# Gender-Specific Machine Learning Analysis of Sarcopenia Risk in Aging Filipinos: Demographic and Lifestyle Perspectives

Maria Maura S. Tinao
School of Library and Information Sciences, University of the Philippines Diliman

Roda L. Tajon
Faculty of Information and Communication Studies, UP Open University

**Abstract**

*This study investigates the demographical and lifestyle risk factors for sarcopenia-specific gender risks to which the aging Filipino population would be predisposed, including age, occupation, smoking history, alcohol use, and existence of co-morbidities. The predictability of physical performance measures—lifting strength, need for assistance with walking, and ability to climb up/downstairs—utilizing machine learning models was assessed. Demographic factors, such as age and gender with the type of occupation, influenced the risk of sarcopenia. In contrast, lifestyle factors, such as smoking and alcohol intake, were not found to be predictive of sarcopenia in this sample. Such a high prevalence among males demands an approach by health intervention differently tailored to be gender sensitive.*

*For the performance of the different machine learning models, this study further gauges the different performance machine learning models and finds that SVM predicts the risk of sarcopenia with 85% accuracy as opposed to other approaches. In terms of recall, SVM did well in the case of prediction of males but underperformed in females and non-binary classifications, which may indicate an area of calibration. This research suggests that integrating predictive modeling into clinical practice can enhance early detection and targeted interventions for sarcopenia. It is advisable to implement. Further studies on sarcopenia using machine learning in the aging Filipino population is essential to determine novel risk factors, improve therapies, and predict disease progression, fostering evidence-based public health policies and better disease management.*

**Keywords**: sarcopenia, gender-specific analysis, aging population, Philippines, machine learning, predictive modeling, health promotion

## INTRODUCTION

The Philippines is experiencing a significant demographic shift as its population steadily ages. According to the Philippine Statistics Authority (PSA), the number of older adults increased from 7.5% in 2015 to 8.5% in 2020 (PSA, 2023). By 2030, projections indicate that older individuals will constitute between 10% and 19% of the population. This transition to an ageing society is largely attributed to declining fertility rates and increasing life expectancy, aligning with global demographic trends (Cruz et al., 2019). Consequently, addressing the health challenges the aging population faces will become a critical focus in the coming years.

Among these challenges, the rising prevalence of sarcopenia—a condition characterized by age-related loss of muscle mass and strength—poses a significant threat to the health, independence, and quality of life of older Filipinos. Sarcopenia has been linked to increased risks of falls, fractures, cognitive decline, and higher healthcare costs (Santilli, Bernetti, Mangione & Paoloni, 2014). Metter et al. (1997) suggest muscle mass and strength may decline in the 40s and 50% by the 80s. Sarcopenia affects 50 million people worldwide. Nearly 13% of 60-70-year-olds will be over 60 by 2050 (Seok, Kim, & Kim, 2023), which threatens health and functional independence, especially in the Philippines' increasingly ageing population.

While global studies show a trend in the escalation of this condition, research specific to the Philippines remains scarce, particularly on gender-specific risk factors and lifestyle determinants (Cruz, et al. 2019). The closest to Sarcopenia is falls as the common health risk for the older population based on the report of the Longitudinal Study of Aging and Health in the Philippines (LSAHP). According to the report, falls are a leading cause of fatal and nonfatal injuries among the older population, which result in hospitalization, reduced ability to perform daily activities, and a diminished quality of life (Natividad, 2019a). Natividad (2019b) highlights that 19% of respondents to the LSAHP reported experiencing at least one fall in the past year, with an average of 1.7 falls during that period. Among those who fell, 15% sustained injuries severe enough to require medical treatment. Natividad (2019a) emphasizes that risk factors include advanced age, reduced mobility, comorbidities, disabilities, psychological distress, impaired balance, polypharmacy, and a history of falls. Importantly, the LSAHP highlights the lack of awareness about sarcopenia and limited access to government programs that could mitigate its effects (Cruz et. al, 2019).

Given these gaps, there is a need to design and implement effective information systems to improve awareness, promote early detection, and enable equitable access to resources for sarcopenia prevention and management. Demographic and lifestyle factors like age, gender, occupation, height, weight, smoking, alcohol intake, and co-morbidities are examined to determine if they affect the risk and occurrence of sarcopenia in aging Filipinos to help assess and promote national health. Pre-sarcopenic physical performance markers including lifting strength, gait assistance, chair-raising, stair-climbing, and fall history will help. By integrating gender-specific predictive models and machine learning tools, such systems can identify at-risk individuals, facilitate targeted interventions, and inform policy decisions. Understanding and mitigating sarcopenia risk factors is crucial to safeguarding the health and functional independence of the Philippines' aging population.

**Research Objectives**

The purpose of this research is to contribute to sarcopenia assessment and health promotion programs in the Philippines through a gendered analysis of the demographic and lifestyle factors, industry of work, smoking, alcohol intake, and the existence of co-morbidities. These will determine the risks and occurrence of sarcopenia among elderly Filipinos. Specifically, the research aims to:

A. Identify demographic factors (age, gender, occupation) contributing to sarcopenia risk;
B. Assess lifestyle factors using lifestyle data (weight, height, smoking habits, alcohol consumption, presence of co-morbidities) influencing sarcopenia;
C. Analyze the data of physical performance indicators (strength in lifting, assistance in walking, rising from a chair, ability to climb stairs, history of falls) as predictors of sarcopenia;
D. Develop gender-specific predictive models for sarcopenia risk among aging Filipinos.

**Conceptual/Theoretical Framework**

This research integrates theoretical perspectives and empirical evidence to understand the gendered interplay of factors influencing sarcopenia risk among aging Filipino men, women, and other gender identities. This study encompasses three main components:

A. *Definition and measurement of Sarcopenia*. Sarcopenia is the age-related loss of muscle mass, strength, and function (Santilli et al., 2014). This framework will adopt international consensus guidelines that consider both muscle mass and function criteria. On the other hand, the Measurement Operational definitions include standardized methods for assessing muscle mass (e.g., bioelectrical impedance analysis, dual-energy X-ray absorptiometry) and muscle strength/function (e.g., grip strength, gait speed, physical performance tests).

