

# **Advanced Algorithms for Predicting Normal and Weak Immune System Among HPV-infected Women**

**Presenter:**

**Ogbolu Melvin Omone**



# Introduction

- **Human papillomavirus(HPV)** is a family of the **Papillomaviridae** that causes infection to humans, through human-to-human contact, especially through sexual intercourse with an infected partner.
- **Human papillomavirus(HPV)**, can cause abnormal modification or transformation in the body that compromises the immune system.
- An infected human immune system, if weakened further can lead to Cervical cancer in women.

# Research Objective

- Aim of the research is to predict different risk factors, given the HPV risk factor (high-risk(hr) or low-risk(lr)) will determine which of the risk have more impact on the women immune system given the scores.
- The scores is used the predict the level of risk on women.

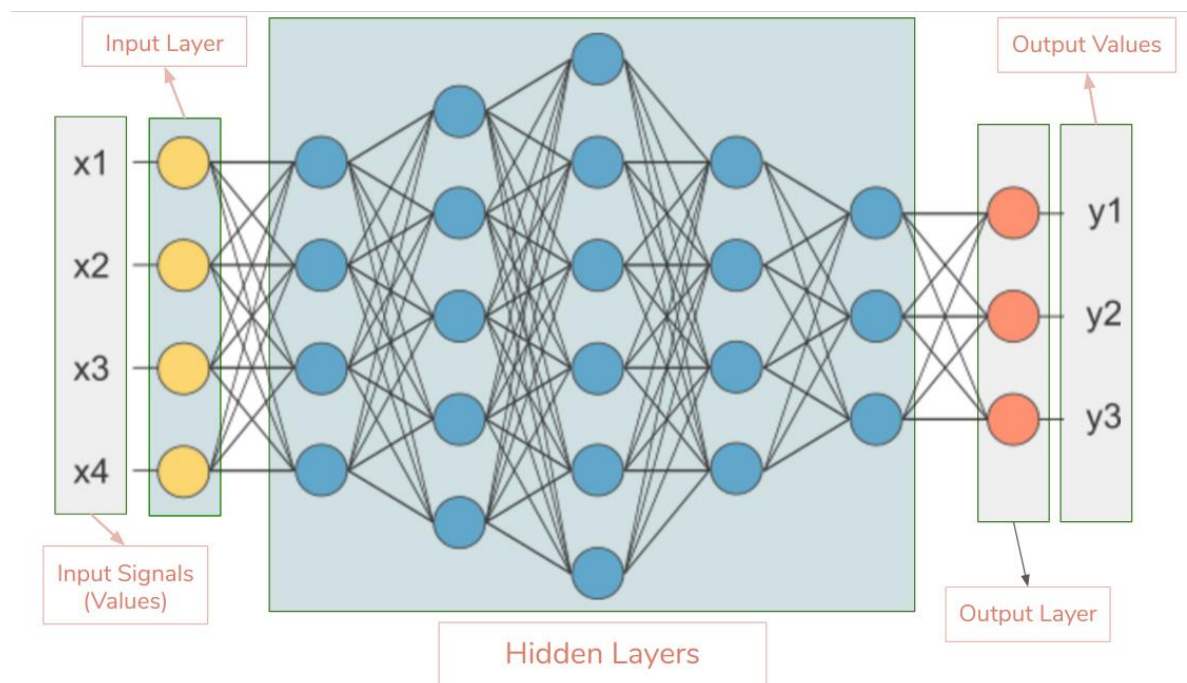
# Material And Method

- A Study was conducted on **1965 female participants** between **ages of 18-60 years**.
- The features selected to perform these prediction are:
  - **women who engaged in protected/unprotected sex,**
  - **women who engage in oral, anus and vaginal sex,**
  - **women who indicated to practice a healthy/unhealthy diet ,**
  - **women who have one/more sexual partners, and**
  - **women who indicate to be smokers.**
- **Human Papillomavirus (HPV) Assessment Test(HAT)** is an online tool developed to access to HPV-related diseases. The tool is designed to collect data that predict progression of cervical cancer in women. the analyze data are then assign scores for  $<20/70$ (lr) and  $>21/70$ (hr) given the data provided by these participants.

- **Data Analysis**

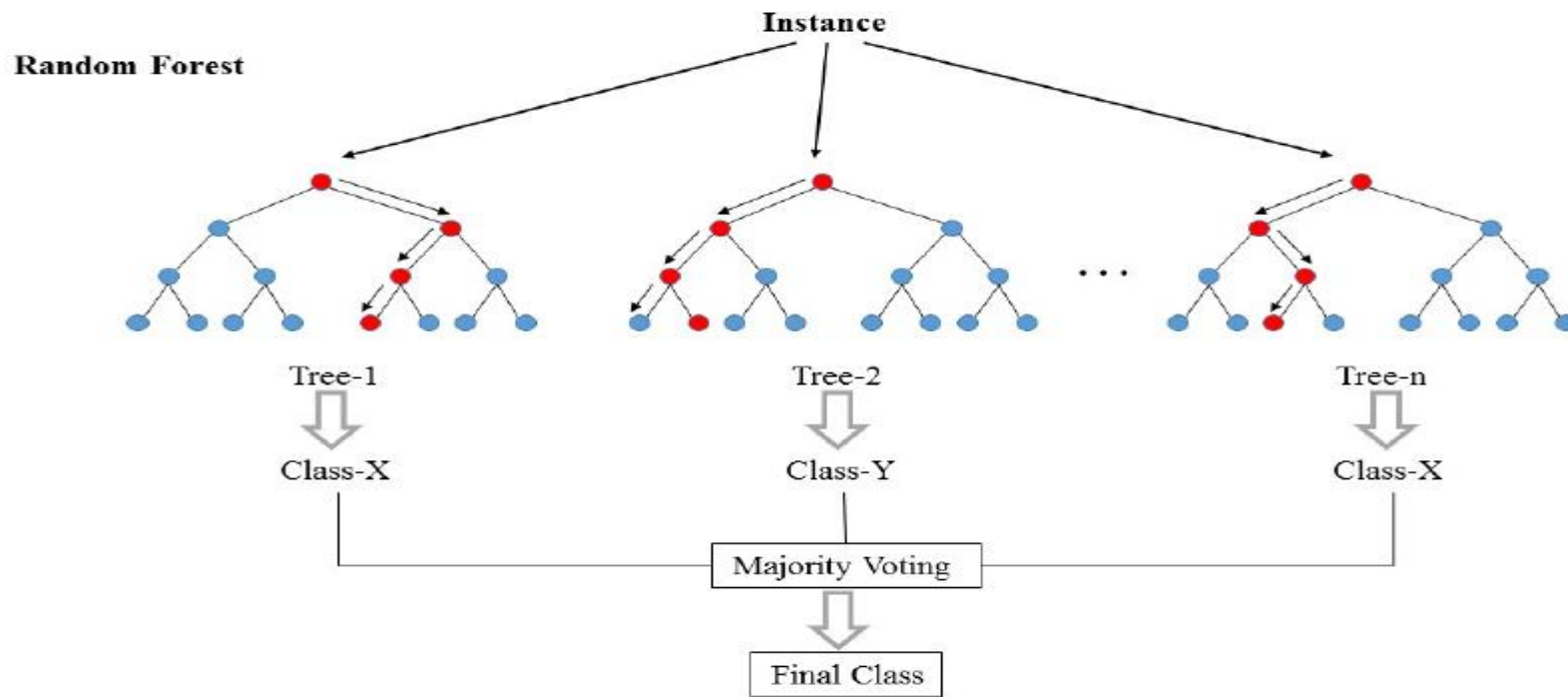
- With the Help of the **HAT** tool, we had access to datasets of HPV related diseases for our data analysis.
- Each participant scored from 8/70 - 61/70.
- Two Machine learning algorithm were implemented to learned the selected features given the scores provided and then grouped into hr and lr factors.
  - lr scores is given as  $<20/70$  and hr factors  $>21/70$ .
  - **Neural Network Algorithm (NNA)** and **Random Forest Algorithm (RFA)** were used to learn the features given the scores.

