Supplementary Material

Misaal Khan[1,2], Mayank Vatsa[1], Richa Singh[1], and Kuldeep Singh[2]

[1] Indian Institute of Technology Jodhpur, India
[2] All India Institute of Medical Science Jodhpur, India
{khan.9, mvatsa, richa}@iitj.ac.in, kulpra@gmail.com

1 AnthroVision Dataset

1. **Equipment and Specifications** Images were captured using a standard smartphone camera, specifically the OnePlus Nord model. The camera specifications included a Sony IMX586 sensor, 48 megapixels, a pixel size of 0.8 µm/48M and 1.6 µm (4 in 1)/12M, a 6P lens quantity, and featured both Optical Image Stabilization (OIS) and Electronic Image Stabilization (EIS), with an aperture of f/1.75.
2. **Image Capture Guidelines** Images were typically captured from a distance of approximately 165 cm and at a height of 50 inches. The lack of strict standardization in distance and height was intentional, as it provided variability in the images, enhancing the generalization capabilities of the algorithm.
3. **Image Collection:** Images are captured from three crucial perspectives: frontal, lateral, and posterior. Additionally, facial images are taken as selfies with the subject's face positioned frontally. The collection includes eight variations per subject to ensure comprehensive coverage: four frontal, one lateral-left, one lateral-right, one posterior, and one selfie.
4. **Anthropometric Measurements:** Precision and accuracy are paramount in recording height, weight, middle-upper-arm circumference (MUAC), head circumference, and waist circumference. These ground truth measurements are taken in the presence and training of trained healthcare practitioners

A graphical user interface (GUI) supports the systematic recording of measurements alongside images, ensuring the dataset's organization for training and validation.

1.1 Statistics and Characteristics

The numerical statistics of key measurements are as follows:

The dataset's potential biases stem from several factors:

- **Geographical Coverage:** The dataset is collected from a specific region in Rajasthan from both Clinical and Community settings.
- **Age Distribution:** The age distribution ranges from 1 year to 18.6 years, with an average age of approximately 11.8 years.

Measurement	Mean	Std Dev	Median
Height	126.62 cm	28.41 cm	137.00 cm
Weight	28.81 kg	17.48 kg	27.80 kg
MUAC	19.26 cm	5.93 cm	18.20 cm
HC	52.07 cm	5.21 cm	51.00 cm
Age	11.82 years	3.73 years	10.50 years
BMI	17.51	2.66	15.04

Table 1. Summary statistics of NutiVision

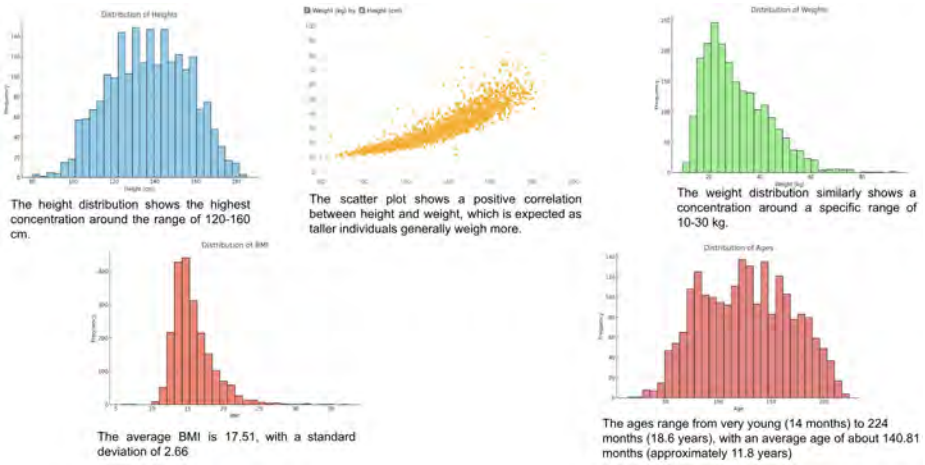

Fig. 1. Visual Representation of Data Statistics

– **Gender Distribution:** The dataset predominantly consists of male subjects across ages.

The data statistics are further outlined in Figure 1 and 2

Model Validation

2 Discretization of labels for Mutual Information Calculation

This conversion enables the calculation of MI for continuous regression tasks alongside discrete classification tasks. The application of the Rice Rule guided the determination of the optimal number of bins for histogram analysis. This decision rule was chosen due to its efficacy in providing a reasonable balance for datasets of varied sizes (train : 1284, test:428, valid:429). The Rice Rule recommends setting the number of bins to $2 \times n^{1/3}$, where n is the number of samples in the dataset. To ensure consistency and robustness in our analysis, we applied the Rice Rule based on the size of our smallest dataset segment.

