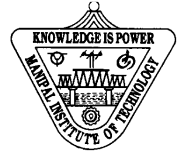# MANIPAL INSTITUTE OF TECHNOLOGY

(A constituent Institute of Manipal University)

## MANIPAL - 576 104, KARNATAKA, INDIA

MARCH 2017

## ******* GENSINI SCORE PREDICTION********

*By*

**Suvimal Yashraj – 140911394**
**Aman Chopra - 140911358**

*Under the Guidance of*

*Miss Akshata K.Naik*

*Miss Anju R.*

## I.  INTRODUCTION

**Role of Metabolic Obesity and body mass index in patients of various age group with coronary artery diseases.**

**Metabolic obesity (Insulin resistance syndrome).**
**Indian subcontinent is highly predisposed to this condition.**
**Prevalence of  Insulin resistance syndrome among Indians (≥30%).**
**Among females is higher than males (50%).**

## II.  PROBLEM DEFINITION

**Train a model capable of predicting the GENSINI score which determines the severity of CAD in the following groups-**
- *Metabolically Healthy Normal Weight (MHNW)*
- *Metabolically Obese Normal Weight (MONW)*
- *Metabolically Healthy Obese (MHO)*
- *Metabolically Abnormal Obese (MAO)*

**Gensini Scoring:**

**It is a scoring system for determining the severity of coronary heart disease.**

**It provides an accurate stratification of patients according to the functional significance of their disease.**
**It provides an opportunity to match  patients with similar degrees of coronary artery disease who are receiving different forms of treatment.**

## III.  OBJECTIVE

**To find the group showing a good association to severity of Coronary artery disease that is which category is more prone to CAD, metabolically obese or phenotypically obese.**

**To find the prognostic markers for CAD among factors like HBA1C, FI, HOMA IR , TC, TG, HDL, LDL and hsCRP and which group shows more association.**

**Several models were used for predicting the GENSINI scores from the given features and using regression and neural networks.**

- *Ridge Regression*
- *LASSO Regression*
- *Neural Networks*

## IV. SCOPE/IMPORTANCE OF PROJECT

There are no study done in India in relation to importance of metabolic obesity and BMI status with severity of Coronary artery disease.

Helpful to find how the Insulin resistance, hsCRP and Lp(a) is associated with the severity of Coronary artery disease.

Effect of Lifestyle modification on Body Mass Index and Waist Circumference in post angioplasty patients.

## V. METHODOLOGY

*Several models were used for predicting the GENSINI scores from the given features-*
- *Ridge Regression*
- *LASSO Regression*
- *Neural Networks*

*Since the* GENSINI score is a continuous value, the metric used for measuring the accuracy of predictions was:
> **Root mean square error**

## PREPROCESSING

### 1) MISSING VALUES

- Out of 15720 values (15*1048), only 2 values (all within the FI column) were missing.
- Since the number of missing values are proportionately very small, the group-wise average value is used to fill them.

  Missing values of FI in Group 1= Average of the FI values available in Group = 35.9386

  Missing values of FI in Group 4= Average of the FI values available in Group 4 = 47.99

## 2) NORMALIZATION

- **Normalization of the data elements was done for the Neural Networks.**

- **It wasn't required for Ridge and LASSO regression as MATRIX ALGEBRA was used for finding the coefficients instead of GRADIENT DESCENT.**

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$$

## 3) DATA SPLIT

- **75% of the data was used for training the model.**
- **The remaining 25% was used for testing it.**
- **Root mean square error was then found out between the actual value and the predicted values of that 25% data.**

## VI. RESULTS

**Obtained Root Mean Square Errors**

Results obtained after applying each model on test data.

| Model | Groups -> | Group 1 | Group 2 | Group 3 | Group 4 |
|---|---|---|---|---|---|
| LASSO Regression | | 14.33 | 8.82 * 10e-15 | 4.79 | 10.55 |
| Ridge Regression | | 21.18 | 24.14 | 510.9 | 20.47 |
| Neural Networks | | 4.11 | 0.14 | 8.81 | 0.55 |

- **For the entire dataset together, Neural Networks gave the best results. The RMSE obtained was 6.170**
- **For Group 2 and 3, LASSO gave the best result.**
- **For Group 1 and 4, Neural Networks gave the best result.**

