Leveraging biometric data to create a digital biomarker for memory task performance

Report by: Kevin Schubert, Zoila Rossi, Amber Jolly, Gavincko Alagesan

ABSTRACT

The goal of the present study is to combine 28 sessions of a subject's biometric and activity data from five timepoints with continuously measured Hexoskin physiological data over a 3-day span to create statistical models (or digital biomarkers) that predict performance on three cognitive memory tasks: Immediate Recognition (IR), Delayed Recognition (DR) and Stroop. Raw physiological and activity data (re-coded by immediate and recent action) was timestamped to coincide with the five timepoints where cognitive task performance was assessed. Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis was used to train three models for predicting each cognitive measure and tuned to minimize mean-squared error (MSE). The recent actions of nothing, reading, and watching TV, as well as the immediate action of intaking sugar (relative to exercise) were significant predictors of Stroop score. The immediate actions of reading and eating sugar and recent actions of reading and nothing (relative to exercise) were significant predictors of IR. The recent actions of nothing, sugar intake, reading and TV (relative to exercise) were significant predictors of DR.

METHODS

Dataset Overview:

Biometric and cognitive performance data was collected over 28 testing sessions from a single individual. In a single 3-day testing session, the subject's cognitive task performance and physiological condition was evaluated at 5 timepoints (in addition to a Hexoskin Smart Shirt continuously monitoring their heart rate, breathing rate and cadence): 11pm on Day 1, 9am, 3pm and 11pm on Day 2 and 11pm on Day 3. Immediately before the three Day 2 measurements, the subject completed one of three possible task permutations consisting of TV or reading, aerobic exercise or no exercise, sugar or no sugar. Taken together, the potential biometric/activity predictors of interest in this study are the activity permutation number, the specific activity performed, systolic (BPS) and diastolic (BPD) blood pressure in mmHg, blood glucose (BG) in mmol/L, heart rate (HR) in beats/minute, breathing rate (BR) in respirations/minute and cadence in strides/minute. These features will be used to predict performance on three memory tasks: IR, DR and Stroop. We will make the assumption that each cognitive task datapoint represents a performance score where a higher score indicates better performance.

Data Cleaning/Pre-Processing:

To create an accurate prediction model, raw measurement data (ie. heart rate, breathing rate, and cadence) from the Hexoskin (sessions 2 to 29) were considered in tandem with the scores and measurements taken at the five designated time points. Within the individual raw data files, timestamps were converted appropriately into date-time format. Cognition scores from the IR, DR, and Stroop tests were manually inputted into the dataframe based on the session, day (ie. 1-3), and exact time. Based on the permutation to be done, as well as the time, new variables were created to show immediate actions and most recent actions. Immediate action displays which action should be occurring based on the permutation. Permutation actions, such as reading or watching TV, are assumed to take a minimum of 30 minutes and a maximum of 60 minutes. Outside of these timeframes, it is assumed that nothing is being done according to the permutation schedule. Most recent action displays the latest permutation action that has been performed but does not consider whether the participant is doing nothing according to the schedule. For example, if the last permutation action was sugar intake, but the participant is currently doing nothing, then the most recent action recorded is sugar intake.

Fitting Models for Predicting Cognitive Performance Measures:

Due to differences in how Braincheck cognitive test scores are processed, as well as the cognitive function they measure (ie. IR/DR for memory, Stroop for executive function), it was deemed most appropriate to create three separate prediction models to determine IR, DR, and Stroop scores. The weighting of these scores relative to one another is unknown; thus there is no particular method to validate a composite score. For convenience, complete-case analysis was used to deal with missing values.

To create each predictive model, we leveraged the machine learning-based regression analysis method LASSO . LASSO shrinks the magnitude of predictor weights (coefficients) towards 0 by penalizing the model (Tibshirani, 2011). This is done to restrict how close the prediction model will be to the training set and thus, provides a more accurate prediction over other regression methods and automates parts of predictor selection (non-relevant predictors are eliminated from the model if their weight is shrunk to 0). For each prediction model, the LASSO - a class of different models based on different lambda values - was first trained using the biometric measurements from the Hexoskin and manually input recent/immediate activities and physiological measures. 10 fold cross-validation was performed to investigate how the model's prediction performance changed with varying values of lambda. The optimally performing prediction model within the class of LASSO models was chosen for each cognitive task based on the model whose lambda value minimized the cross-validation MSE. The cross-validation MSE computes the difference between the observed and predicted cognitive task score values, and thus was used as a measure of the model's performance accuracy. The predictor weights were subsequently estimated for each of the three cognitive task's optimal models.

