Exploring Metabolic Health: Impacts on Body Composition and Immunity via PCA and Regression Analysis

Fernando Pavía-González[1], Jorge Sánchez-Ponce[2], Luis C. Marrufo-Padilla[3], Mateo Yañez-Tanaka[4] and Vidal A. García-López [5]

[1] *Tecnologico de Monterrey, Campus Guadalajara, Data Science and Mathematics Engineering*

Abstract— This study presents a novel approach to understanding the complex relationships between diet, metabolic health, and body composition by integrating advanced statistical techniques with an innovative method for variable reduction. Leveraging data from the National Health and Nutrition Examination Survey (NHANES), Principal Component Analysis (PCA) was applied for dimensionality reduction, retaining components that captured 95% of the total variance. To enhance the analysis, a new method was introduced based on the sum of squared loadings to eliminate variables with minimal contribution to the retained components. Subsequently, Multiple Linear Regression (MLR) models were constructed to correlate anthropometric measures such as Body Mass Index (BMI) and waist circumference to dietary and bloodwork variables. The analysis identified significant predictors, including copper and vitamin D intake, underscoring their critical roles in metabolic regulation and body composition. Additionally, inflammatory markers, including white and red blood cell counts, were positively correlated with Body Mass Index (BMI), highlighting the interplay between immunity and metabolism. This novel combination of Principal Component Analysis (PCA) and squared-loadings-based variable elimination offers a robust framework for simplifying datasets while preserving critical information, advancing the understanding of diet and health interrelations. Future work will refine predictive models through outlier detection and multicollinearity diagnostics to improve accuracy and generalizability.

Keywords— Metabolic health, PCA, Squared loadings, Multiple Linear Regression, BMI, Diet, Immunity.

I. Introduction

Health is the most valuable asset a person can possess, yet it is often taken for granted until it begins to decline. The intricate balance of nutrition, metabolic processes, and physiological function plays a pivotal role in sustaining physical vitality and overall well-being. Poor dietary habits disrupt this harmony, leading to cascading effects that compromise not just physical health but also mental and emotional resilience. This study delves into the connection between nutrition and health, uncovering how key nutrients and metabolic markers influence the body's ability to maintain this delicate equilibrium. By addressing these relationships, we aim to provide actionable insights for improving health and enhancing quality of life.

II. Literature Review

The literature reviewed focused on exploring literature around the relationship between nutrition and its effects on human physical factors, providing a scientific foundation for the central theme of this project. A comprehensive review of studies was conducted, uncovering key insights into how diet influences anthropometric measurements and, in turn, impacts quality of life (QoL).

Anthropometric measurements, which include evaluations such as arm and waist circumference, body mass, and standing height, are essential in understanding the physical dimensions of human health. These measurements are not only descriptive but also serve as diagnostic tools when analyzed alongside nutritional data. For instance, the Body Mass Index (BMI) is a well-established metric used to assess body fat based on a person's weight relative to their height. BMI is calculated using the formula:

$$\text{BMI}\,(\text{kg/m}^2) = \frac{\text{Body mass (kg)}}{\text{Body height}^2\,(\text{m}^2)}$$

BMI is a key indicator that correlates with various domains of QoL, including physical and psychological health. Higher BMI values are often associated with lower QoL scores, highlighting the impact of body composition on overall well-being. However, physical activity can act as a crucial moderating factor in this relationship. Body image, which refers to how individuals perceive and feel about their appearance, also significantly affects QoL. Dissatisfaction with body image negatively influences social relationships and perceptions of personal health [1].

As shown in Table 1, BMI is classified into four weight status categories based on specific ranges.

Table 1: BMI categorisation ranges

Weight status	BMI range
Normal weight	$18\,\text{kg/m}^2 \leq \text{BMI} < 25\,\text{kg/m}^2$
Overweight	$25\,\text{kg/m}^2 \leq \text{BMI} < 30\,\text{kg/m}^2$
Obese	$30\,\text{kg/m}^2 \leq \text{BMI} < 40\,\text{kg/m}^2$
Morbidly obese	$\text{BMI} \geq 40\,\text{kg/m}^2$

Obesity, defined as a condition of having a high BMI, has a

substantial impact on health-related quality of life (HRQoL). Research from the Yorkshire Health Study confirms that individuals with morbid obesity score significantly lower on HRQoL indices compared to those with normal weight. Furthermore, obesity's effects extend beyond physical health, encompassing mental health challenges such as depression, anxiety, and impaired body image. Addressing these factors through interventions focused on weight management, physical activity, and mental health support can substantially improve outcomes. These findings emphasize the necessity of adopting comprehensive health strategies in obesity interventions to enhance QoL across physical, mental, and social domains [2].

Following this research, the methodology used in the project will be explained, along with interpretation of the findings, presentation of results, and, finally, a discussion of conclusions.

III. Methodology

A database on human physiology tests and data was selected and divided into six additional databases, each focusing on specific aspects of human physiology. These include diet (containing data on the vitamins and nutrients consumed by a certain number of people), examinations (providing data on physical tests conducted on the study group), questionnaire (recording responses to a survey administered to the sample), laboratories (offering data from biological tests performed on the sample) and demographic (detailing information about the locations where the sample participants reside and people groups they belong to).

