Modeling the Presence of Brown Fat in Humans

April 12, 2021

Group Roles

Shivi Sharma, Exploratory Data Analysis

William Sinclair, R Markdown File

Afeef Mehta, Background and Significance Research

Wakil Ahmed, Model Selection

Kobe Louis, Discussion and Conclusion

Background and Significance

The human body can hold either white or brown adipose tissue, the latter usually going by brown fat. It's said that this special type of fat can activate and produce heat using its relatively large amount of mitochondria by burning calories in body fat through a process called thermogenesis to resist the potentially harmful metabolic effects of colder temperatures (Townsend, 2012). This ability to increase warmth and maintain body temperature would obviously be useful to animals susceptible to the cold, such as newborn humans and other small mammals. Furthermore, because brown fat processes calories quickly, it could be a possible solution to help treat obesity which itself influences other diseases such as diabetes and cancer. This calorie-burning occurs through using up blood sugar, lipids, and other fat or sugar molecules in the bloodstream which interfere and potentially impair various bodily processes (Townsend, 2012). Although other body fat may be able to provide some warmth to the body through insulation, brown fat is able to actively generate heat by itself through various cellular and metabolic processes.

The motivation for this paper is that it was found somewhat recently that traces of brown fat can be identified and become more prevalent when adult humans are exposed to colder temperatures. The objective is to analyze several factors in relation with the existence and total volume of brown fat to possibly determine an association between them. This is done in this study by building models to help estimate the probability of having brown fat and identify which factors are the most important to having brown fat. The significance of this study is that this research potentially provides insight into determining precursors or causal relationships with brown fat and cancer, allowing doctors to gain an even more complete understanding of the disease and give more accurate diagnoses. It also allows a clearer picture between brown fat and other important body measurements such as BMI, etc. The statistical analyses done on the data consists of fitting a logistic regression model to extract general associations from the data points and constructing graphs to better understand the variables, as well as more complicated model selection and goodness-of-fit testing such as the Hosmer-Lemeshow test.

Exploratory Data Analysis

The data analyzed in this case study comes from the Molecular Imaging Center at The University of Sherbrooke in Quebec, and is a sample of 4842 cancer patients with 20 parameters being considered, each of which are thought to be important to the presence of brown fat in the body. Below is a table providing an overview of the variables considered in this analysis.

```
variable_table = read_excel("variable_meanings.xlsx", col_names = TRUE,
    skip = 1)
```

```
variable_table[is.na(variable_table)] = ""
kable(variable_table[, 1:3])
```

Type of Variable	Variable	Meaning
Basic Background Information	Sex	Sex of the patient.
	Age	Age of the patient, given in years.
	Size	Height of the patient, given in cm.
	Weight	Weight of the patient, given in kg.
	BMI	The body mass index (weight to height ratio) of the patient.
	LBW	Lean body weight, weight of the components of the body that are not fat.
Time of Year	Day	Day of the month.
	Month	Month of the year.
	Season	The season of the year.
	Duration_Sunshine	The amount of time of which the sun is shining, given in minutes.
Temperature	Ext_Temp	The skin temperature of the patient.
	2D_Temp	Average temperature over the last 2 days.
	3D_Temp	Average temperature over the last 3 days.
	7D_Temp	Average temperature over the last week.
	1M_Temp	Average temperature over the last month.
Physiological Measurements	Diabetes	Whether or not the patient has diabetes.
	TSH	Level of thyroid stimulating hormone (NA if test not performed).
	Glycemy	Glycemia, the level of glucose in the blood.
Cancer	Cancer_Status	Whether or not the patient has cancer.
	Cancer_Type	What type of cancer the patient has.
Response Variable (Brown Fat)	BrownFat	Whether the patient has brown fat or not.
	Total_vol	Total volume of brown fat in the patient.

A preliminary data analysis was performed to determine which of the variables being considered are the most important to the presence of brown fat, and it was found that Sex, Age and Diabetes are the most significant predictors through Wald and likelihood ratio testing.