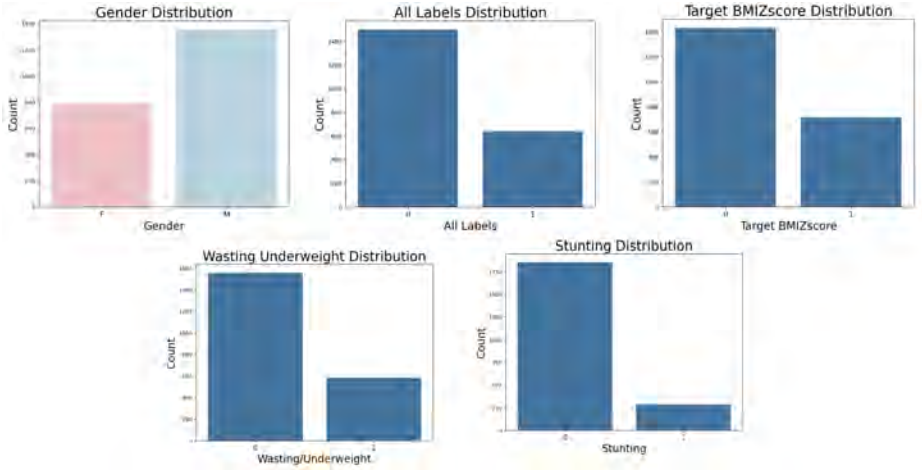

Fig. 2. Class Imbalance in AnthroVision

3 Ablation study to analyse impact of poses on model performance.

Table 2 **A** and **B** explore the ablation of removing each pose one by one to analyse its impact on performance. The removal of *Lateral* pose proved to be most beneficial for improving classification performance while the removal of *Frontal pose* gave the best performance in Regression.

Table 2. Model Performance Ablation by Pose Removal

A. Impact of removing one pose at a time on classification (Precision ↑, Recall ↑, F1 ↑) tasks									
2*Model (Pose Excluded)	**T7**			**T8**			**T9**		
	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑
DomainAdapt (all poses)	0.51 ± 0.02	0.62 ± 0.01	0.54 ± 0.02	0.62 ± 0.05	0.69 ± 0.10	0.64 ± 0.14	0.63 ± 0.01	0.67 ± 0.04	**0.64 ± 0.03**
Frontal	0.46 ± 0.13	0.68 ± 0.21	0.55 ± 0.18	0.54 ± 0.07	**0.73 ± 0.04**	0.62 ± 0.05	0.49 ± 0.04	0.70 ± 0.03	0.58 ± 0.02
Selfie	0.55 ± 0.03	0.59 ± 0.09	**0.57 ± 0.04**	0.54 ± 0.11	**0.73 ± 0.18**	0.62 ± 0.14	0.61 ± 0.37	0.46 ± 0.21	0.47 ± 0.29
Lateral	0.54 ± 0.11	0.68 ± 0.08	0.55 ± 0.05	**0.81 ± 0.04**	**0.73 ± 0.04**	0.62 ± 0.05	**0.72 ± 0.24**	**0.71 ± 0.17**	0.59 ± 0.18
Back	**0.78 ± 0.18**	**0.69 ± 0.13**	0.56 ± 0.09	0.63 ± 0.15	0.67 ± 0.12	**0.65 ± 0.18**	0.49 ± 0.21	0.70 ± 0.12	0.58 ± 0.23

B. Impact of pose elimination on regression (RMSE↓) tasks						
Model (Pose Excluded)	**RMSE T1 ↓**	**RMSE T2 ↓**	**RMSE T3 ↓**	**RMSE T4 ↓**	**RMSE T5 ↓**	**RMSE T6 ↓**
DomainAdapt (all poses)	43.76 ± 1.07	**2.96 ± 0.14**	22.02 ± 0.66	12.44 ± 0.22	3.55 ± 0.24	**5.05 ± 1.07**
Frontal	**41.74 ± 1.02**	3.57 ± 0.15	**21.05 ± 0.67**	11.87 ± 0.23	**3.38 ± 0.18**	6.24 ± 0.56
Selfie	44.54 ± 1.12	4.77 ± 0.21	25.72 ± 0.66	12.24 ± 0.34	3.77 ± 0.11	7.25 ± 0.73
Lateral	42.55 ± 1.07	3.08 ± 0.14	25.41 ± 0.66	**11.81 ± 0.22**	3.52 ± 0.24	7.25 ± 1.07
Back	45.70 ± 1.23	4.58 ± 0.31	29.93 ± 1.02	12.92 ± 0.52	4.65 ± 0.25	10.41 ± 0.67