B. *Determinants of Sarcopenia*. To understand the risks of Sarcopenia, it will also employ an analysis of the Gendered demographic factors such as:

- Sex: assigned at birth and gender identity: Differences in muscle mass and hormonal influences contribute to varying sarcopenia prevalence between men, women, and other gender identities. This would allow the investigators to understand how sarcopenia poses risks to gender identities and map out risks according to other genders.
- Age: Aging is a primary risk factor due to physiological changes in muscle composition and function
- Occupation/s: Physical and strenuous labour with which the participants are engaged that may lead to sarcopenia due to increased muscular activity. This also includes an overview

of the working conditions of the research partnerships that escalate or deter them from using strenuous labor.

- Lifestyle: This covers the height, weight, and physique that influence muscle mass and strength. Besides physique, lifestyle behaviors such as smoking habits and alcohol consumption impact muscle health and function.
- Presence of Co-morbidities: Chronic diseases such as diabetes and hypertension accelerate muscle loss.

C. *Physical Performance Indicators*: These include the following:

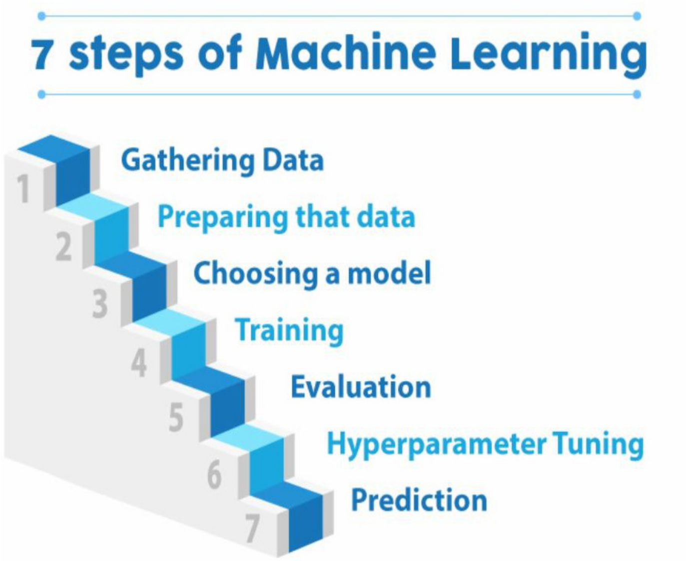
A. Strength in Lifting: Reflects muscle strength and functional capacity;
B. Assistance in Walking: Indicates mobility and independence;
C. Rise from a Chair and Climb Stairs: Assess lower extremity strength and functional ability;
D. History of Falls: Reflects balance, coordination, and overall physical frailty.

The conceptual framework employed in this study involves the use of machine learning classification algorithms to contribute to sarcopenia assessment and health promotion programs in the Philippines through a gendered analysis of the demographic and lifestyle factors.

*Figure 1* displays data collection as the project's machine learning challenges and goal. Through databases, APIs, web scraping, or pre-existing datasets, relevant data is acquired. Next, data preprocessing removes outliers, missing values, and formatting errors. Normalizing and scaling normalize data. Categorical variables are encoded one-hot or label. Next, training, validation, and testing data are separated. Data distribution, relationships, and trends are shown and analyzed using exploratory data analysis (EDA). Also, identify plausible feature-target variable relationships. Feature engineering creates and modifies features to capture data patterns. Selecting model prediction-enhancing characteristics is key. Model selection involves choosing a suitable machine-learning strategy for a problem. Scikit-Learn, TensorFlow, or PyTorch can be used. Next, train the model on the training dataset. Grid search or random search fine-tunes hyperparameters.

In model assessment, accuracy, precision, recall, F1-score, and other metrics are used on the validation dataset to assess model performance. Cross-validation is often utilized to improve the finding's reliability. After validation, model tuning involves changing parameters to increase performance without overfitting via regularization or dropout. These efforts culminate in the final model selection when the model with the best validation results is chosen. Use the test dataset to evaluate the model's real-world performance. Once confident, the well-performing model predicts new, unforeseen facts.

Figure 1 *Machine Learning Classification Algorithm*




The four different machine learning classification algorithms used in this study are as follows: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Logistic Regression and Random Forest. Each algorithm processes the data differently, identifying patterns and creating decision boundaries to analyze or predict older Filipinos' sarcopenia risk using gender-specific predictive models.

## RESEARCH METHOD

This study utilizes quantitative methods – statistical tests and machine learning methods. Descriptive statistics and chi-square tests will be used to explore and compare the characteristics of the sample.

### Research participants

The sample size required for a population of 108,667,043, the Philippines' 2020 population, with a 95% confidence level and a 5% margin of error, is roughly 384 people using the sampling strategy which is the random stratified sampling.

The data source for the data collection method came from a survey designed specifically for this study, which included demographic factors such as age, gender, and industry of work; lifestyle factors such as weight, height, smoking habits, alcohol consumption, and the presence of co-morbidities; and physical performance indicators such as strength in lifting (measured by standardized weight), assistance in walking (self-reported or observed), ability to rise from a chair (time or difficulty level), ability to climb stairs (self-reported or observed), and number of falls experienced.

Sample of the study

Data gathering in this sector was challenging since the senior citizens were very careful who they spoke with. The data source came from only 268 elderly people from different parts of the Philippines randomly selected. They are older relatives and family members of friends and acquaintances. Most of them were from Bataan and Zambales. One in-person interview was conducted while holding the monthly gathering of seniors in Olongapo City, while all others were done through online surveys. Anonymized data was used so there is no need to provide written consent, guaranteeing confidentiality and ethical issues.

Models' Implementation Tools

For the implementation of the machine learning models, the study utilized the Python programming language and the Jupyter Notebook application in Anaconda. Jupyter Notebook is a web-based, interactive computing environment that enables the creation of human-readable documentation while describing the data analysis process. With this powerful tool, the ML models were developed and tested, facilitating the evaluation and interpretation of the results. The Statistical Package for Social Science or SPSS was utilized to compute linear regression.

