IJHSR
ijhighschoolresearch.org
REVIEW ARTICLE

Genetics, Hormones, and Muscle Growth: A Machine Learning Review

Marco Bacares

Meridian World School, 2555 N Interstate Hwy 35, Round Rock, TX 78664, USA; marcobacares67@icloud.com

ABSTRACT: Rising interest in fitness means that more will follow already established programs, but many do not see the same results. AI serves as a way to close the gap between broad programs and individuals, as it can be used to personalize fitness based on the biology of an individual. Through this meta-analysis of various studies of ML models in skeletal and cardiac hypertrophic contexts, it was found that a Generalized Linear Model (GLM) was the most consistently accurate, predicting muscle growth with an AUC of 0.809, while also excelling in transparency compared to other ML models. In applying this, it can be seen that this model is the most efficient in predicting muscle growth, in which a more complete model could include k-means clustering that will allow for the detection of sub-patterns through grouping. These predictions will take in parameters such as protein biomarkers, phenotypic, hormone levels, and genetic data to classify whether a muscle will/won't experience significant muscle growth. It was also found that around 1575 participants would ideally be needed to have enough data for an accurate model that doesn't overfit as well. Through this proposed model/study, machine learning can be utilized universally to progress fitness in individuals.

KEYWORDS: Machine Learning, Computational Biomodeling, Artificial Intelligence, Skeletal Hypertrophy, Algorithms.

Introduction

With the rise of those getting into fitness, it's important to understand goals and what types of movements work best for others. Understanding the type of movements that work better based on their hormones and genetics can help save time and help them progress faster to their goals. For example, the *ACTN3* gene, a genetic polymorphism, determines the muscle-fiber compositions and potential strength of an individual, showing that certain movements may benefit those with certain muscle fiber compositions, showing a greater need for fitness individualization.[1] This demand is urgent as interest in gym/fitness clubs has risen by 8% to 77 million individuals possessing a fitness club membership, with this rate set to progressively increase due to social media shifting gym culture to become more accepting.[2] In addition, the conduction of similar studies has significantly risen post-pandemic, as well, showing a more modern acknowledgement of fitness in everyday life.[3]

My project involves analyzing existing ML models and studies that analyze muscle attributes as well as relationships between hormone levels, phenotypic traits, genetics, and hypertrophy (the bulkening of muscle tissue/muscle growth), the study will present an ideal method that takes into account these variables and models and apply them to an efficient ML model that can accurately predict this specific relationship that specifically takes into account both muscle growth and hormone and genotype, and the dependent relationship on hypertrophy stemming from these factors. This approach is endorsed across other studies that attempt to identify a relationship between phenotypic/external traits and their dependence on genetics, as supported in Libberecht & Noble's study recommending ML to determine gene function.[4] Similar research in tissue-related changes, such as cancer detection, has given similar solutions, such as analyzing CT scan data to detect whether scan images contain the same visual patterns as scans with ovarian cancer, as well as using algorithms to predict its recurrence.[5] This shows that in the context of tissue analysis, machine learning models such as Deep Learning can be used to predict certain indicators of certain events (prognostic biomarkers in the case of disease).[6]

To continue, the following sections of this report will address important terminology in both the fields of ML and hypertrophy, analysis of existing machine learning models in similar domains related to hypertrophy and genetics. Sequentially, a comparison between the covered methods of all projects' attributes will result in a proposed model that best depicts and predicts the relationship between muscle growth and genetics.

Methodology

This paper aims to conceptualize a model and conditions for a study that would accurately create an ML model that predicts hypertrophy based on these genetic factors.

Given the comparative review structure this paper takes on, multiple studies were checked regarding hypertrophy and the algorithms they used to predict for certain genes that signified/were prominent causes for muscle growth, as well as models from studies that predict the muscle thickness/hypertrophic growth itself in addition to predicting whether an individual has the genes or has a high chance of muscle growth in certain areas. Using sites such as Google Scholar, Jstor, and Research Gate to find article files, and searching keywords such as Hypertrophy, Genetics in Hypertrophy, Skeletal Hypertrophy, AI in Hypertrophy, as well as narrow topics specific to the models used in the study, allowed for a more general article in the context of the study. The result was international journals as well as sources on specific topics written by experts in that topic.

DOI: 10.36838/IJHSR816.119

This then let me find journals in the National Journal of Biotech Information, Frontiers, and Scientific Reports. By looking at the data tables they present and comparing them by their AUC across multiple studies, I was able to evaluate them by the same metric of accuracy.

The researcher found universal methods of verifying the accuracy of ML models across multiple studies, finding that strategies such as finding the area under the receiver curve (AUC) were well-established across multiple studies. It is imperative to acknowledge that AUC comparisons were used as a means of accuracy relative to the hypertrophy context, as an algorithm was being conducted on, due to differences in validation strategies, class balances, and varying types of datasets used in the training of different ML models. Due to this, AUCs have been used as a directional accuracy indicator to theoretically show a model's likelihood of being effective in the context that all models are trained in equal, controlled conditions, rather than an objective accuracy ranking.

Through this, ML was also seen to enhance many previously existing predictive models of hypertrophy or genes that support it as well. The aim is to propose a possible ML model geared specifically towards predicting hypertrophy based on genotype, lifestyle habits reflected through calories and sleep hours, and hormone levels. The following models found will be discussed on the metrics of accuracy, transparency in their use, sensitivity, level of data cleaning needed, and self-correction present within each model's application to determine which best fits this application.

Since there were no physical tools or data extracted by personal methods by the researcher, there were little to no ethical considerations regarding the methodology, as all information was gathered from online sources and compiled to create a proposed accurate method for the goal described.

Model Comparison:

Hypertrophic Cardiomyopathy and Genotype Prediction:

Liang *et al.*'s study used combined learning tree-based models with the existing Mayo and Toronto prediction models to predict sarcomeric genotype positivity. As shown in Figure 1, the study used a random forest (RF), which functions by using multiple decision trees when training the model, in which each tree is built on a random subset of the data/muscle tissue used (bootstrapping) and a random subset of the features tested for. Acting in a flowchart-like structure, its internal nodes represent tests on a feature, in which each branch shows the outcome of the test, where the following leaf nodes represent a classification label, such as the patient being positive/negative. The final prediction is decided through aggregating the predictions of all its individual trees via a majority decision for classification. Through this, the variance and overfitting of a model are reduced.[7]

A Gradient Boosted Decision Tree (GBDT), on the other hand, builds trees one at a time, where every new tree made corrects the errors of the previously made trees.[8] It focuses on the residuals/differences between the current and predicted values, combining several weak decision trees into a refined learning model. Through this, accuracy typically increases through iteratively minimizing a loss function through gradient descent.[8] To compare with Table 1, an Adaptive Boosting Decision Tree is another boosting decision tree type that adjusts the weights of wrongly classified records, after a set number of iterations/epochs, creating a specific focus on more obscure/complex cases.[10] The different ways these tree-based methods function have also provided different predictive results for the genotype prediction of hypertrophic cardiomyopathy. Figure 1 illustrates the workflow of the tree-based methods used, indicating the formation of the decision tree in the case of the Random Forest method and the step-by-step formation of the decision tree in the case of the boosting method. Table 1 provides a comparative study of the functioning of these tree-based methods, indicating that the Random Forest method provided the best results in terms of accuracy, with an AUC of 0.92 for the prediction of the hypertrophic cardiomyopathy genotype, using the nonlinear multi-variable function of the genetic traits of the subject. This method is best suited to determine the different aspects of the lifestyle that influence the muscle growth potential of the subject. Gradient Boosted Decision Trees perform best in self-correction, although the chances of overfitting are present. Adaptive Boosted trees perform best when the data is clean and normalized.

