Eric Buehler

**1 Introduction**

High blood pressure, or hypertension is one of the most important risk factors for morbidity and mortality (Paul K. Whelton, et al, 2017). Unfortunately, the measurement of blood pressure is not as straightforward as one might expect due to its inherent variability and inconsistencies in measurement techniques. New approaches have emerged to meet this problem. However, many of them are largely tailored towards privileged groups, ignoring the frequently cited observations that those in rural communities, low-income communities and of minority status have some of the highest risk of hypertension. Thus, the primary goal of this study is to investigate how machine learning might be used in a straightforward and accessible manner to assist healthcare workers in measuring blood pressure. To do this, the study compares newer modeling techniques against traditional OLS regression and leverages these techniques to design a simple tool that can determine if an observed blood pressure is anomalous or not.

**2 Literature Review**

The Lancet reports that uncontrolled hypertension, including those untreated and those inadequately treated, would avert 10 million cardiovascular events worldwide over 10 years (Angell, Sonia Y, et al., 2015). There are many reasons why blood pressure remains difficult to control, however this study chooses to focus on the problems of measurement, and limitations of information and technology. These topics most clearly demonstrate the need for a simple measurement assistant tool that can be used for the detection of anomalies in blood pressure measurements.

While taking BP measurements in office settings is easy, errors are common and can result in a misleading estimation of an individual’s true blood pressure (Paul K. Whelton, et al., 2017). For example, a recent study demonstrated the inability of medical school students to follow blood pressure measurement guidelines and recommended that changes in medical school curriculum be made to emphasize blood pressure measurement (Rakotz, Michael K., et al., 2017). In addition to training issues, in office blood pressure measurement are also subject to white-coat hypertension which is when patients have high blood pressure levels in a physician’s office and normal blood pressure levels at home. White-coat hypertension is thought to be explained by a patient’s anxiety within a physician’s office and in the presence of the physician (Haskard-Zolnierek, Kelly, et al., 2015). The effect of these problems is very real as the NHANES survey found that 18.2%-33.5% individuals who were classified as hypertensive on the first reading were reclassified to lower BP categories (Handler, Joel, et al., 2012). In general, the impact of human error in blood pressure measurement is a well-documented problem, including additional issues such as inaccurate cuff selection and application, incorrect cuff positioning, inadequate rest period, rapid cuff deflation rate, poor observer concentration, digit bias, and lack of repeated measurements (Jones, Daniel W., et al., 2003).

Beyond human based errors, a litany of instrumental errors also affects in office and out of office blood pressure measurements. In terms of in office measurement, the gold standard for in office blood pressure readings has traditionally been the mercury sphygmomanometer paired with the Korotkoff sound technique. However, more recently there has been decreased reliance on this method as mercury is being banned in countries due to environmental concerns (Pickering, Thomas G., et al., 2005). Additionally, systematic errors can occur in sphygmomanometer due to calibration issues and poor maintenance (Parati, Gianfranco, et al., 2006). Aneroid and oscillometric devices are a common alternative to the mercury sphygmomanometer. Oscillometric devices mechanically measure the systolic blood pressure and predict diastolic blood pressure via a company designed algorithm (Whelton, Paul K., et al., 2017). As a result, only devices with a validated measurement protocol are recommended. In general, these devices have been found to be not as accurate as the sphygmomanometer and are still subject to white-coat hypertension (Pickering, Thomas G., et al., 2005).

As an alternative to traditional in office sphygmomanometer and oscillometric readings, more cutting-edge techniques such as Ambulatory blood pressure monitoring (ABPM), and at home monitoring (HBPM). ABPM uses a device to measure blood pressure over a 24 to 48-hour period in or out of office (Seravalle, G., et al., 2018). ABPM is generally accepted as the best measurement of blood pressure, but it’s 24 to 48 measurement period is time consuming and expensive. Additionally, guidelines over best practice have also yet to be established (Whelton, Paul K., et al., 2017). On the other hand, HBPM has a patient intermittently self-report their blood pressure electronically using an oscillometric device (Seravalle, G., et al., 2018). HBPM is often a more practical approach to ABPM as it can be reported at a patient’s convenience and has shown to be similarly effective in measuring blood pressure as ABPM (Whelton, Paul K., et al., 2017; Seravalle, G., et al., 2018). Unfortunately, HBPM self-report style can also be a weakness as patients may fail to report their blood pressure. Since HBPM relies on several measurement, it avoids many of the accuracy issues associated with oscillometric devices (Whelton, Paul K., et al., 2017). Nonetheless, the devices are still subject to miscalibration, which are difficult to detect in home. Overall, the application of either ABPM and HBPM is somewhat situational, but HBPM perhaps provides the best tradeoff between practicality and accuracy.

Despite the promise of ABPM and HBPM, their implementation in low‐resource settings would requires a change in the training of healthcare professionals, regulatory changes regarding the production and sale of BP devices, and technological advances (Seravalle, G., et al., 2018). The irony is that the individuals at the highest risk of hypertension are those, who have limited access to healthcare professionals, technology, and the latest public health evidence and training opportunities (Harris, Jenine K., et al., 2016). No studies have robustly evaluated the cost of implementing HBPM among these high-risk communities (Pickering, Thomas G., 2008). Thus, it seems that one of the most supported new techniques for taking accurate blood pressure measurements, HBPM, generally assumes a privileged audience, and pays little mind to rural and low-income communities who perhaps have the most to gain from more accurate blood pressure readings. In general, the cacophony of instruments and standards used in blood pressure measurements presents the need for useful baseline measurements.

The issue of a privilege audience in new approaches to hypertension control also extends to other novel approaches, specifically those that leverage information technologies. For example, The Kaiser Permanente Northern California (KPNC) has shown that structured, goal-oriented approaches backed up by a large data registry, capable of providing electronic monitoring and target achievements, can greatly improve the control of hypertension (Whelton, Paul K., 2015). NorthShore University Health System uses a unified electronic health record that stores data and makes it accessible across the entire health system, providing consistent information to professional (Merai, Rikita, et al., 2016). In addition, algorithms can be deployed on NorthShore’s health records to accurately identify those at risk of hypertension (Rakotz, M. K., et al., 2014). Telehealth is another common technology-based approach to hypertension management. These approaches use mobile computing and communication technologies to improve the detection and control of hypertension. However, while telehealth strategies demonstrate great promise, they are still a work in progress as they have shown variability and inconsistency (Whelton, Paul K., et al., 2017). Additionally, communication between developers and health professionals on these telehealth technologies has been less than ideal (Burke, Lora E., et al., 2015). The commonality between all these approaches is that they rely on vast data registries, expertise, and cater to an affluent audience.