Results obtained after applying each model on train data.

| Model | Groups -> | Group 1 | Group 2 | Group 3 | Group 4 |
|---|---|---|---|---|---|
| LASSO Regression | | 16.18 | 9.32 * 10e-15 | 3.74 | 2.30 |
| Neural Networks | | 1.27 | 0.17 | 0.46 | 0.83 |

- **Ridge regression model was eliminated because the results obtained with the test data was very inaccurate.**
- **The results obtained for the models selected in each group were not very different from the test model. Hence, we can say that the model is not over fitting the training data.**
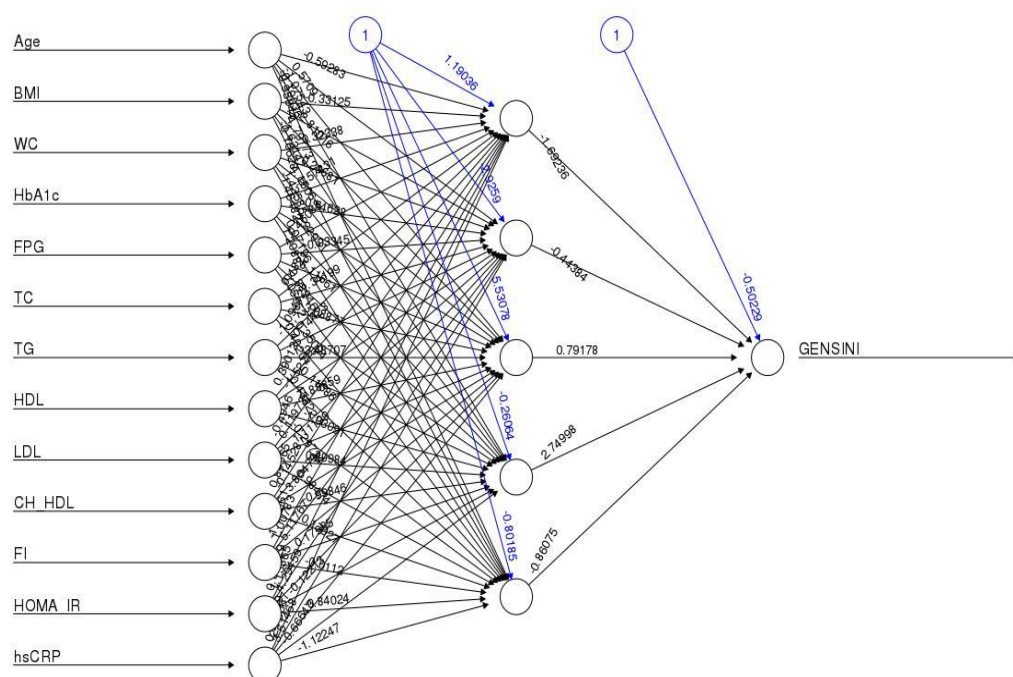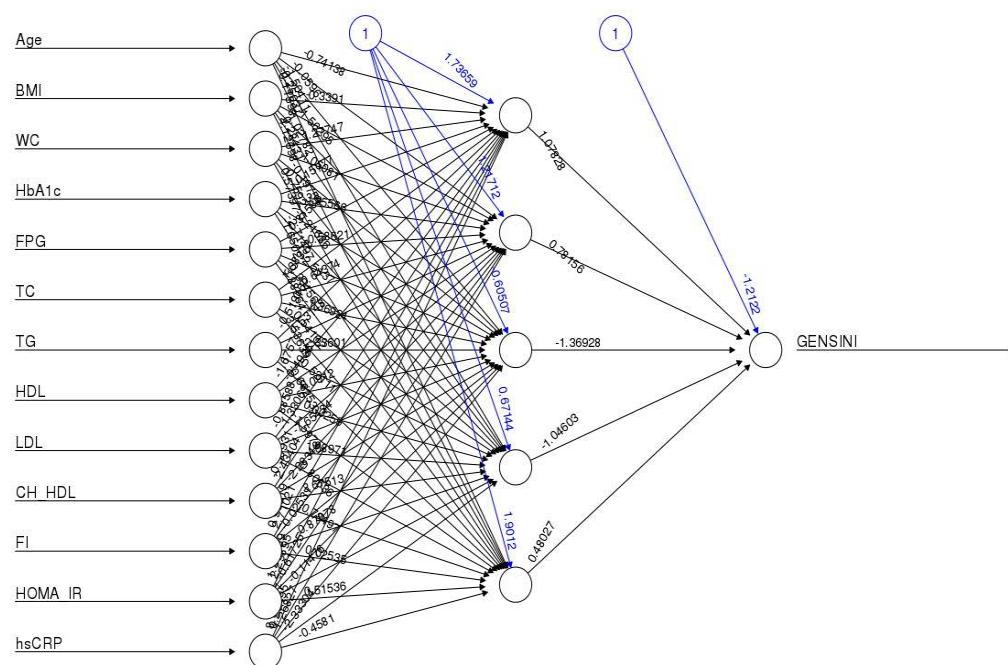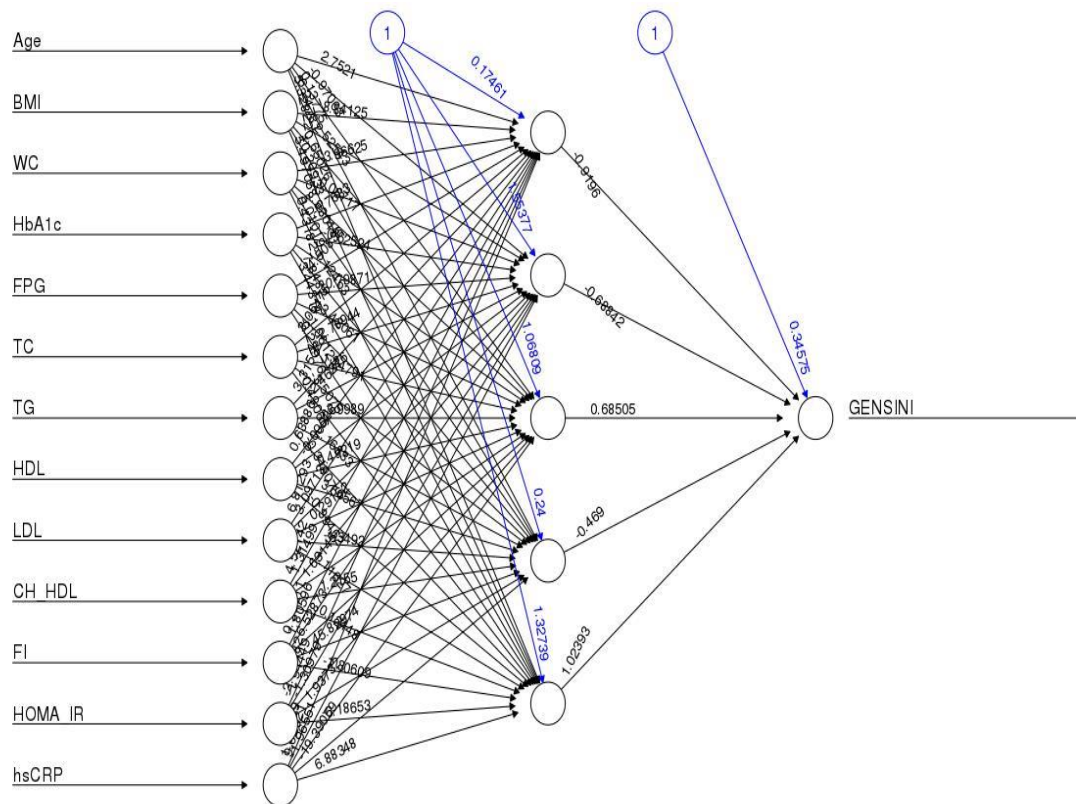
## NEURAL NETWORKS with 1 Hidden Layer