- **Neural Network Algorithm** is an algorithm that imitates the neuron in our brain. The Neural Network Algorithm is supervised learning algorithm, which takes a sequence of input variables features like the female data, our Input is define as  $x$   $\{x_1, \dots, x_n\}$ , it learns dependent variable or output data in our case the scores are our labels  $y$   $\{y_1, \dots, y_n\}$ .



**Neural Network Image**

- **Random Forest Algorithm:** The Random Forest is supervised learning algorithm it can perform as a classification algorithm and a regression algorithm; it is also collection of decision trees. It is built upon features and classes of our dataset using a tree ensemble method or divide and conquer method. It has the same attributes as the decision tree.



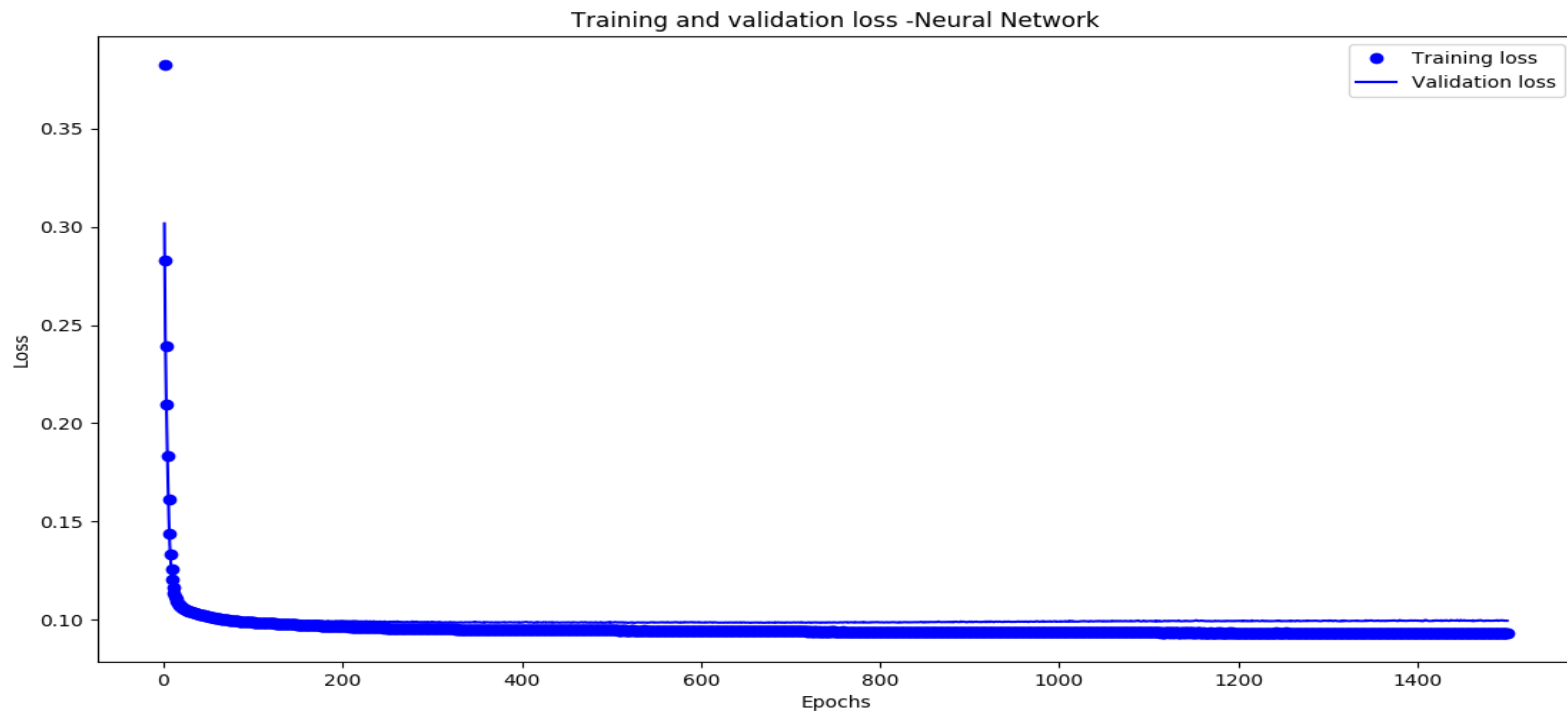