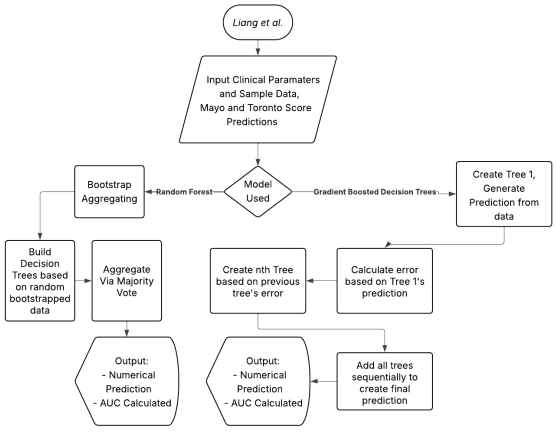

Figure 1: Process for Tree-Based Models in hypertrophic cardiomyopathy prediction: Study on machine learning models in predicting hypertrophic cardiomyopathy risk using Random Forests and Gradient Boosted Decision Trees.[22] Created using Lucid

Table 1: The Random Forest achieved the highest accuracy (AUC: 0.92) when predicting for the hypertrophic cardiomyopathy genotype, while Gradient Boosted models excelled in self-correction while having a greater risk of overfitting.[22]

Model	Function	Accuracy (AUC)	Strengths	Limitations
Random Forest	Creates multiple decorrelated decision trees on random subsets of provided data and features, aggregated by voted majority.	0.92	• High Accuracy • Robust to overfitting • Able to process complex interactions	• Individual predictions of each decision tree aren't as easily interpretable
Gradient Boosted Decision Tree	Creates sequential decision trees, with each tree focused on correcting the previous trees' errors.	0.87	• High Net Reclassification Improvement, as it is able to correct itself more efficiently. • Able to process complex patterns	• More prone to overfitting • Computationally intensive to train each tree sequentially
Adaptive Boosted Decision Tree	Changes the weights of misclassified instances after each epoch, resulting in future iterations focusing on more difficult problems.	0.88	• Effective and accurate for classes with difficult classification	• Unorganized/noisy data could hamper its accuracy, so data needs to be clean/normalized

In terms of study limitations, the sample size was limited, which may have reduced the potential for noise in the data, potentially hindering its general accuracy if applied at a larger scale (Table 1). The baseline data were collected from 104 patients selected from tertiary referral centers, resulting in a smaller data pool and greater bias. This bias stems from the overall younger age of the patients in the groups tested, resulting in increased positive genotype detection across all models.

In all, this source proves itself valuable to applying ML in hypertrophy in relation to genetics through showing random forests to be the most accurate in relation to established tests regarding the precedence of certain types of genotypes within an individual (Figure 1, Table 1). These findings can then be generally applied to genotypes that are explicitly known for increasing muscle growth. While this could likely be applied similarly, the study itself doesn't address events/parameters known to result in changes in muscle tissue, such as exercise and age, in which the types of cells analyzed were exclusively pathogenic sarcomere mutations.

Gene Biomarker Identification Using Machine Learning:

You and Dong used the ML models of Least Absolute Shrinkage and Selection Operator (LASSO) Regression and Support Vector Machine-Recursive Feature Elimination (SVM-RFE). As outlined in Figure 2, an SVM is first trained on a dataset, in which it works by finding an optimal hyperplane in a high-dimensional space that classifies multiple classes with the widest margin. Data points closest to the hyperplane boundary have their distances measured through support vectors, in which RFE ranks all features based on their importance to the SVM model, which was seen through the ordered common genes of *RASD1, CDC42EP4, MYH6, FCN3*, being seen as strong biomarkers. This happens through using the square of the model's coefficients as weights for each attribute, then removing the least important features, in which the SVM is retrained with these remaining features. This allows the model to find the most discriminative subset of genes.[11]

LASSO, on the other hand, is a linear modeling technique that does both variable selection and regularization to improve its accuracy and transparency when being interpreted. It adds a penalty equal to the absolute value of the magnitude of its coefficients via L1 Regularization onto typical linear regression models. This categorizes coefficients' importance, as coefficients seen as less important see their values increase/decrease towards zero, in which they are excluded from the model if they do equal zero. When compared to SVM in Table 2, it prioritizes significant predictive factors in a method that also values interpretability, making it ideal for identifying gene biomarkers when classifying whether an individual did/didn't have hypertrophic cardiomyopathy.[12] Figure 2 shows the performance of SVM-RFE and LASSO in gene selection. The performance of the SVM-RFE method is based on the training of the classifier and the iterative removal of the least important features using the magnitude of the weights. The performance of the LASSO method is based on the implementation of the L1 regularization method. As shown in Table 2, the integration of the two methods resulted in the identification of high-impact genes such as *MYH6, RASD1, CDC42EP4*, and *FCN3* with an accuracy level of 0.993. The advantage of using the LASSO method is the interpretability of the results, unlike the SVM-RFE method, which requires the training of the classifier.

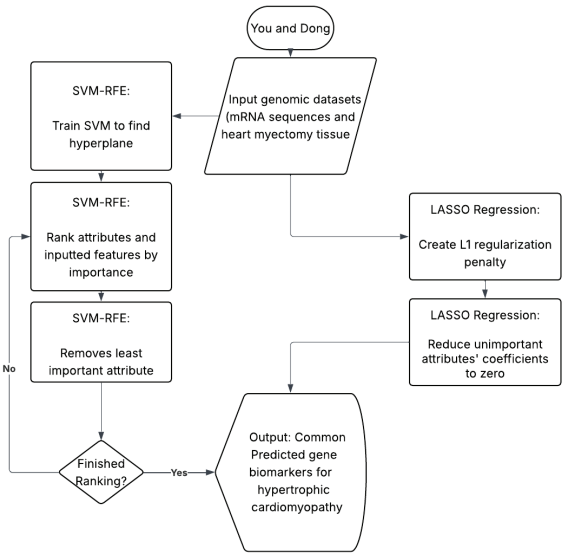

Figure 2: Process for LASSO and SVM-RFE in hypertrophic cardiomyopathy prediction, using machine learning models in predicting gene positivity for hypertrophic cardiomyopathy.[6] Created using Lucid.