The first step taken with the databases was to clean them; this meant dropping every entry that did not belong to the studied demographic, removing all categorical variables, and removing any column and row with less than 80% of their data filled. After that, the K-Nearest Neighbors (KNN) algorithm was applied in order to predict and impute missing values. This approach was chosen as it is more accurate than imputation with the mean. This process took a non-parametric approach with distances between dimensions being calculated so closer neighbors had greater influence on the imputed values. This method leveraged a candidate list ranked by proximity to the missing value, allowing for precise and context-aware imputation. The weight of a candidate neighbor (W_d) is calculated using the formula:

$$W_d(i_m, i_c) = \frac{\Delta(i_m, i_{ck}) - \Delta(i_m, i_c)}{\Delta(i_m, i_{ck}) - \Delta(i_m, i_{c1})}$$

In this formula, i_m represents the instance with missing values, i_c is a candidate neighbor, $\Delta(i_m, i_c)$ denotes the distance between the instance with missing values and the candidate neighbor, i_{ck} is the furthest of the k nearest neighbors, and i_{c1} is the closest neighbor. This formula ensures that closer neighbors (i_{c1}) are assigned greater weights than those further away (i_{ck}), enabling a more precise imputation of missing values by emphasizing proximity. The value imputed is then the weighted average of the KNN [3].

After imputing missing data the mean and variance were obtained and with them the variables were standardized to a variance of 1 and mean of 0 so that those with greater means would not take over the PCA. To do this the following formula was employed:

$$Z = \frac{X - \mu}{\sigma}$$

where μ is the mean of the dataset and σ is the standard deviation. A side effect of this process is that the variables also lost their units, as shown by the dimensionality analysis:

$$\frac{X\ [\text{unit}] - \mu\ [\text{unit}]}{\sigma\ [\text{unit}]} = Z\ [1]$$

The next step was to apply the PCA and analyze its results. PCA was applied through the use of python modules. It was applied to two data frames, the one containing dietary variables and the one containing blood work variables. The first results to be analyzed were those of the data frame with the dietary variables, for which the first analysis was on the loadings of the variables on each of the principal components. The principal components (PC) are described by:

$$Z_m = \phi_{m,1}X_1 + \phi_{m,2}X_2 + \phi_{m,3}X_3 + \ldots + \phi_{m,n}X_n$$

where Z_m represents the PC, ϕ is denoted "loading" and it's a coefficient in the linear combination of the set X which is the set of variables in the space $X = \{X_1, X_2, X_3, \ldots, X_n)\}$.[4] The loadings represent how much of an effect a variable has on each PC, so a heat map was obtained to localize those variables that had the highest effects on each PC

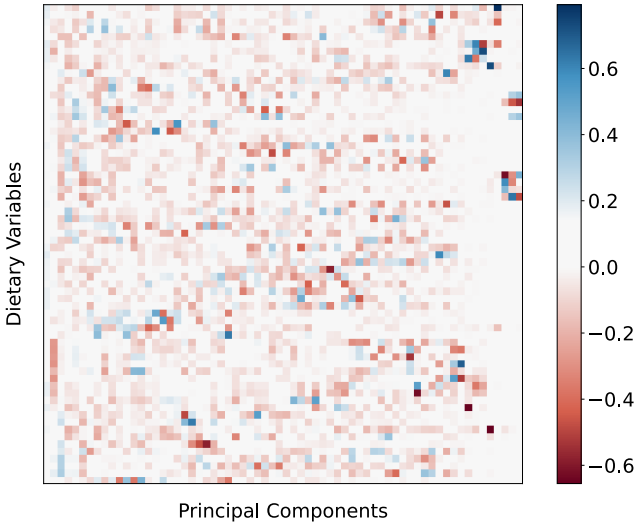

Fig. 1: Heat Map of Component Loadings for Dietary Variables.

From the heat map of the dietary data frame, it can be observed that most variables have a minimal effect on the majority of the principal components, showing strong significance only in a select few.

In order to better understand how the PCs represented the data in the space and how many were necessary; a graph of the cumulative variance explained at each of the PCs was generated, this graph was used to note where the 95% of the variance had been explained.

As the graph shows, 96% of the data had been explained by the 29th PC, thus making the cutoff point PC number 28, therefore resulting in a reduction from 66 to 28 dimensions for this data frame. To evaluate the effectiveness of the dimensionality reduction, the data was reconstructed using

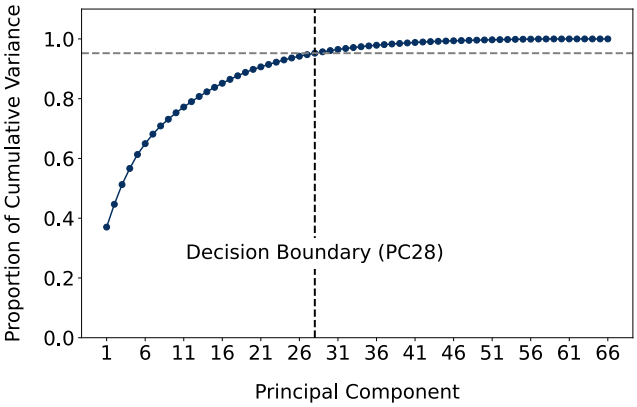

Fig. 2: Graph of Cumulative Variance for Dietary Variables.