Despite appeals from the literature to find hypertension control methods that emphasize high risks groups, these are the common measurement and technology-based approaches being researched and recommended by professionals (Whelton, Paul K., et al., 2017; Whelton, Paul K., 2015; Harris, Jenine K., et al., 2016). While some of these solutions may prove scalable, little of the literature has investigated a detailed approach to implementing these solutions in disadvantaged communities.

Some research has already been done on how machine learning might be used in a straightforward and accessible manner to assist healthcare workers in measuring blood pressure. For example, studies have shown that machine learning can be used to predict the presence of hypertension at a high degree of accuracy using a minimal number of inputs (Golino, Hudson Fernandes, et al., 2014; Latifoğlu, Fatma, et al., 2018). These models provide an accessible way to assist in the measurement of blood pressure if trained on reliable data from appropriate populations. Additionally, MedStar Health has found modest success using a simple system that inserts a default blood pressure goal for all their patients that can be used for providing structured feedback and a reference for in office blood pressure measurements (Merai, Rikita, et al, 2016). Thus, MedStar demonstrates a simpler alternative to the more robust approaches used at NorthShore University Health System and KPNC. These simpler technology-based approaches provide the motivating foundation for this studies goal of proposing a simple tool that helps workers in the detection of anomalies in blood pressure measurements.

**3 Methods**

The primary needs of the proposed study are to predict an expected blood pressure for a given individual based on predictors of readily available medical data, using modern machine learning methods. The prediction will be used as a baseline to determine if an observed blood pressure measurement is anomalous or not based on Mahalanobis distance. Key limitations of this methodological goal lie in the data and prediction process. Data limitations are largely a result of lack of access to longitudinal or experimental data. On the other side, the statistical prediction of any measure introduces a degree of uncertainty that stems from either assumption made about data or limitations of the model being applied. The study also limits the scope of predictors used to common medical data for practical purpose, which further limits the model’s predictive power. The implantation of this methodology and limitations are discussed explicitly below.

**3.1 Data**

Data for this study comes from National Health and Nutrition Examination Survey (NHANES), which is a program of studies designed to assess the health and nutritional status of adults and children in the United States (CDC, 2019). NHANES is an ongoing cohort study that has been conducted every other year since 1999 and data are publicly available up to 2015. The individual units of analysis are individuals eighteen years or older living in the United States. NHANES breaks down into several parts, however this study is only concerned with the Demographic, Examination, Laboratory and Questionnaire portion of the survey. Since NHANES is a cohort study, the data cannot show causality as they come from a cohort study, not experimental or longitudinal. As a result, the proposed tool will likely fall short of true medical standards. However, the research may still serve as a proof of concept.

The primary dependent variables of this study are diastolic and systolic blood pressure. NHANES provides blood pressure using a standardized procedure with sphygmomanometer with five minutes of rest between each measurement (NCHS, 2019a). The standardized procedures for blood pressure measurement is a benefit of the NHANES survey as it controls for some of the problems in blood pressure measurement such as inconsistencies in device use and improper technique (Whelton, Paul K., et al., 2017). However, the measurements still fail to address the problems of white coat hypertension and shortcomings of traditional measurement techniques noted in the literature review (Whelton, Paul K., et al., 2017). While up to five measurements are available, in most cases, only three measurements are available. Thus, the “true” blood pressure can only be estimated from the mean of three single visit blood pressure measurements. This is problematic as more measurements from separate visits is the most desirable for decreasing variability and approximating the “true” blood pressure of an individual (Hughes, Michael D., and Stuart J. Pocock., 1992). Unfortunately, with the current resources available, there is little to correct for these issues. Keeping these limitations in mind, NHANES still presents a large amount of standardized blood pressure data that decently approximates levels in the United States and is reasonably suited to the aims of this study.

Independent variables of interest include those supported by the literature that general practitioner might readily have available in the United States. These variables include age (continuous), race (categorical), gender (binomial), body mass index information (continuous), cholesterol (continuous), diagnosed diseases such as diabetes (binomial) and some basic blood measurements such fasting glucose levels (continuous) (Whelton, Paul K., et al., 2017; Stamler, Jeremiah, et al., 1975; Golino, Hudson Fernandes, et al., 2014;). It is important to note that all diagnosis data is self-report. However, body measurement and other biological factors such as cholesterol were measured and recorded by medical professionals. The full documentation for these variables is publicly available by the CDC (NCHS, 2019a). For now, it will suffice to say that they were collected with similar standards as blood pressure.

By design the study relies only on common place medical data such as BMI and cholesterol and thus excludes more complicated biometric data, which has been shown valuable for precise predictions of blood pressures (Zhang, 2018). Therefore, the study is inherently limited in how much variation it will be able to explain.

**3.2 Ethical Concerns**

Since the data are produced by a Federal agency, they are in the public domain and may be reproduced without permission (NCHS, 2019b). While some NHANES data related to geolocation, STDs and youth participants are limited access and require approval from the NCHS, none of this data is required for this study. According to Denison’s IRB guidelines, studies that use anonymous survey such as NHANES qualify as a category 4 exemption, and thus does not require IRB approval (Denison, 2019). Therefore, the use of NHANES data for this study is in line with the standards of National Center of Health Statistics and Denison IRB.

**3.3 Analytical Approach**

The primary methods for this study will include a variety of predictive models, and outlier detection via Mahalanobis distance measure.

In terms of predictions methods, the study calls for approaches whose predictions are easily interpreted. Thus, methods such as linear regression, decision trees are prime candidates. However, it is possible that while these methods are desirable for their ease of interpretation that they may not have enough predictive power to be realistically applied for the proposed tool. Thus, methods such as random forest will be considered as well.