Table 2: SVM and LASSO Metrics: Functions, Accuracy, Strengths, and Limitations. SVM-RFE and LASSO models identified key gene biomarkers with high accuracies and AUCs up to 0.993. LASSO provides greater interpretability through feature selection.[6]

Model	Function	Accuracy (AUC)	Strengths	Limitations
Support Vector Machine - Recursive Feature Elimination	Ranks attributes/inputs by importance, removing the least important features through every iteration for optimization.	When ranking features, commonly predicted genes between both models had AUCs of: 0.954, 0.993, 0.922, 0.710 Using a logistic regression.	• Identifies and ranks the most discriminative biomarkers from high-dimensional data.	• Requires a separate classifier. • Not viable on its own.
LASSO Regression	Applies an L1 penalty to make the coefficients of less important attributes zero, creating an optimal model to show which attributes primarily contribute to a certain result.	When ranking features, commonly predicted genes between both models had AUCs of: 0.954, 0.993, 0.922, 0.710 Using a logistic regression.	• Highly Interpretable. • Allows for more accurate feature selection	• Assumes a linear relationship between attributes and outputs

14 genes were predicted to be biomarkers as a result. In all, only four common 4 genes (*RASD1, CDC42EP4, MYH6, FCN3*) in both models were seen as strong biomarkers (Figure 2, Table 2). After the probabilities of each sample are calculated, the true positive rate (TPR) and false positive rate (FPR) are plotted at different probabilities, creating the ROC Curve using logistic regression, in which the area below it is the AUC. For the 2019 dataset, the AUCs for these 4 common genes were 0.954, 0.993, 0.922, and 0.710 (Table 2). Biases are present within the setup of the study, as one of the datasets was from heart myectomy samples, implying the HCM cases in the training data were severe.

Like the previous study, factors such as age and gender weren't taken into consideration when training the model as

DOI: 10.36838/IJHSR816.119

well. From this, specific genes such as *MYH6* were seen to reduce hypertrophy when applied therapeutically, and *FCN3* was seen as a possible use in anti-inflammatory treatments as well.

Deep Learning for Genotype Prediction from Imaging:

Deep Convolutional Neural Networks (DCNNs) were used in Morita *et al.*'s study, using a deep learning structure for processing images, which can be represented as grid-like data, in which Figure 3 outlines this classification process. This is seen as DCNNs automate feature extraction as their initial layers contain convolutional filters to detect features such as edges and specific textures. The following layers take these low-level features to detect specific complex patterns, such as myocardial textures, in this context. Pooling layers reduce the complexity, or dimensionality, of the mapped/visualized versions of these features, making feature detection change less often during shifts. All connected layers at the end of the network combine all extracted features to classify an entity, in this case, checking whether it's positive or negative, as seen in the performance metrics of this model in Table 3. This was used in combination with clinical scores developed using logistic regression and analyzed to see which features significantly affected the model using Grad-CAM. Logistic regression functions through modeling the probability of a binary output based on multiple predictor values. DCNNs were also combined with Mayo and Toronto scales, as well as using variables such as age, echocardiographic measurements, and family history.[13]

Figure 3 shows the DCNN model used in the classification of the genotype features. The first layers are responsible for the identification of the low-level features, such as edges, and the subsequent layers are responsible for the identification of the high-level features, such as myocardial texture. The performance metrics are shown in Table 3, indicating that the DCNN method had an AUC value of 0.86 when the Mayo scores were integrated and an AUC value of 0.84 when the Toronto scores were integrated. The results from the Grad-CAM analysis showed the important features, although the lack of precision may result in biases.

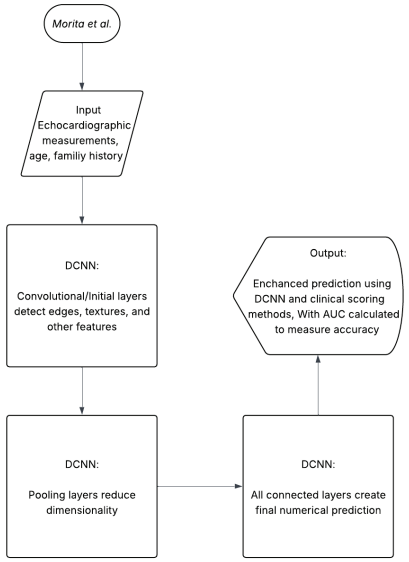

Figure 3: Process for Deep Convolutional Neural Networks in hypertrophic cardiomyopathy prediction.[21] Created using Lucid.

Table 3: Deep Learning Model Performance Characteristics: DCNN Performance Metrics with Clinical Scales. Deep Convolutional Neural Networks (DCNNs) were able to process echocardiographic images relatively accurately for genotype prediction (AUC: 0.86) but were limited by data quality requirements.[21]

Model	Function	Accuracy (AUC)	Strengths	Limitations
Deep Convolutional Neural Network	Extracts and processes attributes and image data using multiple layers. It was combined with clinical scores in the context of this study.	0.86 - DCNN & Mayo 0.84 - DCNN & Toronto	• Excels in processing image data • Automates attribute extraction from data	• Not very transparent, as coefficients aren't as interpretable compared to existing models. • Requires large datasets • Needs decent image quality to be effective

This study was limited by the fact that the sample size of 99 adults is relatively small, which has the potential to reduce its general usability. In addition to this, there is difficulty in understanding their classification reasoning. The coefficients used aren't always clear, making their reasoning and debugging processes increasingly more obscure. Grad-CAM provides some vague insight through deciding which input areas influenced the model's decision the most, but this lack of specificity could create a lack of transparency in the biases within the model's reasoning. The images used in the dataset were high quality, which may not be as generally applicable to hospitals with lower-end imaging. Genetic variant classifications may also change over time.

Machine Learning Models Predicting Muscle Hypertrophy Response:

The study by Yang *et al.* regard the use of three models: Generalized Linear Models (GLMs), Support Vector Machines (SVMs), and Random Forests (RFs). As directed in Figure 4, GLMs are flexible linear regressions that allow the use of response variables with non-normal error distributions, meaning they can process binary, count, and continuous data, as well as form distributions based on unusual error values (error distribution does not result in a normal distribution/bell curve). When deciding whether an individual was a responder/non-responder to Resistance or high-intensity training, a logistic regression GLM was used, functioning by taking a linear combination of input values such as Polygenic Predictor Scores (PPSs), age, and BMI, then using a logistic function to output a binary classifier between 0 and 1, representing the likelihood of someone being a responder, done through a logit link function. A logit link function takes in this linear combination, or sum, and then creates a positive odds value, in which the natural logarithm of all odds values is taken, creating a logit, in which the log of the odds of a class is then turned into a boolean value.[14] GLMs are highly interpretable, with each feature having a coefficient that quantifies its influence on affecting the classification, as positive coefficients mean an attribute increases the probability of a specific classification occurring, and negative coefficients decrease this probability (Table 4).

