# METABOLIC SYNDROME DETECTION AND HEALTHCARE USING GENETIC ALGORITHM

A MAJOR PROJECT REPORT

*Submitted by*

**O. SAI RISHITHA [RA2111003011566]**
**A. SIVA NARAYANA [RA2111003011444]**

*Under the Guidance of*

**Dr. S. SANKARA NARAYANAN**
**Assistant Professor, Department of Computing Technologies**

*in partial fulfillment of the requirements for the degree of*

**BACHELOR of TECHNOLOGY**
**in**
**COMPUTER SCIENCE AND ENGINEERING**

**DEPARTMENT OF COMPUTING TECHNOLOGIES**
**COLLEGE OF ENGINEERING ANDTECHNOLOGY**
**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**KATTANKULATHUR- 603 203**

**MAY 2025**

Department of Computing Technologies

**SRM Institute of Science & Technology**

**Own Work Declaration Form**

| | |
|---|---|
| **Degree/ Course** | **: B.Tech Computer Science and Engineering** |
| **Student Name** | **: O.SAI RISHITHA, A.SIVA NARAYANA** |
| **Registration Number** | **: RA2111003011566, RA2111003011444** |
| **Title of Work** | **: Metabolic syndrome detection and healthcare using genetic algorithm** |

We hereby certify that this assessment compiles with the University's Rules and Regulations r relating to Academic misconduct and plagiarism , as listed in the University Website, Regulations, and the Education Committee guidelines.

We confirm that all the work contained in this assessment is our own except where indicated, and that We have met the following conditions:

- Clearly referenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not our own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that we have received from others (e.g.fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

We understand that any false claim for this work will be penalized in accordance with theUniversity policies and regulations.

**DECLARATION:**

We are aware of and understand the University's policy on Academic misconduct and plagiarism and we certify that this assessment is our own work, except where indicated by referring, and that we have followed the good academic practices noted above.

| | |
|---|---|
| A. SIVA NARAYANA<br>[RA2111003011444] | O.SAI RISHITHA<br>[RA2111003011566] |

# SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
# KATTANKULATHUR – 603 203

## BONAFIDE CERTIFICATE

Certified that **18CSP109L - Major Project** report titled "**Metabolic syndrome detection and healthcare using genetic algorithm** "is the bonafide work of "**O.SAIRISHITHA[RA2111003011566],A.SIVANARAYANA[RA2111003011444]"** who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

| | |
|---|---|
| **SIGNATURE** | **SIGNATURE** |
| **Dr. S. SANKARA NARAYANAN** | **Dr. G. NIRANJANA** |
| **ASSISTANT PROFESSOR**<br>DEPARTMENT OF<br>COMPUTING TECHNOLOGIES | **PROFESSOR & HEAD**<br>DEPARTMENT OF COMPUTING<br>TECHNOLOGIES |
| Internal Examiner | External Examiner |

# ACKNOWLEDGEMENT

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. Leenus Jesu Martin M,** Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout .

We encompass our sincere thanks to **Dr. M. Pushpalatha**,Professor and Associate Chairperson - CS ,School of Computing and **Dr. C.Lakshmi,** Professor and Associate Chairperson – AI, School of Computing , SRM Institute of Science and Technology, for their invaluable project work support.

We are incredibly grateful to our Head of the Department, **Dr. G. Niranjana**, Professor, Department of Computing Technologies, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators , Panel Head, **Dr. S. Saravanan** and Panel Members, **Dr. S. Sankara Narayanan** , **Dr. V. S. Saranya**, Department of Computing Technologies, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisors**, Dr. K. Pradeep Mohan Kumar Associate Professor , Dr. K . Kishore Anthuvan Sahayaraj Assistant Professor** Department of Computing Technologies , SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. S. Sankara Narayanan** , Department of Computing Technologies, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his mentorship. He provided us with the freedom and support to explore the research topics of our interest. His passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff and students of Computing Technologies, School of Computing, SRM Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement

