently in response to changes in the parents' heights, the height of the parent who undergoes these changes more frequently can exert a stronger impact on the child's height. As depicted on the left side of Figure 3-(b) and (c), the range of mother's height from 145cm to over 175cm doesn't appear to have a distinctive influence on the child's height, whereas the child's height appears to vary significantly based on a father's height of approximately 175cm. This leads to the assumption that in Galton's height data, the influence of the father's height on the child's height is more pronounced compared to the influence of the mother's height.

On the other hand, the right side of Figure 3-(b) and 3-(c), which analyze Korean height data, shows different results when compared to the results of the contour plots of Galton's height data. As depicted on the right side of Figure 3-(b), the distribution of son's height does not exhibit a significant change even when the height of the mother changes, except in the case of the father's height being over 175cm. This implies that the father's height plays a more influential role in determining the son's height. In Figure 3-(c), the case of daughter's height has no exceptions, unlike the son's height. Daughter's heights are tall if the mother is tall enough, regardless of the father's height. This indicates that the mother's height has a greater impact on daughter's height compared to son's height.

The 3D contour plots presented in Figure 3 provide an approximate depiction of the relationship between each parental height and the child's height. As demonstrated in Figure 3, the child's height exhibits variations corresponding to alterations in the distribution range of each parental height, both in Galton's and Korean height data. By visually analyzing these distribution ranges of parental height and their combinations, we can gain insights into the potential height range that a child might achieve. This insight forms the fundamental idea behind the deployment strategy pursued to achieve a more accurate AHP in Experiment IV. However, before delving into AHP, it is imperative to validate this observed result. This validation will entail assessing the correlation between parental height and child's height, which will be conducted through the calculation of correlation coefficients in the Correlation Analysis section.

D. CORRELATION ANALYSIS

To validate the results of the visualization analysis of distributions and to comprehend the influence of parental heights on the heights of children, sons, and daughters, we initially calculate the correlation coefficients for both Galton's height data and Korean height data. Recognizing that parental heights exert a complex joint influence on a child's height, we incorporate the concept of Mid-Parental Height (MPH) [31] from both parental heights into the calculation of correlation coefficients. This approach enables us to discern the simultaneous influence exerted by both parents' heights. The calculation equations for MPH are provided below.

Correlation analysis is conducted separately for Galton's height data and Korean height data. The attributes considered in the correlation analysis encompass father's height, mother's height, and MPH. The computation of correlation coefficients between these attributes and child's height is executed separately for each gender of children.

$$Boys(cm) : \frac{(Father's\ Height + Mother's\ Height + 13)}{2} \tag{1}$$

$$Girls(cm) : \frac{\left(Father's\ Height' - 13 + Mother's\ Height\right)}{2} \tag{2}$$

The correlation analysis is based on the method of Pearson. Calculation equations of Pearson correlation analysis are like above three equations.

$$\rho X, Y = \frac{cov(X, Y)}{\sigma X \sigma Y} \tag{3}$$

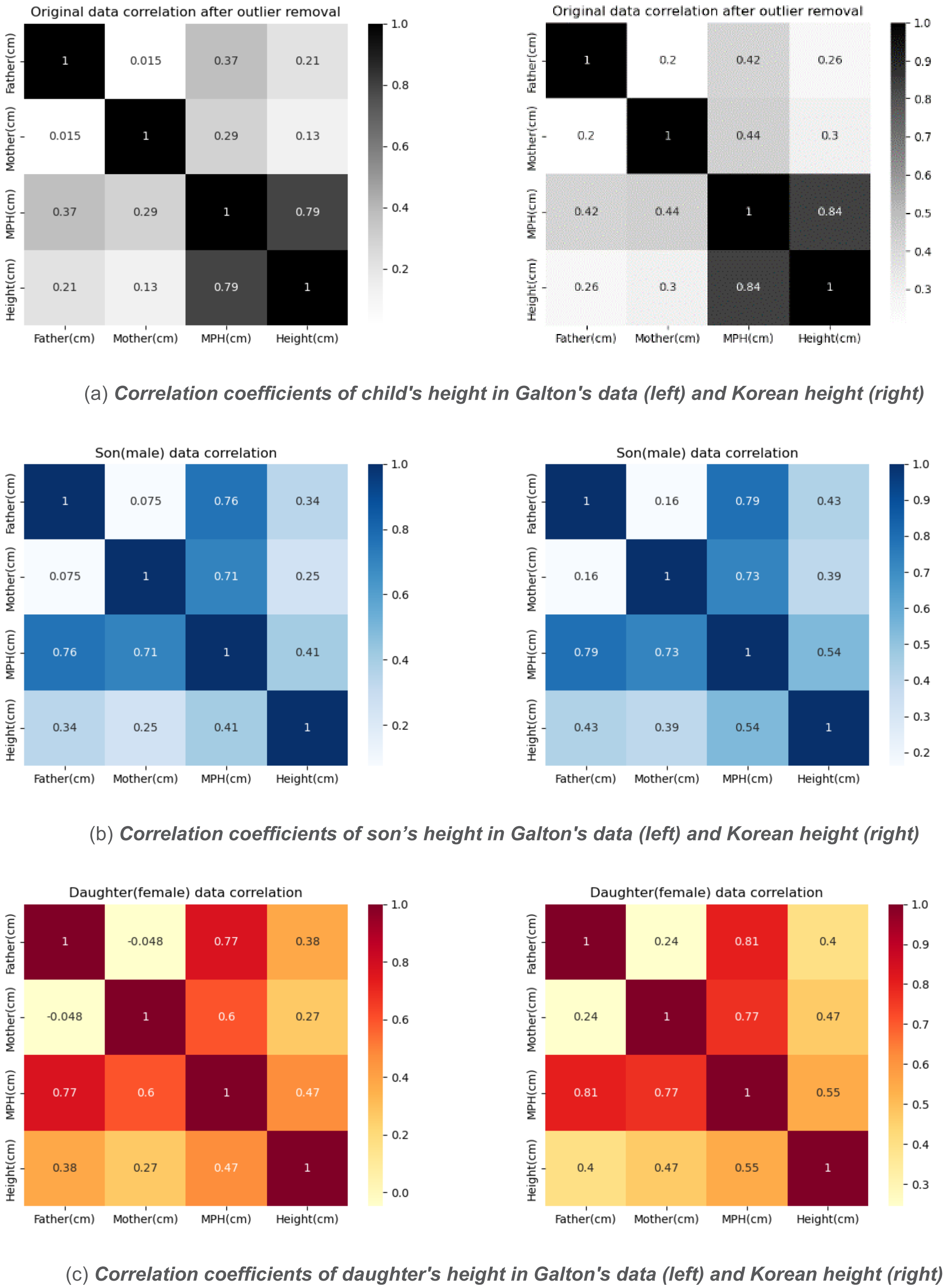
$$Cov(X, Y) = E[(X - \mu x)(Y - \mu y)] \tag{4}$$