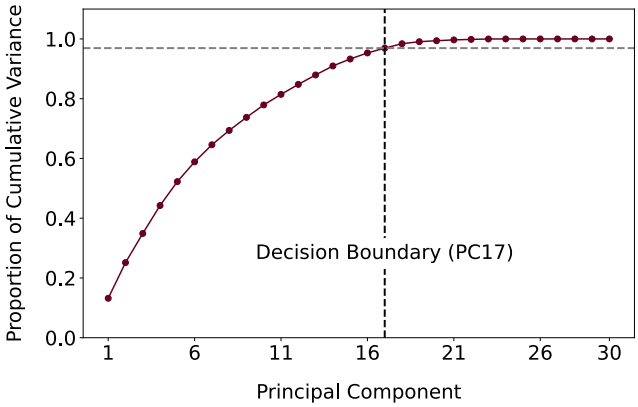

Fig. 4: Graph of Cumulative Variance Explained for Blood Work Variables.

only the 28 PCs and compared to the original dataset. Upon inspection, it was found that the reconstruction was highly accurate.

The results of the blood work data frame were analyzed afterwards. It began in the same way as the aforementioned one, with the analysis of loadings of the variables with each PC, a heat map was derived from this as well to localize the variables with the most effect on the PCs.

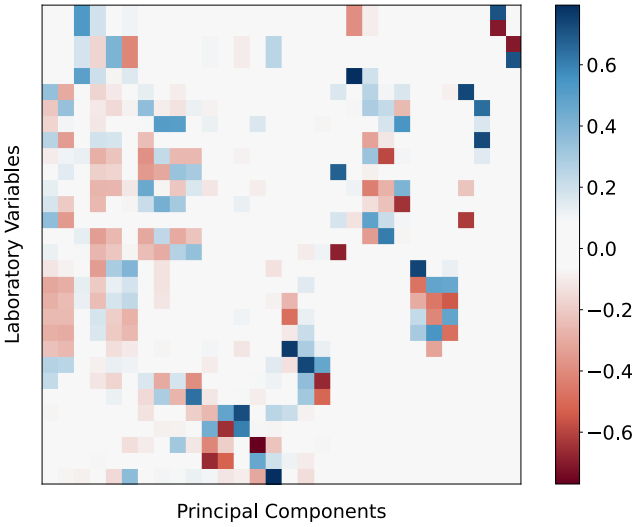

Fig. 3: Heat Map of Component Loadings for the Blood Work Variables.

From this heat map, it can be concluded that all the principal components are only lightly correlated with each of the variables, with most of the loadings being very close to 0. It should also be noted that, although not mentioned previously, both of these processes have been carried out using standard PCA, which implies that:

$$\sum_{i=1}^{n} \phi_{m,i}^2 = 1 \quad \text{and} \quad \sum_{i=1}^{m} \phi_{i,n}^2 = 1$$

This explains why the maximum values of the loadings are around 0.8, while most loadings hover around 0. [4]

Once again, following the same order of steps, the next graph obtained was that of the cumulative explained variance for each PC.

Using this graphic we noted that since 98% of the explained variance accumulated at the 18th PC the cutoff should be PC number 17 which represented a reduction from thirty to seventeen dimensions. The data was again reconstructed using the PCs to confirm their accuracy. Upon inspection, it is confirmed that the reconstructed data closely resembles the original dataset, allowing for the continuation of the variable reduction process.

To reduce the amount of variables to be used in the prediction models the sum of the squared loadings of each variable with respect to the principal components selected for retention was used. This approach was employed because those loadings represent how correlated the variables are to each of the relevant principal components. Therefore, to explain the most variance in the data frame, only the variables that contribute to obtaining the principal components that account for the most variance are considered. The process will began with the diet data frame, generating a bar plot with the sums of the squared loadings with respect to the column they belonged to.

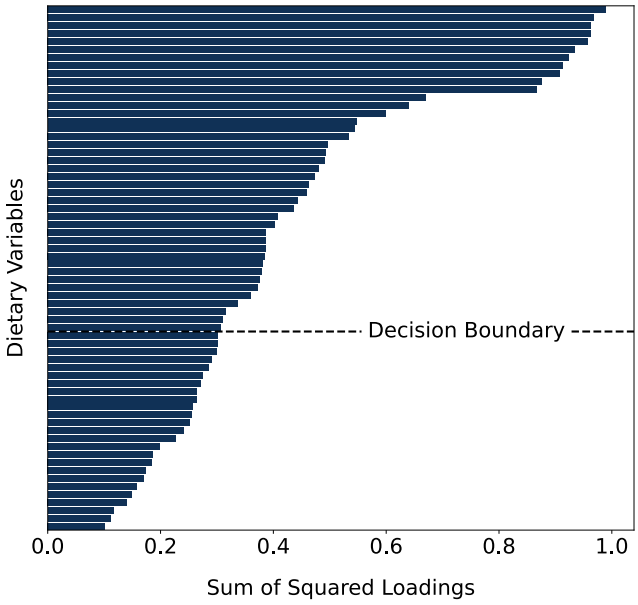

Fig. 5: Sums of Squared Loadings by Dietary Variable.

After examining this graph, a decision was made to exclude those variables from "theamin" onwards, as they do not contribute enough to be considered significant in explaining the variance. However, a few exceptions, identified as irrel-

evant variables suggested by research to be relevant, were retained.

Afterwards, the same process was performed for the loadings of the blood work data frame.