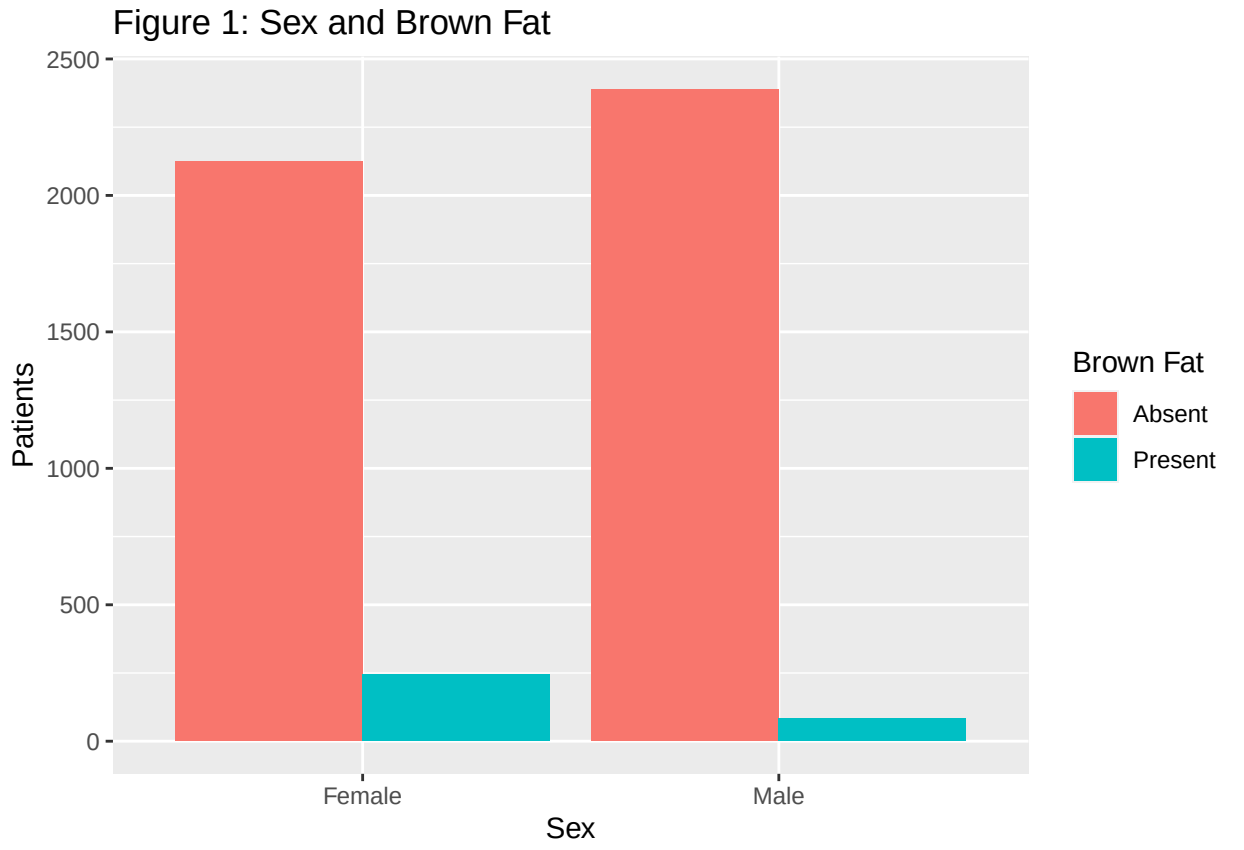
Also note that TSH (thyroid stimulating hormone) was excluded from this analysis due to the overwhelming amount of missing data. To include TSH, most of the patients' data would have to be disregarded, causing issues because brown fat is relatively rare in humans and so drastically reducing the sample size causes all of the other variables to artificially become insignificant predictors.

```
# fitting logistic regression models, one for BrownFat and one for
# Total_vol
presence_of_brownfat = glm(BrownFat ~ Sex + Age + Size + Weigth + BMI +
    LBW + Day + Month + Season + Duration_Sunshine + Ext_Temp + `2D_Temp` +
    `3D_Temp` + `7D_Temp` + `1M_Temp` + Diabetes + Glycemy + Cancer_Status +
    Cancer_Type, family = binomial, data = data)
```

To get a better understanding of the data, contingency tables were analyzed and several plots were made to visualize the data. Figure 1 below shows the relationship between sex and the presence of brown fat; both the contingency table and the graph shows that females appear to be more likely to have brown fat as compared to men. This is confirmed by calculating the odds ratio of the contingency table, which is 0.3013386. This

