In our literature review we tried to find both the most foundational works within the obesity-machine learning field and the most recent works within the field. One of the most useful papers in this regard was the 2018 review of the use of machine learning in obesity research. The review covered how machine learning has been used in recent years in order to better understand obesity and help in its prevention and treatment. Some of the machine learning techniques applied on the vast array of available obesity data were linear and logistic regression, artificial neural networks, deep learning, and decision tree analysis (DeGregory et al., 2018). The data used to test these methods was the National Health and Nutrition Examination Survey which contained data from 1999 to 2006. The variables collected were race, height, weight, blood pressure, body fat levels, waist measurement, education level, age, gender, and BMI. The data set contained about 25,300 records.

Since the data that we are working with has many variables, a Principal Component Analysis could also show useful information. The review states that the reason it is usually not used in obesity studies is due to the lack of variables— especially the lack of unrelated variables (DeGregory et al., 2018). This article, perhaps due to its age, doesn't take into account the data collected from health monitoring devices and apps. The data that we are working with ties more closely to the kind of smart applications receive. We plan to use PCA and DFA in order to uncover different questions in the data that can then be further pursued using machine learning techniques.

In the review paper they state an issue with deep learning on obesity data is that it requires very large training datasets. If the dataset is too small, that might lead to overfitting and poor results. Since working with small datasets requires a lot of hyperparameter optimization, deep learning is not frequently used on obesity datasets (DeGregory et al., 2018). Another paper from 2018 backed up this point. In this paper published in the Asia-Pacific World Congress on Computer Science and Engineering, researchers used machine learning and deep learning techniques on publicly available datasets and the experiments showed promising results on the study. However they noted that traditional machine learning algorithms perform better but require manpower of data scientist experts, robust feature engineering etc. So further research is needed in deep learning models to reach the level of performance of the traditional machine learning algorithms—especially when working with small to medium sized datasets (Kunal, 2018).

Since the review paper didn’t cover the use of SVMs for classification, we broadened our research and found a recent paper published in 2021. In this paper, 1100 data points were collected and 9 different machine learning algorithms were applied such as k-nearest neighbor(k-NN), random forest, logistic regression, Support Vector Machine(SVM) and so on. Based on some performance metric performances of the classifications were measured. From the results, obesity levels of high, medium and low were determined. Logistic regression has the highest accuracy of 97.09% while gradient boosting gave the lowest accuracy (Fariba, 2021).

However, another article from the 44th International Conference on Telecommunications and Signal Processing showed better results for SVM. In this study, researchers used various classification methods, including SVM, to estimate the obesity levels. The results were compared among different machine learning methods. The Cubic SVM method gave 97.8 % accuracy while selecting the appropriate features that were specific to the problem. Similarly, the accuracy value for Bagged trees and artificial neural networks were 95.4% and 96.5% respectively (Celik et al., 2021). The high level of accuracy using the SVM method is proof of concept for our classification system.

A lot of the papers in the field also were surrounding the early detection of obesity—as early as childhood. Childhood obesity is the most common nutritional condition in children and adolescents around the world. Childhood obesity, also known as pediatric obesity, is a global pandemic that requires attention due to the financial burden it places on the healthcare system for both children and adults. Obesity in children and adults is caused by a high-fat diet, a high-sugar diet, and a lack of physical activity. Obesity in children is a powerful predictor of adult obesity. A retrospective study of children found that those who were overweight or obese when they were 5–7 years old had a higher chance of being overweight or obese when they were 15–17 years old. Obesity prevention may be enhanced by interventions in early infancy, according to studies (Singh, B., & Tawfik, H, 2020).

Predictive algorithms that can detect children at high risk of obesity at an early age and allow for focused interventions could thus be useful tools in combating the national obesity pandemic. Seven machine learning models are employed in this work to predict obesity occurrence by the age of seven. Decision Tree, Gaussian Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, Neural Network, Support Vector Machine with Radial Basis Function kernel, and XGB were the machine learning methods that were examined. For model input, demographic variables were chosen. Results show that machine learning models can accurately predict childhood obesity at a young age, suggesting that they could be useful as a clinical decision support tool. A model like this could help clinicians identify youngsters who are at a higher risk of obesity and provide early intervention.

Another research article on childhood obesity observed the development pattern and onset time of childhood obesity. Their dataset included over 860,510 different individuals. They developed a XGBoost model to predict obesity for children at age 2 years whether they will develop obesity in their childhood.They have included new factors such as body temperature and respiratory rather than general factors such as height, weight, etc. The recall for the children aged 24 to 36 months with obesity was 97.63% and the recall for children between the age group 72 to 84 was 48.96% (Pang et al., 2021).