Validation Method

The accuracy metric is a widely used evaluation measure to assess how well a model predicts the correct class labels. The formula for accuracy is as follows:

Accuracy = **(TP + TN) / (TP + TN + FP + FN)**

*where:*

- **TP** (*True Positive*) represents the number of positive instances correctly predicted by the classification model.
- **TN** (*True Negative*) represents the number of negative instances correctly predicted by the classification model.
- **FP** (*False Positive*) represents the number of negative instances incorrectly predicted as positive by the model.
- **FN** (*False Negative*) represents the number of positive instances incorrectly predicted as negative by the model.

**Confusion Matrix**

The confusion matrix is a table that presents a summary of the model's performance by comparing predicted class labels against the actual class labels. It helps visualize the classification errors and correct predictions made by the model.

The confusion matrix typically looks like this:

Table 1 *Confusion Matrix*

| | ACTUAL POSITIVE | ACTUAL NEGATIVE |
|---|---|---|
| Predicted Positive | True Positive (TP) | False Positive (FP) |
| Predicted Negative | False Negative (FN) | True Negative (TN) |

Table 1 demonstrates that the confusion matrix is employed to evaluate the model's performance, depicting the allocation of correct and inaccurate predictions. The main goal is to minimize false negatives and false positives to enhance the model's predictive effectiveness.

It is an important tool for understanding, analyzing, and improving the performance of machine learning models, to make informed decisions and optimizations based on real-world performance metrics rather than relying solely on accuracy.

## RESULTS AND DISCUSSION

Sarcopenia may be among the biggest health threats facing the fast-aging population in the Philippines. Other threats include a higher possibility of falls and fractures, metabolic syndrome, cognitive impairments, disabilities, poorer quality of life, admission and readmission to the hospital, and healthcare costs. This research aims to analyze gender-specific health outcomes using machine learning techniques, forecast sarcopenia risks, and identify key predictors.

## Profile of the Respondents

A. *Demographic Factors*

Table 2 *Age Group and Gender of the Respondents*

| CATEGORY | OPTIONS | FREQUENCY | PERCENT |
|---|---|---|---|
| Age Group | 40-44 | 56 | 21.0% |
| | 45-49 | 65 | 24.3% |
| | 50-54 | 57 | 19.1% |
| | 55-59 | 51 | 14.2% |
| | TOTAL | 268 | 100% |
| Gender | Male | 81 | 30.22% |
| | Female | 181 | 67.54% |
| | Non-binary/third gender | 2 | 0.75% |
| | Others | 4 | 1.49% |
| | TOTAL | 268 | 100% |

**Age Group**: The age group in Table 2 is approximately evenly distributed among the 40- and 65-year-old age brackets. The age cohort of 45-49 years old comprises 24.3% of the sample. The proportion of individuals aged 60-65 is a mere 14.2% at the extreme end. In terms of age distribution, interest in this matter may be primarily among middle-aged or more senior individuals, as the issues being addressed are likely to be related to workforce engagement or aging.

**Gender**: The gender distribution is depicted in Table 2, which indicates that 67.5% of respondents are female, while 30.2% are male. Only 0.7% of respondents identified themselves as non-binary/third gender, while 1.5% did. The sample, therefore, is primarily composed of females, and the perspective or discovery may be contingent upon the subject of the study.

**Industry of work**: As depicted in Table 3, respondents work across many industries, with Government and Public Administration (35.1%) being most dominant, followed by education at 14.9%, and unemployment at 11.9%. Slightly smaller portions are derived from other professional services like consulting, engineering, information, and communication technologies or a variety of other kinds. This distribution suggests a broad strength in the public sector and public administration, education-related fields, and related sectors and services.

B. *Lifestyle Factors*

**Smoking**. The statistics show that most of the respondents are nonsmokers, representing 89.9% of the sample, while only 10.1% reported being smokers. This trend towards nonsmoking may reflect increased awareness of health risks associated with smoking or effective public health campaigns against smoking.

Table 3 *Industry of Work*

| Industry of Work | FREQUENCY | PERCENT |
|---|---|---|
| Agriculture, Forestry, Fishing, and Hunting | 10 | 3.7% |
| Health care and social assistance | 15 | 5.6% |
| Transportation and warehousing | 5 | 1.9% |
| Retail | 9 | 3.4% |
| Education | 40 | 14.9% |
| Government and Public Administration | 94 | 35.1% |
| Manufacturing | 1 | 0.4% |
| Homemaker | 12 | 4.5% |
| Other Professional Services | 22 | 8.2% |
| Retired | 9 | 3.4% |
| Unemployment | 32 | 11.9% |
| Others | 19 | 7.1% |
| TOTAL | 268 | 100% |

**Alcohol Drinker**. Non-drinkers made up the larger percentage, at 87.4%, while 12.6% consume alcohol. The minimal percentage of people consuming alcohol may be suggestive of a well-informed decision or personal choice to abstain from alcohol intake for reasons of health and lifestyle concerns.

**Co-morbidities**. More than half, 52.6%, indicate no co-morbidities; others have diabetes, 13%; cardiovascular diseases, 9.7%; respiratory diseases, 3.7%; or other unspecified conditions, 18.6%. This sample appears relatively healthy, but the presence of chronic illnesses is noteworthy, diabetes and cardiovascular diseases, which merit follow-up.

**Weight**. The weight distribution among the respondents is relatively balanced, though the largest groups are 46.72–60.78 kg (103–134 lbs) at 23.8% and over 70.76 kg (156 lbs) at 20.8%. This variability implies different weight profiles that could lead to different health risk implications within the sample.

**Height**: Most heights of respondents are spread with the majority at 149 cm and below, that is, 16.7% and 152.4 cm, 15.7%. Differences in height will, therefore, imply variations in body composition that may be pertinent in understanding weight-height ratios and health indices.

**Physical Activity or Exercise**: Most of the respondents said they do not engage in physical activity, at 58.8%, while 42.2% did. This limited exercise may raise a red flag, especially regarding the health and well-being implications of exercising, especially among ageing or at-risk populations.

