Predicting Type II Diabetes with Early Symptoms

Rhayoung Park

11/15/20

Diabetes

Diabetes mellitus (hereafter, referred as diabetes) is one of the most largely suffered disease around the world. It can be classified into type I and type II. People with type I diabetes do not produce insulin. On the other hand, people with type II diabetes do produce insulin but do not respond to insulin. This study focuses on type II diabetes.

According to Olokoba et al., it is estimated that 522 million people would have type II diabetes by 2030 (2012). According to American Diabetes Association, the total healthcare cost related to just one specific disease accounts for 1.7% of the total GDP. It is also the 7th leading cause of death in the states, and more people die of diabetes than breast cancer and AIDS combined. USA is not the only country suffering from diabetes. A research conducted by Rahim et al., proved that people suffering from diabetes in Bangladesh increased by 6.8% in five years (2007). Although the population suffering from type II diabetes grows every day, people are still not aware of some early symptoms of diabetes. Understanding early symptoms can lead to early detection of diabetes. This is critical because it can lead to controlling the seriousness, save huge amount of hospital costs and in some cases prevent the disease from developing (Habib, Samira Humaira, et al., 2010). A research showed that one can save $6658 if they detect diabetes at an earlier stage (Habib, Samira Humaira, et al., 2010).

Some early symptoms such as obesity, polyuria, polydipsia and polyphagia are well-known (Olokoba et al., 2012) but other symptoms are easily overlooked or regarded as a symptom for other illness. Among all early symptoms, whether itching is associated with diabetes is still in debate. Here, itching refers to uncomfortable feeling in skin, eventually making people to scratch. Itching was first proposed as one of the early symptoms of diabetes in 1964 (Oakley, 1964) but some research argue that itching occurs after being treated with diabetes or as an allergic reaction to the treatment (Moyes, V., et al., 2006). This study attempts to focus on whether itching is associated with type II diabetes and also its interaction with other early symptoms using logistic regression. In other words, the questions of interest are (i) is there any significant early symptoms for predicting diabetes, excluding the symptoms obesity, polyuria, polydipsia and polyphagia, (ii) is itching a significant early symptom of diabetes, (iii) if yes, how does itching interact with other predictors.

EDA

The data was collected by asking direct questions to the patients of Sylhet Diabetes Hospital in Bangladesh and was approved by a doctor. It could be found in UCI Machine Learning Repository. The data is multivariate having 17 attributes and 520 observations in total. 200 people were not diagnosed with diabetes and 320 people had diabetes. The variable 'diabetes' was used as the response variable and was predicted by predictors such as 'age', 'gender', 'polyuria', 'polydipsia', 'sudden weight loss', 'weakness', 'polyphagia', 'genital thrush', 'visual blurring', 'itching', 'irritability', 'delayed healing', 'partial paresis', 'muscle stiffness', 'alopecia' and 'obesity'. All other predictors except age was factored before starting the analysis. There were no missing values in the data. (Appendix A. Codebook)

A histogram of age, the only numeric variable in the data, was plotted to check the overall distribution. Age ranged from 20 to 65 years, and the distribution resembled a normal distribution. The risk of diabetes seemed to be more visible if age was grouped into categories. Thus, age was grouped into 6 categories. (Appendix B)

A chi-squared test was conducted to find out whether each categorical variable is significant. The variables that were not significant at a 0.05 level were 'itching' and 'delayed healing', and all other variable had chi-squared p-values smaller than 0.05.

Moreover, joint tables were made to check if there are enough data points for each category and also interactions. As shown in the table below, there were not enough data for age under 31, 61-70 and over 71. For all other variables, there were enough data.