RESULTS

Stroop Prediction Model:

10-fold cross-validation determined the optimal LASSO model for Stroop score prediction had a minimal MSE of 52.17 and a lambda coefficient of 1.00×10^{-3}. According to the results from the Stroop score prediction model (**Table 1**), the mean score when all predictors are zero (or categorical variables are at baseline) is approximately 125.48. In other words, this is the mean score we would expect to obtain when all numerical predictors are at zero, the permutation schedule is permutation one, and both the immediate and/or recent actions are exercise. While no predictors were omitted from the model, there are certainly some actions which have greater influence over the Stroop test score. The items with the most influence on the Stroop score included the recent actions of nothing, reading, and watching TV, as well as the immediate action of intaking sugar. The recent action of reading had the greatest positive influence on output score, with a weight of 8.22. In other words, the participant is expected to obtain a score that is 8.22 points higher than when their most recent action was exercising. The recent actions of nothing and watching TV also had a positive influence on cognition, with weights of 4.41 and 3.43 respectively. (**Table 2**) Interestingly, permutation two (ie. reading, aerobic exercise, and sugar intake) as a whole had a negative influence on Stroop test score, reducing the score by nearly two points compared to permutation one. The remainder of the weights left in the model were deemed to have negligible influence and are not discussed further.

Immediate Recognition (IR) Prediction Model:

10-fold cross-validation determined the optimal LASSO model for IR score prediction had a minimal MSE of 172.2 and a lambda coefficient of 6.0×10^{-4}. The predicted mean IR test score when all numerical predictors were 0 and categorical variables were at baseline was 98.048 points **(Table 3)**. Therefore, when a subject has a BPS, BPD, HR, BR and cadence of 0, they completed activity permutation one and their immediate and most recent actions were both aerobic exercise, the model predicts their mean IR score to be 98.408. The only predictor omitted due to shrinkage was the recent action of consuming sugar. All other predictors were retained in the optimal model. The optimal prediction model determined that the immediate actions of reading and eating sugar (relative to aerobic exercise) and recent actions of reading and nothing (relative to aerobic exercise) had the largest weighted effect on predicting mean IR score. The model predicts that a subject who immediately reads (prior to an IR test) will have a mean 2.472 point lower score and those who immediately eat sugar will have a mean 4.470 point lower score than a person who performs aerobic exercise immediately before the test The model also predicts that a subject who recently read will have a mean 3.159 point higher score and a subject who recently did nothing will have a mean 2.179 point higher score on the IR test than a subject who recently did exercise or recently ate sugar **(Table 4)**. The remainder of the weights left in the model were deemed to have negligible influence and are not discussed further.

Delayed Recognition (DR) Prediction Model:

10-fold cross-validation determined the optimal LASSO model for DR score prediction had a minimal MSE of 300.3 and a lambda coefficient of 0.05. According to the results from the DR score prediction model (**Table 5**), the mean score when all predictors are zero (or categorical variables are at baseline) is approximately 67.60. This is the mean score we would expect to obtain when all numerical predictors are at zero, the permutation schedule is permutation one, and both the immediate and/or recent actions are exercise. Overall, recent actions appear to have influence on DR score, as well as specific permutations. Recent actions of nothing, sugar intake, and TV have a positive influence on DR score with respective weights of 4.98, 2.62, and 5.52. Conversely, the recent action of reading significantly reduces the expected score by a mean of 7.74 points compared to having a recent action of exercise (**Table 6**). Overall permutations also had a slight effect on the score outcome relative to permutation one. Specifically, permutation two which involves reading, exercise and sugar reduced the score by a mean 4.32 points. Permutation three, which involves TV watching, no exercise and no sugar, increased the expected score by two points in comparison to permutation one. The other weights left in the model were deemed to have negligible influence and are not discussed further.

DISCUSSION

The Stroop prediction model revealed that recent actions of doing nothing, reading, and watching TV had influence over the final output. The lasting effect of reading had the greatest influence on predicted Stroop score (Weight = 8.22). The Stroop test is a task which measures a subject's executive function, which includes items like processing speed and attention to detail. Depending on the nature of the readings, the ability to decode underlying meanings and flexibility in thinking may have been improved when the subject took the cognitive function test.

