Individualizing Medicine: Using Machine Learning to Predict the Effect of Yoga on Adrenomedullin Levels

Smriti Ramakrishnan

10/29/23

Abstract

Non-communicable diseases (NCDs) are the biggest cause of death globally and have no steadfast vaccine or cure. However, preventative measures such as yoga have been found to prevent NCDs as well as mediate Adrenomedullin (ADM) levels, a biomarker for NCDs, in practitioners. While this relationship has been studied, we plan to study how yoga can aid an individual's efforts to lower their ADM levels. We did this by creating and tuning four regression models that predicted a patient's ADM levels after doing yoga: a Linear regressor, a Random Forest regressor, a Ridge regressor, and an Artificial Neural Network (ANN). We found that the ANN had the smallest mean absolute error (*MAE*) and mean squared error (*MSE*) values, which meant that the ANN's predicted values for ADM levels were closest to the actual values compared to the other models.

Introduction

Non-communicable diseases (NCDs) - illnesses such as cancer, cardiovascular diseases, and other chronic diseases - account for around 74% of all deaths globally ("Non Communicable Diseases", n.d.). For most NCDs, there is no steadfast cure; only preventative measures can be taken ("Non Communicable Diseases", n.d.). However, more and more research is being done on how to predict and track these NCDs. One way this is being done is by using what can be seen as a biomarker for NCDs: Adrenomedullin (Daukantaitė et al., 2018). Adrenomedullin, or ADM, is a vasodilator peptide that lowers blood pressure present and is found in excess in those who are at risk for developing NCDs such as cancers and cardiovascular diseases (Belting et al., 2012). ADM levels are mitigated by yoga (Daukantaitė et al., 2018). But even though this general trend is present, it is still for people to know at an individual level how well a treatment, such as yoga intervention, will work for them. This is what we want to tackle with my research by conducting a study on the efficiency of different Artificial Intelligence (AI)/ Machine Learning (ML) models in predicting how much a person's ADM levels will change after doing yoga for five weeks. By using mean absolute error and mean squared error scores, I will determine which model - Linear Regression, Random Forest Regression, Ridge Regression, or an Artificial Neural Network - will work the best.

Literature Review

Yoga: Applications in Healthcare

Yoga is widely practiced as a therapeutic activity associated with better sleep, reduced anxiety, and an improved quality of life (Moyer, 2023). The focus on breath and the balance between the sympathetic and parasympathetic nervous systems allows control over the body's response to stress, resulting in decreased cardiac stress and an increase in mental and physical health in those who practice it (Woodyard, 2011; Moyer, 2023). This can be seen through evidence from a survey given to over 2300 yoga practitioners by Cartwright et al., 2020, which showed that these practitioners had lower BMIs, lower perceived stress, and improved mental and physical health compared to the UK's average population statistics (Cartwright et. al, 2020).

These health benefits extend to long-term care, with yoga having its use in disease management and prevention. Research on yoga's effect on non-communicable diseases (NCDs) conducted by Verma and Kaur, 2022, showed that yogic asanas (poses) and pranayama (breathing exercises) can improve symptoms of diabetes, obesity, chronic stress, and hypertension by controlling sugar levels, excess fat, stress, and blood flow throughout the body (Verma & Kaur, 2022). Similar ideas are expressed throughout research on yoga and NCDs, corroborating the fact that yoga has been shown to improve symptoms of cardiovascular diseases and diabetes (Salwa & Nair, 2020; Sharma et al. 2019**)**. Yoga also acts as preventative medicine, with a study by Hartley et al., 2014 showing how yoga can decrease blood pressure and cholesterol levels in people at high risk for developing cardiovascular diseases (Hartley et al., 2014).

Adrenomedullin: A Biomarker for Non-Communicable Diseases

While yoga is an important and effective method for preventing and mitigating NCDs and their symptoms, predicting the likelihood of a patient getting an NCD can also shape treatment. Adrenomedullin (ADM) is a vasodilator peptide hormone that lowers blood pressure and may play a role in NCD development and prevention (Geven et al., 2018; Czajkowska et al., 2021; Belting et al., 2012). While a causal relationship has not been proven between ADM and NCDs, it can be seen as a biomarker of NCD development (Daukantaitė et al., 2018). There seems to be a noticeable correlation between elevated ADM levels and the risk of developing cardiac diseases such as hypertension or heart failure (Czajkowska et al., 2021). Similarly, a research study done by Mattias Belting, an oncology professor at Lund University, found that elevated ADM levels in males with cancer were directly correlated with an increased mortality rate (Belting et al., 2012). This pattern has also been found in other studies, with increased ADM levels being correlated with higher severity and mortality rates for individuals with cardiovascular diseases (Wong et al., 2012).