O.SAI RISHITHA[RA2111003011566]

A.SIVA NARAYANA[RA2111003011444]

# ABSTRACT

Metabolic syndrome (MetS) is a severe health condition affected by many factors like lifestyle, genetics, and metabolic irregularities. It increases the risk of cardiovascular diseases and type 2 diabetes. included by many symptoms like high blood pressure, elevated blood sugar levels, obesity, and abnormal cholesterol, MetS is often difficult to diagnose early due to overlapping symptoms. so normal treatment frequently fails to deliver effective outcomes. This project aims to overcome these challenges by identifying optimal feature subsets for the early prediction of MetS. we propose a hybrid feature selection model that combines both filter and wrapper methods. Initially, we use the correlation coefficient (cc) method — a filter technique — to extract the most relevant features from the patient dataset. These selected features are then refined using a Genetic Algorithm (GA), a wrapper method, to boost prediction accuracy while minimizing computational overhead. for prediction, we implement a Bayesian network enhanced by a pre-training phase using autoencoders. Bayesian model is optimized further using a hill-climbing approach and refined with genetic algorithms, ensuring a more accurate and reliable prediction outcome. Additionally, the framework integrates severity classification to evaluate the progression of MetS, enabling more focused interventions. A personalized recommendation system is also developed, offering healthcare, lifestyle, and yoga-based suggestions based on each patient’s profile -promoting a more effective and patient-oriented treatment plan.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| ABBREVIATIONS | FULL FORM |
|---|---|
| **METS** | Metastasis |
| **GA** | Genetic Algorithm |
| **DL** | Deep Learning |
| **ROC** | Receiver Operating Characteristics |
| **EDA** | Exploratory Data Analysis |
| **RF** | Random Forest |

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

Metabolic Syndrome (MetS) is a serious and growing health issue that affects a large portion of the global population. It is not a single disease but rather a group of conditions that occur together—increased blood pressure, high blood sugar, excess body fat around the waist, and abnormal cholesterol or triglyceride levels. significantly raise the chances of developing heart disease, type 2 diabetes, and stroke. Humans following sedentary lifestyles, unhealthy eating habits, the occurrence of MetS is increasing rapidly, making early diagnosis and effective management more important than before. One of the major challenges in dealing with MetS is that its symptoms often overlap with other metabolic disorders. This makes it difficult to detect in the early stages, and as a result, many cases go unnoticed until it becomes serious. In addition, traditional treatment approaches which may not work effectively for every patient due to differences in individual health profiles.

There are several common methods used for feature selection:

1. **Filter Methods:** These are simple and fast. They rank features based on statistical measures like correlation or mutual information. However, they often ignore the interactions between features, which can reduce their effectiveness.

2.**Wrapper Methods:** These use a specific machine learning model to evaluate different subsets of features and choose the best-performing set. Examples include Genetic Algorithms (GA) and Recursive Feature Elimination (RFE). While more accurate, they can be time-consuming.

3.**Embedded Methods**: These integrate feature selection into the model training process itself. Techniques like Lasso Regression or Decision Trees fall into this category. They balance performance and efficiency but still may not capture complex relationships in medical data.

For prediction models, algorithms such as Decision Trees, Random Forests SVMs, Logistic Regression, and Neural Networks are commonly used. these models often lack explainability which is crucial in taking decisions on medical scenarios.

## 1.2 Problem Definition

The main challenge with MetS is that its symptoms often overlap and remain undetected in the early stages, making timely diagnosis difficult. Additionally, most existing diagnostic and treatment methods depends on general practices, which may not meet the needs of every patient. This can lead to poor health outcomes and inefficient treatment plans.