The immediate and delayed recognition tests measure memory function. Immediately consumption of sugar had the greatest negative influence on predicted IR score (Weight = 1.22). Previous studies have found a link between impaired glucose regulation in Type II diabetes patients and cognitive dysfunction (Gluck et al., 2014), suggesting a reason for why heightened glucose impaired cognition in the subject. Interestingly, immediately reading prior to the DR test had the greatest negative impact on predicted test score (W = 7.74 points). This may perhaps be partly due to the primacy effect, where someone is able to better remember the first items they were presented (i.e words they read) than more recent items (i.e words part of the DR test).

The study included several limitations. 'Important' predictors were selected arbitrarily, as no context around the meaning of the BrainCheck cognitive scores was provided and so, the minimally important differences are unknown. A further constraint was the inability to create a single model due to lack of context of how the scores are calculated and weighted relative to one another. As such, a composite score could not be produced nor validated. Inaccuracies may also arise in the predictive models as permutation activities can take anywhere between 30 to 60 minutes to complete; there is no way to tell how long the activity was performed for, and whether the time spent on the activity influenced the final score. Next steps for this analysis include investigating the calculation method of the BrainCheck scores such that a composite score for cognitive function can be computed and a single prediction model for cognitive function can be trained.

REFERENCES

Gluck, M. E., Ziker, C., Schwegler, M., Thearle, M., Votruba, S. B., & Krakoff, J. (2013). Impaired glucose regulation is associated with poorer performance on the Stroop Task. Physiology & behavior, 122, 113–119. https://doi.org/10.1016/j.physbeh.2013.09.001

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. Royal statistical society, 73, 273–282. https://doi.org/10.1111/j.1467-9868.2011.00771.x

APPENDIX:

Lambda (min MSE)	MSE	SE
0.001061	52.17	0.04291

Table 1: Lambda value for minimum mean square error of the Stroop model

Predictors	**Weights**
Intercept	125.4797
Permutation-2	-1.8809
Permutation-3	-0.2786
Immediate_action-Nothing	-0.4183
Immediate_action-Reading	-0.9857
Immediate_action-Sugar	5.7563
Immediate_action-TV	0.4997
Recent_action-Nothing	4.9768
Recent_action-Reading	8.2176
Recent_action-Sugar	0.1493
Recent_action-TV	3.4306
Blood Pressure-Systolic (BPS)	-0.0019
Blood Pressure-Diastolic (BPD)	-0.1306
Heart Rate (HR)	-0.0091
Breathing Rate	-0.1093
Cadence	0.0184