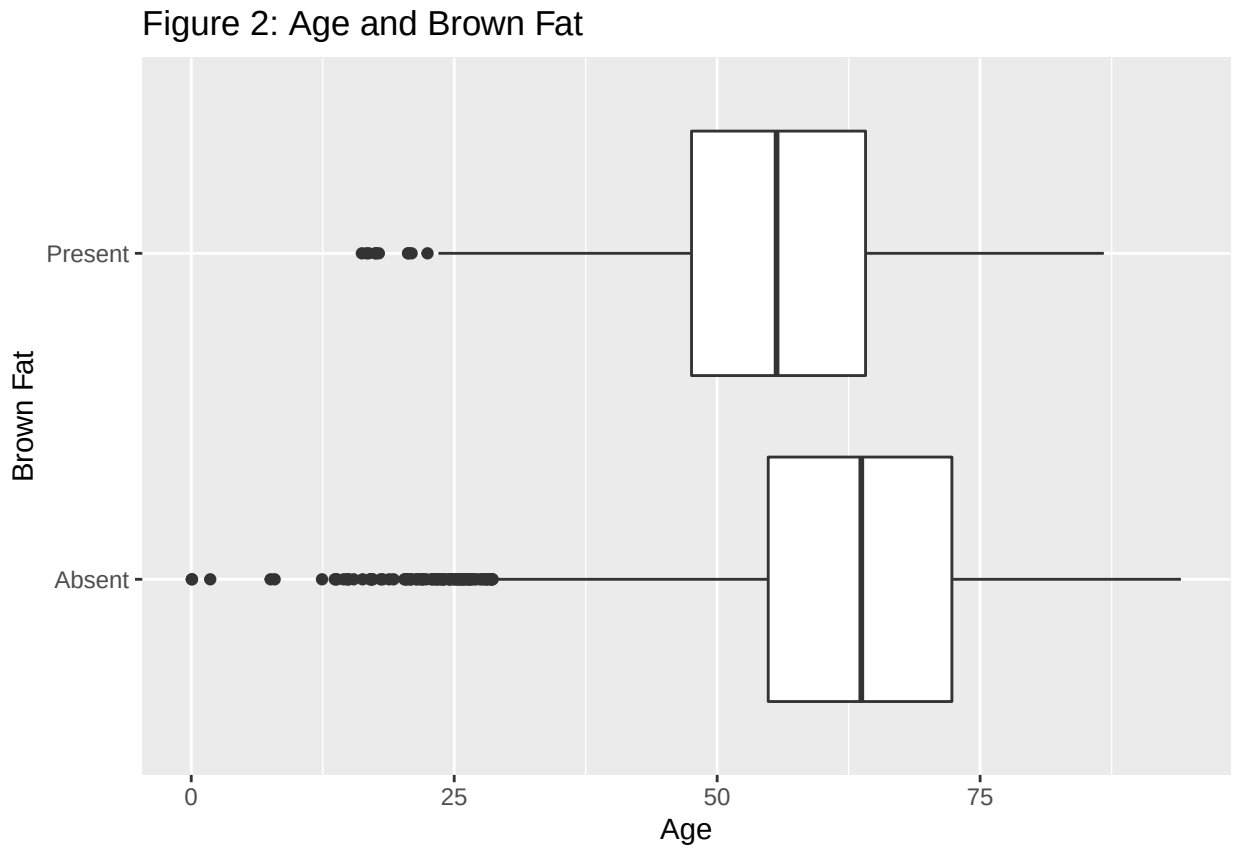
means the odds of not having brown fat for females is 0.3013386 times the odds of not having brown fat for males, or that the odds of having brown fat is higher for a female.

```
ggplot(data, aes(x = Sex, fill = BrownFat)) + geom_bar(position = "dodge") +
    scale_x_discrete(labels = c("Female", "Male")) + labs(y = "Patients") +
    scale_fill_discrete(name = "Brown Fat", labels = c("Absent", "Present")) +
    ggtitle("Figure 1: Sex and Brown Fat")
```



Below is Figure 2, a boxplot showing the relationship between age and the presence of brown fat. It shows that patients without brown fat are slightly older (around 62 compared to around 53) and that the variance between the two groups is comparable, but there are considerably more outliers for those without brown fat.