The exact problem addressed in this project is the lack of an accurate, early prediction model for MetS and the absence of personalized treatment recommendations. To solve this, the project proposes a hybrid machine learning-based framework that combines feature selection and prediction techniques to:

- Identify the most relevant risk factors of MetS,
- Predict the onset of the syndrome with high accuracy, and
- Provide patient-specific healthcare and lifestyle recommendations for better management of the condition.

This approach aims to move beyond one-size-fits-all treatments by offering an intelligent, data-driven, and personalized solution for early Mets diagnosis and care. Based on the patient data, MetS is predicted, after concluding it as positive Mets score is calculated. Based on the MetS score we classify the severity into high, medium, and low. So, the medications are provided based on the severity classifications.

# CHAPTER 2

# LITERATURE SURVEY

## 2.1 SURVEY

Recent research has shown that feature selection has been a critical area of study in enhancing the performance of machine learning models across various domains. Also, the importance of improving accuracy, reducing computational complexity, and addressing high-dimensional datasets has also been highlighted well enough. The authors of [1] explore the application of gene expression data for disease prediction highlighting feature selection methods that would then help increase accuracy. It compares two different feature selection techniques namely Niques-Laplacian Score (a filter method) and Top-Scoring Pair (a hybrid method) to classify leukemia and breast cancer datasets by using an SVM classifier. From the results, it is found that the feature selection significantly improves classification accuracy when compared to using the full dataset where the Laplacian Score is used to achieve the highest accuracy. This underscores the importance of selecting relevant features to enhance disease prediction models and also improve healthcare automation. Another research work, reported in [2], brings to light that optimizing feature selection in gene expression data by using an improved SVM-RFE algorithm will also significantly enhance classification accuracy while reducing the computational time of the process. This method eliminates multiple redundant features per iteration and outperforms traditional approaches. The results highlight that the efficient feature selection helps to improve the model performance while making it more and more suitable for high-dimensional datasets like cancer diagnosis [2].

The importance of using advanced machine learning techniques such as Random Forest (RF) and Support Vector Machine (SVM) to predict diabetes with improved accuracy is also highlighted. Feature selection techniques like step forward and backward selection also play a crucial role in improving the model's performance by identifying the key influencing factors for that model. Dimensionality reduction using PCA also helps simplify the model but its effect is limited to smaller datasets.

The authors of [3] show that the RF outperformed every other model at that time by achieving a staggering accuracy of 83%. These insights highlight how selecting relevant features for better predictive performance and also suggest that proper preprocessing and model selection will lead to improved healthcare solutions for diabetes diagnosis [3]. A novel approach is presented in [4] that highlights the feature selection process in microarray cancer gene expression data done by combining a correlation coefficient filter with a fuzzy rough quick redact algorithm. This method helps to reduce the high-dimensional gene data very effectively while also improving the classification accuracy. The significance of feature selection in handling the "large p, small n" problem where there is a vast number of gene features that are present with limited samples is highlighted well. Finally, the conclusions demonstrate that the proposed approach also enhances classifier performance by selecting the most relevant genes and then eliminating redundancy which makes it a very valuable tool for both cancer diagnosis and treatment [4].

The work presented in [5] evaluates the use of GAs for feature selection in breast cancer diagnosis. By experimenting with various machine learning classifiers like Random Forest and Logistic Regression, the authors conclude that the GA- based approach significantly increases the accuracy of the classifier. This underscores that the GAs are used predominantly to improve diagnostic precision in the detection of breast cancer [5].

A research work presented in [6] showcases how evolutionary algorithms improve the learning phase of bayesian networks, the bayesian network selects the best features to construct an optimal bayesian network that improves the prediction accuracy of the model. This work shows how evolutionary algorithms affect the attribute ordering which affects the overall prediction accuracy of the model. The work presented in [7] discusses how the bayesian network helps in uncovering conditional dependencies among health factors which the linear model misses, in the research the bayesian network models the relationships among several health factors: Waist circumference, HDL cholesterol, triglycerides, blood pressure, and glucose levels. This network is used in actual health checkups to predict risk.

