In order to make a model from the given dataset, we have had to fill in values in order to clean up the dataset.

Our tactics for filling values :

* Left-Right-fill
* Kohl’s Surface Area Volume Fill
  + This fill method uses the surface area and volume measurements to recalculate the missing surface area and volume measurements.
  + 12 Parameters: Torso volume, L/R Arm volume, L/R Leg volume, Total Volume, etc.
* MICE

Multiple Imputation by Chained Equations

With the MICE tactic, we are able to take the original dataset and make replicates where we fill the missing data values using the imputation. Then, we are able to view the replicated, and filled datasets to combine or pool the results to obtain the most accurate filled data values.

Neural Net

* Accuracy
  + The mean difference between the outputs of the Neural Network and the actual values of ALM Adjusted was 1.2108kg. With a mean ALM Adjusted of 18.8604kg, the average relative error of the Neural Network was 6.4200%. The maximum difference between a prediction of the Neural Network and the actual ALM Adjusted was 3.9143kg which resulted in a maximum relative error of 20.7542%. The standard deviation of the differences was 0.8969kg which resulted in a relative standard deviation of 4.7557%.
* Activation Function
  + ReLU
* Loss
  + The loss function we used was L1loss in pytorch
  + The optimizer used was the adam optimizer in pytorch
* Standardization of Data
  + We used the SkLearn Standard Scalar function

**Model 427 Best Model**

**Mean 1.210850872532014; Max 3.914346694946289; Std 0.8969533705234165**

**INTRODUCTION**:

Appendicular Lean Mass (ALM) is the amount of lean tissue present in the arms and legs. ALM measurements are derived from DXA Scans and most commonly used in sarcopenia research related to involuntary loss of strength in skeletal muscle mass. Our project utilized data collected from the Size Stream Scan (SS20) of the LSU Pennington Biomedical Research Center’s Metabolism and Body Composition Laboratory. In order to maintain the high accuracy of the DXA scan while using cost effective measuring method of the SS20 Scan, we design a neural network to predict ALM by training on 47 parameters measured from the SS20 Scan before validating the accuracy of our model based on the maximum, mean and standard deviation difference of the predicted versus actual ALM.

**METHOD**:

1. **Filling Missing Values:**

The SS20 database missing values were filled using (1) Left-Right Method, (2) Surface Area and Volume Calculation, and (3) Multivariate Imputation by Chained Equation (MICE). For the Left-Right Fill Method, we used a person’s left and right measurements to fill for the other if either a left value or a right value was missing. To fill in missing values for any variables pertaining to surface area or volume, we used the total surface area or volume provided in conjunction with the values provided for each of the arms, legs, and torso to fill the missing value. The last method we used to fill any other missing data was a Multivariate Imputer which was trained on only the data that had no missing values. This imputer was then applied to the subjects with missing data and the data was filled. The MICE imputer we used was from the SkLearn package for Python.

1. **Neural Network**

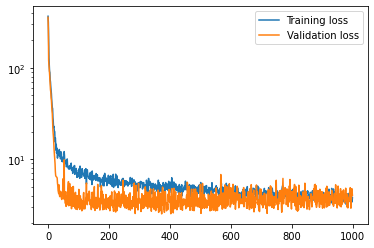
The neural net contains an input layer, 2 hidden layers and an output layer created using the PyTorch package from Python. We use the Adam Optimizer and Rectified Linear Unit (ReLU) Activation Function to train the model at 64 Batch Size and 2000 Epoch and produce outputs if it is positive, otherwise it is zero. To prevent overfitting we standardized our data with SkLearn Standard Scalar function and used the L1 loss function in the PyTorch package.

**RESULTS**

|  | **ALM Differences between Neural Network output and actual ALM value (kg)** | **Relative Error (%)** |
| --- | --- | --- |
| **Mean** | **1.2108**  **ALM Mean at 18.8604 kg** | **6.42** |
| **Standard Deviation** | **0.8969** | **4.7557** |
| **Maximum** | **3.9143** | **20.7542** |

**CONCLUSION**

**Graph: Loss Function of Neural Network**

****

**FUTURE RESEARCH**

In future research, our group recommends classifying subjects dependent on what percentage of their weight is lean muscle and stratifying subjects depending on race, age, and gender equally in the training dataset and validation dataset for the neural network to calculate increasingly accurate ALM prediction.

**Acknowledgment:** Our team would like to thank Dr. Peter Wolenski and Dr. Frederic Marazza for their invaluable guidance and the LSU Pennington Biomedical Research Center’s Metabolism and Body Composition Laboratory for providing the database and sponsoring this project.