The highest accuracy was obtained by the GLM approach with an AUC of 0.809. Moreover, it provided better interpretability. The SVM approach provided an accuracy of 0.792, which is capable of handling non-linear relationships. However, it is less interpretable and requires careful parameter tuning.

The Random Forest approach provided an accuracy of 0.695, which is robust to handling outlying values. However, it is less accurate and less interpretable. Further seen as SVMs, in Figure 4, find the optimal hyperplane to separate classes, handling non-linear relationships implicitly by mapping inputs in the high-dimensional spaces based on feature data, then, without computationally intensive complex equations, use support vectors to find the distance between the closest/accurate values to the boundary plane, which are seen as more important in determining a prediction. RFs create a collection of decision trees, allowing them to handle complex relationships between features without overfitting, but seem less interpretable than GLMs, as compared in Table 4.

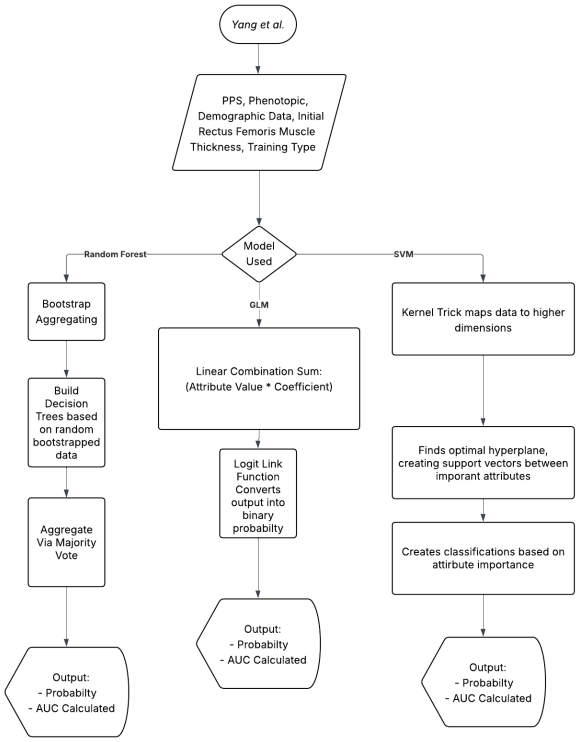

Figure 4: Process for Generalized Linear Modes, Random Forests, and Support Vector Machines in predicting hypertrophy based on exercise type: Study on machine learning models in predicting hypertrophy across GLM, RF, and SVM.[24] Created using Lucid.

Table 4: Summary of Machine Learning Models for Predictive Analysis: Metric Comparison of GLMs, SVMs, and RFs in Hypertrophy prediction. Generalized Linear Models (GLMs) outperformed other models with an AUC of 0.809 for predicting skeletal hypertrophy caused by exercise, balancing accuracy with strong interpretability of feature coefficients.[24]

Model	Function	Accuracy (AUC)	Strengths	Limitations
Generalized Linear Model	Creates a linear combination of inputs and a logit link function to create a boolean probability (between 0 and 1).	0.809	• Most accurate in the study • Highly interpretable coefficients	• Assumes a linear relationship between the logit function's log-odds outputs and input features
Support Vector Machine	Ranks attributes/inputs by importance, removing the least important features through every iteration for optimization.	0.792	• Efficiently handles non-linear relationships well • Useful in high-dimensional spaces	• Less interpretable than GLM • Highly sensitive to parameter and kernel choices
Random Forest	Creates multiple decorrelated decision trees on random subsets of provided data and features, aggregated by voted majority.	0.695	• Handles outliers efficiently	• Least accurate model in the study • Less interpretable/ More "Black Box-like" than GLM.

This study was limited in that the results generated from these models were derived from compound movements such as bench/squat and running intervals, reducing their universal exercise application. The sample size needed to be larger as well, as 450 participants is typically seen as a larger sample size for genome-wide association studies, which typically require millions of SNPs for higher applicability. Bias is present in the population, as they were all young, inactive Chinese adults, creating limitations for possibly identifying changes in hypertrophy across multiple habits and ethnicities. Lifestyle factors such as sleep and calorie intake, which play a large part in muscle growth, weren't measured as well, in addition to hormones associated with muscle growth, such as testosterone. It can be concluded from this study that GLM is the most efficient machine learning model for predicting hypertrophy in this context.

Skeletal Hypertrophy Analysis Using Random Forest Predictors:

Again, Random Forests (RFs) were used in Enzer *et al.*'s study, in which key additions were SHAP (Shapeley Additive ExPlanations) values to explain each attribute variable's influence in determining a prediction, outlined in Figure 5. Based on game theory, SHAP assigns an importance value to an attribute for its contribution to a specific prediction. It calculates an attribute's marginal contribution by considering all combinations of features/attributes and their possible values, allowing for better interpretability to see which features are the most important.[15] The use of SHAP analysis can help provide more clarity to models that typically lack it (contain a black-box nature), allowing them to be more ethically acceptable.[16] These performance metrics are addressed in Table 5, which displays an AUC of 0.744 for the RF model, using SHAP analysis and K-means clustering to increase interpretability and allow for phenotypic subtyping, respectively. The model was further validated to identify this trend within multiple patient subtypes via K-Means clustering, identifying K number of patient subtypes through creating a clustered set of datapoints. The cluster's distance from all points within the dataset is then calculated, creating new clusters based on this calculated distance, assigning it to the nearest centroid, or point within that cluster, creating more clusters until each reassignment presents little to no changes, creating groups based on data. This is seen in this study via attribute data such as race and age, showing a valid method in showing trends among a more specific group, given that the testing data initially presents little biases (Figure 5).[17]

DOI: 10.36838/IJHSR816.119

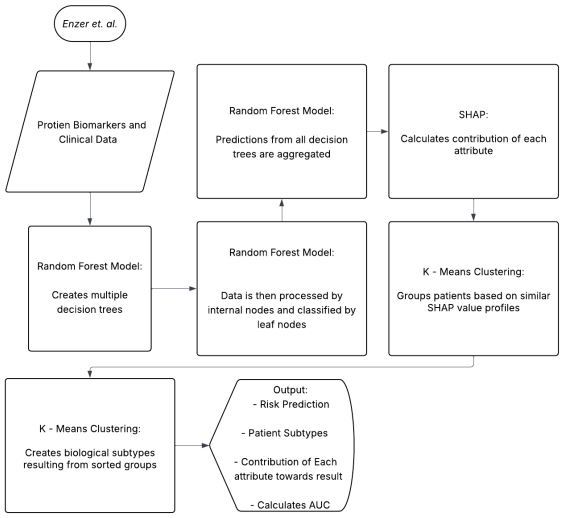

Figure 5: Process for Random Forest Models and Shapely Additive ExPlanations and K-Means Clustering model analysis: Study on machine learning models in predicting hypertrophy using RF models, then being analyzed using SHAP, then validated using K-means Clustering.[25] Created using Lucid.