The authors of [8] proposed a model that helps in identifying the major influential factors of metabolic syndrome in patients with major depressive disorder. The authors mentioned that patients with such conditions often have hidden risk factors of metabolic syndrome that the standard model can't identify.

A bayesian network model is constructed using the clinical and lifestyle data to study the probabilistic relationships among variables like antidepressant use, age, BMI, blood lipids, and blood pressure. The Bayesian network helped clinicians identify which variables directly influenced metabolic syndrome which might be overlooked. The work presented in [9] addresses how high dimensional biomarker datasets make it difficult to select features that contribute meaningfully to classification. In this work the bayesian network is used as a classifier after feature selection. Genetic algorithm is used for optimized feature selection from the data. These subsets of features are then evaluated by constructing Bayesian networks and selecting the subset with best prediction accuracy. The model achieved higher accuracy with reduced dimensionality compared to other ways of feature selection. Another research work presented in [10] focused on improving the efficiency and quality of Bayesian Network structural learning. Bayesian network learning is computationally expensive and is sensitive to variable ordering, so the authors proposed a way to reduce computational cost by GA to evolve and select the best order of variables, leading to fewer edges and better generalizing in the resulting BN.

Study mentioned in [11] explored the use of Metabolic Syndrome Severity score in diverse populations across New Zealand. The study found that usage of MetS score is more effective than traditional ways of MetS classification, the study also found that people with similar categorical MetS diagnosis might have different severity levels, impacting their health outcomes. The authors in [12] challenged the traditional way of binary classification(present/absent)of metabolic syndrome, they proposed a new way of severity-based approach which would offer better predictive power for clinical outcomes like diabetes and cardiovascular diseases. The study also highlighted the limitations of dichotomous diagnosis, such as inability to account for cumulative or partial risk factors.

Another research work presented in [13] validated the application of MetS severity scores among middle-aged and elderly Iranians. It demonstrated how MetS score could reflect the nuanced differences in metabolic health across individuals that categorical MetS failed to capture, the score also helped in identifying the high-risk groups and guiding in early stages. A landmark study presented in [14] developed a regression-based formula to compute a continuous metabolic syndrome score using factors like waist circumference, blood pressure, triglycerides, HDL-C, and fasting glucose.

Based the age and sex the formula is adjusted which helped in increasing the accuracy in risk prediction. This study has been validated against future incidences of type-2 diabetes and cardiovascular conditions, confirming its reliability as a predictive biomarker for metabolic deterioration. The work presented in [15] linked the MetS severity score with mortality in patients suffering from Nonalcoholic Fatty Liver Disease, this study proved that with higher severity scores the subject will be prone to get diagnosed with liver-related complications and premature death. These findings highlight the usage of Severity score in managing liver related diseases and also identifying high-risk individuals who may need intensive care. The work presented in [16] is concerned on designing and evaluating the Health Recommendation System (HRS) using Deep Learning-based Collaborative Filtering.

The proposed solution of the study integrates RBM and CNN for collaborative filtering such that it does the following: Incredible Accuracy, recall, and F-measure from existing systems. Excellent pre-processing with sentiment analysis and patient profile build-up of healthcare datasets. Recommendation systems have been outstanding in resolving information overload problems in different areas, including health. Major innovations in this area include:

- **Collaborative filtering (CF):** Users predict what users prefer based on their similarities to other users or items. Such as in [1] emphasized in their study: This work shows CF's promise applicability in health contexts.
- **Content-based Filtering:** Focusing on item attributes plus user preferences, it is commonly used in recommendation engines for health services.
- **Hybrid Filtering:** Combining CF and content-based methods improves prediction accuracy to solve problems like the "cold-start problem"

  Deep Learning in Recommendation Systems:
- **Improved Personalization:** These include Restricted Boltzmann Machines (RBM) and Convolutional Neural Networks (CNN) that efficiently allow the extraction of features from the complex healthcare datasets.
- **Technology Scale and Accuracy**: Deep learning methods will provide precision and adaptability cost-effectively at large scales along with healthcare data.