**Random Forest Image**

# Results

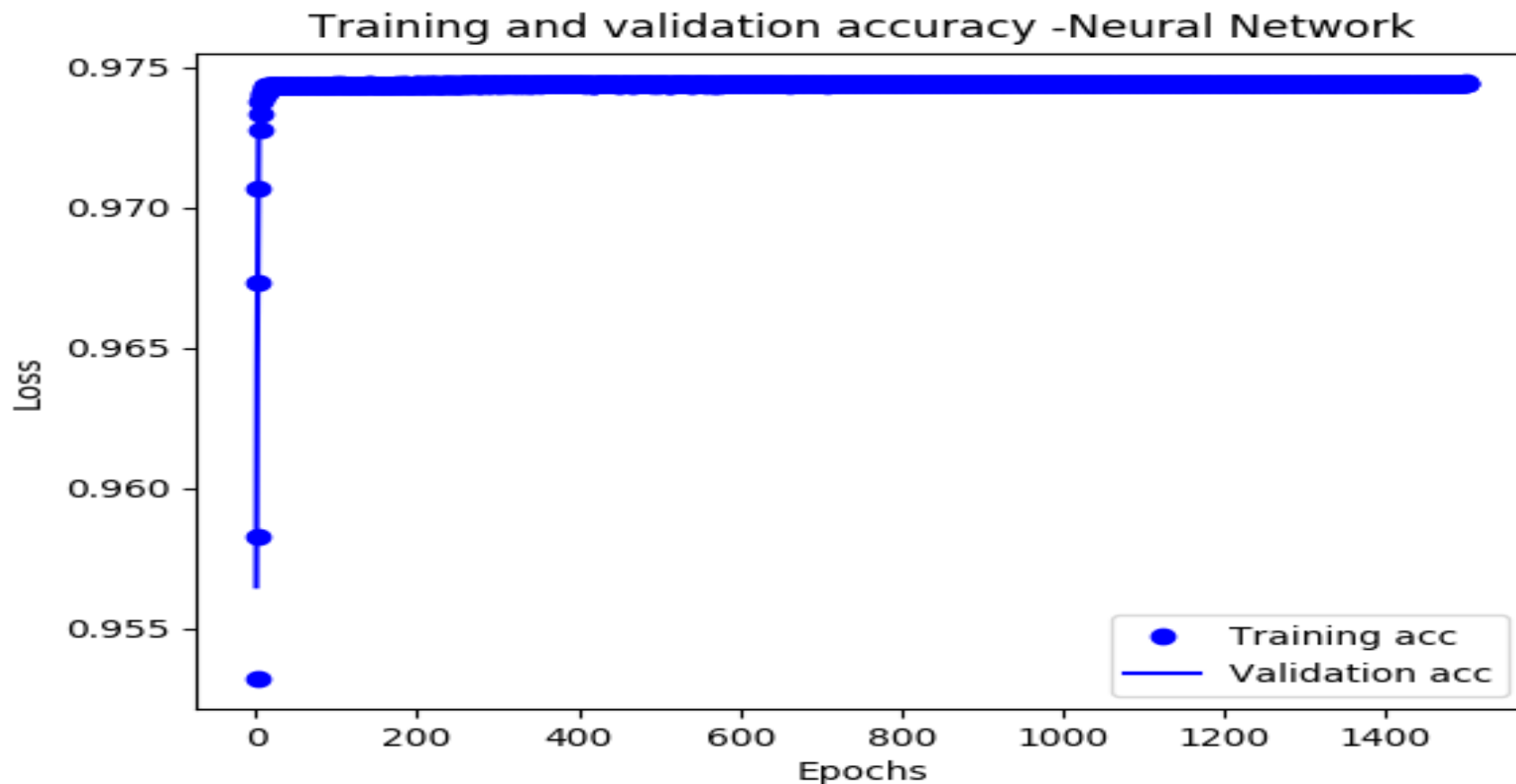
- The results will show the performance of each models that were trained in the course of the research.
- Also the algorithm suited for predicting our classes.
- The Aim of the experiment is to find features that will identify level risk factors that impact the immune system in women.
- The scores of the risk factors, will be grouped into lr factor and hr factor.



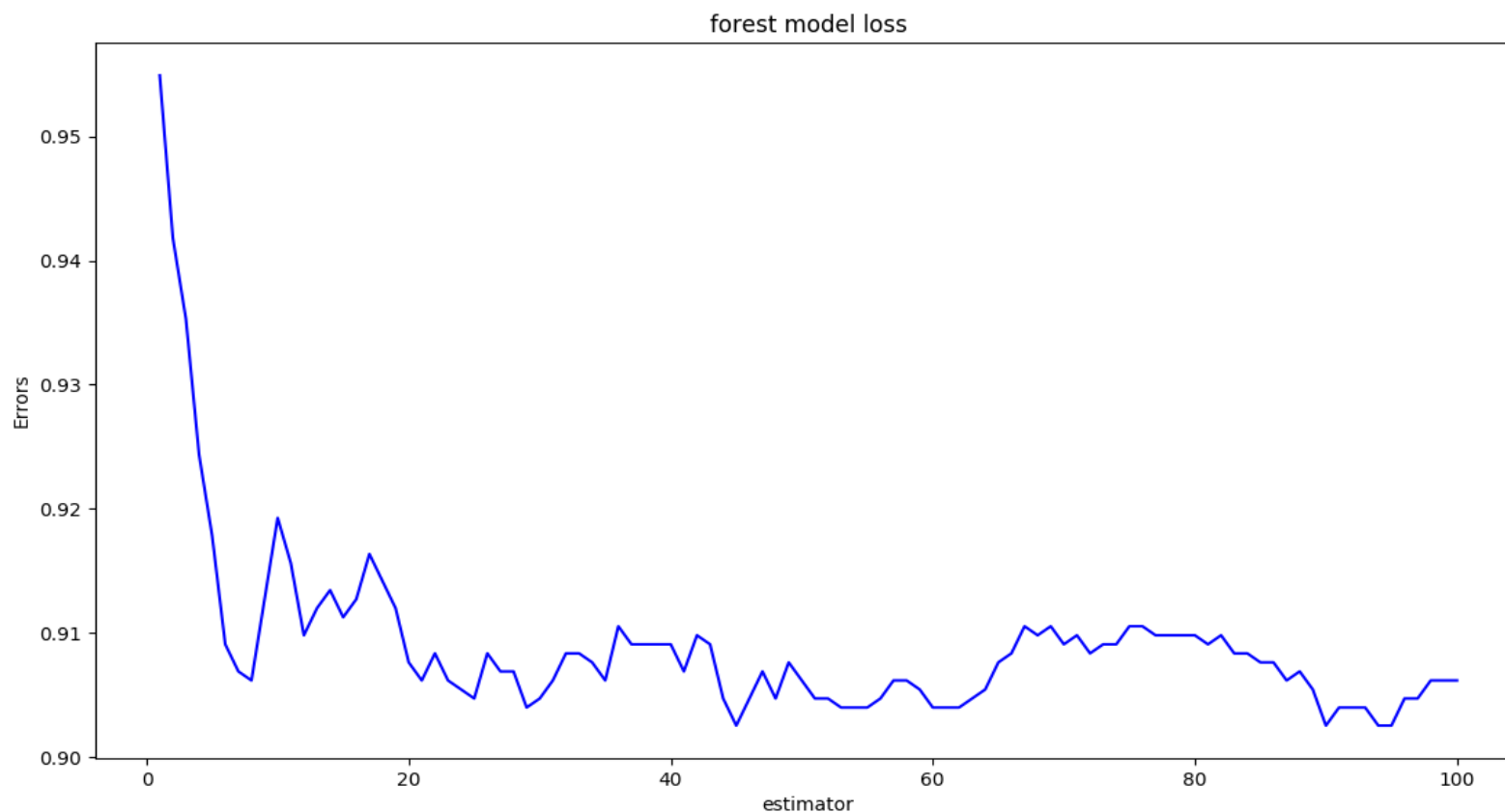
- **Loss function of the Neural Network:** This is a visualization of the training. The loss/cost function, measures the difference between the predicted and observed labels or classes. We know that our model performed well is when the loss values is low or zero, in our case it did. This experiment was performed using 1400 epochs to learn the features correctly.