$$\rho X, Y = \frac{E[(X - \mu x)(Y - \mu y)]}{\delta X \delta Y} \tag{5}$$

Height(cm) in Figure 4 means the height of child. And coefficients in these figures show how strong the linearity between parental height and child's height is. The left-side tables in Figure 4 show the results of correlation analysis using Galton's height data and the right-side tables in Figure 4 show the results of correlation analysis using Korean height data. As shown in Figure 4, all the results of correlation analysis show that composition of parental height, MPH, has much powerful linearity between child's height rather than single parental height. Because MPH itself considers the gender of children and both parental heights.

As shown in the left-side of Figure 4-(a) which is the result of correlation analysis in Galton's height data, the linearity between the height of father and child is stronger than the linearity between the height of mother and child. It has the same results of Figure 1 and left-side plot in Figure3. The influence of father's height is stronger than mother's height on child's height in Galton's height data. And correlation analysis in left-side plots in Figure 4-(b) and 4-(c) show the same result that the influence of father's height on daughter's height is stronger than the influence of father's height on son's height as shown in Figure 1.

The right-side tables of Figure 4 from (a) to (c) show the results of correlation analysis on height of child, sons, and daughters in Korean height data. The coefficient between mother's height and child's height is slightly higher than the coefficient between father's height and child's height as shown in Figure 4-(a) unlike Galton's height data.



(a) *Correlation coefficients of child's height in Galton's data (left) and Korean height (right)*

(b) *Correlation coefficients of son's height in Galton's data (left) and Korean height (right)*

(c) *Correlation coefficients of daughter's height in Galton's data (left) and Korean height (right)*

FIGURE 4. Correlation coefficients of child's height depending on gender in Galton's data (left) and Korean data (right)

As shown in Figure 4-(b), the coefficient between father's height and son's height is higher than that between mother's height and son's height. It means that the case of son's height can be influenced by father's height more than mother's height in both Galton's height data and Korean height data. In addition, the difference in the influence of each parental height on son's height is slightly smaller than the difference in the influence of each parental height on daughter's height. It means that mother's height on son's height has almost similar influence with father's height on son's height.

However, the right-side of Figure 4-(c) shows that daughter's height is more influenced by mother's height more in Korean height. It shows similar results with the right-side table of Figure 4-(a). Mother's height on daughter's height affects more than father's height on daughter's height in Korean height data unlike Galton's height data. As shown in left-side table of Figure 4-(c), it is not because daughter's height and mother's height are similar by their gender. In Galton's height data, father's height on daughter's height affects more than mother's height even though father is male. It means that the height of a single parental height hasn't greater influence on the child's height simply because of the gender of the parent.

Analyzing the results from Figure 4, we can get the same results of above two analysis which use visualized plots with the results of correlation analysis. The difference in the results of correlation analysis with Galton's data and Korean data has possibility to be caused by amount of child's data by gender not racial characteristics or nutritional status and so on. There are 12 more sons than daughters in Galton's height data on the other hand, there are 159 more daughters than sons in Korean height data. In fact, the correlation coefficients of Galton's data are larger than those of Korean data. This means that the linearity of Korean data is stronger than the linearity of Galton's data. In addition, we observed an interesting result that the linearity of daughter's height and each parental height is stronger than the linearity of son's height and each parental height in common with both height data. And this result leads to the assumption that the performance of AHP on daughter rather than son may be higher in the regression analysis using machine learning techniques. It will be validated in Experiment 2.

III. EXPERIMENTS AND RESULTS

The experiments in this section comprise four components as outlined below. The initial two experiments serve to validate outlier removal and investigate the impact of each parental height on child's height, analyzed by gender through regression analysis with several machine learning techniques. For Experiment 3, we employed LIME and SHAP, which are explainable AI methods, also known as XAI, to elucidate the influence of parental height on machine learning techniques for AHP. Through the data analysis with visualization using plots, we figured out that child's heights are clustered in certain range densely depending on both parental heights. We selected and deployed the machine learning model suitable for the gender of child and a specific quartile-based interval of each parental height focusing on the results of observation and Experiment 1-2. And finally, we achieved the goal to predict adult height when child become grown-up more accurately through this ensemble method of machine learning based on interval of parental height.

TABLE 6. Regression results on child's height in Galton's height data before and after outlier removal.

Galton's Height	Original Data	Outlier Removal	Isolation Forest
MPH	5.30	4.7	5.58
Linear Regression	4.77	4.62	5.07
SVR Linear	4.81	4.63	5.76
XGBoost	4.84	4.60	5.73
LightGBM	4.80	4.57	4.88
NeuralNet	6.05	4.74	5.24

Total experiments are conducted like below.

A. EXPERIMENT 1

Regression analysis on Galton's height data and Korean height data after outlier removal

B. EXPERIMENT 2

Regression analysis on Galton's height data and Korean height data by dividing child's height by gender

C. EXPERIMENT 3

Regression analysis with explainable AI (XAI)

D. EXPERIMENT 4

Regression analysis in accordance with the interval of both parental heights

A total of experiments was conducted using MPH and five machine learning-based regression models. The results of these experiments were evaluated using RMSE (Root Mean Square Error). Five machine learning-based regression models used include Linear Regression, SVR (Support Vector-based Linear Regression), XGBoost, Light Gradient Boosting Machine (LightGBM), and Neural Network (Neural Net).