- Study mentioned in [17] The Mediterranean Diet (MedDiet) is underlined primarily for cardiovascular health, as well as to help patients with MetS, and include fruits, vegetables, whole grains, legumes, and olive oil.

Other nutritional plans, such as the DASH diet, New Nordic Diet, and vegetarian diets, have been noted to different extents in ameliorating the components of MetS. Another study mentioned in [18] explains how the PREDIMED-Plus study strongly supports lifestyle changes, namely increased physical activity and adherence to MD, in ameliorating MetS severity; further research is needed to study the longitudinal impact and applicability in different settings.

### Dietary Patterns and Nutritional Factors

Adherence to the Mediterranean diet (MD) inversely correlates with MetS severity. Foods associated with high MD adherence reduce inflammation and enhance metabolic profiles and comprise a high intake of fruits and vegetables, whole grains, olive oil, and moderate amounts of wine. On the contrary, pro-inflammatory dietary patterns with excessive consumption of saturated fats, processed meats, and refined carbohydrates worsen MetS components.

### Innovative Approaches to Measuring MetS

Traditional binary classifications of MetS often fail to encapsulate its nuanced severity. Metabolic Syndrome Severity Score (MetSSS) seeks to address this limitation by quantifying cumulative risk, thus allowing an even more detailed analysis of risk factors and their association with lifestyle and dietary behaviors.

Interrelation Between Depression and MetS

Emerging evidence shows a bidirectional association between MetS and depression, whereby severity increases with elevated depression scores. Shared risk factors such as sedentary behavior, dietary habits, and systemic inflammatory responses influence this association.

work presented in [19] mentioned how the global burden of type 2 diabetes mellitus (T2DM) continues to rise, necessitating valid and personalized healthcare solutions. Exercise helps in the management of T2DM through glycemic control and cardiovascular risk reduction. However, exercise prescription requires the consideration of personal factors such as physical condition, lifestyle, and preferences.

Exercise prescription with FITT principles (frequency, intensity, time, and type) have been found to improve the quality of life of patients with T2DM. It highlights the importance of personalizing exercise according to individual metabolic rates and physical capabilities.

The authors of [20] mentioned about 30% of the global population suffers MetS. It is made up of a number of different morbidities, including obesity, diabetes, CVD, and neurodegenerative conditions. Those are mainly caused by oxidative stress, insulin resistance, and dysbiosis of gut microbiota, which deserve high priority for lifestyle change more particularly dietary intervention and physical activity for effective management of MetS.

### Nutrition

Dietary Polyphenols may be helpful in countering oxidative stress, inflammation, and insulin resistance. Plant polyphenols abundantly found in fruits, vegetables, and whole grains have attracted interest for the potential therapeutic use; however, low bioavailability and absence of long-term human studies remain their main drawbacks.

**Mediterranean Diet**: The MedDiet is rich in unsaturated fats, fiber, and polyphenols that improve lipid profile, decrease inflammation, and reduce MetS incidence. Extra virgin olive oil and fish, its 2 cornerstone foods, enhance CS insulin sensitivity and reduce CVD incidence risk.

- **Gut Microbiota Modulation**: Pre- and probiotics and dietary measures for gut microbiota restoration may act to down-regulate inflammation and enhance glucose metabolism.
- **Exercise**: Moderate to vigorous physical activity is essential in ameliorating MetS, promoting energy balance, and enhancing insulin sensitivity and cardiovascular health.
- **Weight Loss**: Involving the continual maintenance of modest weight loss over some time (5-10%) is of great clinical benefit to MetS components such as glucose level and hypertension.
- **Intermittent Fasting**: Being the newest one of the lot, this appears to be a reasonable alternative for assisting in the management of body weight and other metabolic parameters without the need to alter caloric intake.
- **Antioxidants: A**ntioxidants including carotenoids and vitamins are important in combating oxidative stress and metabolic problems arising thereof, and more so with selenium. Polyphenols, chiefly flavonoids, are very potent in getting rid of oxidative stress and promote cardiometabolic health.