Predictors	P-value	Diabetes Negative		Diabetes Positive	
Gender	2.2e-16	Female: 0.09	Male: 0.91	Female: 0.54	Male: 0.46
Polyuria	2.2e-16	Yes: 0.08	No: 0.93	Yes: 0.76	No: 0.24
Polydipsia	2.2e-16	Yes: 0.04	No: 0.96	Yes: 0.70	No: 0.30
Sudden Weight Loss	2.2e-16	Yes: 0.15	No: 0.86	Yes: 0.59	No: 0.41
Weakness	4.9e-08	Yes: 0.44	No: 0.57	Yes: 0.68	No: 0.32
Polyphagia	1.2e-14	Yes: 0.24	No: 0.76	Yes: 0.59	No: 0.41
Genital Thrush	0.01	Yes: 0.17	No: 0.84	Yes: 0.26	No: 0.74
Visual Blurring	1.7e-08	Yes: 0.29	No: 0.71	Yes: 0.55	No: 0.45
Itching	0.83	Yes: 0.50	No: 0.50	Yes: 0.48	No: 0.52
Irritability	1.7e-11	Yes: 0.08	No: 0.92	Yes: 0.34	No: 0.66
Delayed healing	0.33	Yes: 0.43	No: 0.57	Yes: 0.48	No: 0.52
Partial Paresis	2.2e-16	Yes: 0.16	No: 0.84	Yes: 0.60	No: 0.40
Muscle Stiffness	0.006	Yes: 0.30	No: 0.70	Yes: 0.58	No: 0.42
Alopecia	1.9e-09	Yes: 0.50	No: 0.50	Yes: 0.24	No: 0.76
Obesity	0.13	Yes: 0.14	No: 0.87	Yes: 0.19	No: 0.81

Figure 1: Summary of Chi-Squared Test

	Below 31	Age 31 - 40	Age 41 - 50	Age 51 - 60	Age 61 - 70	Over 71
Has developed diabetes	15	84	87	78	49	7
Has not developed diabetes	30	87	58	49	17	7

Figure 2: Binned Age

Model

Logistic regression was built with all predictors and interactions with itching. An AIC stepwise model selection was done to identify the optimal combination of predictors. As this analysis specifically aims to look at 'itching', the null model included 'itching', and the full model included interactions with itching. Since age did not have sufficient data, interaction between itching and age was excluded. As a result of AIC stepwise selection, interaction between 'itching' and 'visual blurring' was kept and all other interactions were dropped.

For model assessment, binned residual plot were drawn and VIF values were computed to check multicollinearity. In this process, it turned out that binned age had high VIF values of 9.3. Age was then centered to reduce multicollinearity and the final model was built with centering age. Binned residual plots with age centered was similar to the residual plot with binned age. In conclusion, the final model is:

$$logit\hat{y} = \hat{\beta}_0 + \hat{\beta}_a x_{itching} + \hat{\beta}_b x_{polydipsia} + \hat{\beta}_c x_{polyuria} + \hat{\beta}_d x_{gender} + \hat{\beta}_e x_{irritability} + \hat{\beta}_f x_{genitalthrush} + \hat{\beta}_g x_{partialparesis} + \hat{\beta}_h x_{polyphagia} + \hat{\beta}_i x_{agecentered} + \hat{\beta}_j x_{visualblurring} + \hat{\beta}_k x_{itching:visualblurring}$$

Figure 3: Final model

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.2362	0.5012	0.4713	0.6374
itchingYes	-1.752	0.6961	-2.516	0.01186
polydipsiaYes	5.418	0.7716	7.021	2.202e-12
polyuriaYes	4.287	0.6361	6.74	1.581e-11
genderMale	-4.731	0.627	-7.545	4.507e-14
irritabilityYes	2.662	0.5666	4.699	2.619e-06
genitalthrushYes	1.733	0.5645	3.07	0.002138
partialparesisYes	1.151	0.4827	2.386	0.01705
polyphagiaYes	1.235	0.4971	2.484	0.01301
age_centered	-0.07122	0.02345	-3.037	0.002386

	Estimate	Std. Error	z value	Pr(>|z|)
visualblurringYes	2.832	0.9643	2.937	0.003318
itchingYes:visualblurringYes	-3.012	1.095	-2.752	0.005928

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	692.9 on 519 degrees of freedom
Residual deviance:	170.2 on 508 degrees of freedom