## Research Hypotheses

The following three hypotheses have been formulated to

Table 4 *Lifestyle Factors*

| CATEGORY | OPTIONS | FREQUENCY | PERCENT |
|---|---|---|---|
| Smoking | Nonsmoker | 241 | 89.9% |
| | Smoker | 27 | 10.1% |
| | TOTAL | 268 | 100.0% |
| Alchol Drinker | No | 235 | 87.4% |
| | Yes | 34 | 12.6% |
| | TOTAL | 268 | 100.0% |
| Co-morbidities | Cardiovascular (heart) diseases | 26 | 9.7% |
| | Diabetes | 35 | 13.0% |
| | Cancer | 3 | 1.1% |
| | Respiratory diseases | 10 | 3.7% |
| | Kidney diseases | 3 | 1.1% |
| | Others | 50 | 18.6% |
| | None | 141 | 52.6% |
| | TOTAL | 268 | 100.0% |
| Weight (kg/lbs.) | 46.72 - 60.78 / 103 - 134 | 64 | 23.8% |
| | 47.17 - 62.14 / 104 - 137 | 16 | 5.9% |
| | 488.98 - 64.8 / 108 - 143 | 27 | 10.0% |
| | 50.34 - 66.67 / 111 - 147 | 52 | 19.3% |
| | 53.0 - 70.30 / 117 - 155 | 53 | 19.7% |
| | 70.76 and above / 156 and above | 56 | 20.8% |
| | TOTAL | 268 | 100.0% |
| Height (cm/feet) | 149 and below / 4'11" and below | 45 | 16.7% |
| | 152.4 / 5'0" | 42 | 15.7% |
| | 154.9 / 5'1" | 26 | 9.7% |
| | 157.4 / 5'2" | 36 | 13.4% |
| | 160 / 5'3" | 28 | 10.4% |
| | 162.5 / 5'4" | 24 | 8.9% |
| | 165.1 / 5'5" | 25 | 9.3% |
| | 167.7 and above / 5'6" and above | 41 | 15.3% |
| | TOTAL | 268 | 100.0% |
| Physical Activity | Yes | 113 | 42.2% |
| | No | 155 | 58.8% |
| | TOTAL | 268 | 100.0% |

align with the objectives of this research:

**Hypothesis 1: Occupation type is associated with the risk of physical limitations and muscle health decline, including sarcopenia, among elderly Filipino men and women.**

Data analytics on occupational histories, muscle strength, and mobility measures suggest that sedentary occupations are linked to greater physical impairments, including reduced muscle strength and mobility challenges, compared to physically demanding occupations. However, the relationship may vary depending on gender, industry, and the specific physical measures assessed, with some impairments (e.g., stair climbing) showing a stronger association with occupation type than others (e.g., lifting strength or walking support).

Data analysis of physical strength and mobility measures by occupation category (Table 5) reveals gender and industry patterns. Strength (difficulty lifting 10 lbs.), chair rise, stairs, falls, and walking with support were measured. Crosstabulations were used to spread responses across occupational groups to compare physical abilities across industries. Chi-square tests have been performed to examine if industry, gender, and occupation have important effects on physical limitations. A significance level of $p < 0.05$ was used for these correlations.

The only variable that showed a statistically significant association with the type of industry was the ability to climb stairs ($p=0.022$), suggesting that some industries may require a higher level of physical capacity in this area. This would mean that employees in specific fields

Table 5 *Lifestyle Factors Related to the Risk of Sarcopenia among Filipino Aging Individuals*

| Variable | Level of Difficulty | Nonsmoker Count | Smoker Count | Total Count | Square Value | df | p- Value | Notes |
|---|---|---|---|---|---|---|---|---|
| Lifting and Carrinying 10 lbs | None | 121 | 14 | 135 | 0.132 | 2 | 0.936 | 1 cell (16.7%) expected <5 |
| | Some | 98 | 11 | 109 | | | | |
| | A lot or unable | 23 | 2 | 25 | | | | |
| Transferring from Chair to Bed | None | 194 | 21 | 215 | 1.563 | 2 | 0.458 | 2 cells 933.3% expected <5 |
| | Some | 41 | 4 | 45 | | | | |
| | A lot or unable | 7 | 2 | 9 | | | | |
| Climbing 10 steps | None | 151 | 19 | 170 | 2.062 | 2 | 0.357 | 1 cell (16.7%) expected <5 |
| | Some | 82 | 6 | 88 | | | | |
| | A lot or unable | 9 | 2 | 11 | | | | |
| Falls in the Past Year | None | 180 | 20 | 200 | 0.685 | 2 | 0.710 | 1 cell (16.7%) expected <5 |
| | Some | 52 | 5 | 57 | | | | |
| | $\geq$ 4 falls | 10 | 2 | 12 | | | | |
| Walking Difficulty Across Room | None | 197 | 21 | 218 | 2.049 | 2 | 0.359 | 2 cells (33.3%) expected <5 |
| | Some | 39 | 4 | 43 | | | | |
| | A lot or unable | 6 | 2 | 8 | | | | |
| Alcohol Drinker & Lifting 10 lbs. | None | 119 | 16 | 135 | 0.336 | 2 | 0.845 | 1 cell (16.7%) expected <5 |
| | Some | 95 | 14 | 109 | | | | |
| | A lot or unable | 21 | 4 | 25 | | | | |
| Alcohol Drinker & Transfering | None | 191 | 24 | 215 | 4.262 | 2 | 0.119 | 1 cell (16.7%) expected <5 |
| | Some | 38 | 7 | 45 | | | | |
| | A lot or unable without help | 6 | 3 | 9 | | | | |
| Alcohol Drinker & Climbing | None | 147 | 23 | 170 | 0.377 | 2 | 0.828 | 1 cell (16.7%) expected <5 |
| | Some | 78 | 10 | 88 | | | | |
| | A lot or unable | 10 | 1 | 11 | | | | |
| Alcohol Drinker & Falls | None | 178 | 22 | 200 | 9.650 | 2 | 0.008 | 1 cell (16.7%) expected <5 |
| | Some | 50 | 7 | 57 | | | | |
| | $\geq$ 4 falls | 7 | 5 | 12 | | | | |
| Alcohol Drinker & Walking Difficulty | None | 191 | 27 | 218 | 0.080 | 2 | 0.961 | 1 cell (16.7%) expected <5 |
| | Some | 37 | 6 | 43 | | | | |
| | A lot or unable | 7 | 1 | 8 | | | | |

face more challenges in climbing the stairs for the exerting requirements of their job or the physical stress of the type of industry in which they are working. Other physical abilities, such as lifting strength (p = 0.368), getting up from a chair (p = 0.576), fall history (p = 0.675), and assistance in walking (p = 0.507), did not reveal any significant association; hence, these impairments are not significantly different across the different types of occupations. This evidence shows that jobs can influence difficulties in mobility, but overall, most physical limitations are similar regardless of the industry.