The study in [21] presents a hybrid model that combines Genetic Algorithms (GA) with Bayesian ARTMAP to enhance the accuracy of MetS diagnosis.

The GA is used to optimize the network structure and feature selection, resulting in improved classification performance. The model demonstrates robustness and adaptability in handling noisy biomedical datasets. Another study mentioned in the paper [22] introduces a mobile health application designed to monitor risk factors and guide patients with personalized interventions for MetS management. It integrates real-time data collection with rule-based decision-making to offer early alerts and lifestyle recommendations. The system shows promise in enhancing patient engagement and prevenive care.

The authors in [23] propose a GA-based feature selection method to identify the most relevant attributes for diabetes prediction, which shares risk factors with MetS. The method significantly improves the predictive accuracy of multiple classifiers while reducing model complexity.

The approach demonstrates the effectiveness of evolutionary algorithms in medical data analysis. The work in [24] applies deep learning and collaborative filtering to create a personalized health recommendation system. By learning user preferences and health patterns, the model delivers tailored advice for diet and activity. The system showcases high accuracy in personalized suggestion generation, paving the way for scalable digital health platforms.

The authors in [25] utilize Bayesian Optimization to fine-tune machine learning models for predicting MetS risk. The method efficiently explores hyperparameter spaces, yielding higher accuracy and faster convergence. This approach enhances model reliability, especially in clinical decision support systems.

The study in [26] evaluates several machine learning algorithms to predict MetS in primary healthcare settings using demographic and clinical data. The models achieved high performance, with Random Forest and XGBoost being top performers. The study emphasizes ML’s potential in early preventive strategies at the community level.

## 2.2 Motivation / Challenges

**Rising global health burden of MetS**

The increasing prevalence of metabolic syndrome due to sedentary lifestyles, poor dietary habits, and stress has made it a major public health concern. Early diagnosis and prevention can significantly reduce the risk of serious conditions like cardiovascular disease and type 2 diabetes.

**Need for accurate and personalized diagnosis**

Traditional diagnostic methods often apply a one-size-fits-all approach, ignoring individual variations in symptoms and risk factors. This project aims to overcome that limitation by using personalized, data-driven techniques for accurate prediction and tailored interventions.

**High-dimensional medical data challenges**

Medical datasets are often large and complex, making it essential to select the most relevant features for efficient and interpretable model performance. This motivated the use of a hybrid feature selection technique combining statistical and evolutionary approaches.

**Bridging prediction with actionable treatment**

Predicting a condition is not enough; what patients need is a clear, personalized pathway for managing their health. This project goes beyond prediction by offering customized healthcare and diet plans, helping users take proactive steps toward recovery and prevention.

## 2.3 Summary of the Survey and Findings

One of the themes that lit up from the surveyed study is that feature selection has played a very important role in improving the performance of the model in machine learning applications in different fields of health care. Most of the studies address the issue of high-dimensional datasets, such as gene expression data or metabolic health records, and identifying the requirement for creating suitable models that are efficient and interpretable.

Selecting a feature not only increases the classification accuracy but also cuts down the computation overhead required to make it more scalable and robust.

Such studies as [1], [2], and [4] constitute the authority in the contribution of filter and wrapper or hybrid approaches to feature selection.

Improvement witnessed in disease prediction models such as leukemia and breast cancer is as well attributed to techniques like the Laplacian Score, Top-Scoring Pair, and SVM-RFE.

Genetic Algorithms (GAs), as discussed in [5] and [9], have been proven to select most optimal feature subsets for models that markedly improve their accuracy while significantly reducing some redundancy.