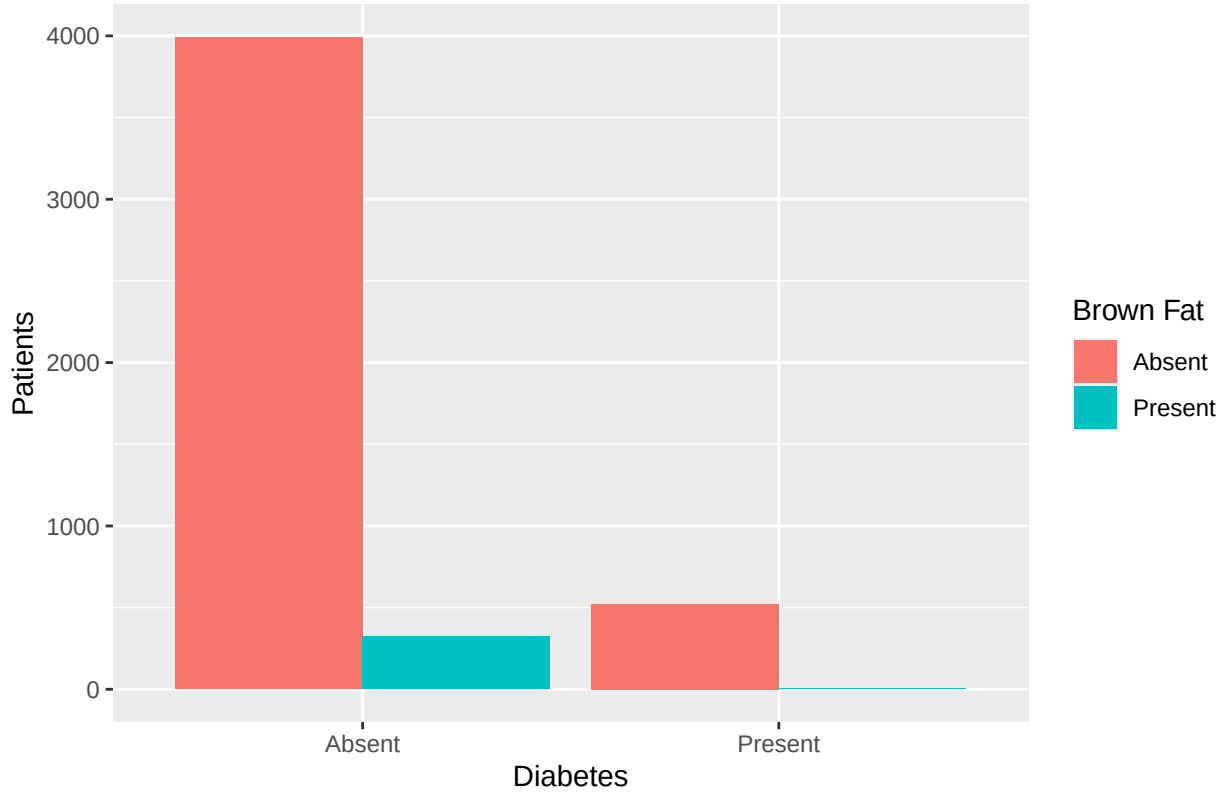
```
ggplot(data, aes(x = Age, y = BrownFat)) + geom_boxplot() + labs(y = "Brown Fat") +
    scale_y_discrete(labels = c("Absent", "Present")) + ggtitle("Figure 2: Age and Brown Fat")
```



Finally, Figure 3 displays the relationship between diabetes and the presence of brown fat; the graph and the contingency table indicate that brown fat is more prevalent in those patients without diabetes. This is supported by the odds ratio for the associated contingency table, which is 0.1421921; this means that the odds of not having brown fat for people without diabetes is 0.14 times the odds of not having brown fat for people with diabetes, or that people without diabetes are more likely to have brown fat.

```
ggplot(data, aes(x = Diabetes, fill = BrownFat)) + geom_bar(position = "dodge") +
    scale_x_discrete(labels = c("Absent", "Present")) + labs(y = "Patients") +
    scale_fill_discrete(name = "Brown Fat", labels = c("Absent", "Present")) +
    ggtitle("Figure 3: Diabetes and Brown Fat")
```