E. RESULTS OF EXPERIMENT 1

Experiment 1 is to compare the results of two heterogeneous groups, Galton's height data and Korean height data and validate the effect of outlier removal to MPH and 5 machine learning based regression models.

As shown in Table 6 and 7, the results of overall machine learning based methods have higher performance rather than the results of MPH in both Galton's height data and Korean height data. All results of regression in Korean height data are well-performed rather than Galton's height data. It is because that the linearity of Korean height data is much stronger than Galton's height data and fewer amounts of data in Galton's rather than Korean height data. In addition, two tables above show that the validity of outlier removal. Outlier Removal is quartile-based removal. Comparison between

TABLE 7. Regression results on child's height Korean height data before and after outlier removal.

Korean Height	Original Data	Outlier Removal	Isolation Forest
MPH	5.10	4.89	5.27
Linear Regression	4.71	4.43	4.92
SVR Linear	4.77	4.23	4.67
XGBoost	4.68	4.36	4.52
LightGBM	4.54	4.28	4.77
NeuralNet	5.05	4.47	4.77

the quartile-based method and Isolation Forest show different results. Isolation Forest is based on density clustering and detects the height anomaly. It is a powerful method of anomaly detection, but it is not proper to both data. Because both data are composed of single cluster each, and they have few anomaly heights to these single clusters. As a result, it shows worse performance than both original data. On the other hand, quartile-based method shows higher performance rather than original data in Galton's and Korean height data. It is because rare cases of extreme high and low parental stature and child's height are worked as noisy for machine learning methods. Finally, we can validate the quartile-based outlier removal method for both data as mentioned in Data analysis and Feature Engineering section. Experiment 2-4 are conducted with both Galton's height data and Korean height data after the quartile-based outlier removal.

F. RESULTS OF EXPERIMENT 2

The Experiment 2 is to figure out the difference between regression results of children's heights by gender. The target heights are composed of full data (both sons and daughters), son's heights only, and daughter's heights only. Table 8 and Table 9 show the results of regression using machine learning techniques in each Galton's height data and Korean height data. All results of regression using ML techniques have higher performance compared to the results of MPH except the case of Neural Net with Galton's height data. Because the amount of Galton's height data is too small for prediction using Neural Net. Among the composition of ML models, there is no huge difference between the performance of regression. In the case of Korean height data, the linear regression model using the support vector machine and the Boosting-based LightGBM model show good performance, and in Galton's height data, the Boosting-based LightGBM and XGBoost show good results. In terms of overall performance, the Boosting-based LightGBM show good performance.

When analyzing the results of Experiment 2, the results of regression when target data is son's height are poor rather than the results when target data is full data or daughter's height. On the other hand, the regression results have high performance when target data is daughter's height. This means that the results of AHP on daughter's height is slightly more exact compared to results AHP on son's height and the linearity of parental height and daughter's height is stronger than the linearity of parental height and son's height same with the results of Data Analysis and Feature Engineering section. As shown in Table 13 and IX, AHP results by gender separately show the same pattern for both Galton's height data and Korean height data. However, in the case of Galton's height data, the difference between upper bound and lower bound is larger than that of Korean height data. And the amount of Galton's height data that can explain the relation between parental heights and child's height is smaller compared to the amount of Korean height data. Because of those reasons, the performance of regression models is lower than that of Korean height data. By these reasons, we can prove that collecting Korean height data was a meaningful work and collecting Korean height data continuously is very important to explain the linearity of parental height and child's height and get better performance on AHP.

TABLE 8. Regression results on child's height by gender in Galton's height data.

Galton's Height	Full Data	Son Only	Daughter Only
MPH	4.7	5.0	4.49
Linear Regression	4.62	4.79	4.46
SVR Linear	4.63	4.83	4.43
XGBoost	4.60	4.78	4.49
LightGBM	4.57	4.64	4.41
NeuralNet	4.74	5.08	4.67

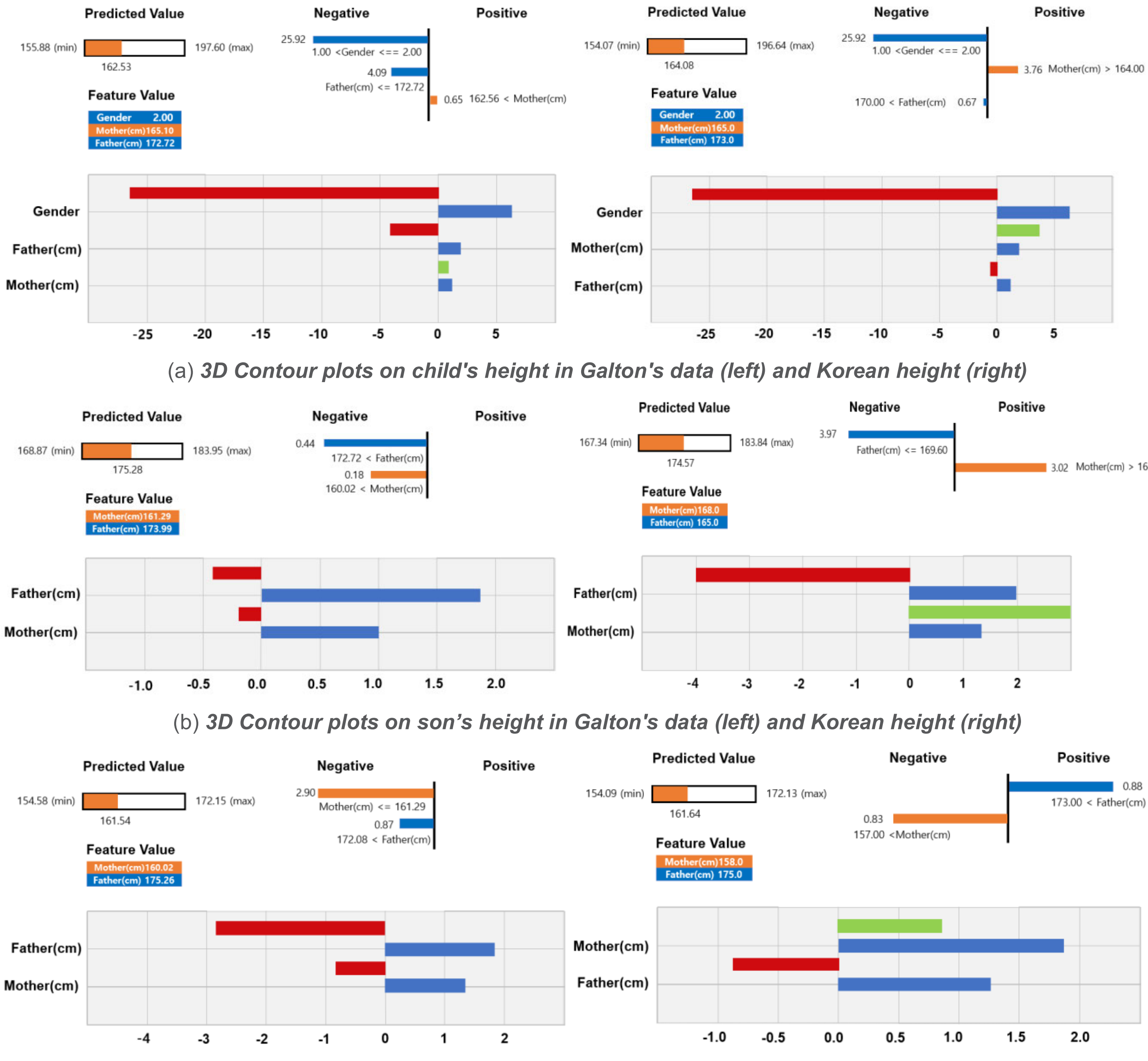
TABLE 9. Regression results on child's height by gender in Korean height data.