Table 2: Stroop score model weights

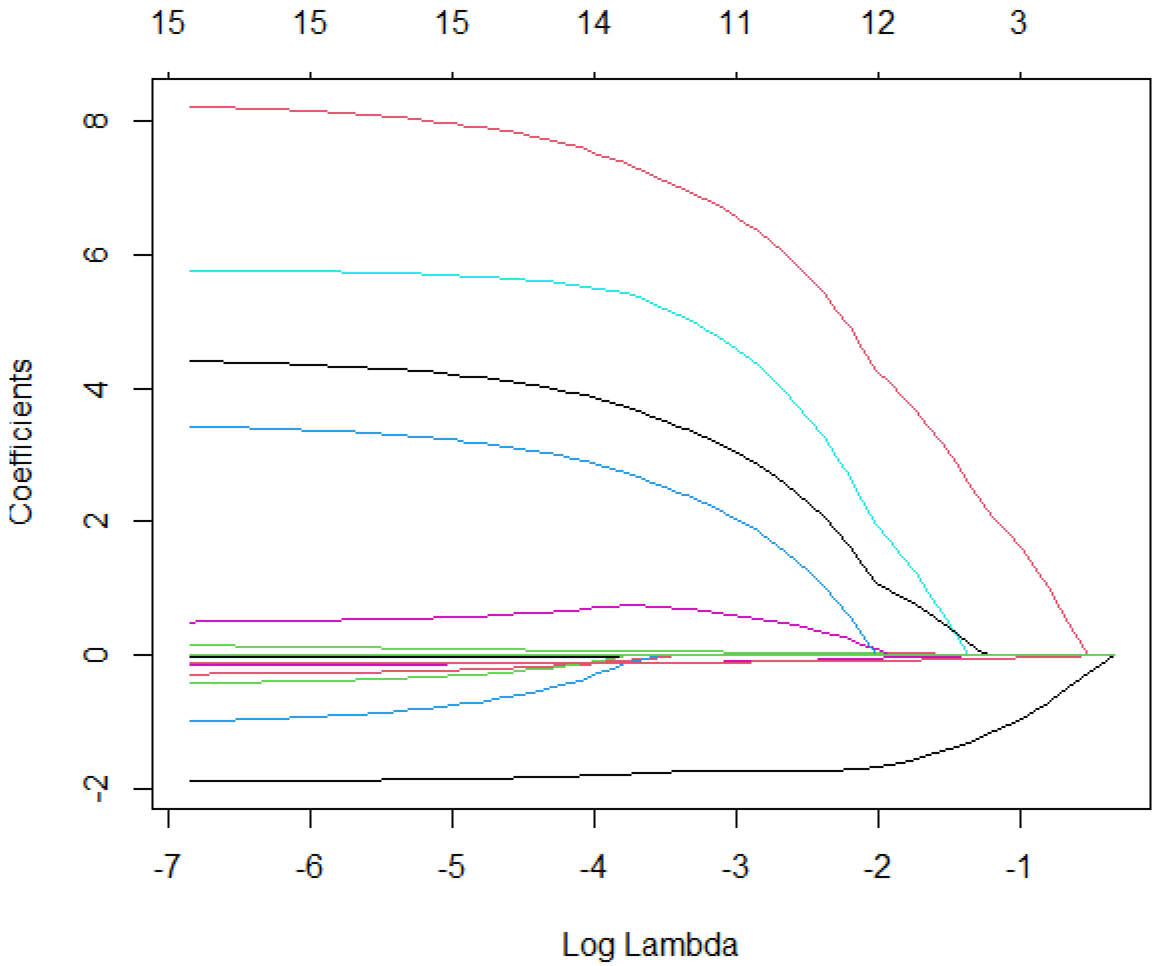

Figure 1: Weights of coefficients for various lambda of Stroop training model

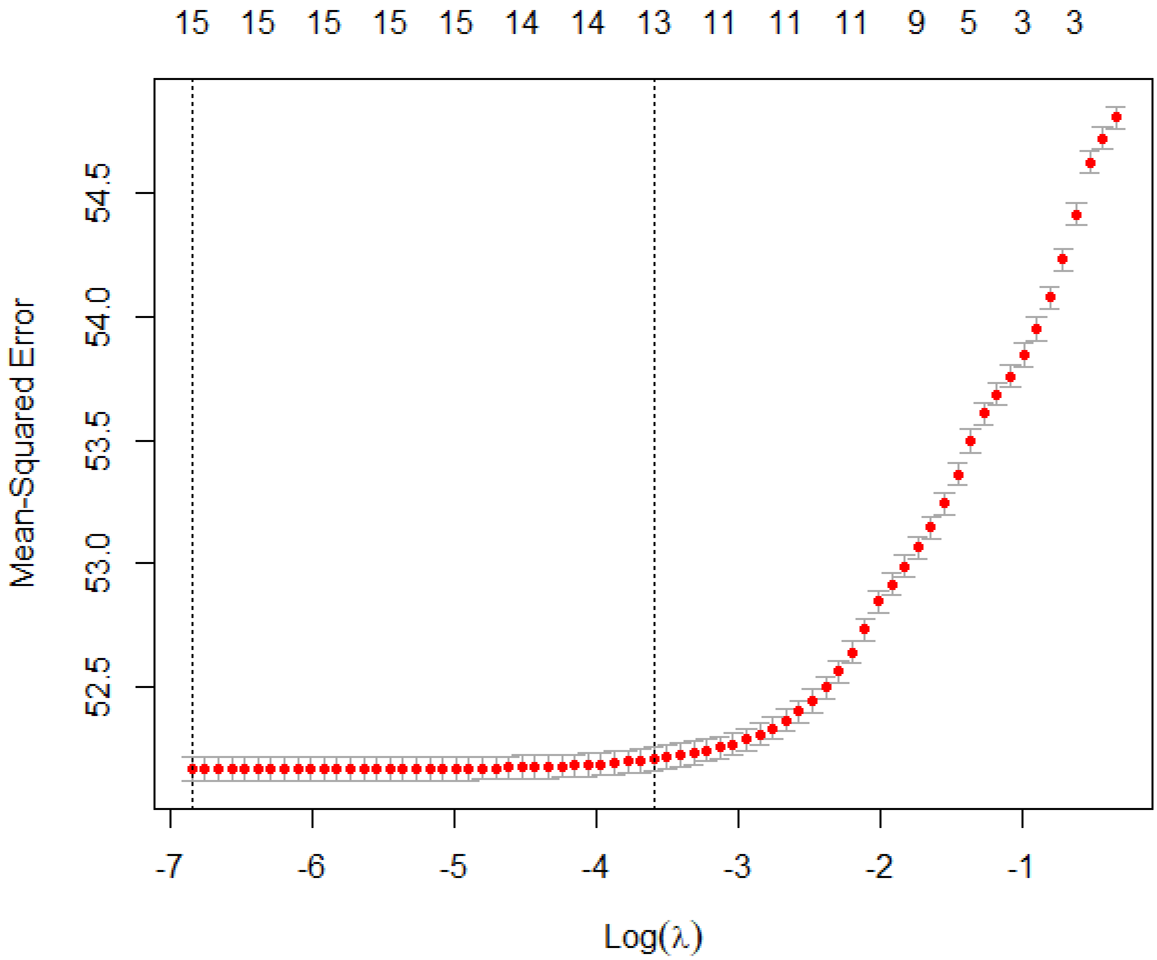

Figure 2: Mean Square Error for various lambda values of Cross-Validation of Stroop model

Lambda (min MSE)	MSE	SE
0.00064	172.2	0.1125

Table 3: Lambda value for minimum mean square error for the IR model

Predictors	**Weights**
Intercept	98.0480
Permutation-2	-0.4110
Permutation-3	-0.8223
Immediate_action-Nothing	-0.1915
Immediate_action-Reading	-2.4721
Immediate_action-Sugar	-4.4700
Immediate_action-TV	-1.0606
Recent_action-Nothing	2.1791
Recent_action-Reading	3.1592
Recent_action-Sugar	No effect
Recent_action-TV	1.1087
Blood Pressure-Systolic (BPS)	-0.0048
Blood Pressure-Diastolic (BPD)	-0.0727
Heart Rate (HR)	0.0033
Breathing Rate	-0.1155
Cadence	0.0155