Figure 3: Diabetes and Brown Fat



Model Selection

The first step taken was to build a logistic regression model with the BrownFat variable as the response and all the predictors with no interaction terms. Then, all of the most significant terms from this model were taken and a new model was built with interaction terms between variables that seemed related. For example, height and weight are usually correlated, so it makes sense to include the interaction of these variables. With this many predictors the model is too complex, so feature selection is necessary. The primary goal is to select a model that is simple and somewhat useful. For this we use backwards elimination based on the chi squared test to reduce the model size and get rid of redundant predictors.

```
twoway <- glm(data = train, formula = BrownFat ~ `7D_Temp` * Ext_Temp +
    LBW * Sex * Weigth * BMI * Age + Diabetes + Cancer_Type, family = "binomial")
backwards <- glm(formula = BrownFat ~ `7D_Temp` + Ext_Temp + LBW + Sex +
    Weigth + BMI + Age + Diabetes + `7D_Temp`:Ext_Temp + LBW:Weigth + LBW:BMI +
    Weigth:BMI + LBW:Age + Weigth:Age + BMI:Age + LBW:Weigth:BMI + LBW:Weigth:Age +
    LBW:BMI:Age + Weigth:BMI:Age + LBW:Weigth:BMI:Age, family = "binomial",
    data = train)

summary(backwards)
```

```
##
## Call:
## glm(formula = BrownFat ~ `7D_Temp` + Ext_Temp + LBW + Sex + Weigth +
##     BMI + Age + Diabetes + `7D_Temp`:Ext_Temp + LBW:Weigth +
##     LBW:BMI + Weigth:BMI + LBW:Age + Weigth:Age + BMI:Age + LBW:Weigth:BMI +
##     LBW:Weigth:Age + LBW:BMI:Age + Weigth:BMI:Age + LBW:Weigth:BMI:Age,
##     family = "binomial", data = train)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1995  -0.4393  -0.2630  -0.1518   3.4911
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -3.287e+01  1.528e+01  -2.152  0.03143 *
## `7D_Temp`           8.579e-04  1.007e-02   0.085  0.93213
## Ext_Temp           -1.775e-02  8.891e-03  -1.997  0.04585 *
## LBW                 9.432e-01  3.580e-01   2.635  0.00842 **
## Sex1               -9.838e-01  3.317e-01  -2.966  0.00302 **
## Weigth             -6.910e-02  3.079e-01  -0.224  0.82246
## BMI                 1.911e+00  8.803e-01   2.171  0.02993 *
## Age                 5.077e-01  2.844e-01   1.785  0.07424 .
## Diabetes1          -1.368e+00  4.260e-01  -3.212  0.00132 **
## `7D_Temp`:Ext_Temp -1.160e-03  4.868e-04  -2.383  0.01719 *
## LBW:Weigth         -4.661e-03  4.400e-03  -1.059  0.28945
## LBW:BMI            -3.629e-02  1.774e-02  -2.046  0.04076 *
## Weigth:BMI         -1.383e-02  9.806e-03  -1.411  0.15837
## LBW:Age            -1.777e-02  7.126e-03  -2.493  0.01265 *
## Weigth:Age          6.209e-03  5.780e-03   1.074  0.28270
## BMI:Age            -3.859e-02  1.599e-02  -2.413  0.01581 *
## LBW:Weigth:BMI      3.277e-04  1.510e-04   2.170  0.02997 *
## LBW:Weigth:Age      7.869e-06  9.378e-05   0.084  0.93312
## LBW:BMI:Age         8.413e-04  3.443e-04   2.443  0.01456 *
## Weigth:BMI:Age      1.460e-04  1.847e-04   0.791  0.42918
## LBW:Weigth:BMI:Age -5.171e-06  3.171e-06  -1.631  0.10288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2247.7  on 4472  degrees of freedom
## Residual deviance: 1942.7  on 4452  degrees of freedom
## AIC: 1984.7
##
## Number of Fisher Scoring iterations: 8
```

It is important to check whether the selected model has a good fit. As we have a binary response variable with ungrouped data, the Hosmer-Lemeshow method makes sense. This method partitions observations into g equal sized groups according to their predicted probabilities, and then calculates a chi-square statistic from the observed and expected frequencies in each of the g quantiles.

```
hoslem.test(backwards$y, fitted(backwards), g = 13)
```

```
##
##  Hosmer and Lemeshow goodness of fit (GOF) test
##
## data:  backwards$y, fitted(backwards)
## X-squared = 12.158, df = 11, p-value = 0.3519
```