Korean Height	Full Data	Son Only	Daughter Only
MPH	4.89	5.15	4.65
Linear Regression	4.43	4.43	4.08
SVR Linear	4.23	4.44	4.07
XGBoost	4.36	4.42	4.01
LightGBM	4.28	4.53	4.12
NeuralNet	4.47	4.77	4.42

G. RESULTS OF EXPERIMENT 3

Experiment 3 is to validate the data analysis performed above. Total analysis is done by exploiting linear regression method with support vector machine, which shows general performance during overall experiment 1 and 2. As shown in Figure 5, figures consist of LIME and SHAP [32], [33] graph and show the influence of parental height on children's height. Through the feature importance calculated by these two XAI models, we can validate the results of data analysis.

As shown in the left-side of Figure 5, the results of both XAI models match results of data analysis exactly in Galton's



(a) *3D Contour plots on child's height in Galton's data (left) and Korean height (right)*

(b) *3D Contour plots on son's height in Galton's data (left) and Korean height (right)*

(c) *3D Contour plots on daughter's height in Galton's data (left) and Korean height (right)*

FIGURE 5. **3D Contour plots on child's height depending on gender in Galton's data (left) and Korean data (right)**

height data except daughter's heights. In Figure 5-(c), feature importance value of LIME shows that mother's heights have more influence on daughter's heights. On the other hand, the case of SHAP shows the opposite result. It is because importance of mother's heights under 161.29 cm is powerful compared to rest of values and LIME considers local properties a lot. But we can think of overall importance of father's heights are more powerful. Then, all results of these experiments are exactly same with those of data analysis. It means we can deduce that the heights of English people in late 19th century were more influenced by father's heights rather than mother's heights.

But as shown in the right-side of Figure 5, they show slightly different results in Korean height data compared to Galton's height data. On the other hand, the results of XAI can validate the results of data analysis. The son's heights are influenced by father's heights more than mother's height, and mother's heights are more important on daughter's heights than father's heights. Finally, overall heights of children are also influenced by mother's heights more than father's heights. But it does not mean that father's heights are not important on children's height. These results are exactly same with data analysis like correlation, and visualization analysis.

The results of data analysis and XAI methods show that there is a difference on child's growth between Galton's height data and Korean height data. The difference between Galton's data and Korean data can be caused by the nutrition status, the race, or genetic reasons. Whatever the causes which make the difference between two data, we need to

TABLE 10. Outliers of each data.

	Mother	Father
1	M < 157.3	F < 170
2	157.3 ≤ M < 160.3	170 ≤ F < 173
3	160.3 ≤ M < 164.61	173 ≤ F < 178
4	164.61 ≤ M	178 ≤ F

separate Galton's data and Korean data for better and clear AHP.

H. RESULTS OF EXPERIMENT 4

Experiment 4 is to enhance the overall performance of AHP which is the eventual purpose in this paper based on previous analysis and experiments. This experiment used Korean height data only because the amount of Galton's height data is too small to divide the interval of each parental height based on quartile.

The main hypothesis of the is experiment is that the combination of each parental height has significant influence on child's height. Additionally, the magnitude of the effect can vary depending on the range of distribution of parent's height. This hypothesis is devised from outlier removal and data analysis section especially visualization analysis in Figure 3. The experiment 4 consists of two works, the categorization and prediction work.

Prior to the prediction work, the categorization work is conducted. In the point of parent, First, we divide parental height into the case of mother and father each. M means mother's height and F means father's height. Second, we divide the distribution interval into 4 based on quartiles as shown in Table 10. With these two variables, total 16 combinations of parental height are possible.

And in the point of child, we divided child into son and daughter. As shown in correlation analysis and Experiment 2-3, there is definite difference between the AHP result of son and daughter. A prediction experiment was conducted for three configuration whole child, son, and daughter.

In the prediction work, regression models with machine learning techniques were used including Linear Regression, SVR Linear, XGBoost, LightGBM, NeuralNet in addition to MPH. It is to experiment the linearity of each combination category of parental height Additionally it is to find and deploy the most adequate machine learning technique for each category. We found and deployed the best machine learning models for each combination of parental height based on their distribution. Eventually, a total of 288 regression analysis were performed for AHP according to each case. Figure 6 shows a simple framework of our deployment strategy and the best models for Korean height prediction.

Table 10 shows the best RMSE results of AHP based on interval of parental height and the best models in accordance with interval of parental height. There was performance improvement when using machine learning techniques comparing to existing method like MPH. The deployment of machine learning models according to the interval of each parental height makes AHP performance improvement going one step further as shown in Table 10.