Table 5: ***Summary of Machine Learning Models and Interpretability Methods:*** RF methods, and features of K-Means Clustering and SHAP Analysis in pattern analysis. While the Random Forest model had a moderate accuracy of AUC 0.744 for predicting sarcopenia development, integrating SHAP Analysis and K-means clustering was vital in improving interpretability and increasing the identification of patient subtypes.[25]

Model	Function	Accuracy (AUC)	Strengths	Limitations
Random Forest	Creates multiple decorrelated decision trees on random subsets of provided data and features, aggregated by voted majority.	0.744	• Handles outliers efficiently	• Least accurate model in the study • Less interpretable/ More "Black Box-like" than GLM.
SHAP Analysis	A Game-theory based method that calculates the contribution of each attribute to a prediction.	None	• Local and global model interpretability • Shows key drivers of prediction	• Computationally expensive
K-Means Clustering	An unsupervised learning algorithm that creates groups/clusters based on the familiarity of SHAP values/data points.	None	• Identifies phenotypic subtypes and patterns within the analyzed population	• The number of clusters (k) can be subjective

Enzer *et al.*'s study utilizes machine learning to predict decreasing rates of hypertrophy in patients or the future development of sarcopenia. Using a Random Forest model, the data used to classify whether a patient had sarcopenia was derived from clinical data and eight protein biomarkers found in blood samples.

The limitations within this model are seen, as in its setup, it was trained on a specific group of smokers from a cohort in an ongoing study at the time, in which measurement also occurred less often due to a single CT Scan image being used to determine the initial muscle measurement and density. Overall, integrating ML appears to be a viable method of detecting changes in hypertrophy, even decreasing changes as seen in the case of sarcopenia, in which the validation methods used can help group smaller patient subclasses as well.

Molecular Pathways and Genes Associated with Muscle Hypertrophy:

Skeletal muscular hypertrophy occurs through microtears that occur in muscle, which then repair through processes such as sleep, increasing in size/thickness. This repair, mechanistically driven by resistance exercise and protein intake, activates the mTORC1 signaling pathway that regulates protein synthesis. Genes such as *Ski, Fst, Acvr2b, AKT,* and *Yap1* are more directly associated with hypertrophy.[18] Illustrated in Figure 6, these molecular relationships are seen through anabolic and catabolic reactions from muscle tearing and their associated genes and pathways. These genes use the following pathways of Igf2-Akt-mTOR axis, myostatin-Smad pathway, and angiotensin-bradykinin pathway to increase anabolic reactions, promoting muscle growth from previously existing muscle, to trigger catabolic reactions to release energy and repress growth, and modulate hypertrophy, occurring the the Igf2-Akt-mTOR axis and the myostatin-Smad pathway respectively.[18,19] Myostatin specifically is a negative regulator, in which its inhibition leads to muscle growth. *IGF1* and *Akt1* genes are also responsible for anabolic reactions in response to hypertrophic training. *Asb15* and *Tpt1* are also seen in skeletal muscle hypertrophy as well.[20]

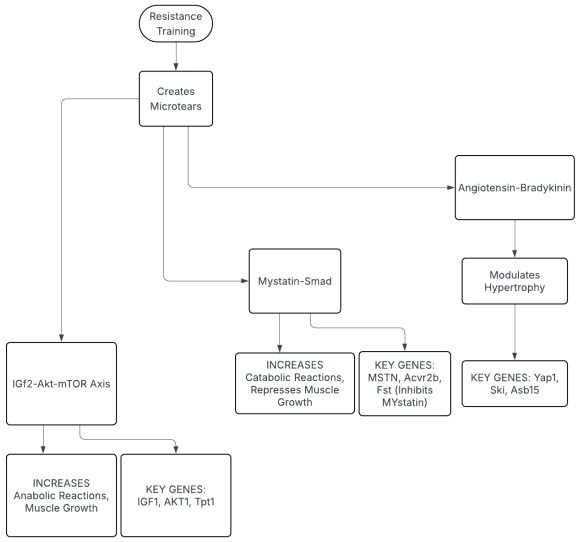

Figure 6: Process for Anabolic and Catabolic reactions from Muscle Tearing, and associated genes: "Gene-pathway diagram illustrating critical molecular pathways involved in skeletal muscle hypertrophy, highlighting genes such as *MYH6, FCN3, Ski, Fst*, and key signaling axes that influence muscle anabolism and catabolism.[18]

■ Discussion

In terms of results presented in each article seen in all studies, the model with the most accuracy in hypertrophy across both cardiac and skeletal hypertrophy is both of You and Dong's models, these being their SVM-RFE and LASSO models, respectively, having AUCs of 0.993 and 0.95, according to Table 6. Conversely, the ML model containing the lowest AUC was the RF model used in Enzer's idea, where its AUC is 0.695. Each study was conducted using a different validation method, as detailed in the Validation Method column of Table 6, including train/test splits and cross-validation, used in addition

to each study's respective machine learning model. Morita *et al.*'s model strengthened pre-existing prediction systems using Deep Convolutional Neural Networks, whereas Enzer *et al.*, on the other hand, used different types of data with the same RF model. In terms of the most efficient models in terms of AUC accuracy for models specializing in predicting skeletal hypertrophy, Yang *et al.* 's GLM had the highest AUC of 0.809. In contrast, its RF model had the lowest AUC of 0.695.

Table 6: ***Inter Model Comparison:*** In all, model performance varies by application, with SVM-RFE/LASSO accurate for gene identification with an AUC of 0.993, while the GLM is the most effective for exercise-induced skeletal hypertrophy with an AUC of 0.809. It further excels in accuracy/interpretability compared to the other models, making it the best for predicting exercise-based muscle growth, allowing for analysis on each attribute's impact in a classification.[6,24]

Primary Focus	Model(s) Used	Key Predictors	Validation Method	AUC/Metrics
HCM Genotype Prediction	RF, GBDT, ABDT	Clinical parameters (LV wall thickness, FHx)	Train/Test Split	RF: 0.92, GBDT: 0.87, ABDT: 0.88
HCM Gene Biomarker ID	SVM-RFE, LASSO	mRNA expression levels	Cross-validated genomics	SVM-RFE (highest detected genes): 0.993, LASSO: 0.95
Genotype from Echo	Deep CNN (+ Mayo & Toronto scores)	Echocardiographic images + clinical models	5-Fold Cross-Validation	DCNN&Mayo: 0.86, DCNN&Toronto: 0.84
Exercise-Induced Hypertrophy	GLM, SVM, RF	PPS, Exercise Type, Baseline Metrics	10-Fold Cross-Validation	GLM: 0.809, SVM: 0.792, RF: 0.695
Sarcopenia Risk Prediction	Random Forest	Proteins (GDF15, EFEMP1), Clinical	Train/Test + Bayesian opt	Combined: 0.744, Biomarker: 0.740, Clinical: 0.646

This meta-analysis emphasises that predicting muscle hypertrophy requires a model that requires genomic information, proteomic and hormonal activity, as well as phenotypic and lifestyle baselines. Given each study, comparing the models used and their strengths, as well as their limitations, allows for a holistic model.