**3.3.1 Linear Regression**

Linear regression performance is well documented in the blood pressure literature (Stamler, Jeremiah, et al., 1975.) While perhaps the most simplistic of the methods, it provides an easy to understand equation and has well documented properties. Thus, it serves as a solid baseline model to compare to others. The short comings of linear regression are largely introduced in its rigorous assumptions of multi-normality, linear independence, no auto-correlation, and homoscedasticity. In terms of blood pressure, linear independence is a particularly troublesome assumption as it is nearly impossible to assume that biological and even demographic variables are completely independent of each other. Homoscedasticity also presents a sizable issue as blood pressure varies unequally across a range of values (quote).

Since blood pressures vary unequally across variables, the stratification of individual’s into different blood pressure groups has been shown to improve model quality (Stamler, Jeremiah, et al., 1975). However, these stratifications typically only rely on a few variables that are known to vary unequally such as race, gender and sex. Alternatively, K-prototype clustering can be used to define blood pressure group based on an entire subset of predictors. Thus, clustering on demographic variables may provide a more holistic method of stratification to account for unequal variance. K-prototype is limited by the fact that the “best” number of clusters is determined by a heuristic that examines within cluster distances. Due to this limitation, it is possible that the clustering offers little improvement over traditional stratification methods.

**3.3.2 Decision Tree**

Decision trees are common alternative to linear regression that are easier to understand and blood pressure’s issue of correlated predictors. The basic idea is that algorithm divides a predictor space into distinct, nonoverlapping regions. It does this by partitioning regions in a way that minimizes the sum of the square of the residuals (RSS). This partitioning allows for interactions of variables as each variable is forced to interact with every variable in subsequent partitions. This is a particularly desirable feature for the prediction of blood pressure, which relies upon dependent predictors. Additionally, decision trees partitioning can be understood in simple tree graph that is even easier to understand than a regression equation, and thus particular desirable for medical predictions. Despite these benefits, decision trees typically overfit data and have high variance. To counteract this, decision trees can be “pruned”, or limited in how many splits they make by setting a threshold for the minimum decrease in RSS, *cp*. However, this weakness may still result in even linear regression outperforming the decision tree.

**3.3.3 Random Forest**

Random forest provides an overall improvement to the decision tree. It does this by aggregating the predictions of many different decision trees. Additionally, at each subspace split the tree is only allowed to consider some number of random predictors, *m*. Thus, one strong predictor need not dictate the entire structure of the tree and new relations of predictors can be considered. The result is a model that decreases the variance of decision trees and that allows for interaction of predictors. However, these benefits come at a cost to interpretability. There is no simple way to visualize a random forest and thus is not ideal in the context of blood pressure predictions. Random forest does provide a way to examine variable importance by looking at the reduction in sum of squared errors whenever a variable is chosen for a split across trees, but this kind of analysis leaves something to be desired considering the linear regression and decision trees.

**Model Evaluation**

Typically, regression models are assessed based on root mean squared error (RMSE), which in this case is the square of the difference between the true blood pressure values and the predicted blood pressure values. Unfortunately, blood pressure isn’t a one-dimensional measure. Rather, it consists of two dimensions: systolic and diastolic blood pressure. A simplistic adaption of RMSE would be to simply take the RMSE of systolic and diastolic blood pressure and compare the model’s performance based on both measures. But what if one model performs better in systolic and worse in diastolic, it is unclear which we would choose as “better”? As a result, comparison by RMSE leaves something to be desired. This problem of comparison can be resolved by Mahalanobis distance, which considers the distance between two multi-dimensional distribution, or in this case the distance between the true systolic and diastolic, and the predicted systolic and diastolic values.

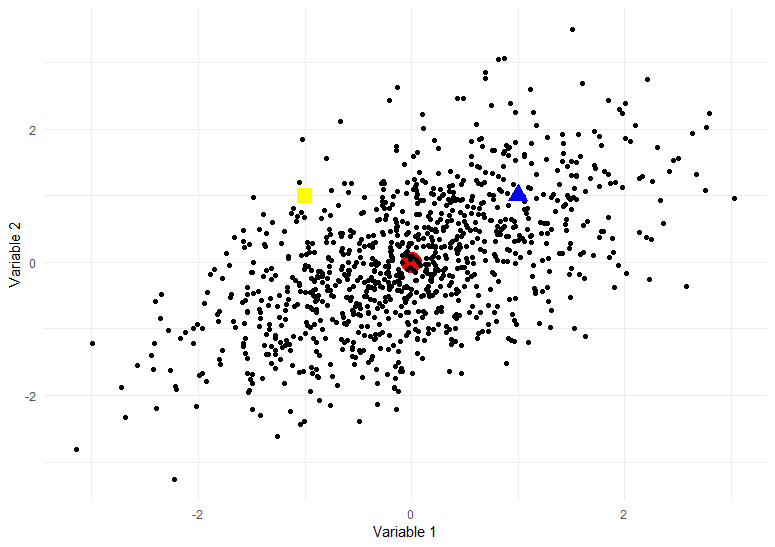
Equation one shows the calculation of the mahalanobis distance between true and predicted blood pressures with represents the matrix of difference between true and predicted systolic and diastolic blood pressure and represents the inverse of the co-variance matrix between the true and predicted blood pressures. Thus, just as with RMSE, a smaller mahalanobis distance between true and predicted blood pressures represents a more accurate model. In short, mahalanobis distance allows for the assessment blood pressure predictions in terms of both its dimensions.

**Outlier Detection**

The primary goal of this study is not to evaluate model performance of models but to leverage the best model for detecting anomalous blood pressures in a novel way. To do this, an observed blood pressure for an individual can be compared to the expected blood pressure generated by the model. If the observed blood pressure is over a certain distance from the blood pressure, it might be considered an anomalous reading as a result of calibration issues with a blood pressure instrument. Otherwise, it might be considered valid.