How Yoga Can Be Used to Mediate Adrenomedullin Levels

With ADM levels having a widely proven correlation with NCDs, ways to manage ADM levels in the body must be studied to prevent the development of such diseases. One way to do so is yoga. Daiva Daukantaitė, an Associate Professor at Lund University, researched the effects of a five-week Yin yoga (calming and meditative yoga postures) and YOMI (Yin Yoga and mental health education) intervention on a group of 97 participants, in which she found that there was a decrease in ADM levels in those who did yoga compared to those who did not (Daukantaitė et al., 2018). However, while there was a significant correlation, some people experienced a greater decrease in ADM levels than others. Like most other studies, Dr. Daukantaitė's study focuses on the general trend. And while that is necessary for studies that are trying to prove a correlation, individualized results are also needed to benefit the population. Where Daukantaitė's study focuses on the pattern, we will be trying to focus on individual people who want to see if yoga is a good method for them to reduce their ADM levels. We will do this by tackling my research question, "To what extent can AI/ML models accurately determine whether yoga therapy would be beneficial for patients in terms of decreasing their ADM levels?"

Dataset

The dataset used for this study was taken from a randomized control trial done by Daukantaite et al. on the effects of yoga intervention on plasma adrenomedullin levels (3). This dataset included 97 participants that recorded their ID number ("ID"), what type of yoga they did - Yin, YOMI, or none - ("Group"), how long they practiced at home ("PractiseHome"), age ("Age"), gender ("Gender"), BMI ("BMI"), their cystatin levels before yoga therapy("T1_cystatin"), the before and after results for ADM levels ("T1_ADM" and "T2_ADM" respectively), the before and after results for stress levels ("T1_PSS" and "T2_PSS" respectively), the before and after results for depression levels ("T1_HADSdepression" and "T2_HADSdepression" respectively), the before and after results for anxiety levels ("T1_HADSanxiety" and "T2_HADSanxiety" respectively), the before and after results for sleep quality (“T1_SleepProblem” and “T2_SleepProblem” respectively), and the difference between them for the following: Adrenomedullin levels (“Diff_ADM”), perceived stress (“Diff_PSS”), anxiety (“Diff_Anxiety”), depression (“Diff_Depression”), and sleep problems (“Diff_SleepProblems”). Participants performing Yin Yoga did a variety of calming and meditative yoga postures. Participants doing the YOMI program did a combination of Yin Yoga and mental health education.

Data Pre-processing

Filling Null Values

To fill the null values in for how long each participant practiced at home, we took an average of the amount of time the participants spent practicing at home and filled in the null values for participants doing Yin yoga and YOMI. For the control participants, we recorded their time as zero. For any null values in the BMI, cystatin, before and after ADM, before and after PSS, before and after perceived anxiety levels, before and after depression levels, and before and after perceived sleep problems data, we took an average of all the available data in each section and filled the null values.

Categorical Feature Creation

We created a new characteristic called “BMI_Category”, where we sorted participant BMI into four categories: Underweight (below 18.5), Normal (18.5 to 24.9), Overweight (25 to 29.9), and Obese (above 29.9). For my categorical features - “BMI-Category”, “Group”, and “Gender”- we assigned a number to each category. In “BMI_Catogory”, Underweight was given a value of 0, Normal was given a value of 1, Overweight was given a value of 2, and Obese was given a value of 3. In the “Group” characteristic, the control group (those who did no yoga) was given a value of 0, those who did the YOMI program were given a value of 1, and those who did

Yin yoga were assigned a value of 2. Finally, in the “Gender” characteristic, Men were given a value of 0, and Women were given a value of 1.

Characteristic Alterations

To properly train the model, we excluded all of the post-therapy characteristics (“T2_ADM”, “T2_PSS”, “T2_HADSanxiety”, “T2_HADSdepression”, “T2_SleepProblem”) and the difference characteristics(“Diff_ADM, Diff_PSS”, “Diff_Anxiety”, “Diff_Depression”, “Diff_SleepProblems”). The “ID” characteristic was excluded as it did not contribute to the patient’s health. The “T1_ADM” and “T1_p_cystatin” characteristics were also excluded as the goal of this model was to make using it as accessible as possible. All other characteristics including age, gender, BMI, stress levels, anxiety levels, depression levels, and sleep problems can be found without the need for third-party tests or equipment.

‘Diff_ADM’ was the characteristic that the model was trained to predict. People seeking to improve their health through their ADM levels should be able to use this model to see whether Yin Yoga or YOMI would benefit them.

Methodology

The goal of this research was to create a model that can effectively predict whether a patient will benefit from a yoga program, as seen through ADM levels.

To understand the data, we conducted quick analyses between the patterns of and relationships between multiple characteristics: BMI frequency, BMI compared to pre-trial ADM levels, time spent practicing compared to stress differences, time spent practicing compared to ADM differences, time spent practicing compared to anxiety differences, and time spent practicing compared to sleep differences. A majority of individuals were either of normal or overweight BMI. There was also a weak negative correlation between ADM levels and the