The studies on hypertrophic cardiomyopathy (HCM), even if not being targeted towards skeletal hypertrophy, are more focused on the predictive power of machine learning for genotype-affected outcomes, in which the models used could be applied to analyzing the predicted effect of several genes in skeletal hypertrophy. This could be seen in Liang *et al.* and Morita *et al.*'s studies, whose models were extremely accurate with AUCs reaching .92 and .88 via Random Forests (RFs) and Gradient Boosted Decision Trees in predicting sarcomeric mutations via clinical data.[21,22] The Support Vector Machine-Recursive Feature Elimination (SVM-RFE) and LASSO regression were found to identify key gene biomarkers such as *MYH6* and *FCN3* accurately; these studies show that genetic architecture is a powerful predictor of muscular phenotypes as well, seen through their respective AUCs of 0.954 and 0.993 for determining the impact of these genes on muscular phenotype.[6] Besides prediction, the models used in previous studies can be used to accurately gauge model growth, as a DCNN can also be used to measure muscle cross-sectional area from MRI scan data, ensuring training data is more accurate, rather than solely relying on muscle thickness.[23] While the application of ML in an HCM context provides direction, in terms of genetic effect on phenotype, the studies don't take into account direct factors that could affect muscle growth, such as exercise, diet, and sleep.

The lack of these aspects is then addressed in an all-around comparison of various machine learning algorithms, in which the Generalized Linear Model (GLM) outdid both the SVM and Random Forest (RF) models. This is due to GLM functioning as a flexible linear regression, where outcomes can also be binary count data based on the number of events, and multiple distribution structures. It assigns weights to multiple predictors, computing a linear combination of predictors to create a score that is then transformed through a logistic function to create a probability between 0 and 1 for that score, which is used to classify an output based on threshold/range conditions for probability.[14] This impressive accuracy is seen in its AUC of 0.809, which, while not being the highest, was the most varied due to performance and interpretability, where its clear coefficients show that many polygenic predictor scores (PPN) and the associated exercise type (Such as resistance training or high interval intensity training) significantly affect predicted muscle growth.[24] It then becomes more eligible to be applied in real-world contexts of nutrition and medical advice. Prediction can be based on the thickness of a particular muscle group, in this case, the rectus femoris. This thickness, while not an exact indicator of strength, is an essential metric in measuring muscle growth itself.[24]

Enzer's study takes a more protein-focused approach in predicting changes in hypertrophy via proteomics. For important biomarkers in showing a decrease in hypertrophy through detecting sarcopenia, their RF model was slightly less accurate than Yang *et al.*'s versatile GLM model (Table 6). This was seen, as it had an AUC of 0.740 for these biomarkers (Table 5), in which it identified eight protein biomarkers associated with muscle loss, such as GDF15 and EFEMP1[25] (Figure 5). Using SHAP analysis helped reduce its black box nature, showing the individual contribution of each predictor. Using k-means clustering on SHAP values to determine separate, smaller groups of patients based on phenotypic subtypes showed how different biologic classes in individuals can lead to similar results. Compared to previous models, the SHAP analysis and K-means clustering, which calculate the distance from the points selected to be the initial cluster/centroid, create more clusters to ensure accurate sorting into groups, which could be useful when taking in a variety of individuals, taking in factors such as age, race, genotypes, and hormone levels. In addition to evaluating significant attributes, ethical appeal is increased through its becoming easily understandable, making this model an Explainable AI, which would allow for easier implementation and stronger model trust in professional medical environments.[26]

Given all these advantages over other algorithms, GLM serves as the most robust architecture, in which inputs can be polygenic risk scores summing the effects of hypertrophy-linked SNPs such as OR7A5, CCND3, and PARD3, as well as relevant hormone levels related to muscle growth, such as testosterone, IGF-1, and myostatin.[18] In addition, initial muscle thickness, BMI, age, sex, sleep duration, and exercise should also serve as important parameters. Important biomarkers in hypertrophy detected in Enzer *et al.*'s study, such as GDF15, EFEMP1, CDON, Lymphotoxin a1/b2, VCAM-1, SPARC, Hat1, and NXPH1, were seen to indicate that dif-

DOI: 10.36838/IJHSR816.119

ferent levels of these biomarkers would either show a higher or lower rate of developing muscle mass, in the low pectoralis muscle area specifically.[25] The consideration of these biomarkers is important, for example, GDF15. This biomarker acts as a stress-induced cytokine that restricts the atrophy, or shrinking of muscle tissue, seen in inflammatory conditions such as sarcopenia, and is increased by exercise. This is further seen as it plays a role as a stress-responsive cytokine released by muscle to respond to mitochondrial challenge. Induced by pathways such as the integrated stress response and mitochondrial unfolded protein response, it aids in regulating metabolism and energy balance.[27] Dividing data into a clear train/test split, implementing k-fold cross-validation, and performing SHAP analysis to reduce the model's unknown nature will increase its general usability and detection of smaller groups. Genotype and hormone levels create the foundational boundaries of possible muscle growth, while aspects such as phenotype and training methods create additional changes to these foundational muscle growth boundaries.