Mahalanobis distance again comes in handy as it allows for the observed blood pressure for an individual to be compared the predicted blood pressure in terms of its two dimensions. This is particularly handy in outlier detection as an observed blood pressure might differ greatly from a predicted blood pressure, but still be in line with the general distribution of blood pressures and still considered valid. Take for example Figure 1, both the triangle and square are roughly the same distance from the circle in the center of the distribution. However, considering these distances in the terms of Mahalanobis distance, the square may be thought to be further from the circle than the triangle.

**Figure 1:**



The only question that remains is how to determine the appropriate threshold for whether an observed point is anomalous or not. It turns out that distribution of Mahalanobis distances for a multivariate normal distribution is chi-square with degrees of freedom equal to the dimensionality, or number of variables, which in the context of this study is two (cite). Thus, a cumulative probability plot can be made using mahalanobis distances plotted against their corresponding chi-square values, and Mahalanobis distances with chi-square values outside of the 95 percentiles of the chi-square distribution can be considered anomalous. Again, the main concern here is determining an acceptable distance of an observed blood pressure to the predicted in order to determine if it may be an artifact of calibration issues with a blood pressure instrument.

**Implementation**

All algorithmic approaches described above will be implemented using R (R Core Team, 2018). Specific packages are still to be determined. All code will be made available on github (Buehler, 2019).

**App-Interface**

The system described in the analytical approach section will be implemented in a Shiny app (Winston Chang et al., 2019). The app will consist of a simple interface that allows users to enter their medical information, which in the background will be used to predict a blood pressure for the individual. The user will then enter their observed blood pressure, which will be compared to the predicted value, using mahalanobis distance. If the blood pressure is over the predetermined cut-off value, it will be considered anomalous. If it is under, it will be considered valid. The overall practicality of this system is subject to critique. However, it is this studies assertion that more equitable applications of machine learning in the hypertension field are worth investigating.

Overall, the methodology aims of the research is to compare predictive methods based on both systolic and diastolic blood pressures using NHANES to determine if more modern machine learning methods can provide an update to standard linear regression. And subsequently, propose a simple tool that leverages the best predictive model for detecting anomalous blood pressure readings from instrument error or white coat hypertension, and doesn’t rely on vast electronic health records or cater to a vastly affluent audience.

**Pre-Processing**

Each variable of interest in the NHANES survey was downloaded manually from the CDC’s website from years 1999-2015 in .XPT format (CDC, 2019). The files were easily read into R and combined into a single data frame using SASxport package in R (SASxport, 2019). All files included in the study can be found in the NHANES\_DATA folder in this study’s github repository (Buehler, 2019). Fortunately, due to the uniformity of NHANES, there were few discrepancies in the variables from year to year with one exception. From years 1999-2003, total cholesterol and HDL cholesterol (good cholesterol) were combined in the same .XPT file. While year’s 2005-2015, they were in separate .XPT files. This discrepancy was easily resolved by making sure rows bound correctly in the pre-processing stage.

Missing values were then removed, and the data was split into 75% training data and 25% testing data. The resulting training data set contained 4950 observations of 20 variables, while the test set contained 1204 observations of 20. Table 1 shows a snapshot of the resulting training set:

**Table 1:**

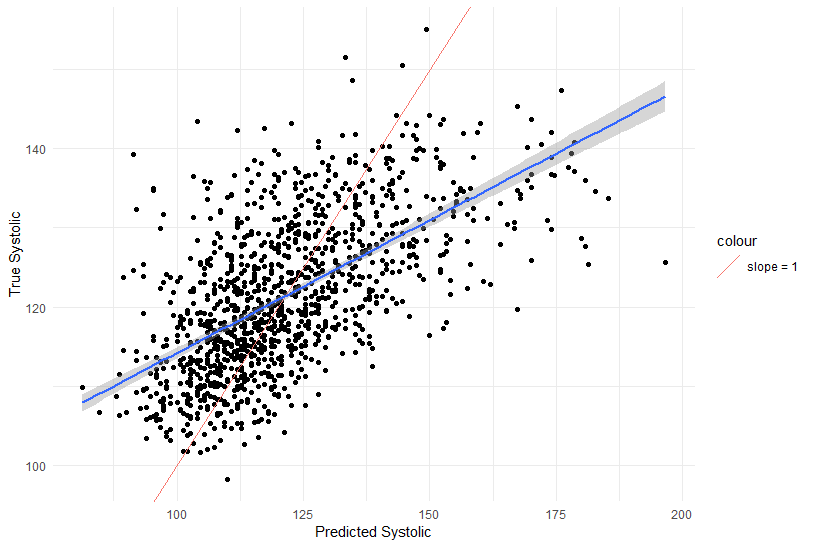
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Estimated Diastolic (mmHg)** | **Estimated Systolic (mmHg)** | **Weight (kg)** | **BMI (kg/m^22)** | **Gender**  **(male =1 / female = 0)** | **Age (yrs.)** | **Hypertensive (yes/no)** | **Hypertension Medication (yes/no)** |
| 55.33333333 | 93.33333333 | 74.4 | 25.4 | 1 | 22 | 0 | 0 |
| 80 | 128 | 62.9 | 24.3 | 0 | 42 | 0 | 0 |
| 59.33333333 | 124 | 63.1 | 24.8 | 0 | 56 | 0 | 0 |
| 80 | 122 | 79.5 | 28.4 | 1 | 24 | 0 | 0 |
| 75.33333333 | 126 | 72.8 | 26.2 | 1 | 51 | 0 | 0 |
| 44.66666667 | 110.6666667 | 53.8 | 17.9 | 1 | 22 | 0 | 0 |