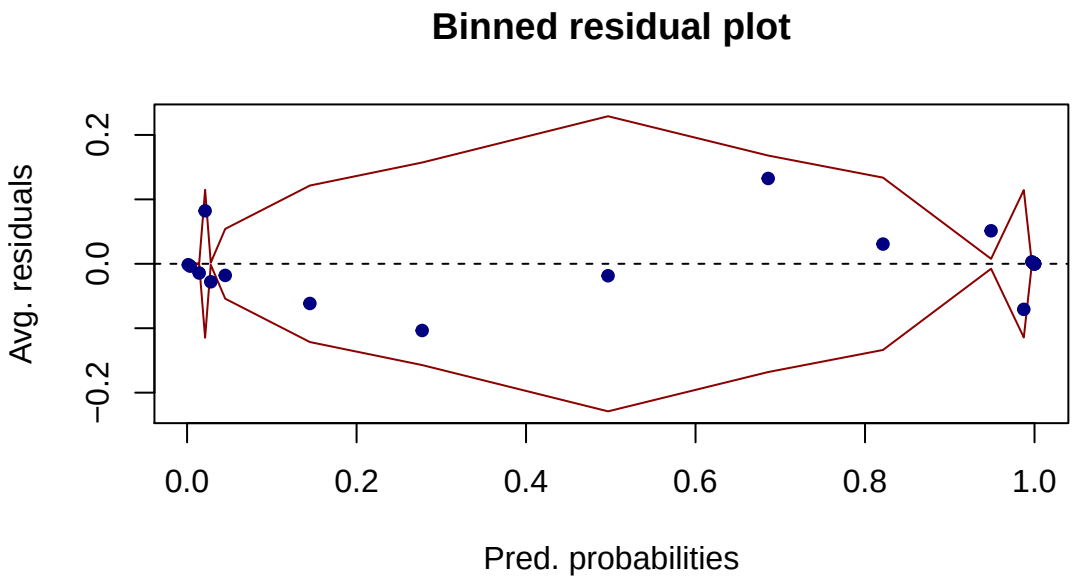
The intercept in this model represents a female who is 48 years old and does not have symptoms of itching, polydipsia, polyuria, polyphagia, irritability, genital thrush, partial paresis and visual blurring. Itching was a significant predictor of diabetes. Holding everything else constant, people who experienced itching has 83% less odds of having diabetes, compared to people who did not have itching. If they had experienced both itching and visual blurring, the odds of having diabetes were 99% less than people who experienced only itching.

Polydipsia, polyphagia and polyuria were significant predictors of type II diabetes, which coincides with earlier research. This finding is supported by the research conducted by Pawar et al., which demonstrated that these three symptoms are significant predictor of diabetes as all of them have p-vale less than 0.0001 (2017).

Sex was a significant predictor in this model. Holding everything else constant, if a person is male, then the odds ratio of him being diagnosed with diabetes is 99% less compared to a female. This result contradicts the research conducted by Nordström et al., which concluded that men were twice more likely to have type II diabetes compared to women (2016). One of the reasons for the discrepancy could be attributed to the difference in demographics of the participants. In the research done by Nordström et al., the participants were all 70 years old northern Europeans whereas this analysis includes participants from 20 years old to 65 years old Bangladeshis. Moreover, Nordström et al. observed that higher prevalence of type II diabetes in male are related to visceral fat (2016), a predictor that was not included in this analysis.

Holding everything else constant, if a person has genital thrush, he/she has 464% more odds of developing diabetes compared to a person who does not have genital thrush. If a person has partial paresis, he/she has 216% more odds of developing diabetes compared to somebody without partial paresis. If a person experience irritability, he/she will have 1388% more odds of having diabetes compared to a person who didn't have irritability. If a 48 years old person gets one year older, then the odds of having diabetes decreases by 6.7%. Holding everything else constant, if a person has visual blurring, he/she has 1598% more odds of developing diabetes compared to someone without visual blurring.