**Hypothesis 2: Lifestyle factors such as smoking and alcohol consumption may contribute to sarcopenia risk in the ageing Filipino population. However, smoking does not show a significant association with physical limitations or sarcopenia in this study. The role of alcohol consumption, or other non-lifestyle factors, in sarcopenia risk may be more prominent, and further analysis is needed to clarify these relationships. Gender-specific prevalence rates suggest higher sarcopenia rates among men, but the contribution of lifestyle factors to these differences remains unclear.**

*Table 5* indicates that the analysis of smoking as a lifestyle factor revealed nonsignificant relationships with any of the physical performance measurements in aging Filipinos; hence, it may not be a predictor of sarcopenia risk in this population. Specifically, the Chi-Square tests for strength p = 0.936, getting up from a chair p = 0.458, climbing stairs p = 0.357, falls p = 0.710, and assistance in walking p = 0.359 indicate that smoking status has no association with these activities. This suggests that smoker and nonsmoker difficulties are similar along all these measures, and smoking per se does not appear to increase problems with strength, mobility, or risk of falls.

Sarcopenia prevalence was higher in men, but smoking did not appear to contribute to physical difficulties associated with sarcopenia in ageing. Investigating various dimensions of performance, no differences were found between smokers and nonsmokers. These results suggest that lifestyle factors other than smoking, possibly alcohol consumption, or factors unrelated to lifestyle are most strongly related to the risk of sarcopenia among Filipino aging individuals.

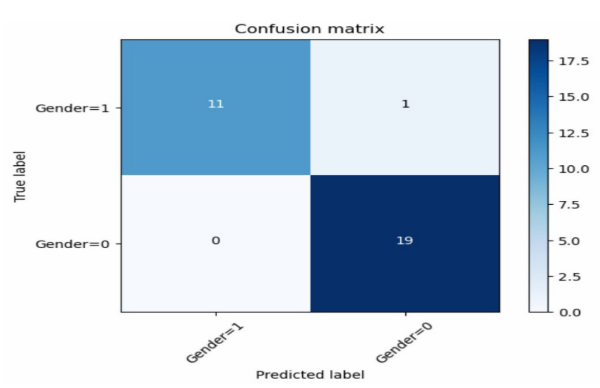
Hypothesis 3: Co-morbidities (diabetes, hypertension, and cardiovascular disease) are a major predictor of sarcopenia, as modeled by predictive analytics and decision-support systems that can be used to predict the interaction between co-morbidities and sarcopenia risk.

In this study, the model showed a high prediction rate for the presence of sarcopenia in both sexes. It gives a high accuracy in correctly predicting 30 out of 31 cases. It is a great predictive value since co-morbidities are indeed some of the strong predictors for sarcopenia in elderly Filipinos. For males (Gender 0), the model was correct in 11 of 12 cases. One error, a single false positive, does not strongly reduce the overall accuracy for this population. The model was 100% accurate for women or Gender 1, meaning all 19 cases were correctly classified. In other words, the performance of the model may be said to be perfectly reliable in detecting sarcopenia presence among women.

The model was perfect for women; no false negatives and false positives were observed, suggesting a particularly robust association between co-morbidities and sarcopenia risk in elderly Filipino women. This finding supports the hypothesis that co-morbidities might be more predictive of sarcopenia for women compared to men. -In males, the model functioned fine but still showed a solitary misclassification, which indicates the connection of co-morbidities with sarcopenia is not as strong and straightforward as that in women. It falls right in place within the notion that co-morbidities exerted a higher influence over the muscle wastage and sarcopenia risk, at least to women who are associated with the biological, hormonal, or lifestyle factors.

The log loss for this model was reasonable. The score of 1.1369 is not so very high but shows that, generally, there is still much room to improve the confidence of the model in making predictions, especially in scenarios where probabilities close to the decision boundary may have been assigned. This aspect is highly relevant for the refinement of the model in such cases that may fall under "gray areas" in between sarcopenia and not. However, this moderate log loss score does not mask the high classification accuracy seen in the confusion matrix (Figure 2).

Figure 2 *Confusion Matrix*



Consequently, healthcare professionals will focus on sarcopenia screenings and therapies for older Filipino females with comorbidities, as they are at elevated risk and the model demonstrates optimal accuracy in identification. Programs for lifestyle and health management customized for sarcopenia and associated disorders could significantly benefit that demographic.

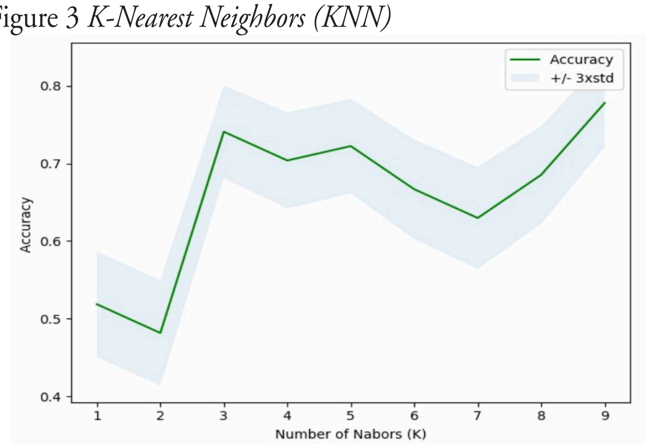
In elderly Filipino men, although co-morbidities continue to predict risk for sarcopenia, perhaps the association may not be that strong. Therefore, factors beyond co-morbidities, such as the level of physical activity or nutritional intake may need to be considered when assessing risk for sarcopenia in men.