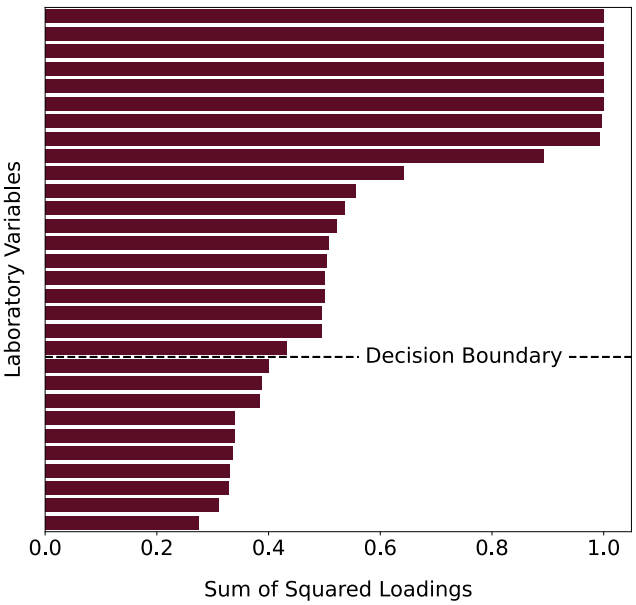

Fig. 6: Sums of Squared Loadings by Blood Work Variable.

After examining this graph, a decision has been made to drop variables from "mean cell hemoglobin concentration" onwards, as they do not significantly contribute to the principal components of this data frame. This process severely reduced the amount of variables in each of the data frames, going from 66 to 44, and 32 to 21 respectively.

With the data frames now reduced, they were rescaled back to their original units in order to generate predictive models, for this we used

$$X = \sigma Z + \mu$$

Before generating predictive regression models an attempt was made to find linear correlations within and between three data frames: the diet data frame, the blood work data frame (Labs) and the anthropometric measures data frame (Examinations), this was done through the use of Pearson Correlation scores, which were plotted in the following heat map.

After finding no simple linear correlations it was decided that MLR should be tested for the prediction of anthropometric variables, for this process, a null hypothesis H_0 was established for each variable, stating that the variable was not useful for predicting the person's anthropometric measure. An algorithm iteratively created models using all variables and eliminated the one with the highest p-value in each iteration. This process continued until the p-value of every variable was lower than the critical value for the acceptance region of H_0, which in this case was set at $\alpha = 0.05$ for a 95% confidence level. Six models were created in this way, however, only 3 of those showed significant adjusted R^2 values. Those were the models for BMI, arm circumference and waist circumference.

For the BMI , the model appeared to perform very well. A reasonable number of variables were retained, and the adjusted R^2 was found to be 0.945, indicating that the model

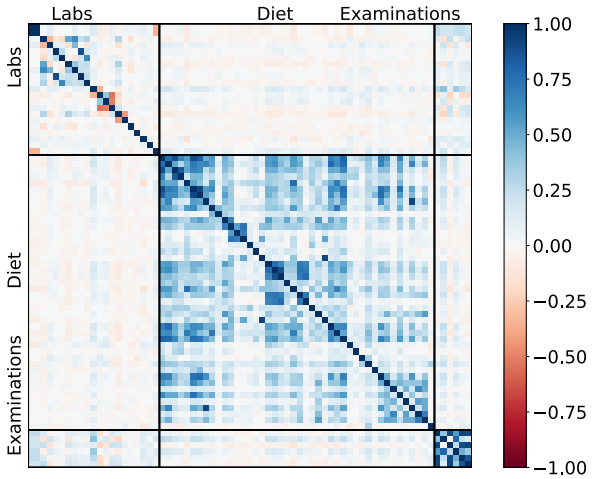

Fig. 7: Heat Map of Pearson Correlation Coefficients, Including Labs, Examinations and Diet Variables.

may explain up to 94.5% of the variance in the dataset. The model was described by:

$$BMI = \mathbf{w}^\top \mathbf{x}.$$

where

$$\mathbf{x} = \begin{bmatrix} \text{urine creatinine (mg/dl)} \\ \text{serum creatinine (mg/dl)} \\ \text{white blood cell count (cells/uL)} \\ \text{absolute monocyte count (cells/uL)} \\ \text{red blood cell count (million cells/uL)} \\ \text{mean corpuscular volume (fL)} \\ \text{red cell distribution width (\%)} \\ \text{platelet count (cells/uL)} \\ \text{mean platelet volume (fL)} \\ \text{hemoglobin a1c (\%)} \\ \text{hepatitis b surface antigen} \\ \text{urine volume (mL)} \\ \text{vitamin d} \\ \text{copper (mg)} \\ \text{moisture (g)} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 2.038 \times 10^{-6} \\ 0.0002 \\ 0.7260 \\ -4.7044 \\ 0.7970 \\ -0.1172 \\ 0.7653 \\ 0.0236 \\ 1.0979 \\ 1.0856 \\ 1.2895 \\ 0.0059 \\ -0.0656 \\ -0.7303 \\ 0.0005 \end{bmatrix}.$$

As a result, the residuals of the model were analyzed to determine whether it was overpredicting or underpredicting.

It can be observed that the model tended to underpredict, as the residuals distribution was skewed to the right, and the QQ-plot indicated a disproportionate amount of negative residuals.

For this model the MSE ≈ 37, which is considered quite good as it implies an actual error of approximately ± 6.1. In terms of the BMI values contained in the dataframe, this does not represent significant variation. Based on this, it can be concluded that this model is a somewhat good fit for predicting BMI, however, most interestingly the variables in this model began a trend that would show in the other two models, of copper, vitamin D and metabolic health markers, such as monocyte count and absolute white blood cell count contributing to these measures in different ways.