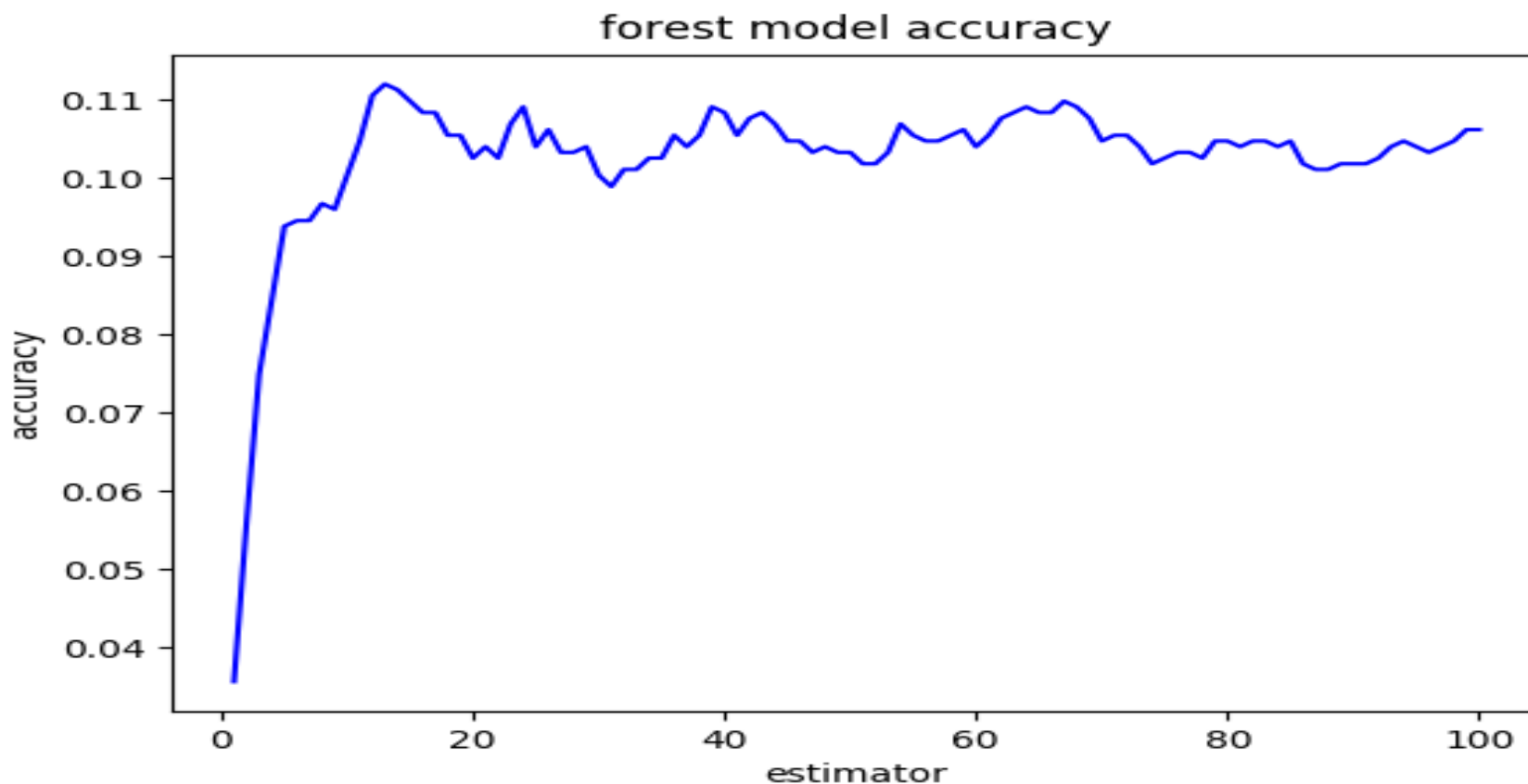
- **Neural Network accuracy Graph:** The Graph shows the accuracy of how the model performed during the training. Neural Network model accuracy after the training is 97.5 percent.



- **Random Forest Loss Graph:** The result shows both loss and accuracy of the Random forest algorithm. The random forest uses the Out of the Bag loss function to measure the loss of each generated trees in the random forest.