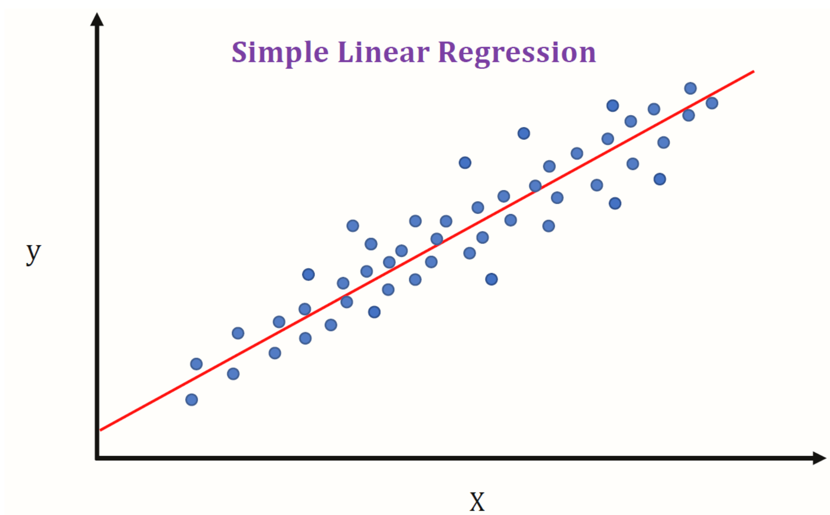
"PractiseHome" characteristic. However, there was no significant correlation between any of the other compared characteristics (see Appendix A).

We created four regression models to see which one performed the best. We created regression models as the models' goal was to predict a value, rather than a category. We ran each model five times and took an average of the resulting mean absolute errors and mean squared error scores to get an accurate view of the model's predictive abilities.

Linear Regression

Figure 1

Simple Linear Regression Visualization



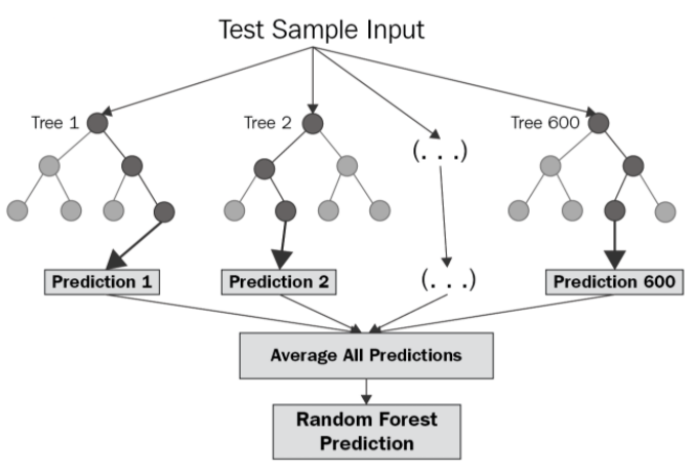
Note. X and y are placeholders for the two variables that are being compared in the regression model. Image sourced from M. Tin, 2020.

The first regression model implemented was a Linear Regression model, as illustrated in Figure 1. This machine-learning model observes the relationship between a dependent and an independent variable. We created this model using a 75-25 train-test split- 75% of the data was used to train the model, and 25% would be used for testing. In tuning the Linear Regression model hyperparameters (variables whose values control how the model learns), The hyperparameters we tried to change were "fit_intercept", "copy_x", "n_jobs", and "positive". All of these hyperparameters were boolean (True or False) variables. Ultimately, we changed the value of "fit_intercept" to False, which resulted in the y-intercept of the linear regression not being used in the prediction.

Random Forest Regression

Figure 2

Random Forest Regression Visualization



Note. The top node of each major tree represents a split dataset, while the bottom nodes represent predictions. Image sourced from A. Raj, 2020.

The second regression model we created was a Random Forest Regressor, as represented by Figure 2. A Random Forest Regressor splits the dataset into smaller sections and trains decision trees to make predictions based on these subsets. It then takes an average of these results to get a more accurate result. We created this model using a 75-25 train-test split. In tuning the Random Forest Regressor, we modified the hyperparameters “n_estimators”, “max_depth”, and “criterion”. The “n_estimators” hyperparameter is an integer (whole number) variable that determines how many decision trees should be used and averaged. The “max_depth” variable is also an integer variable but determines how many levels each decision tree should have. The ‘criterion’ hyperparameter is a variable to measure a split (“Sklearn.ensemble.RandomForestRegressor”, n.d.). But the hyperparameter modifications that benefitted the model most were a “max_depth” value of 1 and a “criterion” value of “abolute_error” (see Appendix B).

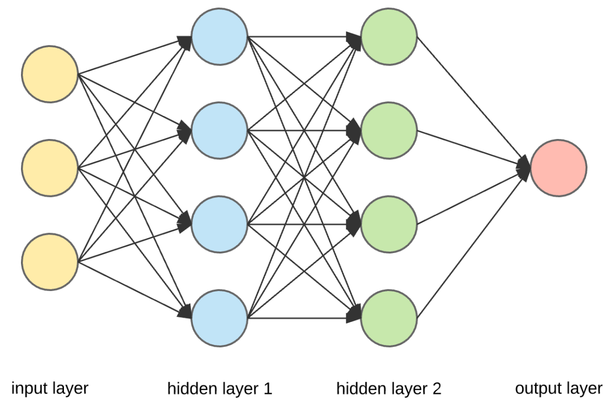
Ridge Regression

The third regression model we created was a Ridge regressor model. Ridge regression can help minimize variance and error in models with multiple independent variables, such as the adm_dataset, and correlations between independent variables/characteristics. We created this model using a 75-25 train-test split. The hyperparameters that we attempted to change were the float (decimal) hyperparameter “alpha”, which determines how the L2 norm affects the weights of the characteristics and the regularization, and the float hyperparameter “tol”, which determines the details of convergence criteria. The only hyperparameter that remained changed was alpha, which was adjusted to a value of 44 (see Appendix C).