- **Group 1**
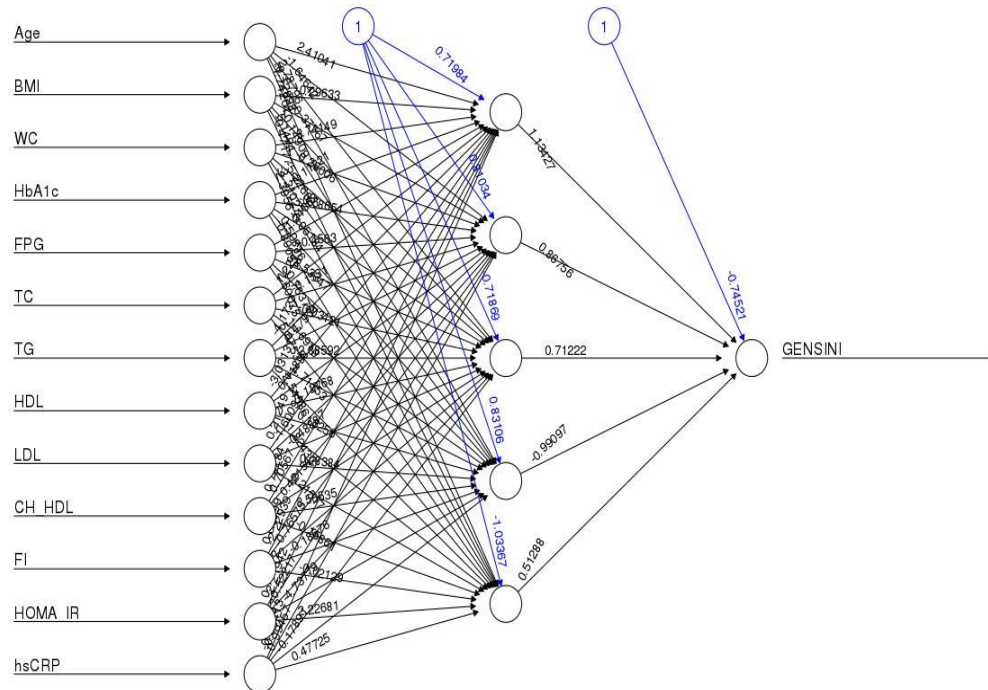


- **Group 2**
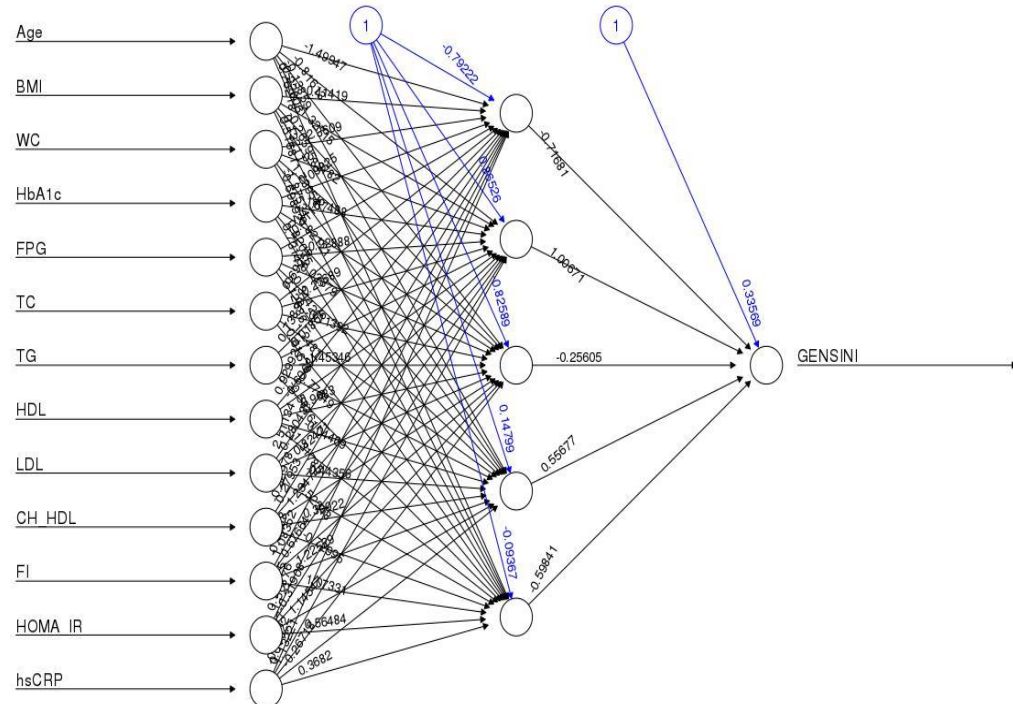
- **Group 3**



- **Group 4**

- **Entire Dataset**



- **Ridge Regression**

    Ridge Regression was used as the initially for regression. The initial idea to ensure that over-fitting does not occur was to  give a penalty to the cost function for every unit increase of the parameter values.