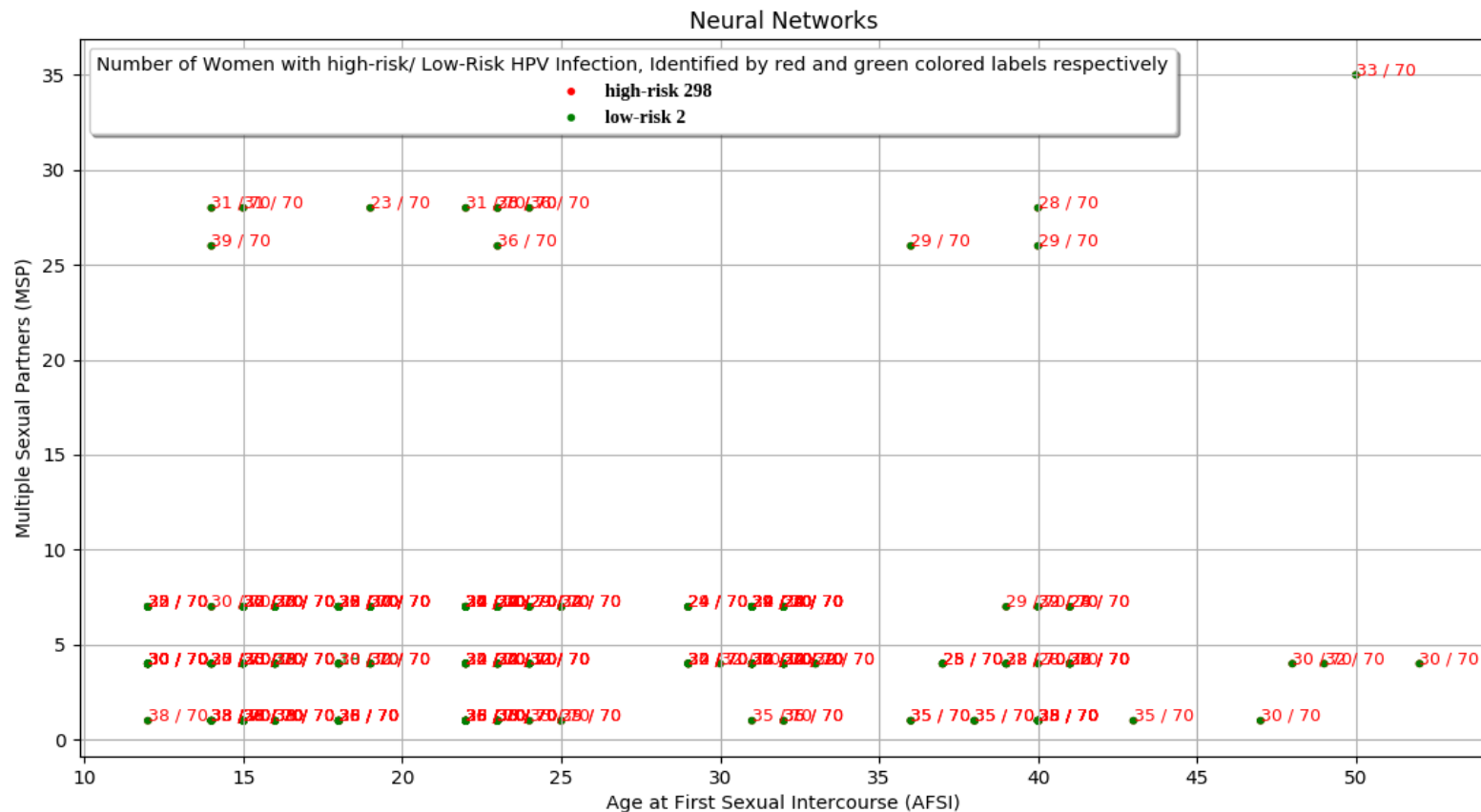


- **Random forest Accuracy Graph:** The Graph shows the performance of the random forest model. The random forest performed with an accuracy 0.1. Though we have a poor performance during the training, given that we have more classes than features to learn from.