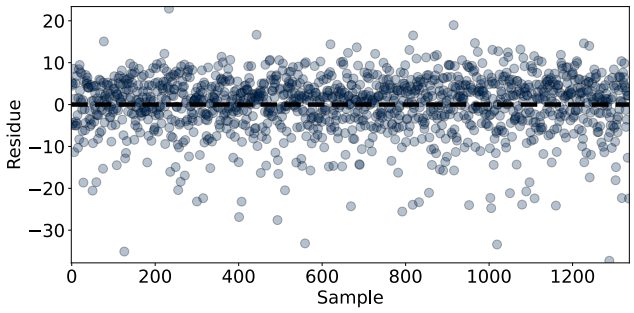

Fig. 8: Unordered Residues Generated by the BMI Prediction Model.

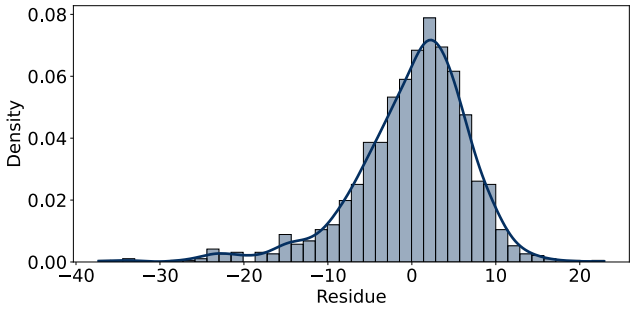

Fig. 9: Distribution of Residues Generated by the BMI Prediction Model.

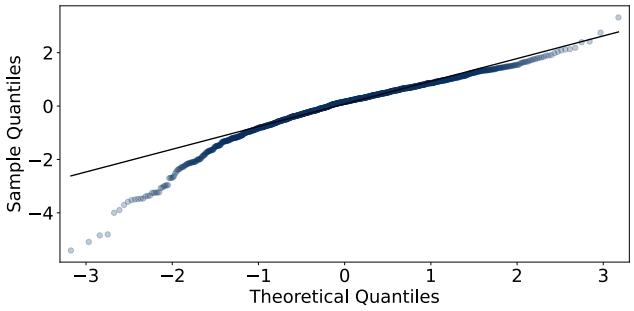

Fig. 10: Q-Q plot of Residues Generated by the BMI Prediction Model.

The model for arm circumference was identified as another strong fit. The adjusted R^2 has a value of 0.978, indicating that the model can predict up to 97% of the variance in the data. The model was described by:

$$\text{Arm Circumference} = \mathbf{w}^\top \mathbf{x}.$$

where

$$\mathbf{x} = \begin{bmatrix} \text{urine creatinine (mg/dl)} \\ \text{serum creatinine (mg/dl)} \\ \text{white blood cell count (cells/uL)} \\ \text{absolute monocyte count (cells/uL)} \\ \text{red blood cell count (million cells/uL)} \\ \text{red cell distribution width (\%)} \\ \text{platelet count (cells/uL)} \\ \text{mean platelet volume (fL)} \\ \text{hemoglobin a1c (\%)} \\ \text{hepatitis b surface antigen} \\ \text{urine volume (mL)} \\ \text{folate fortified (μg)} \\ \text{moisture (g)} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 1.9090 \times 10^{-6} \\ 0.0002 \\ 0.2556 \\ -2.2249 \\ 2.4322 \\ 0.4905 \\ 0.0136 \\ 0.4767 \\ 1.1375 \\ 0.9595 \\ 0.0075 \\ -0.0027 \\ 0.0005 \end{bmatrix}.$$

Further analysis of the residuals was conducted to determine whether the model was overpredicting or underpredicting.

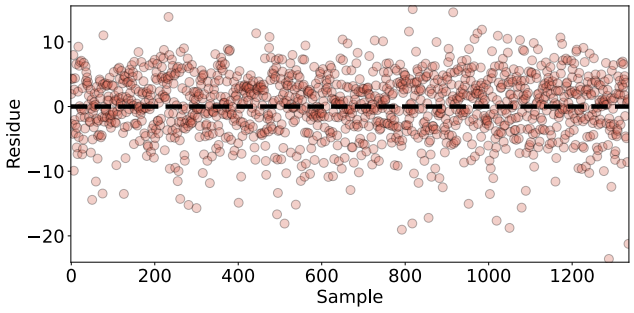

Fig. 11: Unordered Residues Generated by the Arm Circumference Prediction Model

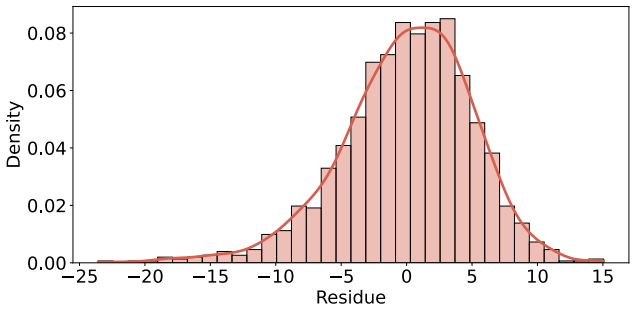

Fig. 12: Distribution of Residues Generated by the Arm Circumference Prediction Model.