Bayesian Networks (BNs) have emerged as one of the strongest tools in dependency modeling among health-related variables, as pointed out in studies [6], [7], and [8]. BNs

help integrate clinical and lifestyle data while exposing some hidden risks that may not be reflected or accurately considered in traditional models; so they are very useful in conditions like metabolic syndrome and depression.

The other finding in the research indicates that evolutionary algorithms can also further enhance BN learning-theory by optimizing variable order and structure to increase predictive performance.

Traditional binary classification from Metabolic Syndrome (MetS) becomes a scoring based on its severity-the emergence of which can be seen in [11], [12], and [14].

According to such studies, MetS Severity Scores (MetSSS) were better at predicting conditions like Type 2 Diabetes Mellitus (T2DM), cardiovascular diseases, and Non-Alcoholic Fatty Liver Disease (NAFLD). Severity scoring captures more subtle variations in metabolism across individuals and thus has implications for early intervention and personalized medicine.

Nutritional and lifestyle measures, like the Mediterranean Diet (MedDiet), as seen in [17] to [20], and based on FITT principles (Frequency, Intensity, Time, Type) physical activity prescriptions, were also effective at reducing MetS severity.

Interventions such as short-term fasting, gut microbiota modulation via diet, and polyphenol-rich diets also showed promise in reducing the inflammation, insulin resistance, and obesity contributing to MetS.

Finally, advances in healthcare recommender systems toward deep learning, such as RBM and CNN-based collaborative filtering ([16]), give good indications on the extent to which decision support systems would benefit from artificial intelligence.

In small terms, the survey reveals that feature selection, Bayesian modeling, MetS severity grading, and other diversified approaches powerfully hold the future for advancing precision medicine.

## 2.4 – Objectives of the Work

**Early Prediction of Metabolic Syndrome (MetS):**

To create a model for predicting the individuals who are at risk of developing MetS at an early stage so that preventive measures can be taken in time, thereby minimizing the chances of coming up with the serious health complications later.

**Optimal Feature Selection:**

Introduce a hybrid strategy for feature selection that implements filter and wrapper methods to find the best features. This is needed to improve model accuracy and reduce computation costs as it only considers essential variables.

**Personalized Lifestyle Recommendation:**

It should provide tailored recommendations regarding diet and exercise and other lifestyle changes through prediction outcomes in the management or prevention of MetS through the personalized intervention mechanisms.