- **LASSO**

    Ridge Regression cannot zero out coefficients ; thus either all coefficients are included or none. But LASSO does both parameter shrinkage and variable selection automatically , and hence was thought of as a nice approach to avoid over fitting.
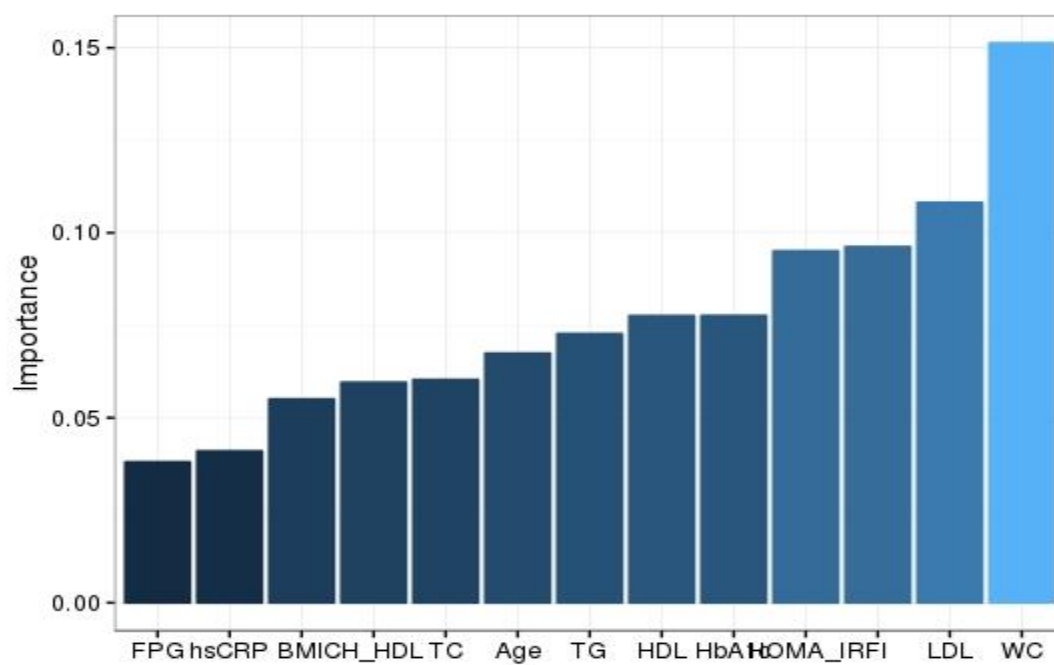
- **Neural Networks**