Unfortunately, binned residual plot showed specific trend. As seen below, there is a downward trend and then an upward trend and a downward trend again. Three residuals were outside the significance line, and five were right on top of the significance line. It could be said that there are a few outliers in the model.



VIF table for the final model is shown below. The VIF value for visual blurring and the interaction between itching and visual blurring is high. Interactions that include categorical variables are correlated in nature and thus it might have relatively high VIF scores compared to other predictors. Although there were some predictors that was notable, most of the predictor's VIF value was below 10. Therefore, it was kept in the model.

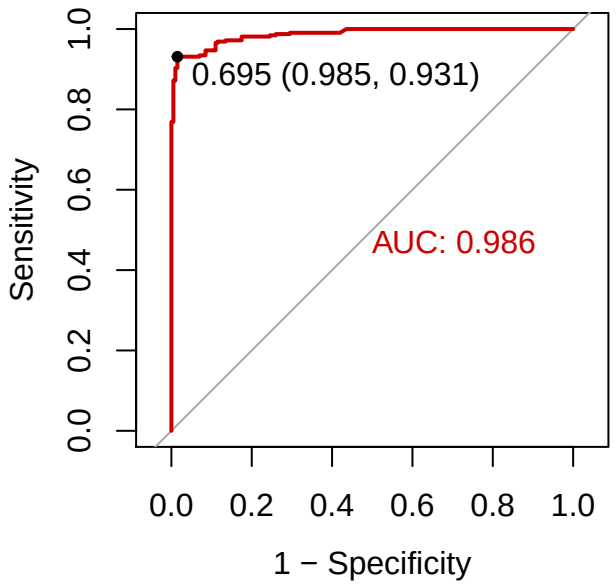
	VIF		VIF
Itching	3.16	Has genital thrush	1.61
Male	2.10	Has partial paresis	1.24
Has polydipsia	1.92	Age (centered)	1.99
Has polyuria	2.07	Has visual blurring	5.81
Has polyphagia	1.44	Has itching and visual blurring	6.84
Has irritability	1.42		

Figure 4: VIF

Moreover, ROC curve was plotted to check the performance of the final model. The AUC value was 0.986. Considering that AUC value of 1 means that the model predicts diabetes correctly 100% of the time, our final model's AUC value is very high. Above table shows the confusion matrix with accuracy, sensitivity and specificity values. All three values were very high as it is close to 1. Accuracy indicate the overall performance of the model. From the value of 0.92, one can say that the model's performance is good. Sensitivity of 0.90 indicates that the model predicts 90% the people who will have diabetes but generates false-negatives 10% of the time. Specificity of 0.94 means that 94% of the time, the model will be correct in predicting people who will not have diabetes.