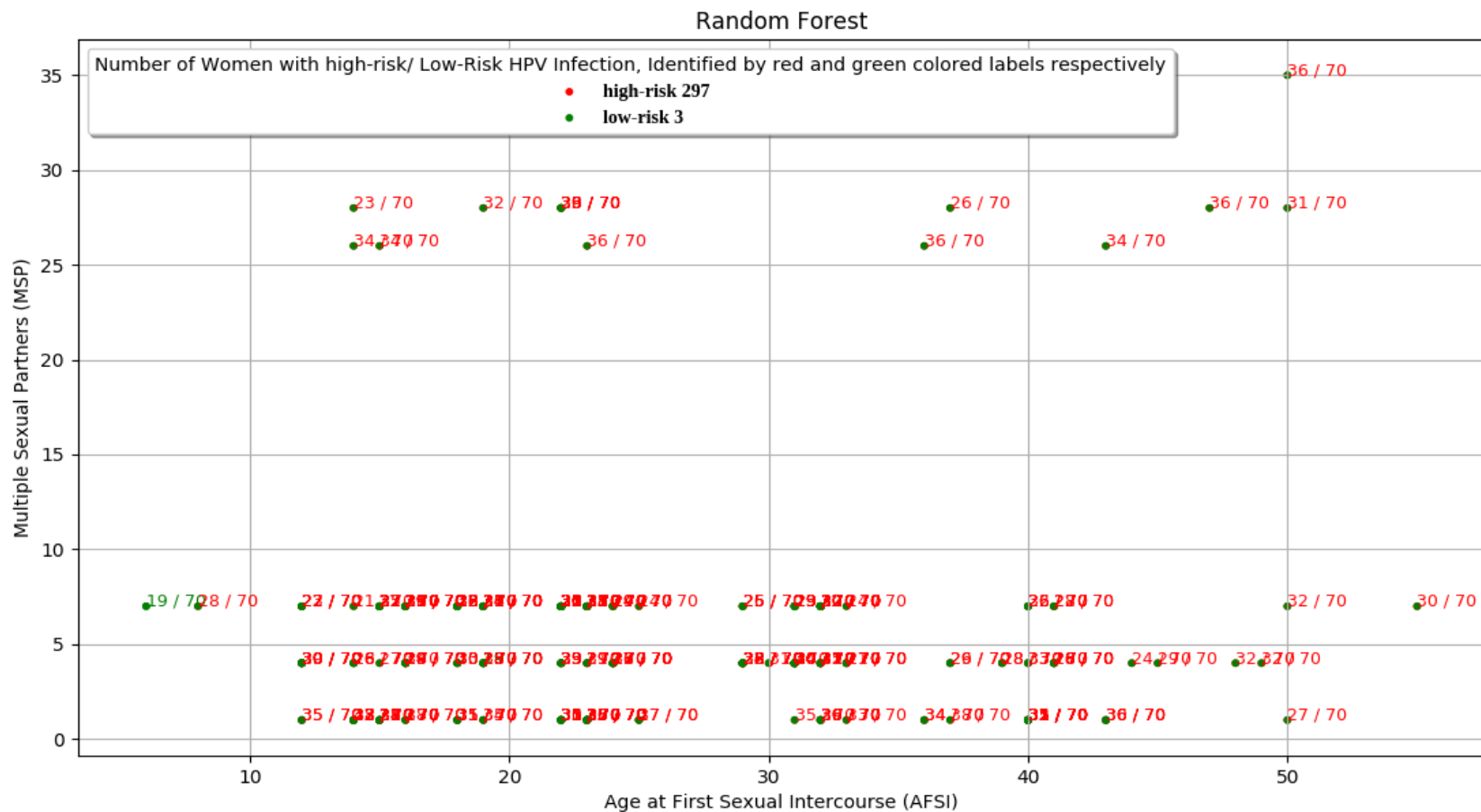


- During the training we perform prediction pairing two different features in our two models.
- The features predicted using both **Random Forest Algorithm (RFA)** and **Neural Network Algorithm (NNA)**, are:
  - Classification of MSP and AFSI using NNA and RFA.
  - Classification of Engage in oral sex, anal sex, and vaginal sex and Engage in unprotected sex using NNA and RFA.
  - Classification of Unhealthy diet and Smokers using NNA and RFA.

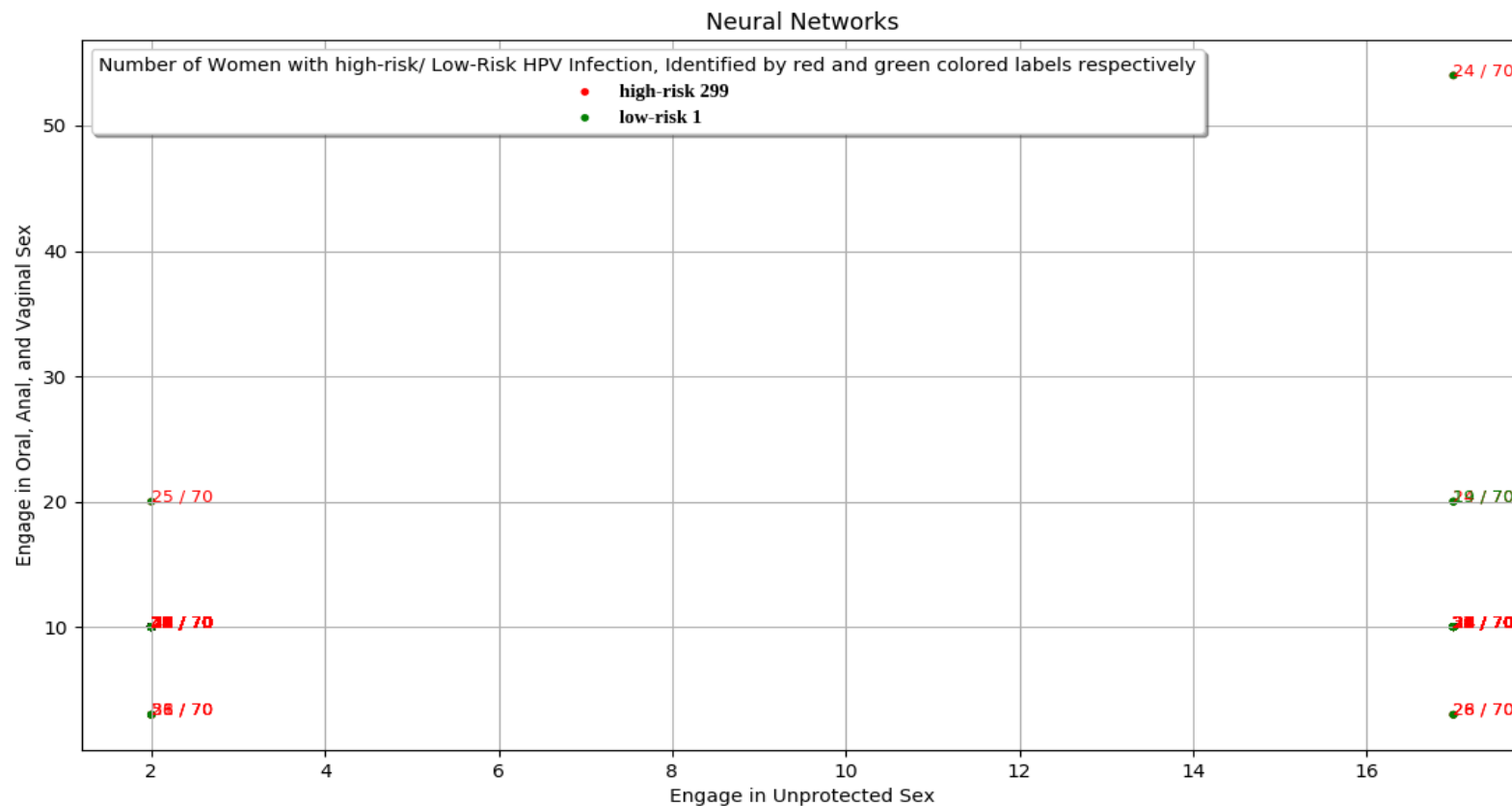
- Classification of Multiple Sexual Partners (MSP) and Age at First Sexual Intercourse (AFSI) using NNA.** During classification, using Neural Network, we got **298** participants who have a hrHPV and **2** lrHPV infection.



- Classification of Multiple Sexual Partners (MSP) and Age at First Sexual Intercourse (AFSI) using RFA.** During classification, using RFA, we got **297** participants who have a hrHPV and **3** a lrHPV infection.