**Machine Learning Models**

Leveraging the potential of machine learning, this research also aims to assess and compare the capabilities of four distinct algorithms - K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), and Logistic Regression - in the domain of sarcopenia risk.

1. **K-Nearest Neighbors (KNN)**: KNN is a non-parametric classification algorithm that makes predictions based on the majority class of its k-nearest neighbors. The model indicated an ability to moderately predict risk when it came to the data set for the training group, with accuracy levels going up to 0.693, whereas accuracy levels for test groups stood at 0.667, which portrayed poor generality. A training accuracy of approximately 69% suggests that this model can capture and generalize on the patterns related to sarcopenia risk based on the training data alone. The decrease down to accuracy on the test set was around 67%, which represents a slight reduction in predictive performance, rather typical with KNN models when the selected features provide no substantial separability of classes. This would indicate that though the model does capture some underlying relationships in the data, other factors may still play a role in contributing to variability in sarcopenia risk that it fails to capture well.

Figure 3 *K-Nearest Neighbors (KNN)*




As related to gender, if incorporated as a feature, the model's accuracy implies that gender might contribute meaningfully to the prediction of sarcopenia risk, as men often show higher prevalence and risk factors associated with sarcopenia. However, it may also be the case that other predictors are important, such as the level of physical activity, nutrition, or lifestyle factors, including smoking or alcohol use, which would improve the performance of the model. Moreover, KNN's sensitivity to the scaling and relevance of the features implies that factors such as changes in muscle mass with age or gender could contribute to the risk of sarcopenia and be supplemented with other information to achieve better accuracy.

Overall, gender may play a role in prediction but is unlikely to be an adequate determinant of high-accuracy risk prediction in an aging population without additional, more specific health metrics.

2. **Random Fores**t: Random Forest is a commonly used machine learning algorithm that combines the output of multiple decision trees to reach a single result. The confusion matrix indicated by the Random Forest model showed an overall of 42 correctly classified instances of the negative class, not-at-risk, while there are 57 classifications for the positive class at risk of sarcopenia with an over-classified result as indicated by being at wrong at 58 negative cases and 43 positive class instances. This gives an overall accuracy of 0.49 indicating that the model is far from being able to really classify the risk of having sarcopenia within that sample population.The precision of the classification report depicts 0.49 of precision for the negative class denoted as (0), and 0.5 for the positive class or (1), with rates of recall standing at 0.42 and 0.57, respectively. This implies that despite being a bit more accurate in the right classification of at-risk patients for sarcopenia, class 1, the model has a long way to go to achieve its optimal best. Negative class F1 scores 0.45 and the positive class F1 score 0.53 mean that precision and recall tend to balance against each other but still indicate that the predictors or features need to be made stronger to build their effectiveness.

With respect to gender as a risk predictor for sarcopenia, the performance of the Random Forest model implies that gender could be one of the features that might influence the predictions, but its direct contribution is not quantified from the given feature importance data. Feature importances show that feature_7, feature_3, and feature_5 are the most important predictors, though the exact nature of these features remains unspecified. If gender falls among these features, its relation to the risk of sarcopenia may prove to be critical because, according to earlier studies, men are generally at higher risk than women, said to be due to less muscle mass

Figure 4 *Random Forest*

```
Confusion Matrix:
[[42 58]
 [43 57]]

Classification Report:
              precision    recall  f1-score   support

           0       0.49      0.42      0.45       100
           1       0.50      0.57      0.53       100

    accuracy                           0.49       200
   macro avg       0.49      0.49      0.49       200
weighted avg       0.49      0.49      0.49       200


Feature Importances:
     Feature  Importance
7  feature_7    0.111995
3  feature_3    0.109996
5  feature_5    0.107357
9  feature_9    0.102804
1  feature_1    0.099662
4  feature_4    0.097453
6  feature_6    0.096647
8  feature_8    0.092085
2  feature_2    0.091137
0  feature_0    0.090864
```

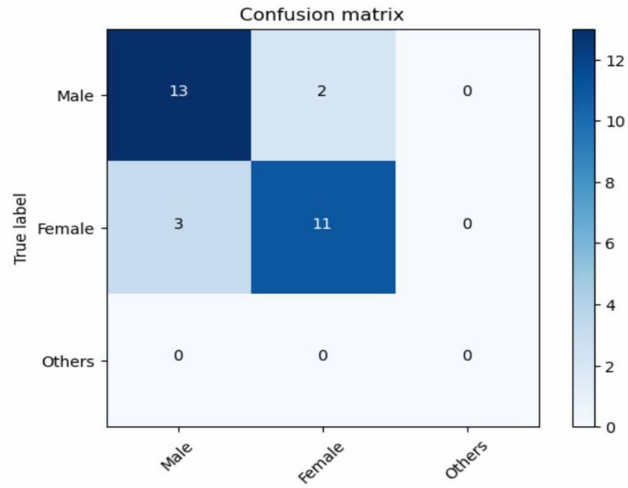
Figure 5 *Support Vector Machine (SVM)*

```
              precision    recall  f1-score   support

           0       0.88      1.00      0.94        22
           1       0.81      0.76      0.79        17
           2       0.85      0.73      0.79        15

    accuracy                           0.85        54
   macro avg       0.85      0.83      0.84        54
weighted avg       0.85      0.85      0.85        54

Confusion matrix, without normalization
[[13  2  0]
 [ 3 11  0]
 [ 0  0  0]]
```




and differences at the level of physical activity compared to women. Low overall accuracy for the model indicates that although gender could play a role, it probably is not the most important determining factor in the prediction of risk for sarcopenia, indicating that other lifestyle, health, and demographic factors must be integrated to enhance predictive ability. These analyses suggest that the present form of the model is too superficial and requires further elaboration and investigation of more essential features that capture all critical relationships influencing sarcopenia risk in the different ageing genders.

3. **Support Vector Machine (SVM)**: SVM is a powerful algorithm that aims to find the optimal hyperplane for separating classes in a high-dimensional space.