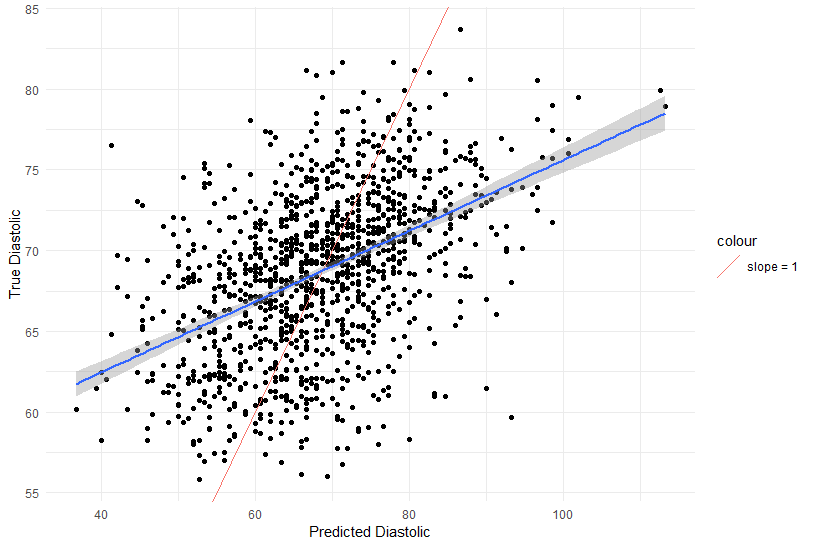
As noted in the data overview, the estimated diastolic and systolic blood pressure were calculated from three observed readings using the mean of three single visit blood pressure measurements from a mercury sphygmomanometer, which introduces a degree of variance between readings that cannot fully be controlled for.

**Results**

The predictive baseline for blood pressure was established by fitting an OLS regression model to both systolic and diastolic blood pressure. Interaction terms were added to the model based on support from the literature and looking at two-way interactions using ANOVA. Assumptions of homoscedasticity, normality and multicollinearity were checked and corrected for as best as possible using transformations (log, squares, etc.). Overall, these assumptions were met rigorously and appear in the appendix section.

The systolic and diastolic linear models achieved R-squares of 33% and 21%, respectively. Figure 1 shows the true systolic blood and diastolic blood pressures compared to the linear predicted diastolic and systolic blood pressures. The blue line with error bars represents the predicted fit of the models, while the red line without error represents a line with a slope of one. As demonstrated by there divergent nature, there is considerable room for improvement over the traditional model, which perhaps is not surprising considering the R-square values. For example, a predicted systolic value of 125 may have an underlying true value ranging anywhere from 115-138 mmHg, roughly speaking.

**Figure 2: True Blood Pressure compered to Linear Predictions**

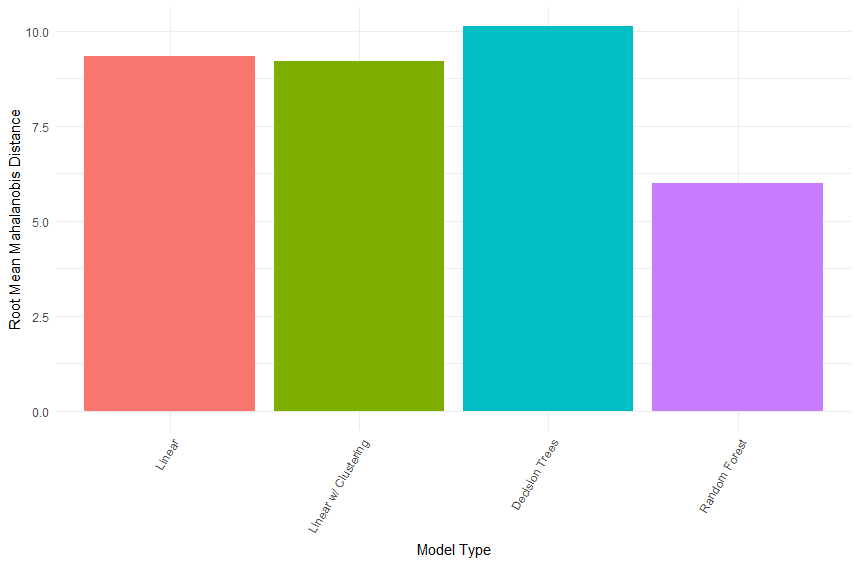


Outside of this somewhat arbitrary example, the RMSE of the systolic and diastolic models are 191 and 92, respectively, which roughly that on average systolic predictions are off by 13 mmHg, while diastolic predictions are off by 9 mmHg. The average Mahalanobis distance is 9, which means on average the linear model is off by 9 mmHg. The RMSE and Mahalanobis distance are not ideal but shows that on average the model does possess some predictive power and at least maintains a semblance of a moderate linear relation between the true and predicted values. Many of the model’s shortcoming are likely due to the unequal variance in blood pressure values across age and likely other variables (see appendix).

Fortunately, the baseline model can be considerably improved upon as shown in Figure 3. Random Forest outperforms the other models by a significant margin with a mahalanobis distance of 5.9, which means that on average the random forest model’s blood pressure predictions were only off by about 6 mmHg. In terms of RMSE, systolic was off 7.8 on average and diastolic, 4.3 on average. Thus, random forest not only outperforms the other models in terms of joint blood pressure, but also systolic and diastolic considered separately.

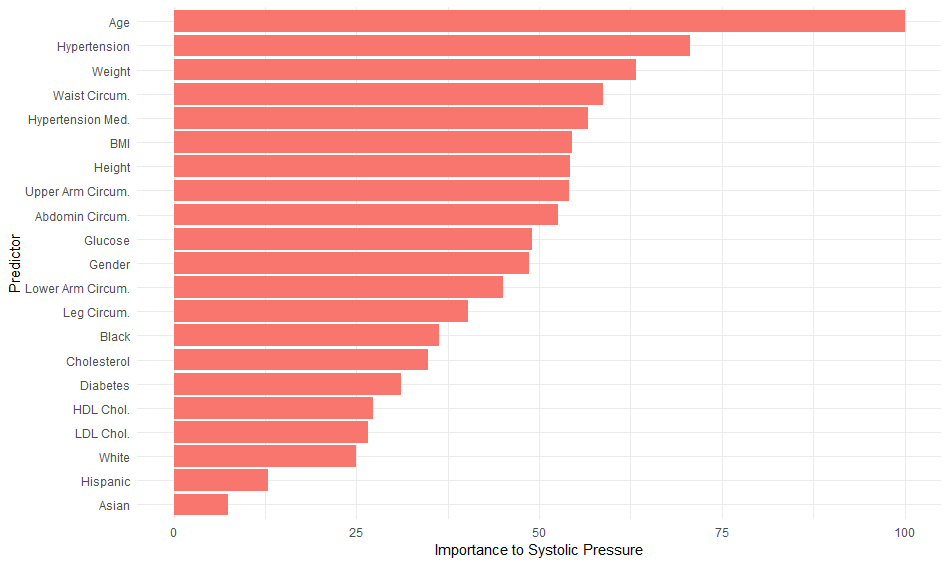
It’s worth noting here that clustering on demographic variables did little to nothing to improve the linear model’s performance, which is the opposite of the expected. This may be a result of poor separation between the clusters in terms of blood pressure. Thus, it may be possible for a better linear model to be constructed using more traditional approaches to stratification.