Table 4: Immediate Recognition (IR) model weights

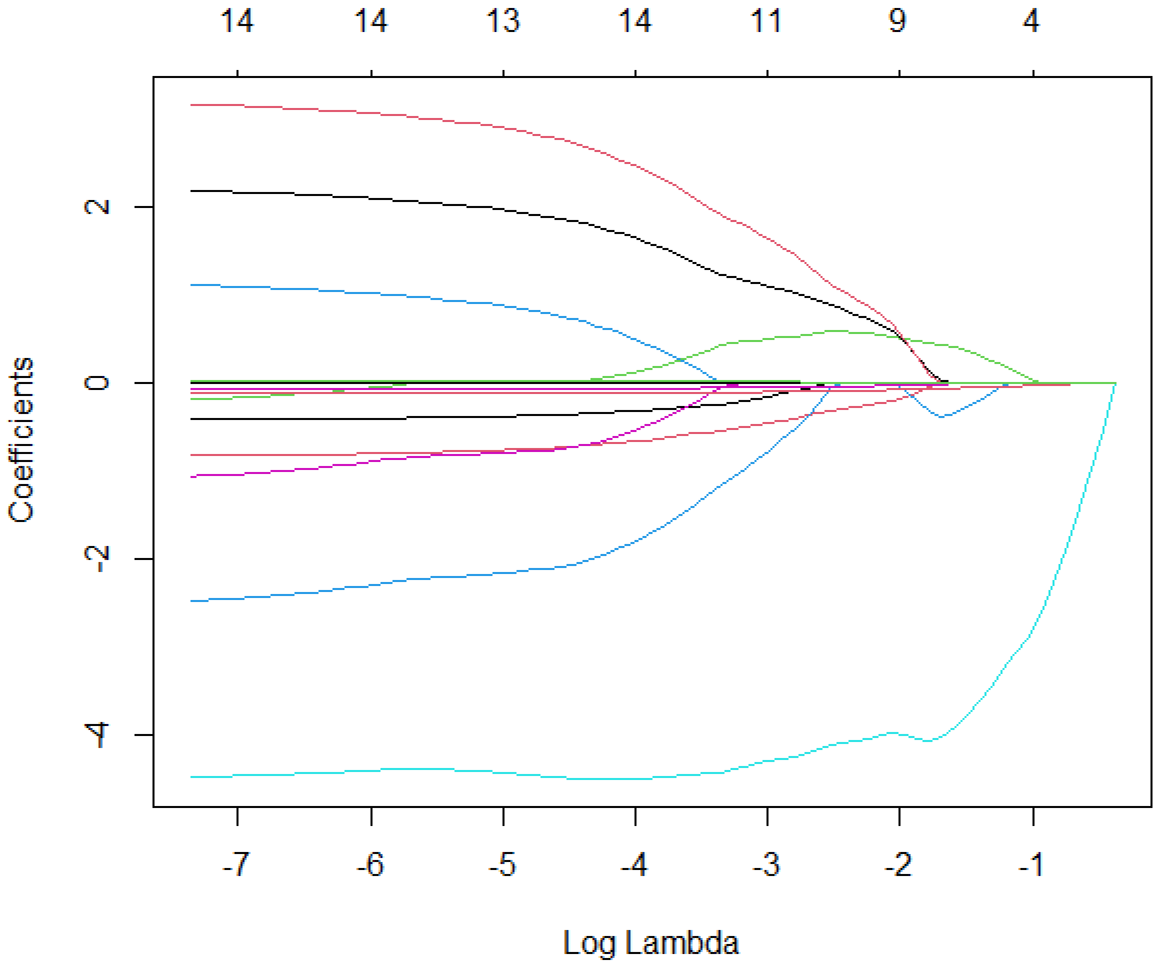

Figure 3: Weights of coefficients for various lambda of IR training model

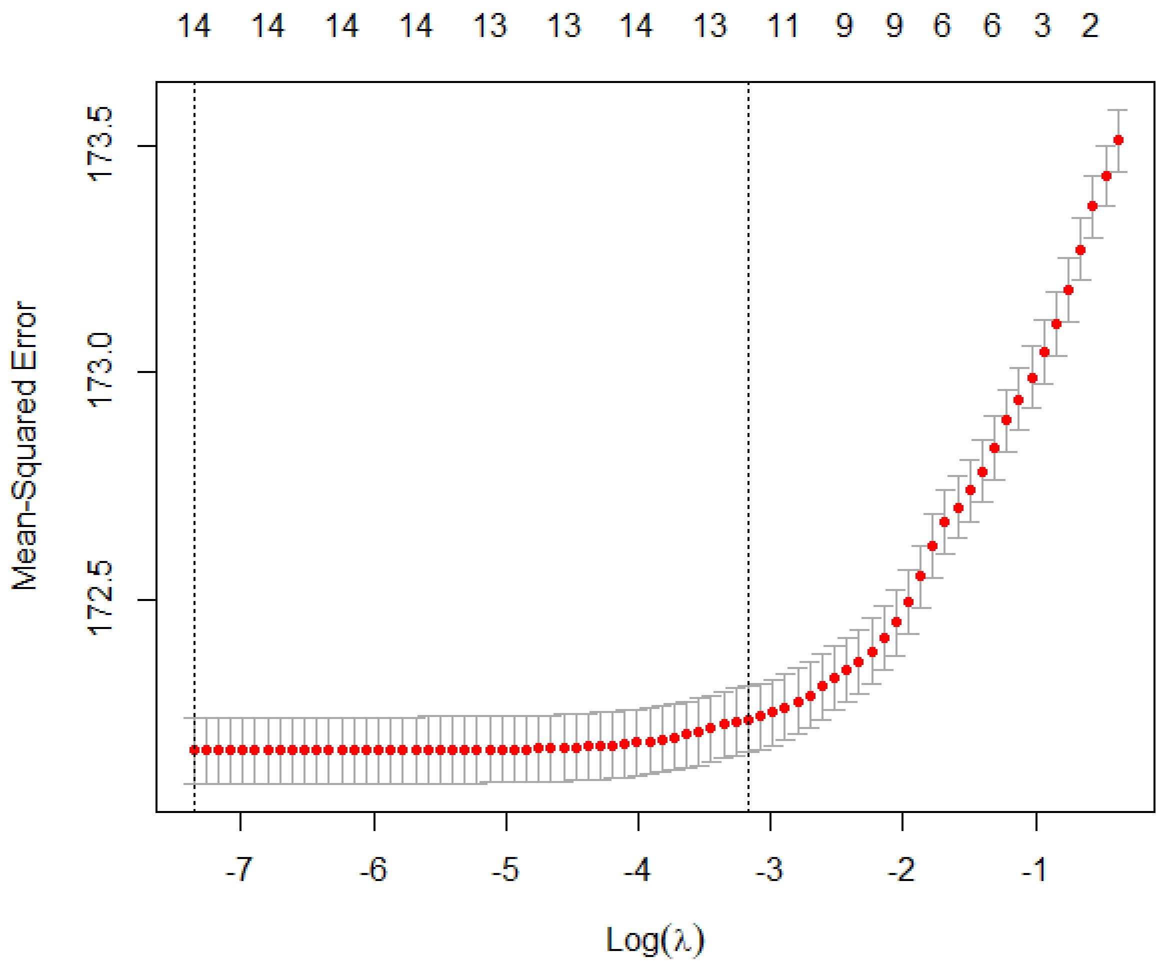

Figure 4: Mean Square Error for various lambda values of Cross-Validation of IR model