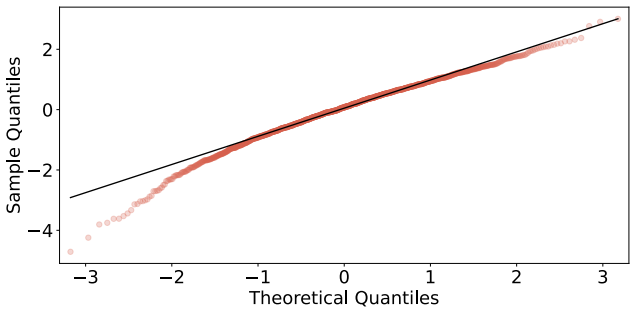

Fig. 13: Q-Q plot of the Residues Generated by the Arm Circumference Prediction Model.

As observed from these graphs, the model tended to underpredict, as the residuals were slightly skewed to the right. The QQ-plot confirmed this observation, as the residuals deviated from the normal quantiles, with more data concentrated in the negative or lower quantiles.

For this model $MSE \approx 19$ which means that the actual error will be of approximately ± 4.35 which in terms of the arm circumference does not represent much variation in this data frame. Again making the model a somewhat good fit for the prediction of the arm circumference, however as was mentioned before, the most interesting thing obtained from this model was the relationship between metabolic markers and anthropometric measures.

The final model was the one for waist circumference which was identified as another strong fit. The adjusted R^2 had a value of 0.973, indicating that the model could predict up to

97% of the variance in the data. This model was described by:

$$\text{Waist Circumference} = \mathbf{w}^\top \mathbf{x}.$$

where

$$\mathbf{x} = \begin{bmatrix} \text{urine creatinine (mg/dl)} \\ \text{serum creatinine (mg/dl)} \\ \text{white blood cell count (cells/uL)} \\ \text{absolute monocyte count (cells/uL)} \\ \text{red blood cell count (million cells/uL)} \\ \text{mean corpuscular volume (fL)} \\ \text{mean cell hemoglobin concentration (g/dL)} \\ \text{red cell distribution width (\%)} \\ \text{platelet count (cells/uL)} \\ \text{mean platelet volume (fL)} \\ \text{fasting hours before lab visit} \\ \text{hemoglobin a1c (\%)} \\ \text{hepatitis b surface antigen} \\ \text{urine volume (mL)} \\ \text{copper (mg)} \\ \text{moisture (g)} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 4.798 \times 10^{-6} \\ 0.0004 \\ 1.7260 \\ -7.2601 \\ 3.1252 \\ -0.2984 \\ 0.9573 \\ 1.4615 \\ 0.0486 \\ 1.8274 \\ 0.1938 \\ 2.0210 \\ 3.7205 \\ 0.0147 \\ -1.4439 \\ 0.0011 \end{bmatrix}.$$

The residuals of the model were analyzed to determine whether it was overpredicting or underpredicting.

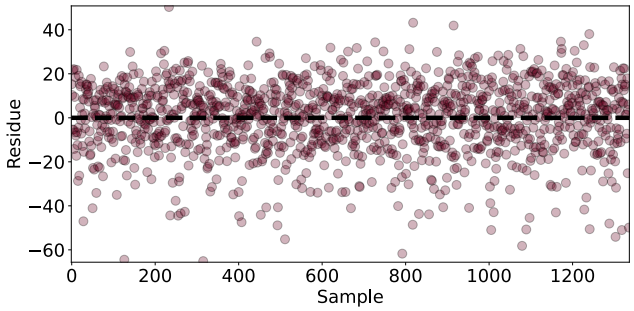

Fig. 14: Unordered Residues Generated by the Waist Circumference Prediction Model

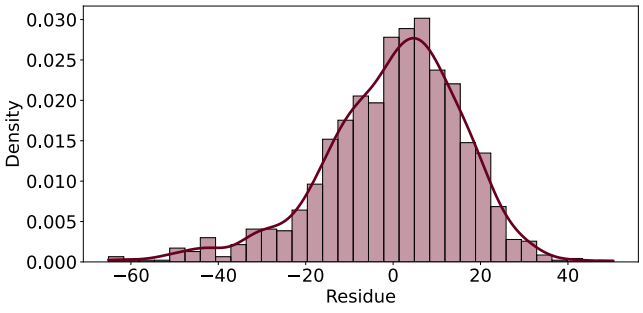

Fig. 15: Distribution of Residues Generated by the Waist Circumference Prediction Model.

The distribution of the residuals for this model was also skewed to the right, indicating that the model tended to underpredict. The QQ-plot further confirmed this observation, showing a disproportionate amount of data in the lower quantiles.

For this final model the MSE ≈ 219, which means that the actual error was approximately ± 15. In terms of waist circumference, this does not represent significant variation in the dataframe. Based on this, it can be concluded that this

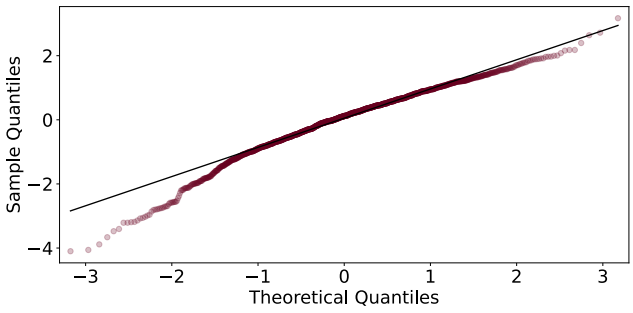

Fig. 16: Q-Q plot of the Residues Generated by the Waist Circumference Prediction Model.

model is a somewhat good fit for predicting waist circumference. But again the most interesting thing obtained from this model was the highlight of the relationship between dietary intake of copper, metabolic health markers such as monocyte count and white blood cell count and this anthropometric measure, following the trend set by the model for BMI.