Artificial Neural Network

Figure 3

Neural Network Visualization



Note. This Neural Network diagram has one node at the end, which means that the result is one value, similar to my research. Image sourced from G. Ognjanovski, 2019.

The final model that we created was an Artificial Neural Network (ANN), as visualized in Figure 3. An ANN takes in data from an input layer, processes and finds patterns in the data in the hidden layer(s), and creates an output based on the desired result. We created this model using an 83-17 train-test split. The ANN has one input layer with six neurons and the activation function ‘softmax’. It then has one hidden layer for each characteristic (nine hidden layers total) with thirty neurons each and the activation function ‘relu’. Finally, it has one output layer with one neuron since the model is trying to predict one number based on all of the input with an activation function of ‘linear’. The model was run for five epochs, with the optimizer function ‘sgd’, the mean squared error loss function, and the mean absolute error metrics function.

Results and Discussion

The metrics we measured to test the accuracy of the models were mean absolute error (*MAE*) and mean squared error (*MSE*).

$$MAE = \frac{\sum_{i=1}^{n} \left| y_i - x_i \right|}{n} \tag{1}$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n} \tag{2}$$

y_i = predicted value

x_i = actual value

n = total number of data points

MAE determines the average absolute value difference between the actual and predicted values, while *MSE* determines the average value of the squared difference between the actual and predicted values.

The measured *MAE* for the linear regression model was around 0.257 and the measured *MSE* was around 0.109. The measured *MAE* for the random forest regression model was around 0.234, and the *MSE* was around 0.106. The measured *MAE* for the ridge regression model was around 0.260, and the *MSE* was around 0.111. Finally, the measured *MAE* for the artificial neural network was around 0.215, and the *MSE* was around 0.0773.

Out of the four models, the ANN performed the best, with the results for both the *MAE* and *MSE* remaining relatively low and performing better than most of the models. The low scores indicate that it had the highest accuracy in predicting the ADM difference before and after the trial among all models. This means that it could most effectively help patients and yoga practitioners determine whether or not yoga can reduce their ADM levels.

However, in the context of the values for the 'Diff_ADM' characteristic ranging from -1.087 to 0.97, the ANN's *MAE* of 0.215 could lead to inaccurate estimates of a person's ADM level difference. One source of this error could be that many of the values on the dataset were estimated, due to incomplete data. Some values in the "Diff_ADM", "T1_DM", "T1_PSS", "T1_HADSdepression'" "T1_HADSanxiety", "T1_SleepProblems", and "PractiseHome" were null, and had to be filled using means of the characteristic. This means that the models didn't have accurate results by which to make predictions. Another potential source of error could be in the size of the dataset. With only 97 rows of data, some of them having null values that were then assigned mean values, the models might have had inconsistent or limited data to make their predictions, resulting in higher *MAE* and *MSE* values.

Future steps that can be taken to limit such errors include using a larger dataset with a limited number of null values.

Future Improvements

Future steps for this model include creating a User Interface (UI) or application where patients considering yoga as a viable alternative treatment for reducing their risk for NCDs can input their age, gender, BMI, what type of yoga they want to practice, how long they want to practice it for, and their test scores for anxiety, depression, sleep problems, and stress into a program. It can then run the data through the ANN and output an estimate of their ADM level increase or decrease. Since the models made predictions excluding ADM and cystatin levels, people can easily input their data into a UI/application and find out their results.

Conclusion

AI/ML models can be used to predict a person's future ADM levels relatively well. The best model that was created was the ANN, which had an *MAE* score of 0.215 and an *MSE* score of 0.0773. However, further research needs to be done not only to improve the accuracy of the models but also to make them easier to use and more accessible to a wider population of people trying to improve their ADM level and reduce their risk for NCDs.

Works Cited

Belting, M., Almgren, P., Manjer, J., Hedblad, B., Struck, J., Wang, T. J., Bergmann, A., & Melander, O. (2012). Vasoactive peptides with angiogenesis-regulating activity predict cancer risk in males. *Cancer Epidemiology, Biomarkers & Prevention*, *21*(3), 513–522. https://doi.org/10.1158/1055-9965.epi-11-0840

Cartwright, T., Mason, H., Porter, A., & Pilkington, K. (2020). Yoga practice in the UK: A cross-sectional survey of motivation, health benefits and behaviours. *BMJ Open*, *10*(1). https://doi.org/10.1136/bmjopen-2019-031848

Czajkowska, K., Zbroch, E., Bielach-Bazyluk, A., Mitrosz, K., Bujno, E., Kakareko, K., Rydzewska-Rosolowska, A., & Hryszko, T. (2021). Mid-regional Proadrenomedullin as a new biomarker of kidney and cardiovascular diseases—is it the future? *Journal of Clinical Medicine*, *10*(3), 524. https://doi.org/10.3390/jcm10030524

Daukantaitė, D., Tellhed, U., Maddux, R. E., Svensson, T., & Melander, O. (2018). Five-week Yin Yoga-based interventions decreased plasma adrenomedullin and increased psychological health in stressed adults: A randomized controlled trial. *PLOS ONE*, *13*(7). https://doi.org/10.1371/journal.pone.0200518

Geven, C., Kox, M., & Pickkers, P. (2018). Adrenomedullin and Adrenomedullin-targeted therapy as treatment strategies relevant for sepsis. *Frontiers in Immunology*, *9*. https://doi.org/10.3389/fimmu.2018.00292

Hartley, L., Dyakova, M., Holmes, J., Clarke, A., Lee, M. S., Ernst, E., & Rees, K. (2014). Yoga for the primary prevention of cardiovascular disease. *Cochrane Database of Systematic Reviews*, *2014*(6). https://doi.org/10.1002/14651858.cd010072.pub2

Moyer, M. W. (2023, April 17). Yoga for skeptics. *The New York Times*. Retrieved October 18, 2023, from https://www.nytimes.com/2023/04/17/well/move/yoga-benefits.html.