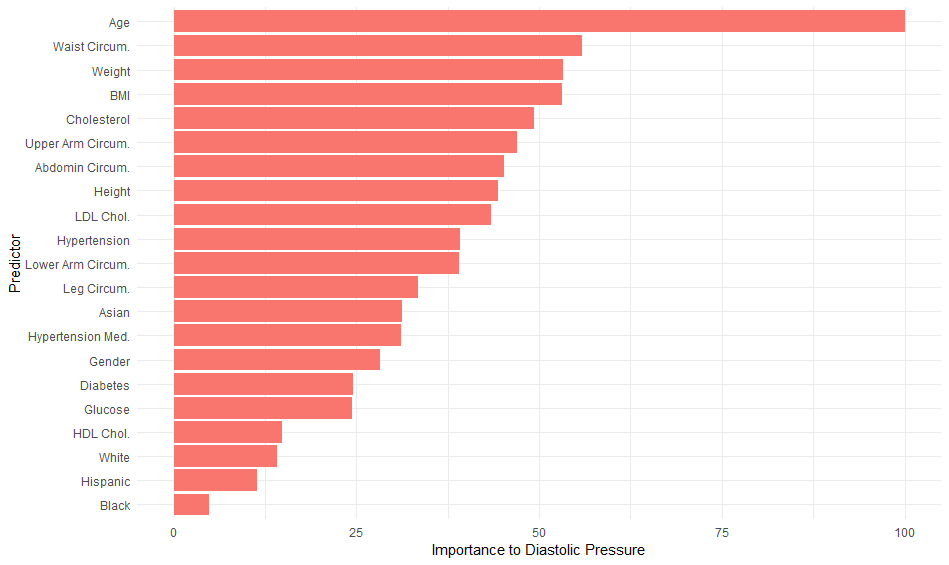
**Figure 3: Comparison of Models in Predicting Blood Pressure**



The final random forest models considered a subset of two predictors at each split and was averaged over on-thousand trees. Figure 4 and 5 on the following page show the variable importance of predictors for systolic and diastolic predictions in the models, respectively.

**Figure 4: Variables Importance in Predicting Diastolic Pressure for Random Forest**

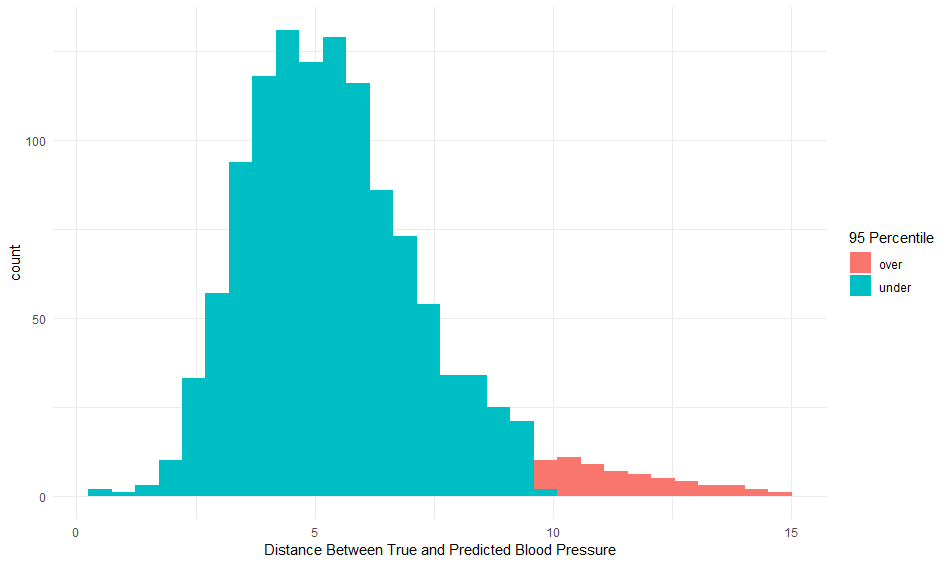
****

**Figure 5: Variables Importance in Predicting Diastolic Pressure for Random Forest**

The graphs make intuitive sense in terms of the literature with variables such as age, weight and waist circumference being among the top five variables for both systolic and diastolic pressures. Additionally, the graphs give a good picture for how the two dimension of blood pressure differ. For example, hypertension condition seems considerably more important in the prediction of systolic blood pressure than systolic, as both the hypertension diagnosis and medication variables appear in the top five most important predictors for systolic, but barely make the top ten predictors for diastolic. On the other hand, cholesterol seems more important for diastolic pressure with total cholesterol appearing in the top five predictors for diastolic pressure, while all three types of cholesterol appear in the bottom ten predictors for systolic pressure. However, here in lies one of the major drawbacks of this random forest model. Variable can only be discussed in terms of their relative importance. There is no equation that tells us the variables coefficients, or visual of a tree that explains how each decision is made. Thus, while the model’s performance is impressive, it lacks what is often needed in medical support tools: explanatory power.

Overall, the random forest’s vast improvement in predictions is most likely due its ability to decrease the variance of prediction, particularly when compared to the other methods and its ability to allow for many interactions between variables, which is particularly beneficial for correlated predictors.