Especially, the best performance of daughter's height prediction was RMSE 2.34 when mother's height is over 164.61cm (M4) and father's height is under 173cm and over 170cm (F2). And XGBoost model was the best model, compared to RMSE 4.56 of MPH, for this combination. And the best performance of son's height prediction was RMSE 2.48 when mother's height is under 157.3cm and under 160.3cm (M2) and father's height is over 178cm (F4). Linear Regression model was the best model for this combination compared to MPH 4.06. It was a noticeable performance of AHP improvement comparing to previous works.

As shown in Figure 6, The machine learning models which show the best performance in accordance with the interval of parental height by gender are deployed and work in an ensemble method. The overall performance comparison is shown in Table 12. As shown in Table 12, the proposed method of machine learning model deployment outperformed all existing single regression machine learning models. And this experimental result validated that AHP by child's gender separately is an effective way to predict height of child when they become grown-up more accurately. The noticeable results of Experiment 4 are that our proposed method, deployment of machine learning models according to child's gender and the interval of parental height, outperformed existing single regression methods and our proposed method is the only method which achieved RMSE under 4.0. Furthermore, our proposed method is the only method which made RMSE results for child's height prediction by gender less than 3.5.

IV. SUMMARY OF EXPERIMENTS

In this paper, there are data analysis part and 4 experiments conducted. Galton's height data and Korean height data got analyzed separately for the comparison. Through the results of data analysis, we can figure out that the influence of parental heights on children's adult heights, son's heights, and daughter's height. First, Visualization of heights with 2D, 3D plots, make us enable to intuitively grasp the influence of parental heights. Second, it becomes possible to quantify the influence of parental heights through calculation of Pearson's correlation coefficient.

Total experiments 1-4 are based on the regression analysis exploiting machine learning techniques and check the validity of influence of parental heights on children's heights.

Experiment 1 shows that removal of outliers in certain distribution is to validate for ML techniques. After outliers' removal, overall performance of AHP models gets better because the outliers can be a noisy for ML regression models.

Experiment 2 is to check the validity of influence by gender. As many of studies analyzed growth status by gender. That's because there are differences by gender in factors like hormone, bone density and so on which can affect heights of

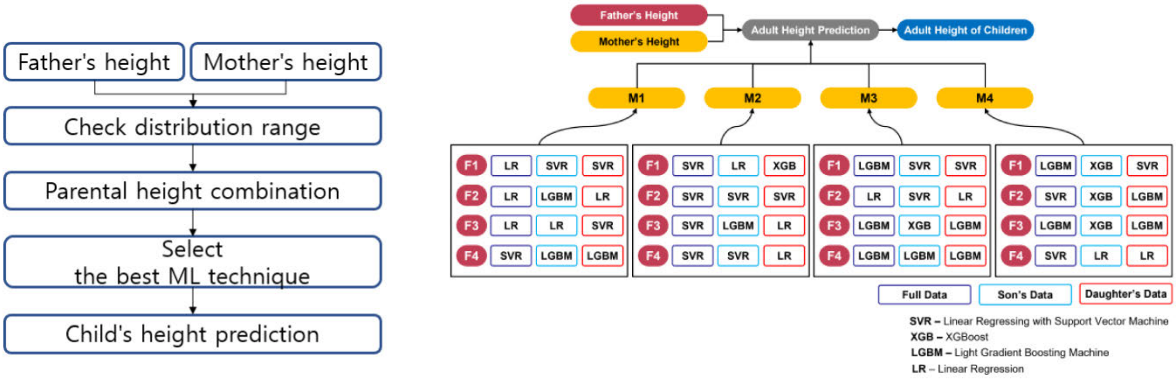


FIGURE 6. A simple framework of deployment strategy and the best machine learning models deployed for AHP depending on the intervals of Korean parental heights

children directly. And we can figure out there are differences on AHP by gender with ML regression techniques.

Experiment 3 is to validate the results of data analysis using XAI techniques. As shown in the results of the experiments 1-2, AHP on Korean height data is better than Galton's height data, but there is no such huge difference. However, when analyzing data and the results of experiment 3, There are quite different to each other. In a case of Korean height data, heights of children are more influenced by mother's heights, although heights of son are opposite. In a case of Galton's height data, heights of fathers are more influential on all of children's heights. This result is same with the results of correlation analysis with both data.

Experiment 4 enhance the performance of AHP. The hypothesis of this experiment is that the distribution of parental heights affects adult heights of children. Even though, when dividing parental heights into 4 groups each, total 16 cases, it is too small to test on Galton's height data, it shows enhanced performance on AHP. Of course, there are some studies on AHP with ML but the deployment of the best models on each parental height distribution, makes better performance.

V. DISCUSSION

Through the many of studies, we know that many factors of the heredity and the environment have influence on child's height. However, we also know that it is impossible to figure out the factors of the heredity and environment which have positive and negative effect on child's height simultaneously. That's why over all analysis and experiments were conducted with an assumption that statistical distribution of child's height are the results caused by the heredity of parents and many environmental reasons.

We prepared two heterogeneous data groups called Galton's height data and Korean height data. During the process of Data Analysis and Feature Engineering, it was confirmed that the influence of parental height on child's height was different between Galton's height data and Korean height data. In Galton's height data, the influence of father's height comparing to mother's height is stronger relatively. In Korean height data show opposite result that the influence of mother's height is stronger comparing to the influence of father's height. Of course, these results may be led by the fact that the amount of both data samples are too small, but there also will be the possibility that the different results between two data are caused by the factors of era, region, race, and lots of reasons. It is very dangerous to conduct the experiments using both data together without figuring out the factors on child's height because it can lead to the wrong results. That's why Korean height data we collected by ourselves is very valuable to predict Korean child's height and the more Korean height data, we can get better performance on AHP of Korean.

Since many studies are based on longitudinal data quite different from our data, there is no direct comparator of our studies except [25], [26]. They try to use a single machine learning technique for AHP, and we used this method for comparison. Only using a single machine learning technique cannot achieved RMSE under 4.