Ognjanovski, G. (2019). *Neural Network Architecture*. Medium. Retrieved October 19, 2023, from https://towardsdatascience.com/everything-you-need-to-know-about-neural-networks-and-backpropagation-machine-learning-made-easy-e5285bc2be3a.

Raj, A. (2020). *Random Forest Regression Visual*. Medium. Retrieved October 16, 2023, from https://towardsdatascience.com/a-quick-and-dirty-guide-to-random-forest-regression-52ca0af157f8.

Salwa, H., & Nair, P. M. K. (2020). Raising burden of non-communicable diseases: Importance of integrating yoga and naturopathy at primary care level. *Journal of Complementary and Integrative Medicine*, *18*(2), 271–278. https://doi.org/10.1515/jcim-2017-0019

Sharma, R., Shahi, V., Khanduri, S., Goyal, A., Chaudhary, S., Rana, R., Singhal, R., Srikanth, N., & Dhiman, K. (2019). Effect of ayurveda intervention, lifestyle modification and yoga in prediabetic and type 2 diabetes under the National Programme for Prevention and control of cancer, diabetes, cardiovascular diseases and stroke (npcdcs)–ayush integration project. *AYU (An International Quarterly Journal of Research in Ayurveda)*, *40*(1), 8. https://doi.org/10.4103/ayu.ayu_105_19

Sklearn.ensemble.RandomForestRegressor. scikit-learn. (n.d.). https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Tin, M. (2020). *Simple Linear Regression*. Medium. Retrieved October 16, 2023, from https://miro.medium.com/v2/resize:fit:1358/0*wB_oQO7aqtCUWu3p.png.

Verma, N., & Kaur, N. (2022). Role and Impact of Yoga in Prevention and Management of Non Communicable Diseases(NCDs). *World Journal of Pharmaceutical Research*, *11*(3). https://doi.org/10.20959/wjpr20223-23284

Wong, H. K., Cheung, T. T., & Cheung, B. M. (2012). Adrenomedullin and cardiovascular diseases. *JRSM Cardiovascular Disease*, *1*(5), 1–7. https://doi.org/10.1258/cvd.2012.012003

Woodyard, C. (2011). Exploring the therapeutic effects of yoga and its ability to increase quality of life. *International Journal of Yoga*, *4*(2), 49. https://doi.org/10.4103/0973-6131.85485

World Health Organization. (n.d.). *Non communicable diseases*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases#:~:text=Noncommunicable%20diseases%20(NCDs)%20kill%2041,-%20and%20middle-income%20countries

Appendix A

Data Visualization

This appendix contains figures of different characteristics in my dataset. The figures were used to identify patterns in the data and correlations between characteristics.

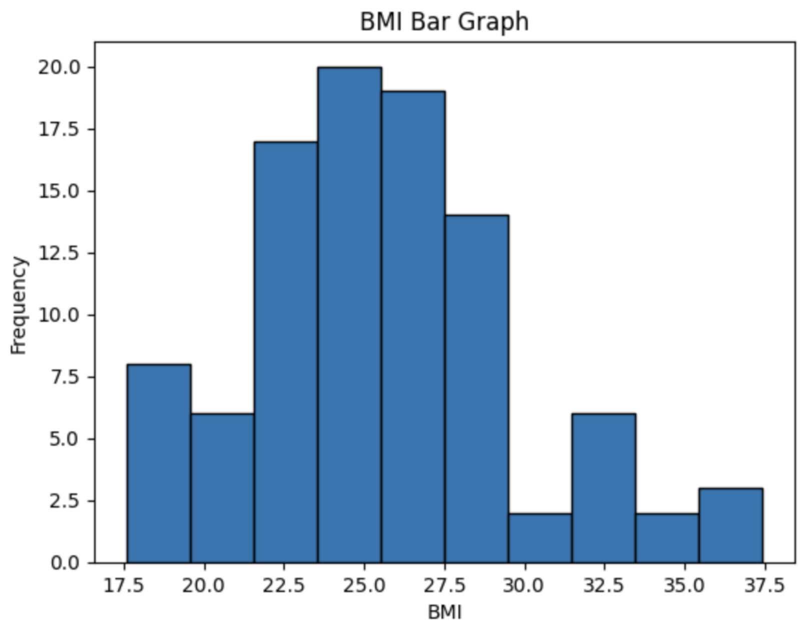


Figure A1. Distribution of BMIs in the dataset. Most BMIs were found to be within the range of 21 and 29.