After comparing the distance between true and predicted blood pressure values to the chi-square distribution, a cut-off mahalanobis distance of 9 was considered reasonable (see Appendix Figure 2), which translates to the top 2.5% of distance between true and predicted values as shown in Figure 6. Additionally, the figure also highlights the strength of the predictive model as it shows that only in the top 2.5% of cases is the model predicting blood pressures that are 9 mmHg or more away from the true value. While the approach for selecting a cut-off value is reasonable approach, it is difficult to have full confidence in its validity without testing it in actual practice, highlighting the key limitation in this approach.

**Figure 6: Determining an Outlier Cut-off**

**Discussion**

The goal of this study was to identify weaknesses in the literature concerning the measurement and prediction of blood pressure and design a suggested solution to the issue. This study finds that there are various competing standards and devices in the measurement of blood pressure, which complicate the problem of detecting hypertensive patients. And, the novel technology driven approaches design to resolve this measurement issue have largely ignored their scalability to low-income, rural communities, who suffer from some of the highest levels of uncontrolled hypertension. The presented research outlines a suggested design for a tool that could possibly help bring technology driven solutions for this measurement issue to low-income, rural communities. The design work by providing baseline estimates for individuals in the US that can be used to detect if an individual’s measured blood pressure is due to calibration issues. The results show that a random forest model addresses inherent issues in the prediction of blood pressure such as the dependent nature of blood pressure variables and can be fitted to cohort data from NHANES to obtain accurate measurements with blood pressure prediction only being off on average by 6 mmHg. The study also finds that an observed blood pressure of 9 mmHg or greater away from a blood pressure predicted by the random forest model could be considered anomalous. An app-interface demonstrating such a tool will soon be available on this studies Github (Buehler, 2019).

The question, however, remains, would such a proposed tool be feasible and useful in practice? The question speaks directly to the key limitations of the study, concerning the data used and overall medical standards. The data used for the study comes from NHANES, which was not designed to serve in the use of medical support systems. Experimental, or actual medical data is naturally the standard for such systems. However, the proposed tool doesn’t aim to make any life changing decisions for patients, but rather detect calibrations in a blood pressure device. Since NHANES has been found to be representative of the US, using it for this task may not be completely inappropriate. Additionally, it is a methodological restriction of this study that the tool accessible without relying on the medically resource rich. Nonetheless, this is the most this study can assert is that is not completely inappropriate to use cohort data in the proposed tool. Perhaps, it may be said that the way in which this study attempts to show how machine learning models trained on more reliable data might be deployed in technologies that can help a broader and not just those within a specific health system. Getting to such a point may be far off due to great technical hurdles and the sensitivity of medical data, however, in this studies opinion, it is this studies opinion that it is important to consider how medical machine learning might be used in democratizing ways as opposed to ways that simply benefit the resource rich.

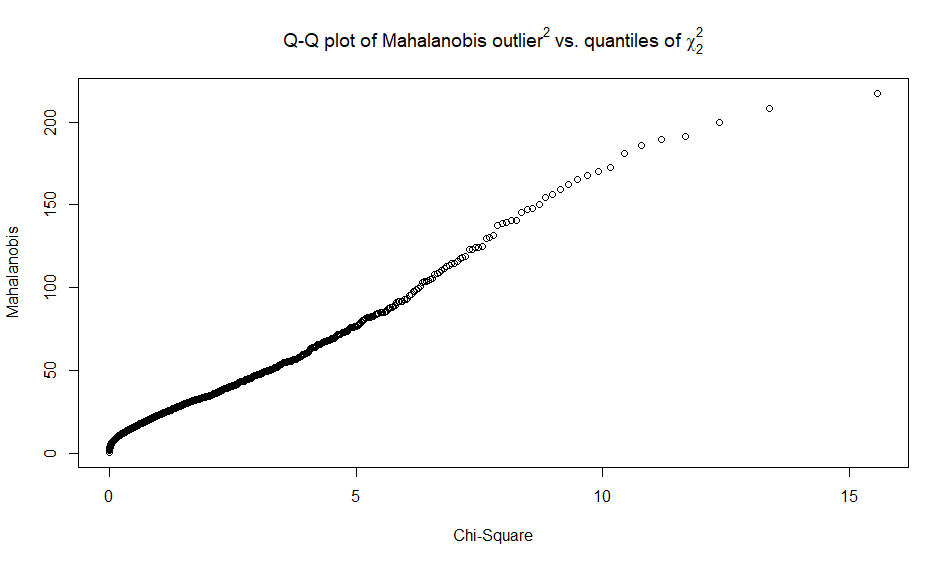
Beyond data limitations, the study is also limited in the sense that it is completely unclear if this tool would work in practice. For example, what if a baseline prediction within 6 mmHg of the true value is not good enough for detecting outliers? Is 9 mmHg a suitable cut-off value? What if it does not detect any miscalibrations or detects too many false-positives miscalibrations? As reasonable as the number and methods might sounds, the fact remains that the tool might not work in practice and this study has no way of answering this question without testing the tool rigorously. Even beyond these concerns, assuming the system does work, the question remains of whether medical professionals or even individuals would find this tool useful? It very well may be the case that the system adds too much complexity for its provided benefit. For example, a simple in-office chart based on a few predictors of high blood pressure may prove more useful in practice than this device. These are the hardest hitting question of this study, but even if the negative response proves true for all them this study still asserts that machine learning technologies are largely being developed in a way that caters to the resource rich. Ultimately, we hope that this studies serves as an example for how we might re-think the application of technology to important problems in health such as hypertension control to serve a broader audience.