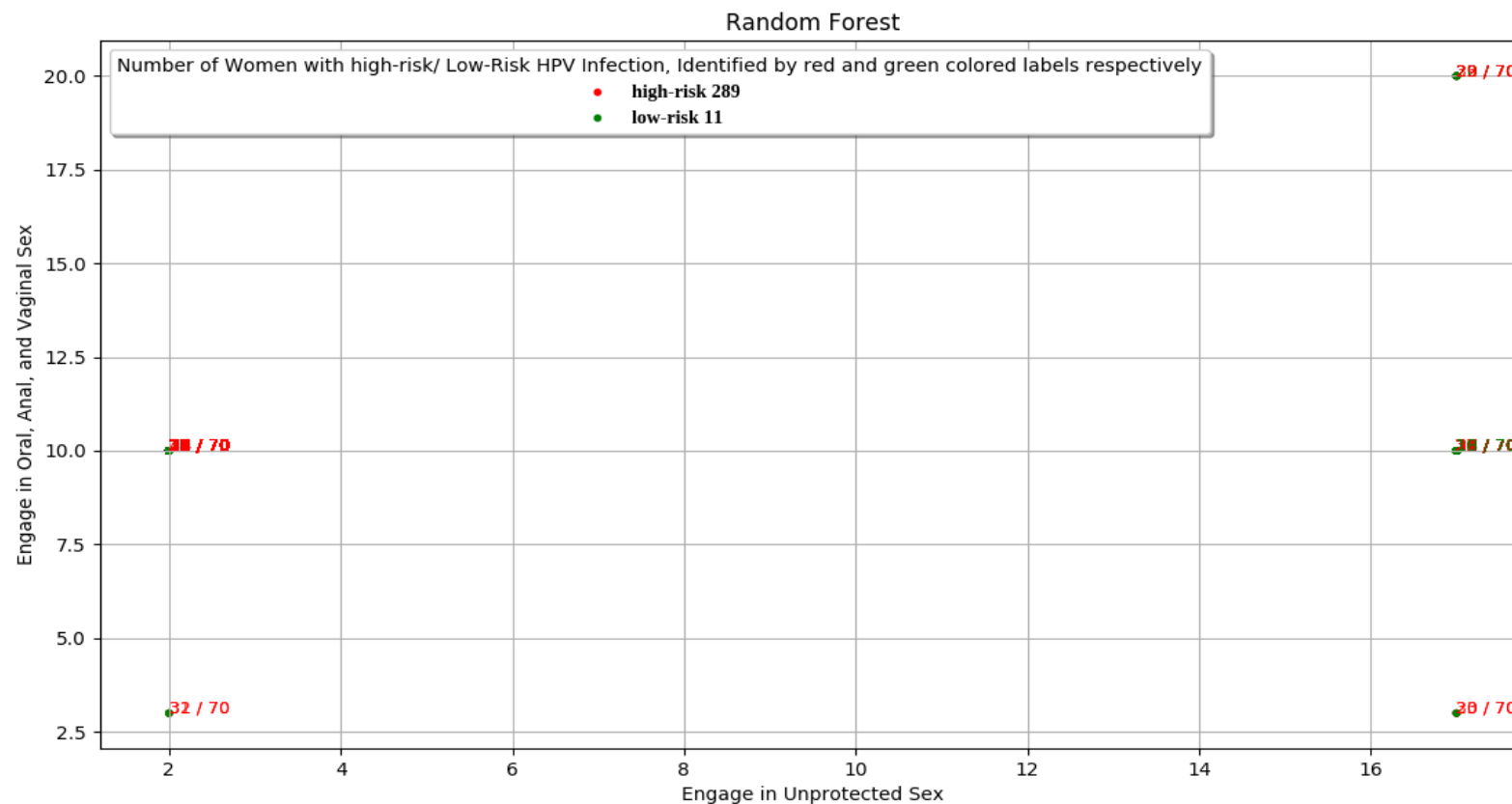


- **Classification of women who Engage in oral sex, anal sex, and vaginal sex and Engage in unprotected sex using NNA. During classification, using NNA, we got 299 participants who have a hrHPV infection and 1 lrHPV infection.**