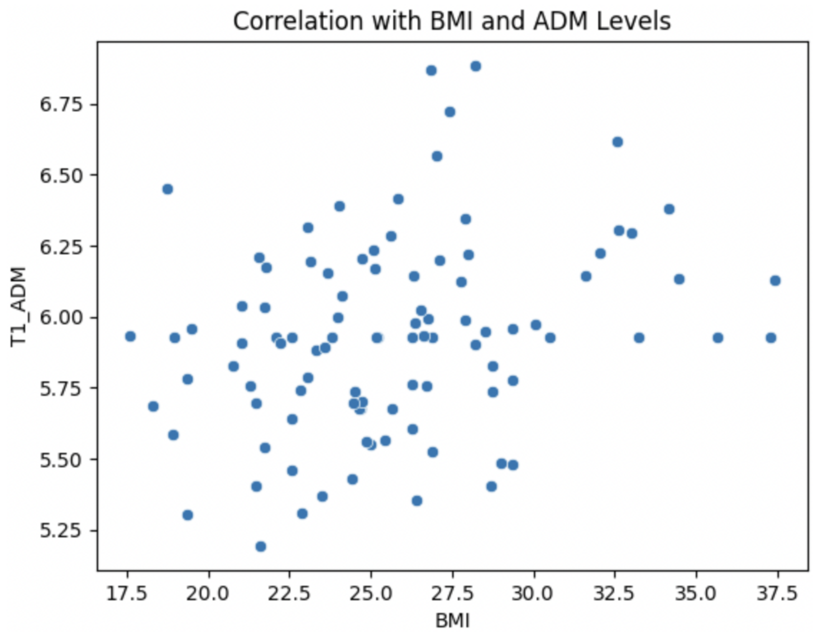


Figure A2. How BMI affects ADM levels in subjects in the dataset. There was no meaningful correlation found between the two.

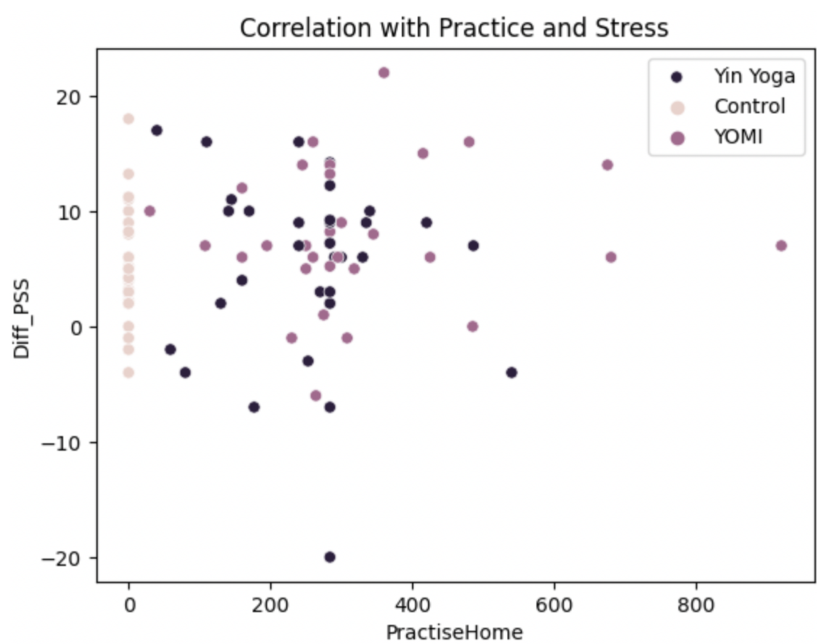


Figure A3. How stress levels were affected by how long the participants practiced yoga on their own. There was no meaningful association found between the two characteristics.

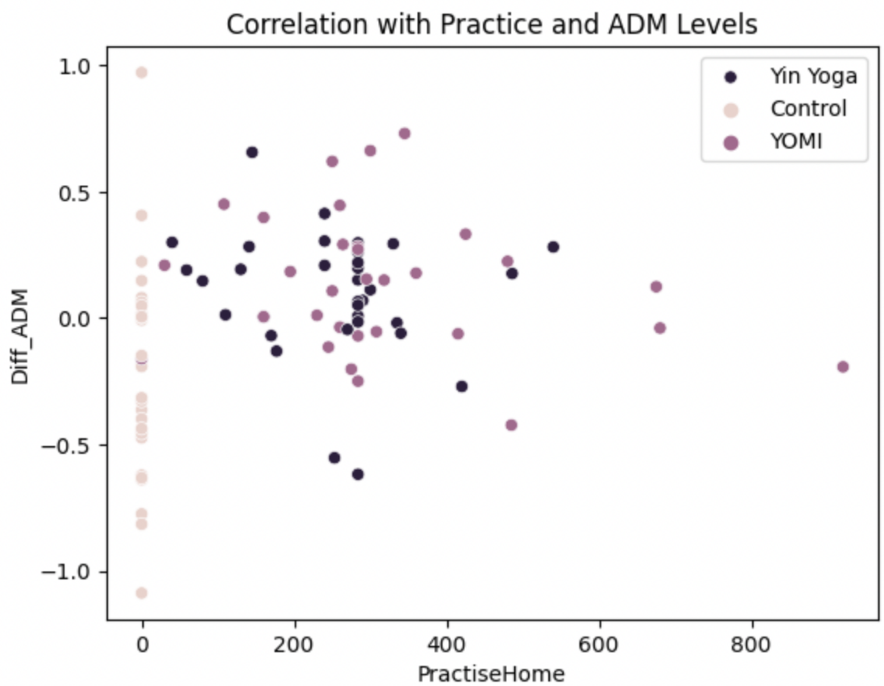


Figure A4. How ADM levels were affected by how long the participants practiced yoga. There was a weak negative correlation found between the characteristics