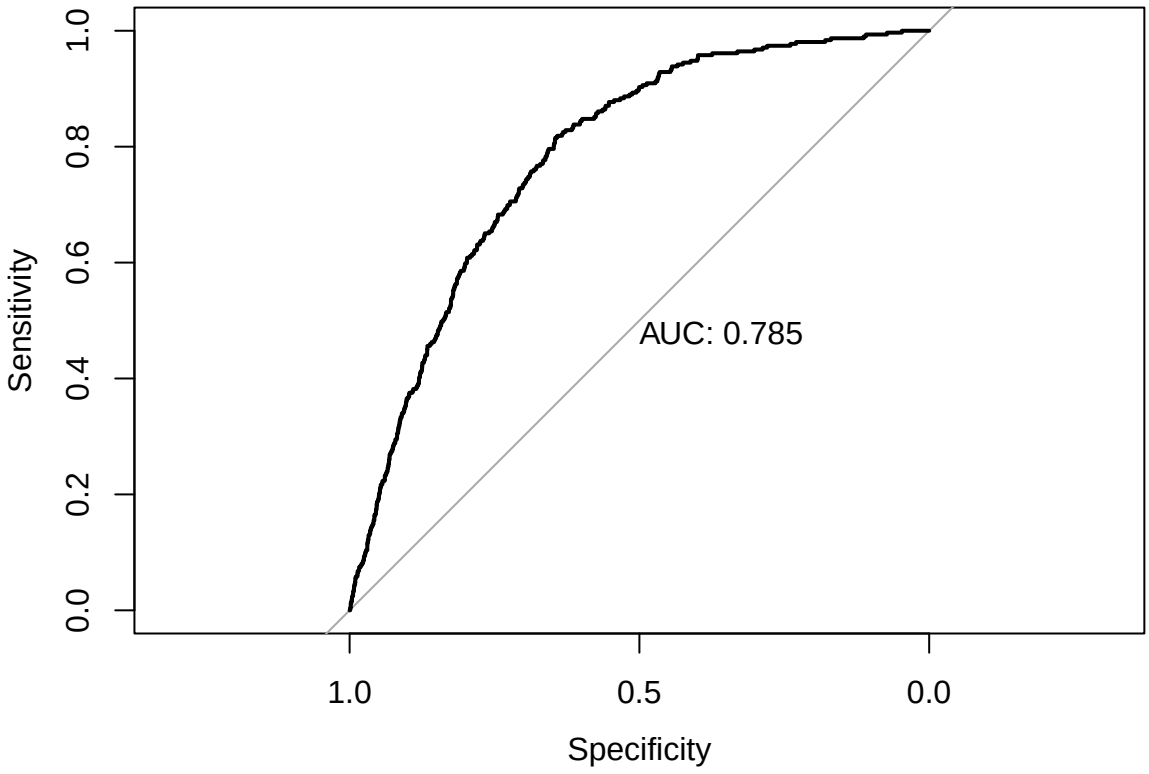
The null hypothesis here is that the observed frequencies are equal to the expected frequencies. Picking g can be arbitrary but it is recommended that g>p+1 where p is the number of covariates in the model. With g = 13 in this case, the p-value of this test is p = 0.3519 > 0.05, which means that the current model fits the data well. Plotting the pairs of sensitivity versus one minus specificity on a scatter plot provides a ROC

(receiver operating characteristic) curve.

```
test_ROC = roc(backwards$y ~ fitted(backwards), plot = TRUE, print.auc = TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```



The area under the curve (AUC) provides an overall measure of goodness of fit of the model by providing the probability that the positive class will have a higher fitted value which means a higher predicted probability. With our model we have AUC = 0.785, which indicates the model selected is moderate to good.

Conclusion

After analyzing all the factors in relation with the existence and total volume of brown fat, the variables external temperature, sex, BMI, and the presence of diabetes were discovered to be the most reliable predictors after discarding the TSH variable due to a lack of data. Interactions which were also reliable were between average seven day temperature and external temperature, LBW and BMI, LBW and age, BMI and age, LBW, weight, and BMI, and LBW, BMI, and Age. Additionally, the effects of some of these variables on the existence of brown fat were determined using boxplots and contingency tables. When it came to sex, men were three times more likely to not have brown fat than women. As for age, the average age for people with brown fat, 53, was 9 years younger than people without brown fat at 62. Finally, for the presence of diabetes, people with diabetes were 7 times more likely to not have brown fat than people with diabetes. These findings create the implication that brown fat occurs more commonly in healthier individuals, especially those that are female. Seeing as this is a topic that little is known about, such findings shed light on good predictors of people with brown fat. These findings also help researchers focus their studies of the topic to healthier individuals, saving time and money. Despite this, researchers should bear in mind limitations, such as that the sample consists of cancer patients and that this is only a single study, so these findings may not coincide

with brown fat findings on other populations.