However, since our research uses data that was collected from adults, we focussed our research around obesity in adults. There were many studies that used machine learning and data analysis techniques to establish ties between certain variables and obesity. For example, this 2016 study observing the relationship between BMI and psychological measures, uses a unique method based on machine learning algorithms such as gradient boosting, KNN, SVM classification and regression tree, random forest and many other algorithms to anticipate unobserved BMI values using psychological characteristics such as depression as predictors (Mitchell T. & Steele R., 2016). These algorithms used psychological characteristics to predict BMI levels as well as BMI status. BMI was used as both a categorical and a continuous variable. The dataset used is made up of psychological variables. When the various machine learning algorithms were trained on positive and negative psychological variables separately, the results revealed that positive psychological variables have no effect on BMI, whereas negative psychological variables do. This shows us that ML approaches can help us enhance our understanding of the disease and our ability to forecast it more accurately.

Another study looked into the relationship between obesity and physical activity; this was investigated using classification methods such as naive Bayes, radial basis function, k-nearest neighbors, classification via regression, random subspace, decision table, and others. These machine learning techniques offer a novel way to examine multi-factorial data, which may then be used to develop predictions about the complex interrelationships that are likely to influence obesity risk (Cheng, X et al , 2021).. The findings show that physical activity particularly moderate to vigorous intensity is a key risk factor for being overweight or obese. As long as energy consumption is not compensated by calorie intake, physical exercise boosts a person's energy expenditure and helps them maintain their energy balance or even lose weight.

A more recent paper on obesity looked into the relationship between obesity and mortality rates for COVID-19. This article points out that obesity is likely associated with increased mortality as a result of COVID infection (Dubnoc-Raz et al., 2021). This is a trend that was also observed during the H1N1 influenza. Obesity is associated with many attributes of poor health, especially poor pulmonary function. The greater the level of obesity, the higher the risk. Our research hopes to create a classification system that will identify obesity level based on the various available features. The ability to differentiate key differences between obesity classes can help personalize medicine and create more efficient treatments for all bodies.

References:

Dubnov-Raz, G., Maor, S., & Ziv-Baran, T. (2021). Pediatric obesity and body weight following the COVID-19 pandemic. *Child: Health, Care, Development*. https://doi.org/10.1111/cch.12939

DeGregory, K. W., Kuiper, P., DeSilvio, T., Pleuss, J. D., Miller, R., Roginski, J. W., Fisher, C. B., Harness, D., Viswanath, S., Heymsfield, S. B., Dungan, I., & Thomas, D. M. (2018). A review of machine learning in obesity. *Obesity Reviews : An Official Journal of the International Association for the Study of Obesity*, *19*(5), 668. <https://doi.org/10.1111/OBR.12667>

Mitchell, T. B., & Steele, R. G. (2016). The Effect of Body Mass Index, Negative Affect, and Disordered Eating on Health-Related Quality of Life in Preadolescent Youth. *Journal of Pediatric Psychology*, *41*(7), 768–776. https://doi.org/10.1093/JPEPSY/JSV163

la Othman Barzinji, C. M. (2021, September). A Machine Learning Approach to Predict the Trend of Obesity Prevalence at a Global Level. *Big Data, Cloud Computing, Data Science & Engineering (BCD)*.

Faria Ferdowsy, K. S. (2021, November). A machine learning approach for obesity risk prediction. *Current Research in Behavioral Sciences, 2*.

Kunal Rajput, G. C. (2018, december). Obesity and Co-Morbidity Detection in Clinical Text Using Deep Learning and Machine Learning Techniques. *Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*.

Paul Wiechmann, K. L. (2017, November). Identifying Discriminative Attributes to Gain Insights Regarding Child Obesity in Hispanic Preschoolers Using Machine Learning Techniques. *International Conference on Tools for Artificial Intelligence (ICTAI)*.

Xueqin Pang, C. B.-S. (2019, December). Understanding Early Childhood Obesity via Interpretation of Machine Learning Model Predictions. *International Conference on Machine Learning and Applications (ICMLA)*.

Yaren Celik, S. G. (2021, Jun 10th). Obesity Level Estimation based on Machine Learning Methods and Artificial Neural Networks. *International Conference on Telecommunications and Signal Processing (TSP)*.

Delnevo, G., Mancini, G., Roccetti, M., Salomoni, P., Trombini, E., & Andrei, F. (2021). The Prediction of Body Mass Index from Negative Affectivity through Machine Learning: A Confirmatory Study. *Sensors*, *21*(7), 2361. <https://doi.org/10.3390/s21072361>

Cheng, X., Lin, S. Y., Liu, J., Liu, S., Zhang, J., Nie, P., Fuemmeler, B. F., Wang, Y., & Xue, H. (2021). Does Physical Activity Predict Obesity-A Machine Learning and Statistical Method-Based Analysis. *International journal of environmental research and public health*, *18*(8), 3966. https://doi.org/10.3390/ijerph18083966