    We tried using Neural networks with varying number of inner layers. The best results obtained for the test data were from using only a single hidden layer. Including more layers caused the model to over fit.
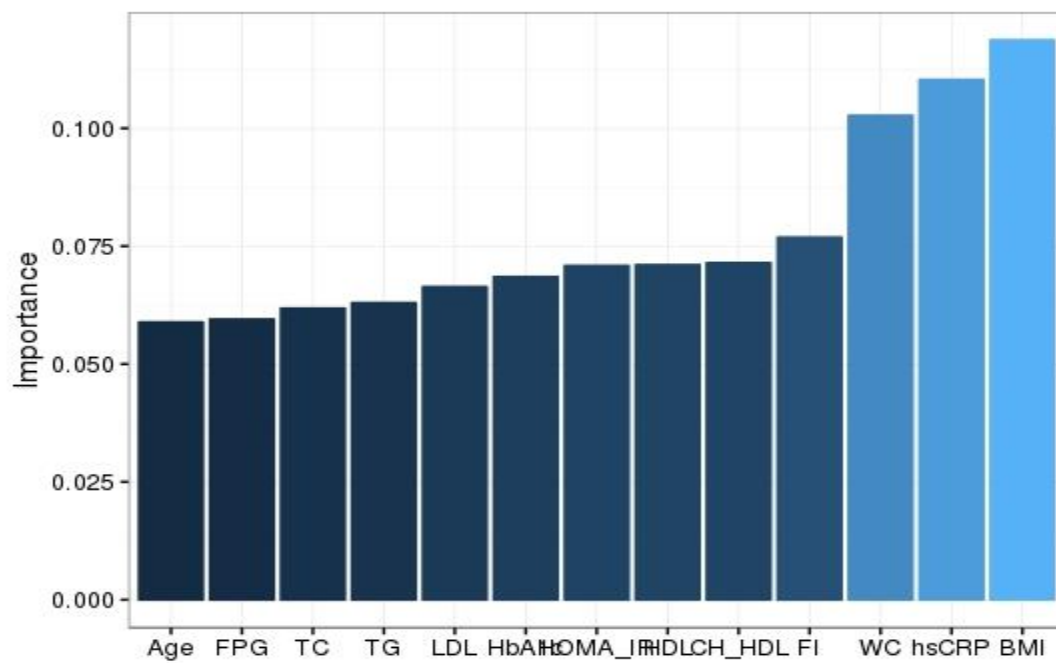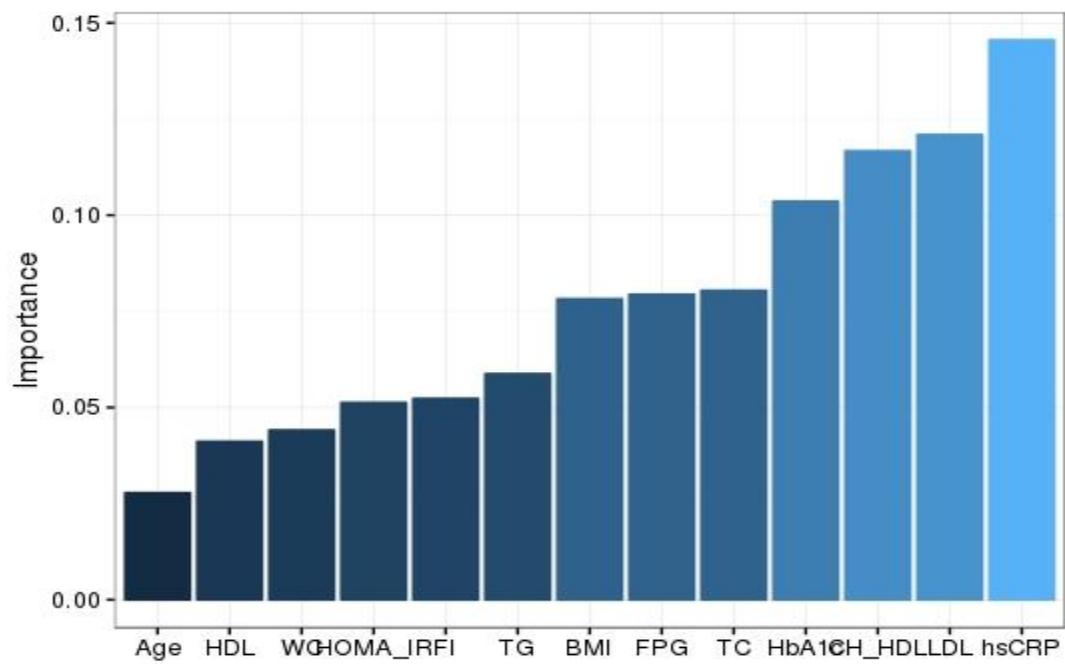
# FEATURE IMPORTANCE
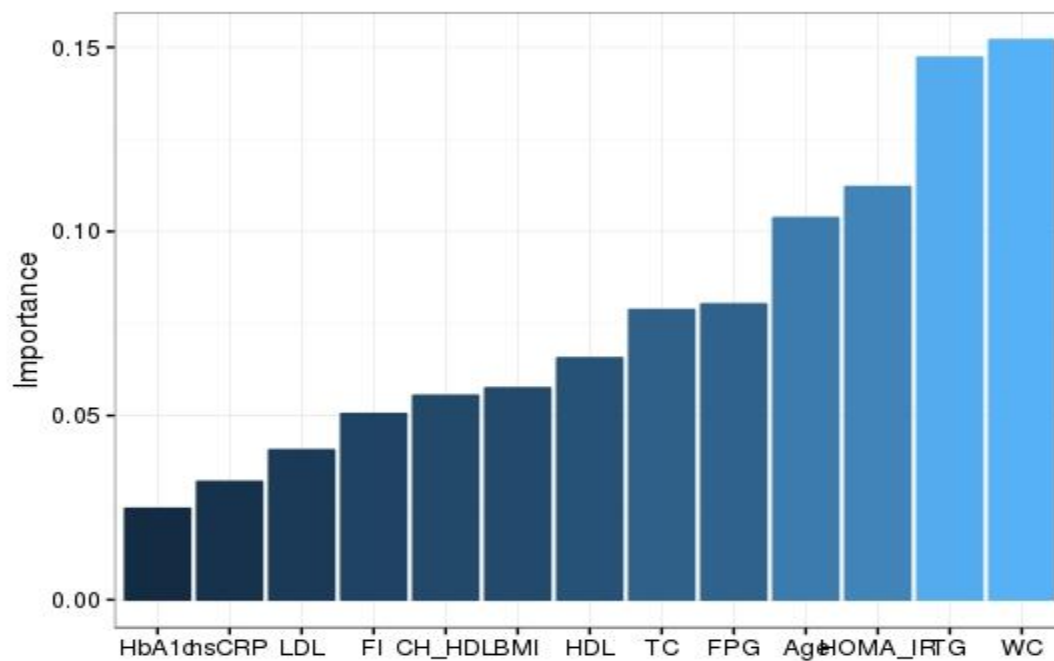
- **Group 1**