The deployment of machine learning models according to the interval of each parental height enhance the performance of Korean AHP. In this part, we set quartile values of each parental height as the cut-off values for the interval of each parental height. It is based on the thought that child's height based on the interval of each parental height clustered closely to each other and the proper machine learning models for AHP are different according to the interval of parental height. We validated this hypothesis by conducting Experiment 4.

As a result, we got RMSE values under 4 which a single machine learning model cannot achieve using our proposed method. Furthermore, we achieved RMSE values under 3 in the specific interval of each parental height. Unfortunately, Experiment 4 with Galton's height data was impossible because of lack of data. Throughout the total results of analysis and experiments, we can validate importance of our Korean height data and validity of our proposed method finally.

Even though we have achieved AHP better. There is a significant trade-off for this experiment. As we focus on the

TABLE 11. Regression results on child's height by gender in Korean height data.

Mother Quartile	Father Quartile	Regression Model	RMSE		
			Full Data	Son	Daughter
Mother under Q1	Father under Q1	MPH	5.76	5.79	5.73
(M1)	Full Data length: 536	Linear Regression	4.34	4.09	4.07
	Son Data length: 238	SVR Linear	3.92	4.05	3.85
	Daughter Data length: 298	XGBoost	4.10	4.56	4.15
	(F1)	LightGBM	4.14	4.18	4.16
		NeuralNet	4.58	4.37	4.84
	Father between Q1 and Q2	MPH	5.09	5.53	4.81
	Full Data length: 241	Linear Regression	3.98	4.30	3.90
	Son Data length: 89	SVR Linear	4.20	4.57	4.70
	Daughter Data length: 152	XGBoost	4.34	4.00	4.21
	(F2)	LightGBM	4.25	3.27	4.25
		NeuralNet	5.00	4.57	4.43
	Father between Q2 and Q3	MPH	5.03	5.80	4.25
	Full Data length: 152	Linear Regression	3.11	3.69	3.69
	Son Data length: 71	SVR Linear	3.97	4.00	3.56
	Daughter Data length: 81	XGBoost	3.90	3.98	3.62
	(F3)	LightGBM	3.43	3.81	3.66
		NeuralNet	5.26	4.76	4.04
	Father over Q3	MPH	4.85	4.64	5.11
	Full Data length: 68	Linear Regression	4.49	4.01	3.92
	Son Data length: 39	SVR Linear	2.94	3.61	3.78
	Daughter Data length: 29	XGBoost	4.12	3.17	4.10
	(F4)	LightGBM	3.76	3.11	2.59
		NeuralNet	5.87	3.86	3.86
Mother between Q1	Father under Q1	MPH	5.40	5.69	5.11
and Q2	Full Data length: 291	Linear Regression	4.57	3.89	3.58
(M2)	Son Data length: 141	SVR Linear	3.91	4.13	3.96
	Daughter Data length: 150	XGBoost	4.38	4.83	3.43
	(F1)	LightGBM	4.00	4.14	4.06
		NeuralNet	5.33	4.66	3.65
	Father between Q1 and Q2	MPH	5.11	5.21	4.99
	Full Data length: 140	Linear Regression	3.59	3.17	3.97
	Son Data length: 72	SVR Linear	3.48	2.74	3.31
	Daughter Data length: 68	XGBoost	4.97	4.43	3.79
	(F2)	LightGBM	4.33	3.77	3.66
		NeuralNet	4.13	3.98	3.99
	Father between Q2 and Q3	MPH	4.59	4.94	4.42
	Full Data length: 111	Linear Regression	4.04	4.46	3.44
	Son Data length: 35	SVR Linear	3.38	4.15	3.56
	Daughter Data length: 76	XGBoost	4.25	3.81	3.62
	(F3)	LightGBM	4.36	2.77	4.20
		NeuralNet	5.66	3.91	3.87
	Father over Q3	MPH	5.04	5.65	4.06
	Full Data length: 73	Linear Regression	4.59	4.18	2.48
	Son Data length: 42	SVR Linear	3.01	3.68	2.78
	Daughter Data length: 31	XGBoost	4.33	3.87	3.21
	(F4)	LightGBM	4.29	3.77	3.65

TABLE 11. ***(Continued.)*** **Regression results on child's height by gender in Korean height data.**

		NeuralNet	5.83	4.24	3.39
Mother between Q2 and Q3 (M3)	Father under Q1 Full Data length: 216 Son Data length: 101 Daughter Data length: 115 (F1)	MPH	5.18	4.89	5.43
		Linear Regression	4.50	4.01	4.10
		SVR Linear	4.54	3.58	3.25
		XGBoost	4.64	4.25	4.07
		LightGBM	4.06	4.17	3.52
		NeuralNet	5.15	4.28	5.37
	Father between Q1 and Q2 Full Data length: 175 Son Data length: 81 Daughter Data length: 94 (F2)	MPH	4.90	5.04	4.78
		Linear Regression	3.76	4.06	2.85
		SVR Linear	3.88	3.35	3.09
		XGBoost	4.09	4.10	4.11
		LightGBM	4.58	3.64	4.28
		NeuralNet	5.70	3.73	3.88
	Father between Q2 and Q3 Full Data length: 127 Son Data length: 62 Daughter Data length: 65 (F3)	MPH	4.78	4.56	4.98
		Linear Regression	5.17	4.11	4.29
		SVR Linear	4.69	3.54	3.75
		XGBoost	4.29	2.58	3.13
		LightGBM	3.57	3.18	3.12
		NeuralNet	4.99	3.25	4.30
	Father over Q3 Full Data length: 73 Son Data length: 31 Daughter Data length: 42 (F4)	MPH	4.79	5.39	4.29
		Linear Regression	4.11	5.17	3.87
		SVR Linear	4.43	4.58	3.62
		XGBoost	4.78	4.30	3.54
		LightGBM	4.08	3.28	3.49
		NeuralNet	5.22	4.88	3.95
Mother over Q3 (M4)	Father under Q1 Full Data length: 102 Son Data length: 59 Daughter Data length: 43 (F1)	MPH	5.18	4.67	5.81
		Linear Regression	4.62	4.19	3.63
		SVR Linear	4.34	3.96	3.17
		XGBoost	4.27	3.67	4.07
		LightGBM	3.99	3.90	3.49
		NeuralNet	4.95	4.49	4.28
	Father between Q1 and Q2 Full Data length: 83 Son Data length: 41 Daughter Data length: 42 (F2)	MPH	4.39	3.14	5.34
		Linear Regression	3.89	2.82	4.14
		SVR Linear	3.49	2.76	4.08
		XGBoost	5.58	2.34	3.74
		LightGBM	4.25	2.47	3.19
		NeuralNet	4.49	2.39	4.52
	Father between Q2 and Q3 Full Data length: 68 Son Data length: 30 Daughter Data length: 38 (F3)	MPH	4.82	4.56	5.01
		Linear Regression	4.07	3.26	3.54
		SVR Linear	4.34	3.37	4.04
		XGBoost	4.78	2.44	4.48
		LightGBM	3.43	3.35	3.32
		NeuralNet	6.84	3.20	4.22
	Father over Q3 Full Data length: 80 Son Data length: 37 Daughter Data length: 43 (F4)	MPH	4.90	5.03	4.78
		Linear Regression	3.75	3.23	2.81
		SVR Linear	3.68	3.47	4.41
		XGBoost	5.66	3.71	4.19
		LightGBM	3.89	3.53	3.30
		NeuralNet	5.72	3.91	3.64