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

## Setting direction: controls < cases
```




```
##
## Call:
## roc.default(response = dt$diabetes, predictor = fitted(again),     plot = T, print.thres = "best", l
##
## Data: fitted(again) in 200 controls (dt$diabetes 0) < 320 cases (dt$diabetes 1).
## Area under the curve: 0.9865
```

Metric	Values
Accuracy	0.92
Sensitivity	0.90
Specificity	0.94

Figure 5: Summary

Conclusion

There are also some limitations to this analysis. As demonstrated above, binned residual plots did not show randomness. Also, potentially significant variables such as BMI index, physical activity and socioeconomic status was not provided in the data. With these predictors, the model could have been more powerful in predicting diabetes.

In conclusion, itching was a significant predictor for diabetes. People who experienced itching has 83% less odds of having diabetes, compared to people who did not have itching. Other significant predictors were polydipsia, polyuria, gender, irritability, genital thrush, partial paresis, polyphagia, age, visual blurring. Also, the interaction between itching and visual blurring was statistically significant. It was observed that the odds of developing diabetes were 99% lower for people who had both visual blurring and itching compared to people who only had itching, holding everything else constant.

Reference

Anna Nordström, Jenny Hadrévi, Tommy Olsson, Paul W. Franks, Peter Nordström, Higher Prevalence of Type 2 Diabetes in Men Than in Women Is Associated With Differences in Visceral Fat Mass, The Journal of Clinical Endocrinology & Metabolism, Volume 101, Issue 10, 1 October 2016, Pages 3740–37 46.

Maria O. Christensen, Anna K. Berg, Karen Rytter, Eva Hommel, Jacob P. Thyssen, Jannet Svensson, and Kirsten Nørgaard.Diabetes Technology & Therapeutics.Apr 2019.215-221.

Olokoba, Abdulfatai B et al. "Type 2 diabetes mellitus: a review of current trends." Oman medical journal vol. 27,4 (2012): 269-73. doi:10.5001/omj.2012.68

Oakley, Wilfrid. "Itching and Diabetes." British Medical Journal vol. 2,5412 (1964): 818.

Boyle JP, Honeycutt AA, Narayan KM. et al. Projection of diabetes burden through 2050: impact of changing demography and disease prevalence in the US. Diabetes Care.2001;24:1936-1940

Habib, Samira Humaira, et al. "Cost-effectiveness analysis of medical intervention in patients with early detection of diabetic foot in a tertiary care hospital in Bangladesh." Journal of Diabetes and its Complications 24.4 (2010): 259-264.

Moyes, V., et al. "Insulin allergy in a patient with type 2 diabetes successfully treated with continuous subcutaneous insulin infusion." Diabetic medicine 23.2 (2006): 204-206.

Pawar, Shivshakti D., et al. "The accuracy of polyuria, polydipsia, polyphagia, and Indian Diabetes Risk Score in adults screened for diabetes mellitus type-II." Medical Journal of Dr. DY Patil University 10.3 (2017): 263.

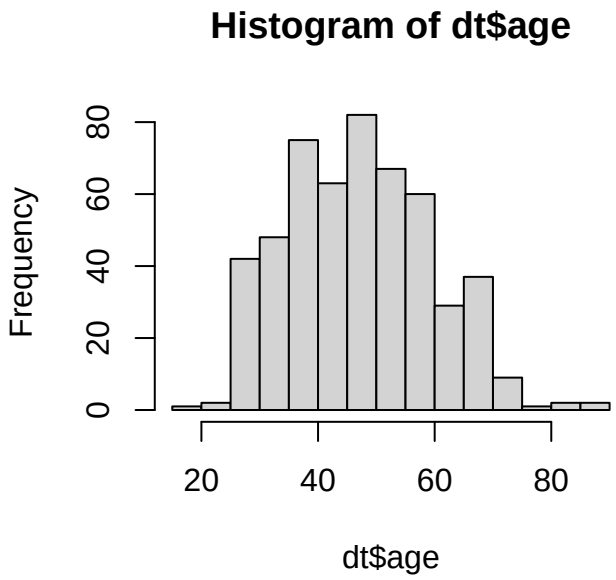
Rahim, M. A., et al. "Rising prevalence of type 2 diabetes in rural Bangladesh: a population based study." Diabetes research and clinical practice 77.2 (2007): 300-305.

Appendix A. Codebook

	Attributes	Value		Attributes	Value
1	Age	20 – 65 years	10	Itching	Yes:1, No:0
2	Gender	Male:1, Female:0	11	Irritability	Yes:1, No:0
3	Polyuria	Yes:1, No:0	12	Delayed healing	Yes:1, No:0
4	Polydipsia	Yes:1, No:0	13	Partial paresis	Yes:1, No:0
5	Sudden weight loss	Yes:1, No:0	14	Muscle stiffness	Yes:1, No:0
6	Weakness	Yes:1, No:0	15	Alopecia	Yes:1, No:0
7	Polyphagia	Yes:1, No:0	16	Obesity	Yes:1, No:0
8	Genital thrush	Yes:1, No:0	17	Diabetes (response)	Positive:1, Negative:0
9	Visual blurring	Yes:1, No:0			

Figure 6: Summary

Appendix B. Histogram for Age


Histogram of dt$age
Frequency
0
20
40
60
80
20
40
60
80
dt$age