- **Group 2**

- **Group 3**



- **Group 4**
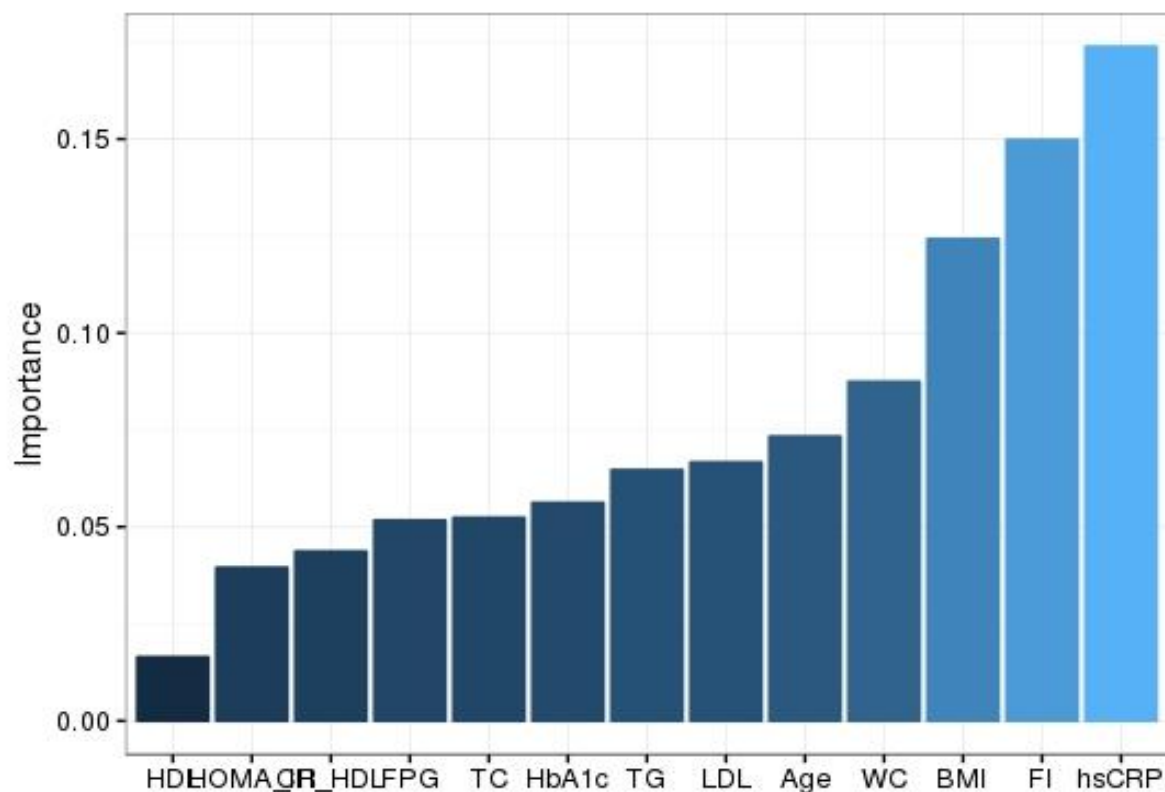
- **Entire Dataset**



| | Feature Importance |
|---|---|
| Group 1 | WC > TG > HOMA_IR > AGE |
| Group 2 | WC > LDL > FI > HOMA_IR |
| Group 3 | BMI > hsCRP > WC > FI |
| Group 4 | hsCRP > LDL > CH_LDL > HbA1c |
| **Entire Dataset** | **hsCRP > FI > BMI > WC** |

## VII. CONCLUSION

**Relationship between Obesity type and CAD**

- 95% confidence interval was found out for the GENSINI scores within each group

| Groups | 95% Confidence Intervals |
|---|---|
| *Metabolically Obese Normal Weight* | 8.848 to 19.818 |
| *Metabolically Healthy Obese* | 2.504 to 15.629 |
| *Metabolically Abnormal Obese* | 8.483 to 23.342 |
| *Metabolically Healthy Normal* | 1.945 to 15.018 |

- It means that the GENSINI scores will lie between the given ranges with 95% probability in the population.

- This means that the severity of CAD varies as follows-
    **Phenotypic obese < Metabolic obese < Metabolic and Phenotypic obese**

- The GENSINI score for the groups can be accurately predicted using-
    - **LASSO for MHNW and MAO**
    - **Neural Networks for MONW and MHO**
    - **Overall, if a single model is to be used with the group number as a parameter then Neural Networks give the best results.**

- With a shift from **phenotypic to metabolic obesity, the severity of CAD increases**.

| Factors | Correlation | p-Value | Statistical Significance ($\alpha$ = 0.05) |
|---------|-------------|---------|------|
| hsCRP | 0.109 | 0.2791 | No |
| FI | -0.0133 | 0.8950 | No |
| BMI | -0.224 | 0.0249 | Yes |
| WC | 0.0228 | 0.8214 | No |
| AGE | 0.4165 | 0.0000162 | Yes |
| HBA1C | 0.3430 | 0.000475 | Yes |
| HOMA IR | 0.09821 | 0.3309 | No |
| TC | -0.0732 | 0.4688 | No |
| TG | 0.05696 | 0.5556 | No |
| HDL | 0.0735 | 0.4658 | No |
| LDL | -0.14321 | 0.1551 | No |

- **BMI, AGE and HBA1C can be used as prognostic markers** to determine the severity of CAD.
- **HBA1C and AGE have positive correlations.**
- **BMI has a negative correlation.**

## VII. REFERENCES

Mulhall BP, Ong JP, Younossi ZM. Non-alcoholic fatty liver disease: an overview. J Gastroenterol Hepatol 2002;17(11): 1136–43.
Reaven GM. Banting lecture 1988. Role of insulin resistance in human disease. Diabetes 1988;37(12):1595–607.
Sims EA. Are there persons who are obese, but metabolically healthy? Metabolism 2001;50(12):1499–504.
KarelisAD, Faraj M, Bastard JP, et al. The metabolically healthy but obese individual presents a favourable inflammation profile. J Clin Endocrinol Metab 2005;90(7): 4145–50.
St-Pierre AC, Cantin B, Mauriege P, et al. Insulin resistance syndrome, body mass index and the risk of ischemic heart disease. CMAJ 2005;172(10):1301–5.