TABLE 12. Comparison between regression models and our deployment method.

Korean Height	Full Data	Son Only	Daughter Only
MPH	4.89	5.15	4.65
Linear Regression	4.43	4.43	4.08
SVR Linear	4.23	4.44	4.07
XGBoost	4.36	4.42	4.01
LightGBM	4.28	4.53	4.12
NeuralNet	4.47	4.77	4.42
Deployment method	3.73	3.45	3.43

influence of parental height, we cannot explain the difference of siblings. To make our experiment perfect, we need cohort data of children and their parent both. And the growth curve prediction is our final goal based on experiment 4. Since both the child's height data and the growth curve for the parental height are required causing a lot of time and cost, the next experiment is under planning.

VI. CONCLUSION

For this study, we collected 2,687 Korean heights data and it becomes more than 3 thousand of heights data including their parent's heights. And we compared these data to Galton's height data collected in the late 19th. The comparison analysis between these two heterogeneous data groups is so meaningful that it is possible to confirm that it is difficult to apply Galton's height data directly to modern Korean AHP. Between the two data groups, there are clear differences on feature importance of parental heights. Furthermore, we can reduce the RMSE which means the difference between the predicted heights and actual heights under 4.0 using only Korean height data using our proposed method by deploying the different machine learning model according to the quartile-based interval of parental height. And we can predict adult's height of child even under RMSE 3.0 in some interval of parental height. It is very valuable because only using parental heights and information of child's gender makes the performance of AHP better, not requiring the results of complex examination like examination of bone mineral density. Eventually, collecting Korean height data leads to more accurate Korean height prediction through our proposed method, and improved AHP can be applied and help in many ways to children's height growth in health care fields.

REFERENCES

[1] M. de Onis and A. W. Onyango, "WHO child growth standards," *Lancet*, vol. 371, no. 9608, p. 204, Jan. 2008.

[2] M. de Onis, A. Onyango, E. Borghi, A. Siyam, M. Blössner, and C. Lutter, "Worldwide implementation of the WHO child growth standards," *Public Health Nutrition*, vol. 15, no. 9, pp. 1603–1610, Sep. 2012.

[3] M. de Onis, C. Garza, A. W. Onyango, and E. Borghi, "Comparison of the WHO child growth standards and the CDC 2000 growth charts," *J. Nutrition*, vol. 137, no. 1, pp. 144–148, Jan. 2007.

[4] J. H. Kim, S. Yun, S.-S. Hwang, J. O. Shim, H. W. Chae, Y. J. Lee, J. H. Lee, S. C. Kim, D. Lim, S. W. Yang, K. Oh, and J. S. Moon, "The 2017 Korean national growth charts for children and adolescents: Development, improvement, and prospects," *Korean J. Pediatrics*, vol. 61, no. 5, p. 135, 2018.

[5] F. Galton, "Regression towards mediocrity in hereditary stature," *J. Anthropological Inst. Great Britain Ireland*, vol. 15, pp. 246–263, Jan. 1886.

[6] T. J. Cole, "Galton's midparent height revisited," *Ann. Hum. Biol.*, vol. 27, no. 4, pp. 401–405, Jan. 2000.

[7] S. Senn, "Francis galton and regression to the mean," *Significance*, vol. 8, no. 3, pp. 124–126, Sep. 2011.

[8] A. Holmgren, A. Niklasson, A. F. M. Nierop, L. Gelander, A. S. Aronson, A. Sjöberg, L. Lissner, and K. Albertsson-Wikland, "Estimating secular changes in longitudinal growth patterns underlying adult height with the QEPS model: The grow up Gothenburg cohorts," *Pediatric Res.*, vol. 84, no. 1, pp. 41–49, Jul. 2018.

[9] A. Holmgren, A. Niklasson, A. F. M. Nierop, G. Butler, and K. Albertsson-Wikland, "Growth pattern evaluation of the Edinburgh and Gothenburg cohorts by QEPS height model," *Pediatric Res.*, vol. 92, no. 2, pp. 592–601, Aug. 2022.

[10] R. M. Malina, A. D. Rogol, S. P. Cumming, M. J. C. E. Silva, and A. J. Figueiredo, "Biological maturation of youth athletes: Assessment and implications," *Brit. J. Sports Med.*, vol. 49, no. 13, pp. 852–859, Jul. 2015.

[11] R. M. Malina, T. P. Dompier, J. W. Powell, M. J. Barron, and M. T. Moore, "Validation of a noninvasive maturity estimate relative to skeletal age in youth football players," *Clin. J. Sport Med.*, vol. 17, no. 5, pp. 362–368, Sep. 2007.

[12] R. L. Mirwald, A. D. Baxter-Jones, D. A. Bailey, and G. P. Beunen, "An assessment of maturity from anthropometric measurements," *Med. Sci. Sports Exerc.*, vol. 34, no. 4, pp. 689–694, 2002.