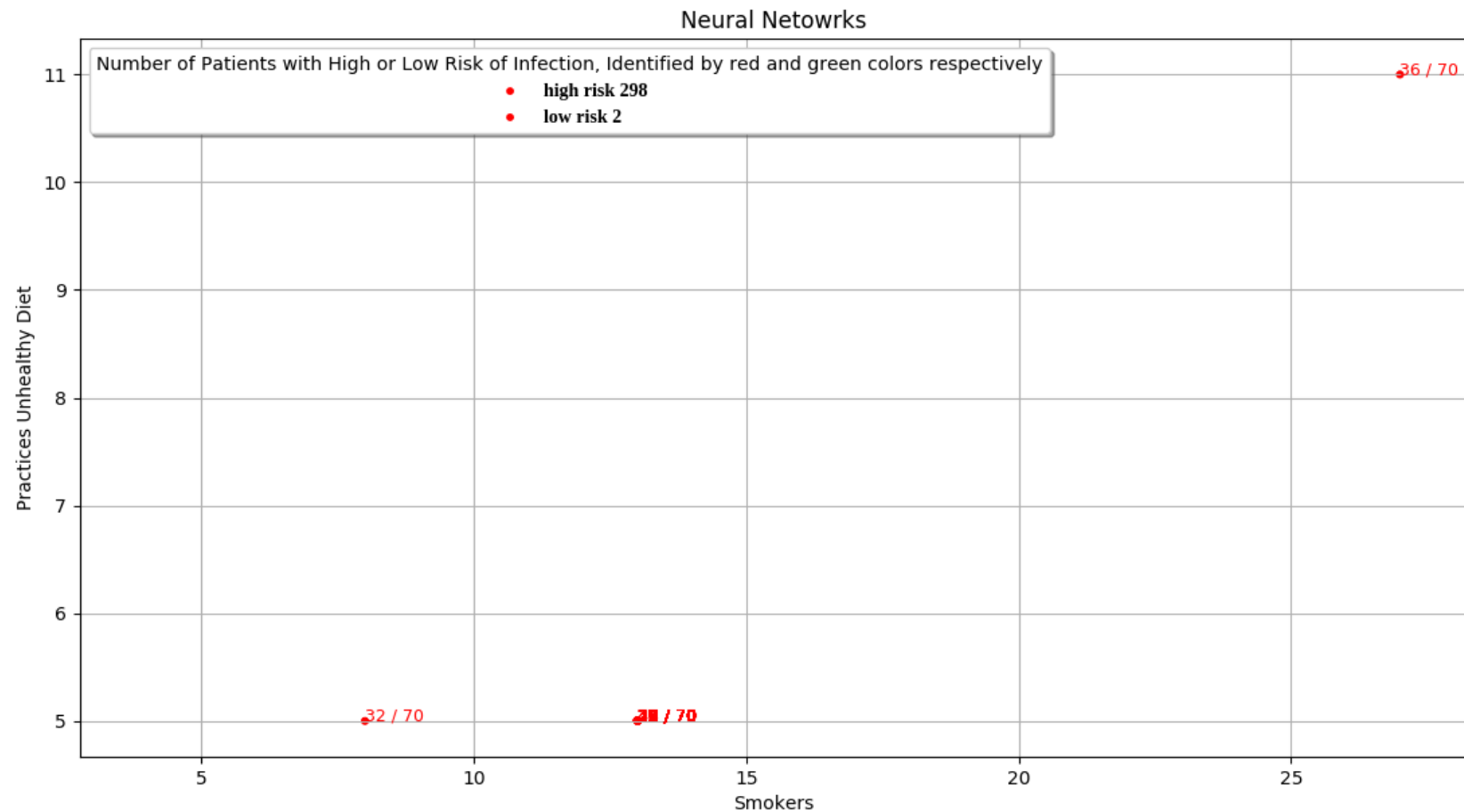




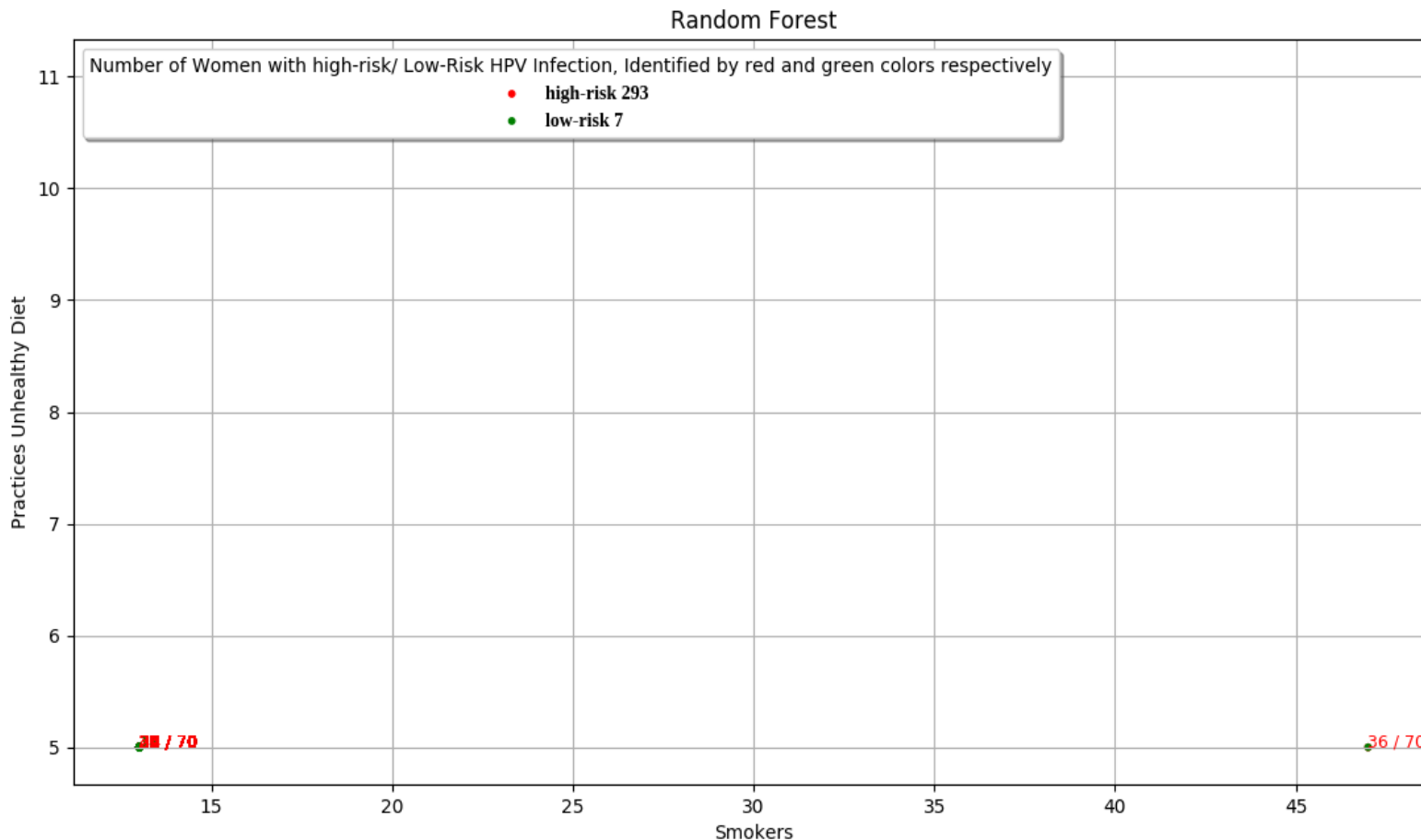
- **Classification of women who Engage in oral sex, anal sex, and vaginal sex and Engage in unprotected sex using RFA. During classification, using RFA, we got 289 participants who have a hrHPV and 11 lrHPV infection.**



- **Classification of Unhealthy diet and Smokers using NNA.** During classification, using NNA, we got **298** participants who have a hrHPV and **2** lrHPV infection.



- **Classification of Unhealthy diet and Smokers using RFA.** During classification, using RFA, we got **293** participants who have a hrHPV and **7** lrHPV infection.



# Conclusion

- In conclusion, the analysis carried out proves positive in using machine learning algorithm in learning features data and in classifying these features into their various classes.
- The features of the collected dataset were classified into **38 score-labels**, which were grouped into two to identify women with hrHPV and lrHPV infection.
- RFA and NNA had some issues detecting some labels as a result of the datatype contained in the dataset. During analysis an imbalanced classification problem was encountered because there existed more classes than features in the dataset.
- The aim of the analysis was achieved for classifying women with normal immune system and weak immune system. Wherefore, there were more women with high-risk HPV infection than women with low-risk HPV infection.

**Thank You!**

**QUESTIONS?**