It can be depicted from the classification report and confusion matrix that SVM model is well performing in classifying among different gender categories, including male, female, others, for assessing the risk of sarcopenia. The accuracy of the model is 0.85, meaning it correctly classified the instances of the data set by a huge margin.

The precision for each class is depicted in Figure 5; 0: male = 0.88, 1: female = 0.81, and 2: others = 0.85. This indicates that the model can make an exact prediction of a given class very accurately, particularly so for males where the score is the highest. Precision would be defined as the ratio of true positives to the total number of predicted positives. It shows, therefore, that the model picks most instances of a gender correctly without many false positives.

Recall scores are the measures of how good the model is at correctly identifying actual instances of each class. In this case, recall for males is 1.00, which means all male instances in the test set were classified correctly. For females and others, the recall is lower at 0.76 and 0.73, respectively, indicating some misclassification within these categories. Recall is a measure of the proportion of true positives to actual positives and thus indicates the model does well at recognizing males but has an issue in classifying females and other gender identities accurately.

Classes show the balance between precision and recall. The F1-score for males was very good at 0.94, while for females and others, it was only at 0.79 and 0.79 respectively. The F1-score is a combination of precision and recall, so these variations indicate that proper classification strategies are required to improve them.

The performance of the model can also be understood from the confusion matrix:

- True Class Label "Male" 13 Correct predictions, 2 Misclassifications, are females

- True Class Label "Female" 11 correct predictions with 3 being Misclassified, males
- "Others" No predictions (i.e., the algorithm didn't recognize any instance belonging to this class as such in the prediction task.

Sarcopenia risk prediction findings show that the SVM model can tell males and females but not "others"—which may reflect data inadequacies in this class. Sarcopenia risk is heavily correlated with gender through biological and lifestyle factors. Hence, good accuracy for males indicates successful extraction of key group traits. However, more training or feature refinement may increase this model's ability to distinguish females and other gender identities, making it a better sarcopenia risk prediction in a more diverse population. The overall research stresses that gender and other demographic parameters should be incorporated in sarcopenia health outcome models since gender-specific therapies may be needed at certain thresholds.

4. **Logistic Regression**: Logistic Regression is a linear classification algorithm that models the probability of the binary outcome. In this study, the model showed a high prediction rate to predict sarcopenia presence of both sexes. It gives a high accuracy in that it predicts correctly 30 out of 31 cases. It is a great predictive value since co-morbidities are indeed some of the strong predictors for sarcopenia in the elderly Filipinos. For males (Gender 0), the model was correct in 11 of 12 cases. One error, a single false positive, does not strongly reduce the overall accuracy for this population. The model was 100% accurate for women or Gender 1, which meant that all 19 cases were correctly classified. In other words, the performance of the model may be said to be perfectly reliable in detecting sarcopenia presence among women.

The model was perfect for women, with no false negatives or positives, demonstrating a strong link between co-morbidities and sarcopenia risk in old Filipino women. This suggests that co-morbidities may predict sarcopenia more in women than men. In men, the model worked well but showed one misclassification, indicating that co-morbidities and sarcopenia are not as strongly linked as in women. It supports the idea that co-morbidities may have had a greater impact on muscle wastage and sarcopenia risk in women with biological, hormonal, or lifestyle variables.

The log loss for this model was reasonable. The score of 1.1369 is not so very high but shows that, generally, there is still much room to improve the confidence of the model in making predictions, especially in scenarios where probabilities close to the decision boundary may have been assigned. This aspect is highly relevant for the refinement of the model in such cases that may fall under "gray areas" in between sarcopenia and not. However, this moderate log loss score does not mask the high classification accuracy seen in the confusion matrix. Based on the results, for example, healthcare providers will target sarcopenia screenings and interventions for elderly female Filipino with co-morbidity since these are at high risk and the model is most accurate in its identification. Lifestyle and health management programs tailored to conditions of both sarcopenia and concomitant conditions could go a long way for that population.

In elderly Filipino men, although co-morbidities continue to predict risk for sarcopenia, perhaps the association may not be that strong. Therefore, factors beyond co-morbidities, such as the level of physical activity or nutritional intake may need to be considered when assessing risk for sarcopenia in men.

## CONCLUSION

1. Sarcopenia risk is heavily influenced by demographics. These factors include age, gender, and occupation. Apart from physical performance criteria, the Chi-Square test showed significant results for all demographic variables; thus, elevated male prevalence percentages compared to females emphasize the need for gender-sensitive health programs.
2. Contrary to common assumptions that lifestyle poses risk factors, alcohol and smoking may be insignificant risk factors for sarcopenia. Smoking and alcohol consumption cannot predict sarcopenia based on p-values because most of the p-values related to various physical performance indicators were quite high. While these risk variables are ubiquitous and associated with many bad health outcomes, they cannot predict risk in this elderly Philippine population. More data on lifestyle factors or health behaviors may be needed to attribute these factors to sarcopenia.
3. Patient sarcopenia risk classification accuracy varies by predictive model. Compared to KNN, 69.3%; Random Forest, 49%; and Logistic Regression, not available, SVM had 85% accuracy. Thus, proper model selection is crucial; in this situation, the SVM is accurate and may be useful for gender-based sarcopenia risk discrimination.
4. Gender classification is also good using SVM. Due to its recall value of 1.00, SVM categorized all males accurately. SVM underperformed for females and "others". Compared to other classes, this model

needs to be adjusted to accurately characterize both genders for health prediction.