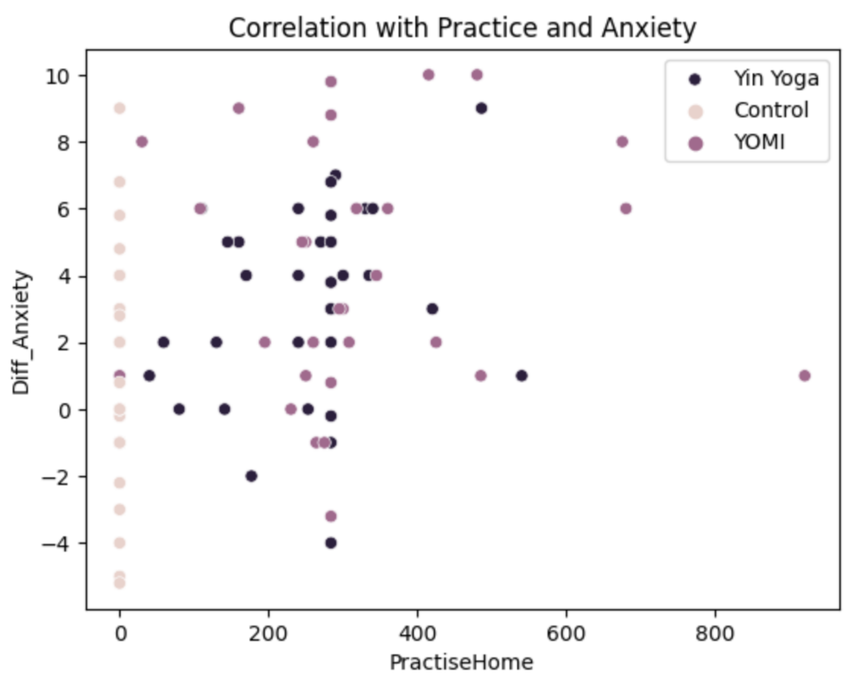


Figure A5. How anxiety levels were affected by how long participants practiced yoga. There was no meaningful correlation found.

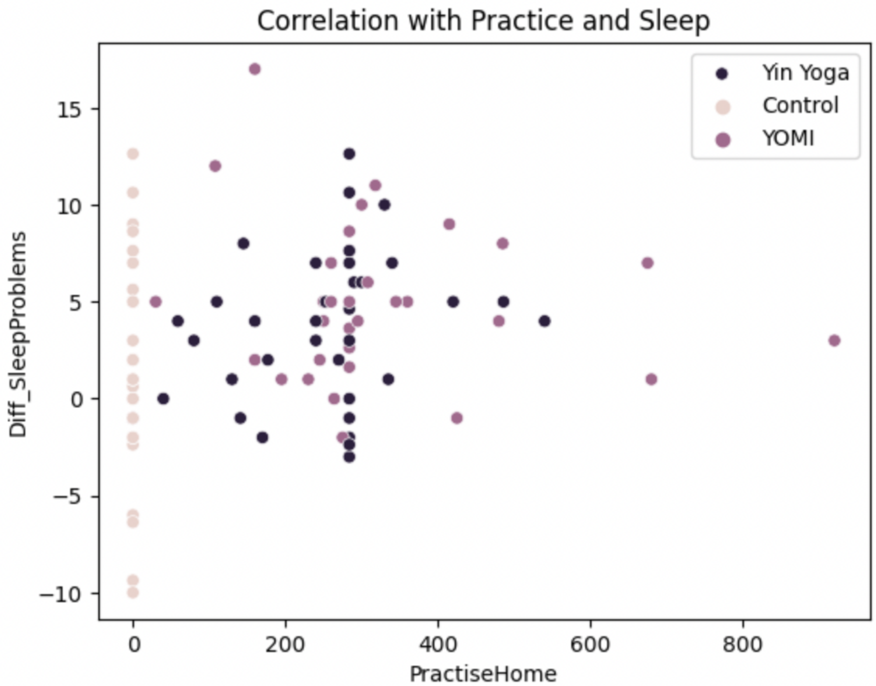


Figure A6. How sleep quality was affected by how long the participants practiced yoga. We found no meaningful correlation.

Appendix B

Random Forest Regressor Hyperparameter Tuning

This appendix contains figures that visualize the hyperparameter tuning done for the edited hyperparameters in my Random Forest Regressor.