Furthermore, the model needs to be able to account for the fiber type and tissue specificity. It must take a specific approach tailored to a specific muscle/muscle group, in which the mass of the muscle shouldn't be the only differentiating factor in determining strength. Still, also functional outcomes for a more applicable model.[25] Additional relevant parameters should include robust phenotypic data, such as the fiber cross-sectional area, fiber type distribution, and functional strength to show whether growth improves movements. These requirements, focusing more on transparency in how it functions, may be difficult to include, given that the current majority of models, such as the ones evaluated, have a black-box nature, meaning their coefficients may not be as clear. Given this, Yang *et al.*'s study fulfils these requirements, focusing on a specific muscle group (rectus femoris) and using models that are highly transparent when compared to other models, such as GLM.[24]

In terms of the studies seen and the models used, there is a greater need to include data relating to the direct hormonal effect on hypertrophy. Despite these hormones being important regulators of protein synthesis, the literature reviewed focuses more on genetics, proteomics, and cardiac imaging. This is likely due to the nature in which hormonal data is collected, as it requires constant sampling and complications in training models due to high biological variability. In terms of how these studies were set up, the participants were all in similar age/sex demographics, showing minimal hormone fluctuations, meaning the relationship between hormones and hypertrophy likely wasn't greatly prioritized in the scope of these studies. In terms of the direct impact of the inclusion of hormones such as testosterone levels and the IGF-1 gene on hypertrophy rates, as seen in Figure 6, these molecules result in anabolic reactions that then increase muscle growth through the Igf2-Akt-mTOR pathway. IGF-1 and increasing testosterone levels increase mechanical load-induced growth seen in skeletal hypertrophy, meaning that accounting for the rate at which hypertrophy is affected based on the involvement of key hormones and genes could help formulate a more complete gauge of one's potential muscle growth.

There are limits present in the actual execution of this model, as it raises ethical concerns. Initially, concerns are present in the individual's privacy, regarding whether the biological information being extracted from people can be kept safe, as well as the control certain experiments/procedures may have on an individual's life. This can be seen in ensuring someone follows a specific exercise, diet, and sleep routine daily over a prolonged period of time, in which extensive consent will be needed. While this model doesn't deal with PII such as credit cards and bank statements, attributes such as age and sex will be taken into account, meaning the use of this model should comply with legal guidelines. Taking into account AI in this context, the use of these ML models should entail descriptive processes of how each model within each study, or the proposed model used in this case uses patient data, highlighting a level of transparency within each model is important to highlight how the actual algorithms function, as well as how the algorithms specifically handle this PII and biodata as well. Any of these models shouldn't replace human expertise in treating or advising in these hypertrophy-related conditions, as they should be used as tools to help direct this assistance.

Study Limitations

Due to this study being a meta-analysis, it couldn't engage in empirical data collection, as it lacked all the data needed to ensure an accurate model. This is due to a lack of access to the technology needed to extract biodata from participants, as well as concerns surrounding a proper sample size. The latter is decided in the Events per Variable Rule (EPV), as according to Peduzzi *et al.*'s frequently cited study, *A simulation study of the number of events per variable in logistic regression analysis*, where the amount of data should range from the amount of events, or the necessary amount of rarer classes in this case, this being the amount of individuals with substantial response to hypertrophic actions, in which the amount of hypertrophic change/muscle thickness will decide how participants are classified.[28]

There should be enough events for every variable being tested: these variables include phenotypic attributes, hormone levels, biomarkers, age, sex, as well as baseline muscle thickness and BMI. The amount of data needed can then be seen through the typical number of events (15), then multiplied by the 35 attributes/variables present, meaning a minimum of 525 events is needed.[29] This number of events is then multiplied by the amount of the rarer class's prevalence, this being 45% for those experiencing great positive change in hypertrophy, meaning that around 1575 (35*45) participants are needed to train a model, in the intention not to overfit data, meaning that a model isn't solely accurate for the training data.[5,13] Given this great magnitude of individuals needed, the resources to acquire such a number of events aren't present, as well as low feasibility in convincing a diverse group of 1575 individuals. With the possible implementation of empirical data, the meta-analysis's findings could have been represented and applied more directly.

Conclusion

In all, taking into account several concepts and strengths from all the models analyzed in different hypertrophic contexts, a new framework is proposed based on the strengths and transparency present in the GLM model, as well as necessary parameters such as phenotypic data, genotypes, levels of protein biomarkers, initial muscle mass, and hormone levels. The model will use validation measures such as k-means clustering to identify smaller groups and subpatterns within the individuals assessed, with the intention of being geared specifically for separate muscle groups, meaning data gathered will have to be based on hypertrophy of a specific muscle. It was also discussed that this model should be used as a tool in addition to human guidance, as well as needing to be very transparent in its data and follow existing policies regarding individual health data, such as HIPAA. Its implementation stems further than simply collecting data, as when implemented directly, models trained on data of one demographic would overfit for that dataset, being inaccurate for broader use (algorithmic bias), in which the use of these overfitting models would create a large responsibility gap, as biased predictions may drive serious fitness decisions.[30]

Furthermore, this study could be extended through the use of larger datasets, such as the application of said ML models discussed above in the context of muscle growth in multiple areas besides the lower pec and thigh areas, allowing for the accounting of more specific factors pertaining to hypertrophy trends in each muscle given certain activity. The inclusion of lifestyle variables such as alcohol/drug consumption, and how these additional substances play a part in regulating hypertrophy, could be addressed to account for dynamic changes an individual could experience while training, allowing for more educated control on the actions one may take in life.

What this study implies is that while various workflow processes are automated by existing AI technology, organic bodily processes that require both physical and mental effort can't be fully driven by AI or any rising technology for that matter. Increasing fitness interests across many signifies a need to accelerate the rate at which certain training plans are created, in which AI serves to bridge the gap between general fitness plans and users, as it can be used to optimize the specificity needed for efficient workout routines pertaining to the genetic makeup of an individual. Modern fitness needs to leverage AI's capabilities in creating optimized routines and lifestyle habits to promote an increasingly more athletic and well-developed population.

Acknowledgments

Special thanks to Dr. Monica Sava, Professor Kay Diaz, and Professor Virgel Torremocha for assisting me on my research journey. Thank you to Jothsna Kethar for giving me this research opportunity.