**Appendix**

**Table A1:**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Systolic Model** | **Diastolic Model** |
| (Intercept) | 116.81 \*\*\* | 66.22 \*\*\* |
|  | -0.65 | -0.33 |
| Arm Circumference (cm) | `-1.97\*\*\* | `-1.66 \*\*\* |
|  | -0.52 | -0.35 |
| Sagittal Abdominal Diameter (cm) | 5.48 \*\*\* | 1.16 \*\*\* |
|  | `-0.70 | `-0.34 |
| Standing Height (cm) | `-1.51 \*\*\* |  |
|  | -0.4 |  |
| Waist Circumference (cm) | `-2.95 \*\*\* |  |
|  | -0.82 |  |
| Weight (kg) | 2.28 \* |  |
|  | -0.93 |  |
| Gender (0=female/1=male) | 5.52 \*\*\* | 1.99 \*\*\* |
|  | -0.59 | -0.31 |
| Age (yrs.) | -0.42 | 15.22 \*\*\* |
|  | -1.23 | -0.91 |
| Age^2 | 6.81 \*\*\* | `-15.17 \*\*\* |
|  | -1.23 | -0.96 |
| Hispanic | -0.87 |  |
|  | -0.64 |  |
| White | `-1.85 \*\* | 0.64 |
|  | -0.62 | -0.35 |
| Black | 3.34 \*\*\* | 1.28 \*\* |
|  | -0.7 | -0.41 |
| Diabetes (0=no/1=yes) | -0.34 | 0.33 |
|  | -1.06 | -0.46 |
| LDL-Cholesterol (mg/dL) | `-2.63 \*\*\* |  |
|  | -0.51 |  |
| log(Total Cholesterol) (mg/dL) | 3.12 \*\*\* |  |
|  | -0.5 |  |
| log(Fasting Glucose) (mg/dL)) | 1.61 \*\*\* |  |
|  | -0.33 |  |
| Hypertensive (yes/no) | 7.26 \*\*\* | 3.83 \*\*\* |
|  | -0.95 | -0.68 |
| Hypertension Medication (yes/no) | 1.06 | 0.91 |
|  | -1.05 | -0.78 |
| Arm Circumference:Age | `-0.98 \*\*\* |  |
|  | -0.23 |  |
| Gender:LDL | 1.35 \*\*\* | 0.56 \* |
|  | -0.4 | -0.28 |
| Gender:Fasting Glucose | -0.64 |  |
|  | -0.39 |  |
| SAD:Diabetes | `-1.86 \*\* |  |
|  | -0.644337149 |  |
| Gender:Age | `-2.48 \*\*\* |  |
|  | -0.42 |  |
| SAD:Hypertension Medication | `-1.97 \*\*\* |  |
|  | -0.53 |  |
| Age:White | `-2.31 \*\*\* | `-0.59 \* |
|  | -0.41 | -0.29 |
| Age:Diabetes | 1.55 |  |
|  | -0.86 |  |
| Age:LDL-Cholesterol | 1.11 \*\*\* |  |
|  | -0.21 |  |
| `log(Weight)` |  | 1.46 \*\* |
|  |  | -0.44 |
| Asian |  | 2.98 \*\*\* |
|  |  | -0.49 |
| `log(LDL-Cholesterol)` (mg/dL) | | `-1.00 \*\* |
|  |  | -0.34 |
| Total Cholesterol (mg/dL) | | 1.73 \*\*\* |
|  |  | -0.33 |
| SAD:Age | | `-1.01 \*\*\* |
|  |  | -0.16 |
| Age:Hypertension Medication | | `-2.88 \*\*\* |
|  |  | -0.48 |
| N | 4950 | 4950 |
| R2 | 0.33 | 0.21 |
| \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. | | |

**Figure A1:**

****

**Figure :**

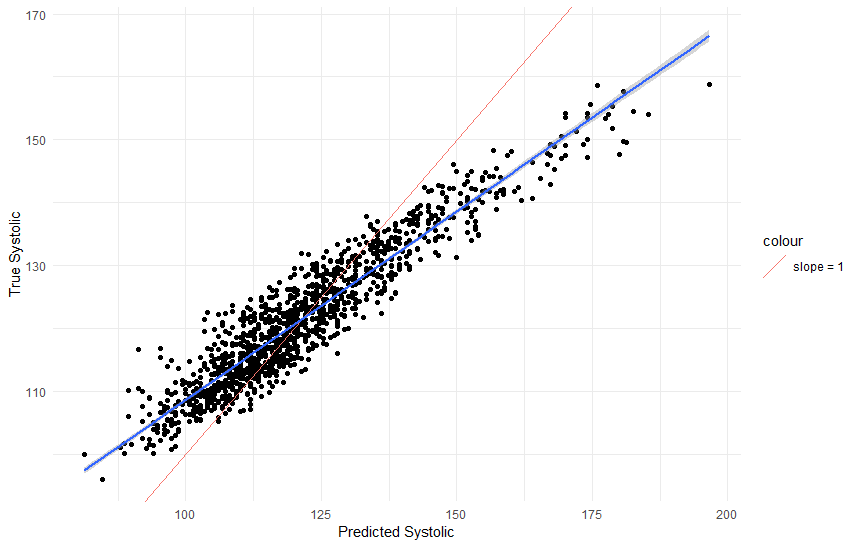
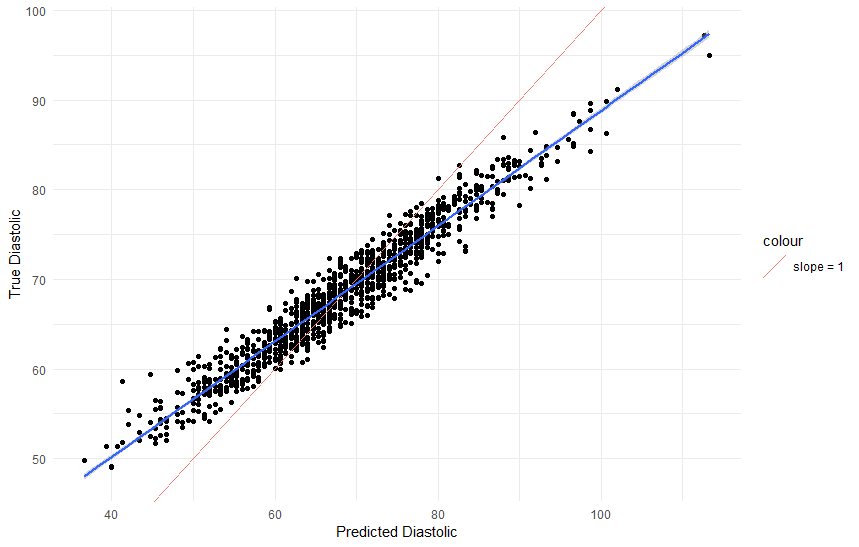
****

Figure :