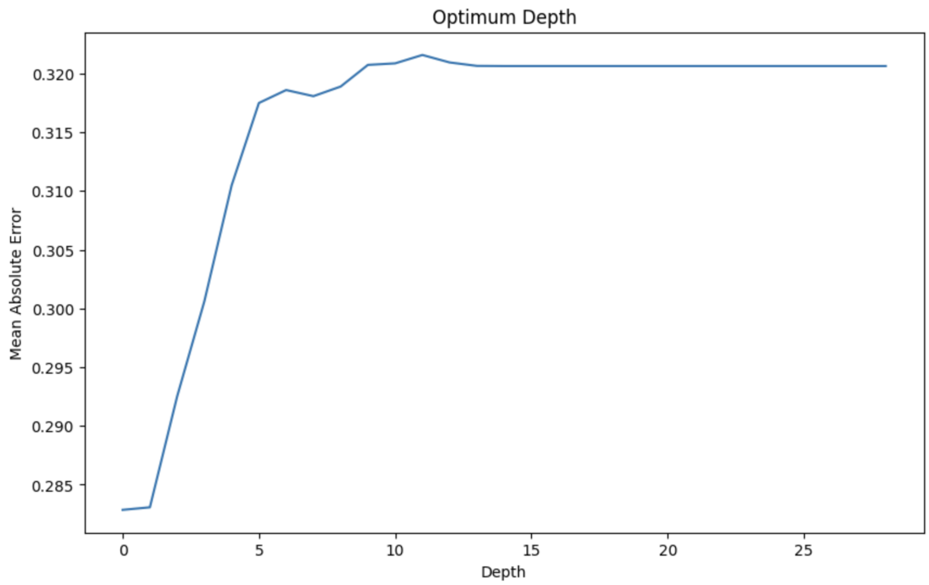


Figure B1. Visualization for best optimum depth for “max_depth” hyperparameter based on *MAE* score. Index 0 on the graph refers to a depth of 1, with the pattern continuing for the rest of the data points.

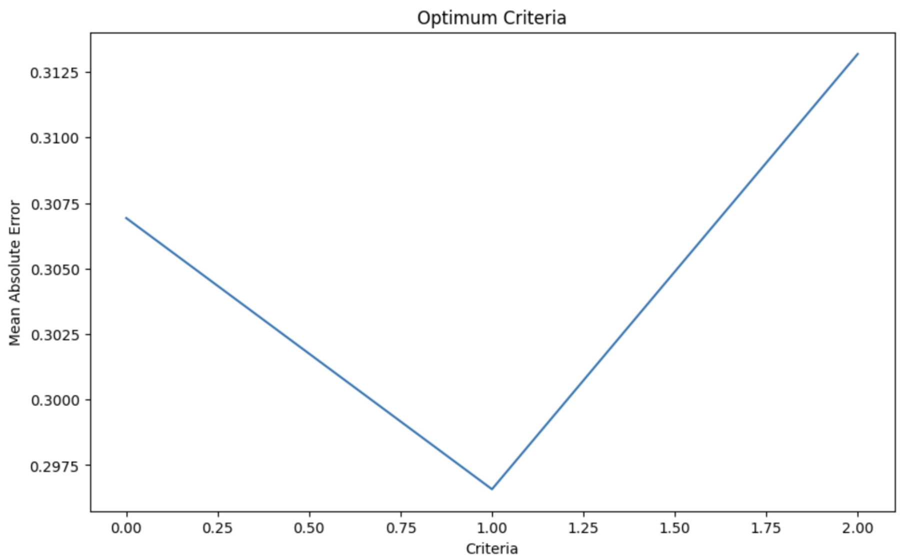
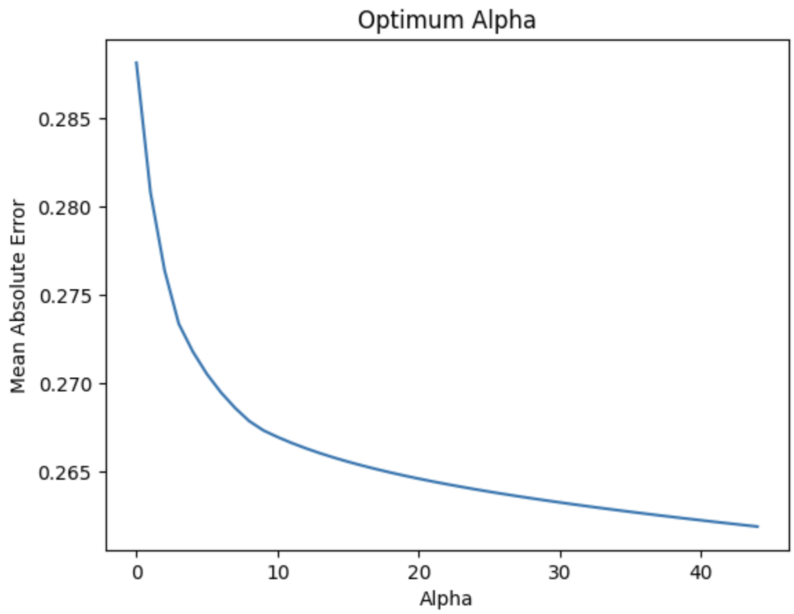


Figure B2. Visualization for best criteria based on *MAE* score: 0 refers to the “squared_error” criterion, 1 refers to the “absolute_error” criterion, and 2 refers to the “friedman_mse” criterion.

Appendix C

Ridge Regressor Hyperparameter Tuning



Appendix C. Visualization for the best value for hyperparameter "alpha" based on lowest *MAE*.