References

1. Pickering, C.; Kiely, J. ACTN3: More than Just a Gene for Speed. *Frontiers in Physiology* **2017**, *8* (29326606). https://doi.org/10.3389/fphys.2017.01080.
2. Johnson, K. *State of the Fitness Market: 2025 Edition - Lincoln International LLC*. Lincoln International LLC. https://www.lincolninternational.com/perspectives/articles/state-of-the-fitness-market-2025-edition/
3. Beyond Definition. *SFIA's Topline Report Shows 247.1 Million Americans Were Active in 2024*. Sports and Fitness Industry Association. https://sfia.org/resources/sfias-topline-participation-report-shows-247-1-million-americans-were-active-in-2024/.
4. Libbrecht, M. W.; Noble, W. S. Machine Learning Applications in Genetics and Genomics. *Nature Reviews Genetics* **2015**, *16* (6), 321–332. https://doi.org/10.1038/nrg3920.
5. Xu, Z.; Wang, X.; Zeng, S.; Ren, X.; Yan, Y.; Gong, Z. Applying Artificial Intelligence for Cancer Immunotherapy. *Acta Pharmaceutica Sinica B* **2021**, *11* (11), 3393–3405. https://doi.org/10.1016/j.apsb.2021.02.007.
6. You, H.; Dong, M. Prediction of Diagnostic Gene Biomarkers for Hypertrophic Cardiomyopathy by Integrated Machine Learning. *Journal of International Medical Research* **2023**, *51* (11). https://doi.org/10.1177/03000605231213781.
7. Breiman, L. Random Forests. *Machine Learning* **2001**, *45* (1), 5–32.
8. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **2001**, *29* (5), 1189–1232.
9. Angeli, C.; Wagers, S.; Harkema, S.; Enrico Rejc. Sensory Information Modulates Voluntary Movement in an Individual with a Clinically Motor- and Sensory-Complete Spinal Cord Injury: A Case Report. *Journal of Clinical Medicine* **2023**, *12* (21), 6875–6875. https://doi.org/10.3390/jcm12216875.
10. Subasi, A. *adaptive boosting - an overview | ScienceDirect Topics*. www.sciencedirect.com. https://www.sciencedirect.com/topics/computer-science/adaptive-boosting.
11. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene Selection for Cancer Classification Using Support Vector Machines. *Machine Learning* **2002**, *46* (1/3), 389–422. https://doi.org/10.1023/a:1012487302797.
12. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58* (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
13. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. https://doi.org/10.1038/nature14539.
14. Clark, D. R. GLM for Dummies (and Actuaries). *CAS E-Forum* **2023**, *Summer*. https://doi.org/10.1080/00401706.2020.1791959;citation_issn=0040-1706;citation_journal_abbrev.
15. Lundberg, S.; Allen, P.; Lee, S.-I. *A Unified Approach to Interpreting Model Predictions*; 2017. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.
16. Ponce-Bobadilla, A. V.; Schmitt, V.; Maier, C. S.; Mensing, S.; Stodtmann, S. Practical Guide to SHAP Analysis: Explaining Supervised Machine Learning Model Predictions in Drug Development. *Clinical and Translational Science* **2024**, *17* (11). https://doi.org/10.1111/cts.70056.
17. Imad Dabbura. *K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks | Towards Data Science. Towards Data Science*. https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e-644b48a/.
18. Verbrugge, S. A. J.; Schönfelder, M.; Becker, L.; Yaghoob Nezhad, F.; Hrabě de Angelis, M.; Wackerhage, H. Genes Whose Gain or Loss-Of-Function Increases Skeletal Muscle Mass in Mice: A Sys-

DOI: 10.36838/IJHSR816.119

tematic Literature Review. *Frontiers in Physiology* **2018**, *9*. https://doi.org/10.3389/fphys.2018.00553.
19. Atherton, P. J.; Smith, K. Muscle Protein Synthesis in Response to Nutrition and Exercise. *The Journal of Physiology* **2012**, *590* (5), 1049–1057. https://doi.org/10.1113/jphysiol.2011.225003.
20. Chambers, T. L.; Murach, K. A. A History of Omics Discoveries Reveals the Correlates and Mechanisms of Loading-Induced Hypertrophy in Adult Skeletal Muscle. 2024 CaMPS Young Investigator Award Invited Review. *American Journal of Physiology-Cell Physiology* **2025**, *328* (5), C1535–C1557. https://doi.org/10.1152/ajpcell.00968.2024.
21. Morita, S. X.; Kusunose, K.; Haga, A.; Sata, M.; Hasegawa, K.; Raita, Y.; Reilly, M. P.; Fifer, M. A.; Maurer, M. S.; Shimada, Y. J. Deep Learning Analysis of Echocardiographic Images to Predict Positive Genotype in Patients with Hypertrophic Cardiomyopathy. *Frontiers in Cardiovascular Medicine* **2021**, *8*. https://doi.org/10.3389/fcvm.2021.669860.
22. Liang, L. W.; Fifer, M. A.; Hasegawa, K.; Maurer, M. S.; Reilly, M. P.; Shimada, Y. J. Prediction of Genotype Positivity in Patients with Hypertrophic Cardiomyopathy Using Machine Learning. *Circulation: Genomic and Precision Medicine* **2021**, *14* (3). https://doi.org/10.1161/circgen.120.003259.
23. Agosti, A.; Enea Shaqiri; Paoletti, M.; Solazzo, F.; Niels Bergsland; Colelli, G.; Savini, G.; Muzic, S. I.; Santini, F.; Xeni Deligianni; Diamanti, L.; Monforte, M.; Tasca, G.; Ricci, E.; Stefano Bastianello; Pichiecchio, A. Deep Learning for Automatic Segmentation of Thigh and Leg Muscles. *Magnetic Resonance Materials in Physics Biology and Medicine* **2021**, *35* (3), 467–483. https://doi.org/10.1007/s10334-021-00967-4.
24. Yang, X.; Li, Y.; Mei, T.; Duan, J.; Yan, X.; McNaughton, L. R.; He, Z. Genome-Wide Association Study of Exercise-Induced Skeletal Muscle Hypertrophy and the Construction of Predictive Model. *Physiological Genomics* **2024**, *56* (8), 578–589. https://doi.org/10.1152/physiolgenomics.00019.2024.
25. Enzer, N. A.; Chiles, J.; Mason, S.; Toru Shirahata; Castro, V.; Regan, E.; Choi, B.; Yuan, N. F.; Diaz, A. A.; Washko, G. R.; McDonald, M.-L.; San, R.; Ash, S. Y.; Hanania, N. A.; Atik, M.; Bertrand, L.; Aladin Boriek; Monaco, T.; Dharani Narendra; Polverino, F. Proteomics and Machine Learning in the Prediction and Explanation of Low Pectoralis Muscle Area. *Scientific Reports* **2024**, *14* (1). https://doi.org/10.1038/s41598-024-68447-y.
26. Adadi, A.; Berrada, M. Peeking inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160. https://doi.org/10.1109/access.2018.2870052.
27. Johann, K.; Kleinert, M.; Klaus, S. The Role of GDF15 as a Myomitokine. *Cells* **2021**, *10* (11), 2990. https://doi.org/10.3390/cells10112990.
28. Peduzzi, P.; Concato, J.; Kemper, E.; Holford, T. R.; Feinstein, A. R. A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. *Journal of Clinical Epidemiology* **1996**, *49* (12), 1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3.
29. Austin, P. C.; Steyerberg, E. W. Events per Variable (EPV) and the Relative Performance of Different Strategies for Estimating the Out-of-Sample Validity of Logistic Regression Models. *Statistical Methods in Medical Research* **2014**, *26* (2), 796–808. https://doi.org/10.1177/0962280214558972.
30. Char, D. S.; Shah, N. H.; Magnus, D. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *New England Journal of Medicine* **2018**, *378* (11), 981–983. https://doi.org/10.1056/nejmp1714229.

Author

Marco Bacares is a senior at Meridian World School whose research interests blend with his passions for weightlifting and machine learning. Beyond his academic work, he spends his time running, playing soccer, and reading manga. His long-term goal is to pursue a career as a security analyst.