A detailed explanation of the data preprocessing steps, PCA application, and regression analysis can be found in the supplementary material, available at: [5].

IV. Discussion

The regression analysis conducted in this study revealed several significant predictors of anthropometric measures, including copper, vitamin D, monocyte count, white blood cell count, and red blood cell count. These variables were strongly associated with body composition metrics such as BMI, arm circumference, and waist circumference, highlighting their critical roles in metabolic regulation and overall health. The discussion that follows explores the implications of each of these components, providing a detailed examination of their physiological and metabolic contributions to the observed relationships.

Copper is essential for many enzymatic and biochemical processes, such as antioxidant defense, metabolism, and lipid regulation. It plays a crucial role in physiological and metabolic functions by maintaining copper homeostasis. Dietary copper deficiency and excess in the diet negatively influence metabolism, leading to an energy imbalance. Specifically, healthier liver function and reduced lipid deposition are observed with optimal copper intake. In addition, copper interacts with other essential minerals, such as magnesium and zinc, further affecting metabolic health. These findings provide a basis for understanding the inverse relationship between copper intake and markers of body composition, such as BMI and waist circumference, since sufficient copper intake supports metabolic stability and lipid regulation. [6]

Vitamin D plays a key role in the regulation of glucose, lipid, and energy metabolism, which are essential for maintaining a healthy body weight. Studies consistently demonstrate an inverse correlation between body fat mass and serum vitamin D levels, reporting a 4% decrease in vitamin D concentrations for every 10% increase in BMI. This inverse relationship is due to reduced dietary vitamin D intake and reduced sunlight exposure in obese individuals. These findings suggest that higher Vitamin D intake may reflect healthier dietary patterns and lifestyle choices, contributing

to improved body composition.

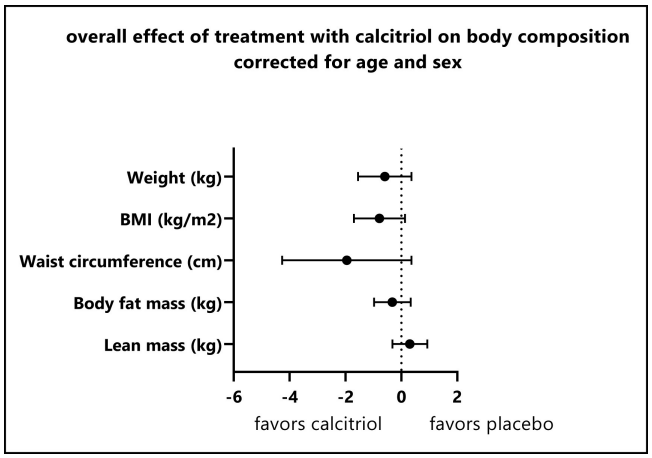

Fig. 17: Combining intervention estimates from calcitriol supplementation uncorrected (A) and corrected (B) for age and sex. Reproduced from *"The effect of active vitamin D supplementation on body weight and composition: A meta-analysis of individual participant data"* by Sabrina M. Oussaada et al., published in *Clinical Nutrition*, Volume 43, Pages 99–105, 2024, under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Available at ScienceDirect.

These findings align with the trends observed in Figure 17, which illustrates the combined intervention estimates for calcitriol (the active form of vitamin D) supplementation and its effects on body weight and composition. The figure highlights how age and sex adjustments refine the understanding of vitamin D's role in metabolic health and body composition, further supporting its significance as identified in our regression analysis. [7]

Monocytes, a type of white blood cell involved in immune responses, play a critical role in metabolic regulation and body composition. Research highlights that monocyte-related biomarkers, such as MCP-1, are positively correlated with BMI. Reduced monocyte activity can reflect improved metabolic health, associated with lower BMI and waist circumference. These findings align with the observed inverse relationship between Absolute Monocyte Count and body composition metrics in our study, highlighting the complex interactions between immune function and metabolic regulation. [8]

White blood cell count (WBC) and red blood cell count (RBC) were positively associated with BMI and body size measures. Elevated WBC levels are related to chronic inflammation caused by adipose tissue (body fat), and these inflammatory responses are also associated with metabolic disturbances. Meanwhile, the increase in RBC count with higher BMI might be due to the reduced oxygen availability in the body caused by excess adipose tissue. This condition leads to higher production of red blood cells as the body adapts to improve oxygen transport. Together, these mechanisms highlight the observed correlations between blood parameters and body composition metrics such as arm and waist circumferences.

As shown in Figure 18, the mean white blood cell (WBC) count varies significantly according to obesity status. This visual reinforces the positive association between BMI and WBC levels identified in our analysis, reflecting the inflammatory response triggered by excess adipose tissue. The figure also provides additional context for understanding how these biomarkers differ across populations. [9]

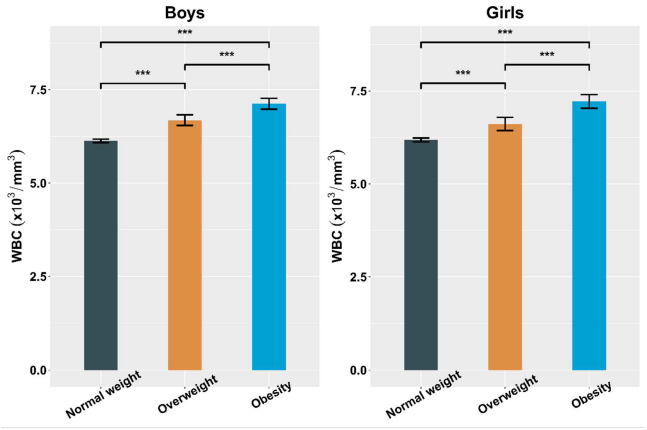

Fig. 18: (a) Mean white blood cell (WBC) according to sex and obesity. Reproduced from *"Positive Associations between Body Mass Index and Hematological Parameters, Including RBCs, WBCs, and Platelet Counts, in Korean Children and Adolescents"* by Jeong, H.R., Lee, H.S., Shim, Y.S., and Hwang, J.S., published in *Children*, Volume 9, Article 109, 2022, under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Available at MDPI.