[13] S. A. Moore, H. A. McKay, H. Macdonald, L. Nettlefold, A. D. Baxter-Jones, N. Cameron, and P. M. Brasher, "Enhancing a somatic maturity prediction model," *Med. Sci. Sports Exerc.*, vol. 47, no. 8, pp. 1755–1764, 2015.

[14] A. Courtiol, M. Raymond, B. Godelle, and J.-B. Ferdy, "Mate choice and human stature: Homogamy as a unified framework for understanding mating preferences," *Evolution*, vol. 84, pp. 2189–2203, Apr. 2010.

[15] H.-L. Tao, "Gender-role ideology and height preference in mate selection," *Econ. Hum. Biol.*, vol. 39, Dec. 2020, Art. no. 100927.

[16] J.-R. Park and S.-H. Park, "A comparison of the preference by gender on the height of males & females and the female body," *J. Korean Soc. Clothing Textiles*, vol. 34, no. 3, pp. 437–447, Mar. 2010.

[17] P. F. Collett-Solberg, G. Ambler, P. F. Backeljauw, M. Bidlingmaier, B. M. Biller, M. C. Boguszewski, P. T. Cheung, C. S. Y. Choong, L. E. Cohen, P. Cohen, and A. Dauber, "Diagnosis, genetics, and therapy of short stature in children: A growth hormone research society international perspective," *Hormone Res. Paediatrics*, vol. 92, no. 1, pp. 1–14, 2019.

[18] V. Ekbote, A. Khadilkar, S. Chiplonkar, Z. Mughal, and V. Khadilkar, "Enhanced effect of zinc and calcium supplementation on bone status in growth hormone-deficient children treated with growth hormone: A pilot randomized controlled trial," *Endocrine*, vol. 43, no. 3, pp. 686–695, Jun. 2013.

[19] E. Lin, S.-J. Tsai, P.-H. Kuo, Y.-L. Liu, A. C. Yang, M. P. Conomos, and T. A. Thornton, "Genome-wide association study in the Taiwan biobank identifies four novel genes for human height: NABP2, RASA2, RNF41 and SLC39A5," *Hum. Mol. Genet.*, vol. 30, no. 23, pp. 2362–2369, Nov. 2021.

[20] Y. J. Oh, B. K. Yu, J. Y. Shin, K. H. Lee, S. H. Park, K. C. Lee, and C. S. Son, "Comparison of predicted adult heights measured by Bayley-Pinneau and Tanner-Whitehouse 3 methods in normal children, those with precocious puberty and with constitutional growth delay," *Clin. Exp. Pediatrics*, vol. 52, no. 3, pp. 351–355, 2009.

[21] K.-Y. Kang, J.-K. Han, and Y.-H. Kim, "The study on correlationship between parent's height and adult height prediction according to TW3 method," *J. Korean Oriental Pediatrics*, vol. 26, no. 3, pp. 46–54, Aug. 2012.

[22] H. B. Khazri, S. C. Shimmi, and M. T. H. Parash, "A multivariate analysis to propose linear models for the stature estimation in the Sabahan young adult population," *PLoS ONE*, vol. 17, no. 8, Aug. 2022, Art. no. e0273840.

[23] J.-H. Park, M. Lee, D. Kim, H.-W. Kwon, Y.-J. Choi, K.-R. Park, S. Park, S.-B. Park, and J. Cho, "Estimating adult stature using metatarsal length in the Korean population: A cadaveric study," *Int. J. Environ. Res. Public Health*, vol. 19, no. 22, p. 15124, Nov. 2022.

[24] D. L. Kuh, C. Power, and B. Rodgers, "Secular trends in social class and sex differences in adult height," *Int. J. Epidemiol.*, vol. 20, no. 4, pp. 1001–1009, 1991.

[25] J. R. Cordeiro, O. Postolache, and J. C. Ferreira, "Child's target height prediction evolution," *Appl. Sci.*, vol. 9, no. 24, p. 5447, 2019.

[26] M. Shmoish, A. German, N. Devir, A. Hecht, G. Butler, A. Niklasson, K. Albertsson-Wikland, and Z. Hochberg, "Prediction of adult height by machine learning technique," *J. Clin. Endocrinol. Metabolism*, vol. 106, no. 7, pp. 2700–2710, Jun. 2021.

[27] M. Mlakar, A. Gradišek, M. Lustrek, G. Jurak, M. Sorić, B. Leskošek, and G. Starc, "Adult height prediction using the growth curve comparison method," *PLoS ONE*, vol. 18, no. 2, Feb. 2023, Art. no. e0281960.

[28] C. Il-Gyo and J. Chi-Hyuck, "A prediction method combining clustering method and stepwise regression," in *Proc. Korean Oper. Manag. Sci. Soc. Conf.*, 2003, pp. 949–952.

[29] X. Yang, H. Yang, F. Zhang, L. Zhang, X. Fan, Q. Ye, and L. Fu, "Piecewise linear regression based on plane clustering," *IEEE Access*, vol. 7, pp. 29845–29855, 2019.

[30] A. Bemporad, "A piecewise linear regression and classification algorithm with application to learning and model predictive control of hybrid systems," *IEEE Trans. Autom. Control*, vol. 68, no. 6, pp. 3194–3209, Jun. 2023.

[31] J. M. Tanner, "Use and abuse of growth standards," *Hum. Growth*, vol. 3, pp. 95–109, Jan. 1986.

[32] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[33] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016.

JI-SUNG PARK received the B.S. and M.S. degrees in computer engineering from Hanyang University, South Korea, in 2017 and 2020, respectively, where he is currently pursuing the Ph.D. degree in applied artificial intelligence (major in bio artificial intelligence). His current interests include artificial intelligence, natural language processing, and data analysis.

DONG-HO LEE received the B.S. degree from Hongik University, in 1995, and the M.S. and Ph.D. degrees in computer engineering from Seoul National University, South Korea, in 1997 and 2001, respectively. From 2001 to 2004, he was with the Software Center, Samsung Electronics Company Ltd., where he was involved in several research projects for next-generation digital appliances. He is currently a Professor with the Department of Artificial Intelligence, Hanyang University ERICA Campus, South Korea. His research interests include system software for flash memory, embedded DBMS, and multimedia systems.