5. The Random Forest model's confusion matrix showed 49% accuracy and zero correct gender predictions for "others." This implies low precision and recall for both males and females, suggesting this approach may not be appropriate for diagnosing sarcopenia risk. Reevaluating feature selection and tweaking parameters in the Random Forest model may improve prediction accuracy.
6. Sarcopenia risk prediction variables were identified by Random Forest model feature importance analysis. Features_7 and feature_3 were the most important, suggesting additional study of their implications in sarcopenia risk. This suggests targeted actions on these key variables could lower sarcopenia risk.
7. Needs inclusive physical performance intervention programs. The study recommends holistic health therapies that address sarcopenia-risk physical performance metrics. Since predictive models are not perfect, they must be updated and validated to provide accurate predictions. To minimize sarcopenia, community health programs should support data-driven insights to target interventions that enhance strength, mobility, and physical activity among aging Filipinos, with a specific focus on high-risk demographics. By integrating predictive analytics and health informatics, these programs can identify at-risk populations, monitor progress through digital health tools, and design personalized, evidence-based interventions that address the unique needs of diverse demographic groups.
8. Sarcopenia risk is influenced by physical performance indicators like weightlifting, chair transfer, and stair-climbing. The high associations between these metrics and demographics in Chi-Square testing demonstrate that these physical performance measurements predict sarcopenia. Sarcopenia-prone populations may benefit from periodic physical performance indicator examinations for early detection and intervention. It emphasizes data-driven and gender-sensitive health strategies.
9. Sarcopenia prevalence disparities between genders highlight the necessity for gender-specific health initiatives. Sarcopenia is more common in men, possibly because of variations in physical activity, health-seeking, and socioeconomic situations.
10. Possible clinical predictive modeling integration. The SVM model's promising results suggest predictive modeling for sarcopenia risk forecasting in clinical settings. Healthcare professionals can stratify patients by risk and personalized interventions backed by data-driven insights. Machine learning algorithms can be integrated into clinical workflows, predictive models can analyze patient data to identify high-risk individuals early and can enable timely and targeted interventions for the aging population.

## RECOMMENDATION

To support the study methodologically and medically, the following recommendations aim to assist researchers and healthcare professionals specifically interested in Sarcopenia and public health. There is still a need to conduct further studies to support the development of an information framework and predictive models for the condition:

1. Train healthcare professionals to use predictive modeling tools such as the SVM model to assess a patient's risk of sarcopenia. Its practice application will boost patient stratification and streamline personalized treatment plans specifically designed for older adults.
2. Increase sarcopenia awareness and research through an information framework that supports popular sarcopenia education programs for healthcare providers, caregivers, and the community.
3. Determine other gendered factors on sarcopenia, such as sexual orientation and gender identity. It is important to stress these factors as they are connected to other intersections such as lifestyle and socio-economic status, among others.
4. Promote a multidisciplinary approach to caring for the elderly. Physicians, dietitians, occupational therapists, and information professionals should collaborate to develop a data-driven information model on the assessment and management of sarcopenia, considering varying factors such as physical, socio-economic, nutritional, psychological, and gendered factors.
5. Further study on Sarcopenia using machine learning in the aging Filipino population of varying contexts is needed to identify novel risk factors, effective therapies, and long-term outcomes for better disease understanding. Machine learning can be utilized to predict disease progression, evaluate the effectiveness of interventions, and encourage evidence-based public health policies in the long run.

## DECLARATION OF CONFLICTING INTEREST

The author declared no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

## DECLARATION ON SOURCES OF FUNDING

The author received no financial support for the research, authorship, and publication of this article.

## REFERENCES

Asian Development Bank. (2021). *Strengthening health and social protection systems for ageing populations in Asia and the Pacific*. Asian Development Bank. https://www.adb.org/publications/strengthening-health-social-protection-systems-aging-populations-asia-pacific

Cruz, G. T., Cruz, C. J. P., & Saito, Y. (Eds.). (2019). *Ageing and health in the Philippines. Economic Research Institute for ASEAN and East Asia (ERIA*). https://www.drdf.org.ph/wp-content/uploads/2023/04/Full-Report-Ageing-and-Health-Philippines-0208.pdf

Metter, E. J., Conwit, R., Tobin, J., and Fozard, J. L. "Age-Associated Loss of Power and Strength in the Upper Extremities in Women and Men." *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 52*, no. 5 (1997): B267–B276. https://doi.org/10.1093/gerona/52A.5.B267.

Natividad, J. (2019a). Health status. In G. T. Cruz, C. J. P. Cruz, & Y. Saito (Eds.), *Ageing and health in the Philippines* (pp. 47–75). Economic Research Institute for ASEAN and East Asia (ERIA).

Natividad, J. (2019b). Health care and healthcare utilization. In G. T. Cruz, C. J. P. Cruz, & Y. Saito (Eds.), *Ageing and health in the Philippines* (pp. 89–104). Economic Research Institute for ASEAN and East Asia (ERIA).

Philippines Statistics Authority. (2023). *Philippine population projections 2020–2050: Key findings on the ageing population*. Philippines Statistics Authority. https://www.psa.gov.ph/node/1684064140

Santilli, V., Bernetti, A., Mangone, M., & Paoloni, M. (2014). *Clinical definition of sarcopenia. Clinical Cases in Mineral and Bone Metabolism, 11*(3), 177–180. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4373271/

Seok, Min, Woo Kim, and Jin Kim. *"Machine Learning for Sarcopenia Prediction in the Elderly Using Physical Activity and Obesity Information."* Healthcare 11, no. 21 (2023): 2881. https://doi.org/10.3390/healthcare11212881.

## AUTHOR BIOGRAPHIES

**Maria Maura S. Tina**o has been with the University of the Philippines since 1993. She is a full-time faculty at the School of Library and Information Studies (SLIS) at the University of the Philippines Diliman. A certified IBM Data Science and Data Analyst Professional, her expertise lies in data science, machine learning, predictive analytics, and data analysis. Her research spans breast cancer detection, sarcopenia risk prediction, and AI adoption among the aging population, demonstrating her commitment to projects that bridge technology and meaningful societal impact.

Email: *em@slis.upd.edu.ph*

**Roda L. Tajon** is an assistant professor at the Faculty of Information and Communication Studies (FICS), UP Open University, where she teaches multimedia theory, popular culture, and gender studies at both undergraduate and graduate levels. She holds an MA in Women and Development Studies from UP Diliman. Before joining the academia, Tajon worked as a development worker with various non-government organizations, focusing on disaster risk reduction and management (DRRM), children's rights, and Indigenous peoples' rights. She also volunteers for Gantala Press, a feminist small press that publishes work by and for women in communities.

Email: *rtajon@upou.edu.ph*

This work is licensed under a
Creative Commons Attribution 4.0 International License.

The PhJLIS is published by the
School of Library and Information Studies,
University of the Philippines Diliman.
ISSN 2719-0471 (Online)