Lambda (min MSE)	MSE	SE
0.00858	300.3	0.1271

Table 5: Lambda value for minimum mean square error value of the DR model

Predictors	**Weights**
Intercept	67.5994
Permutation-2	-4.3281

Permutation-3	2.0906
Immediate_action-Nothing	0.1048
Immediate_action-Reading	-1.0734
Immediate_action-Sugar	0.0881
Immediate_action-TV	-0.5438
Recent_action-Nothing	4.9768
Recent_action-Reading	-7.7436
Recent_action-Sugar	2.6239
Recent_action-TV	5.5255
Blood Pressure-Systolic (BPS)	0.1322
Blood Pressure-Diastolic (BPD)	0.0292
Heart Rate (HR)	-0.0222
Breathing Rate	-0.1308
Cadence	0.0621

Table 6: Delayed Recognition (DR) model weights

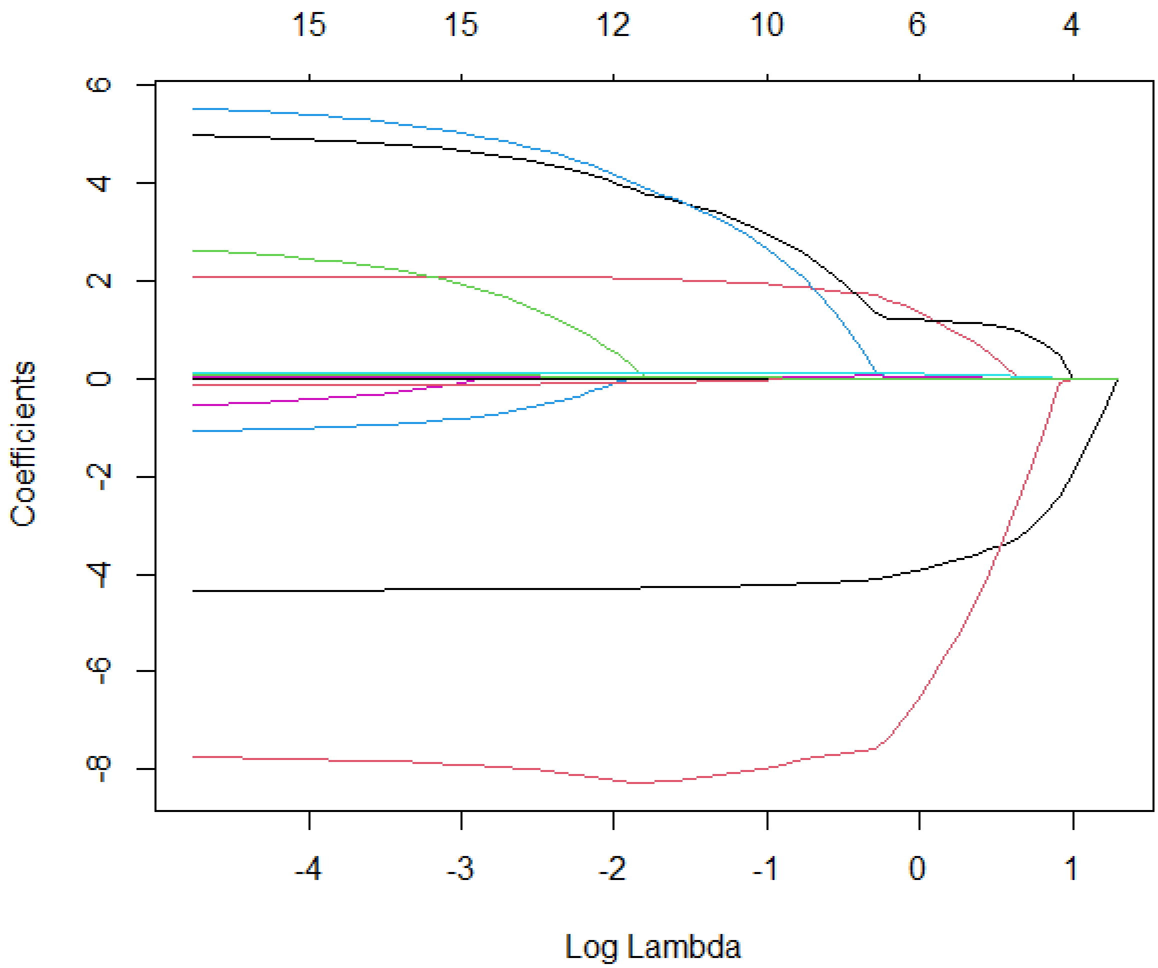

Figure 5: Weights of coefficients for various lambda of DR training model

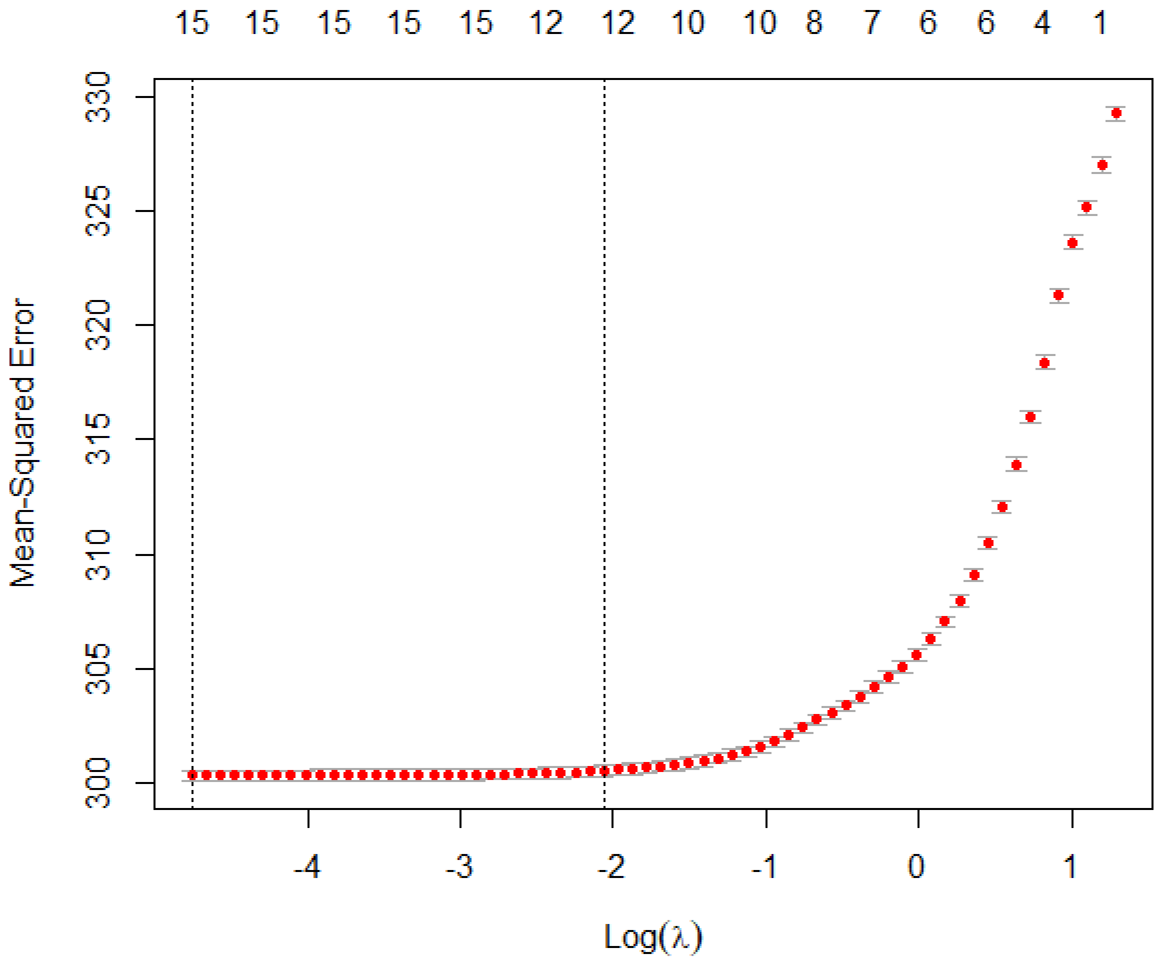

Figure 6: Mean Square Error for various lambda values of Cross-Validation of DR model