V. Conclusions

The data analyzed in this project demonstrates significant dimensionality reduction. PCA was applied using cumulative explained variance as the selection criterion. This resulted in a reduction from 66 to 28 principal components for the dietary data frame and from 30 to 17 principal components for the blood works data frame. This threshold was chosen to ensure sufficient information retention while reducing model complexity and computational demands, highlighting the effectiveness of PCA in simplifying feature spaces without significant loss of information.

Multiple linear regression models, conducted with a significance threshold of $\alpha = 0.05$, identified copper as a significant predictor of anthropometric measures, emphasizing its potential role in influencing physical body metrics. Additionally, these regression models demonstrated a strong relationship between metabolic health and anthropometric measures, as evidenced by adjusted R^2 values of 0.945, 0.978, and 0.973 across the models. All predictor variables were statistically significant, with p-values below 0.05, the largest being 0.03. Copper emerged as a common, significant, and interesting dietary predictor, alongside other variables included in the models. These results underscore the robustness of the models and the interconnected nature of metabolic health and anthropometric measures, highlighting their relevance to health outcomes.

To enhance the predictive models developed in this study, it is recommended to systematically identify and remove outliers using robust statistical methods, as these may distort the models' performance. Preliminary analysis also indicated potential multicollinearity among certain variables, suggesting the need for rigorous assessment using diagnostic tools

such as the variance inflation factor (VIF) to ensure that predictors contribute independently to the models. Implementing these improvements is expected to refine model accuracy, reliability, and interpretability.

However, the findings are limited to the specific dataset analyzed, and further validation on external datasets is recommended to confirm the generalizability of these results.

REFERENCES

[1] L. Milanović, D. Živković, A. Đošić, P. Mitić, B. Cicović, T. Purenović-Ivanović, J. Nedeljković, V. Cicović, and S. Pantelić, "Bmi, body image, and quality of life—moderating role of physical activity," *Applied Sciences*, vol. 12, no. 14, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/14/7061

[2] J. Stephenson, C. M. Smith, B. Kearns, A. Haywood, and P. Bissell, "The association between obesity and quality of life: a retrospective analysis of a large-scale population-based cohort study," *BMC Public Health*, vol. 21, no. 1, p. 1990, 2021. [Online]. Available: https://doi.org/10.1186/s12889-021-12009-8

[3] Y. Yang, J. Darmont, F. Ravat, and O. Teste, *Dimensional Data KNN-Based Imputation*. Springer International Publishing, 2022, p. 315–329. [Online]. Available: http://dx.doi.org/10.1007/978-3-031-15740-0_23

[4] J. A. Rodrigo, "Análisis de componentes principales (pca) con python," n.d., disponible con licencia CC BY-NC-SA 4.0. [Online]. Available: https://www.cienciadedatos.net/documentos/py19-pca-python.html

[5] L. C. M.-P. M. Y.-T. V. A. G.-L. Fernando Pavía-González, Jorge Sánchez-Ponce, "Supplementary material for "exploring metabolic health: Impacts on body composition and immunity via pca and regression analysis"," https://drive.google.com/drive/folders/1TvE-7xxw8a_6v4Xj8l07CMGOGPEqMQFZ?usp=sharing, 2024, accessed: 2024-11-28.

[6] C.-C. Zhong, J. Ke, C.-C. Song, X.-Y. Tan, Y.-C. Xu, W.-H. Lv, Y.-F. Song, and Z. Luo, "Effects of dietary copper (cu) on growth performance, body composition, mineral content, hepatic histology and cu transport of the gift strain of nile tilapia (oreochromis niloticus)," *Aquaculture*, vol. 574, p. 739638, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S004484862300412X

[7] S. M. Oussaada, I. Akkermans, S. Chohan, J. Limpens, J. W. Twisk, C. Winkler, J. Karalliedde, J. C. Gallagher, J. A. Romijn, M. J. Serlie, and K. W. ter Horst, "The effect of active vitamin d supplementation on body weight and composition: A meta-analysis of individual participant data," *Clinical Nutrition*, vol. 43, no. 11, pp. 99–105, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0261561424003091

[8] H.-H. Huang, L.-Y. Chen, K.-Y. Chen, Y. chi Lee, C.-Y. Tsai, and C.-Y. Chen, "Increased monocyte chemoattractant protein-1 and nitrotyrosine are associated with increased body weight in patients with rheumatoid arthritis after etanercept therapy," *Neuropeptides*, vol. 84, p. 102100, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0143417920301189

[9] H. R. Jeong, H. S. Lee, Y. S. Shim, and J. S. Hwang, "Positive associations between body mass index and hematological parameters, including rbcs, wbcs, and platelet counts, in korean children and adolescents," *Children*, vol. 9, no. 1, 2022. [Online]. Available: https://www.mdpi.com/2227-9067/9/1/109