# CHAPTER 3

## PROPOSED WORK

### 3.1 Architecture of the System

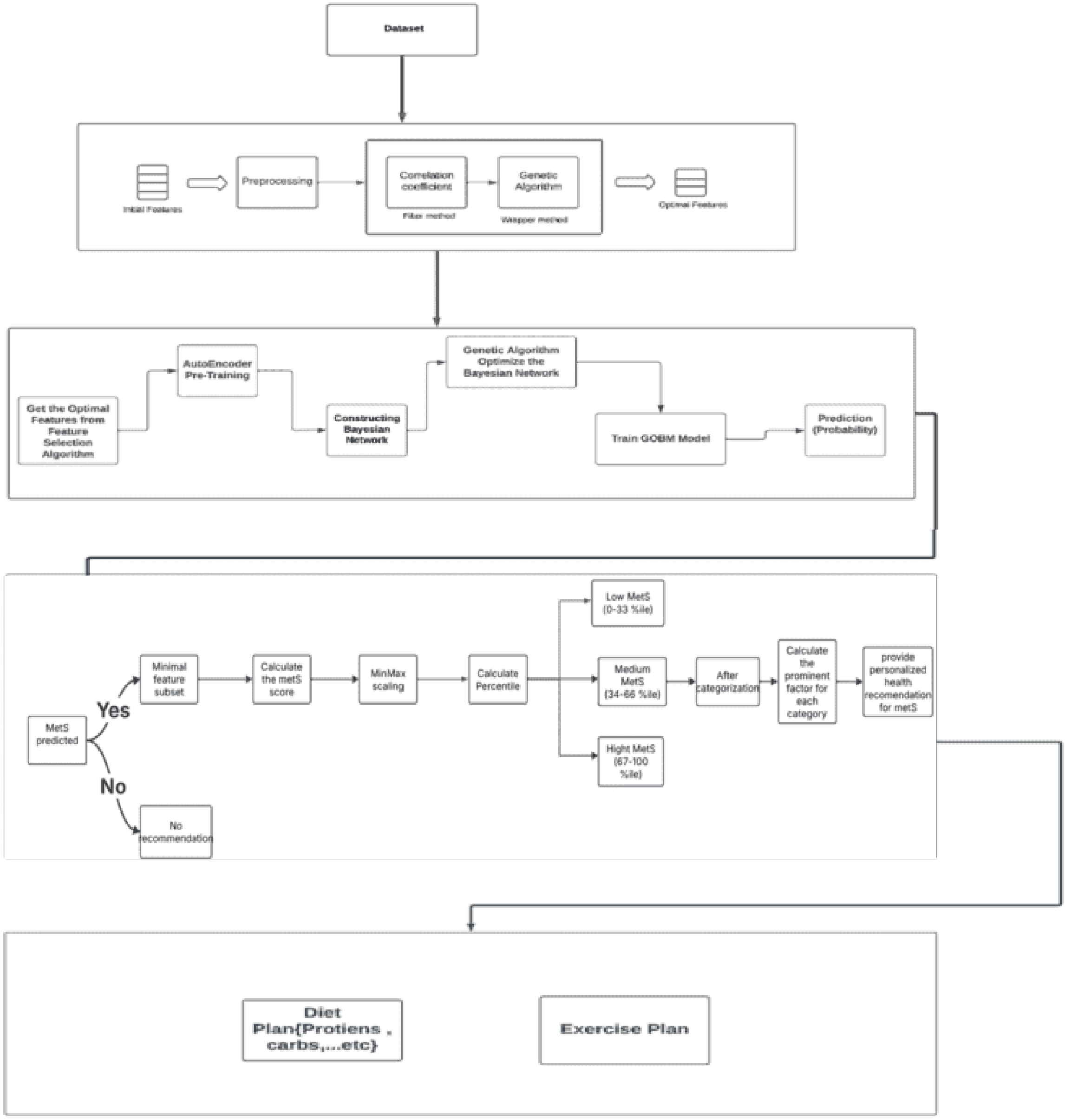


**Fig 3.1**: Architecture of the system

### 3.1.1 Phases of the Architecture

In our proposed architecture (Fig 3.1) we have four phases of solving our problem statement

1. Feature Selection Phase
2. Prediction of Metabolic Syndrome Phase
3. Mets_Score calculation and Severity Classification Phase
4. Personalized Healthcare and Dietary Plan Recommendation Phase

To find solution for our problem statement, we will be having initial dataset which has all the features related to metabolic syndrome, from that we will start our process

**Phase -1:**

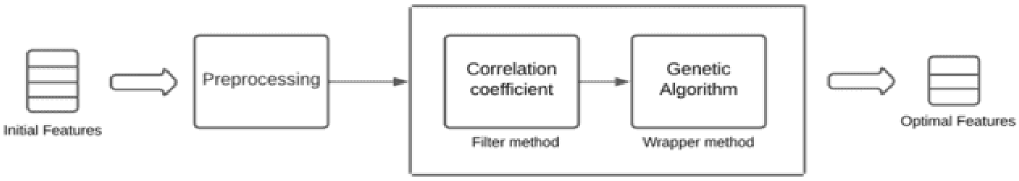


**Fig 3.2:** Feature Selection Architecture

The architecture of this phase is depicted in fig 3.2, in this phase the entire dataset features will be sent and we will get the reduced subset of features which are most contributing to metabolic syndrome

**Phase -2:**

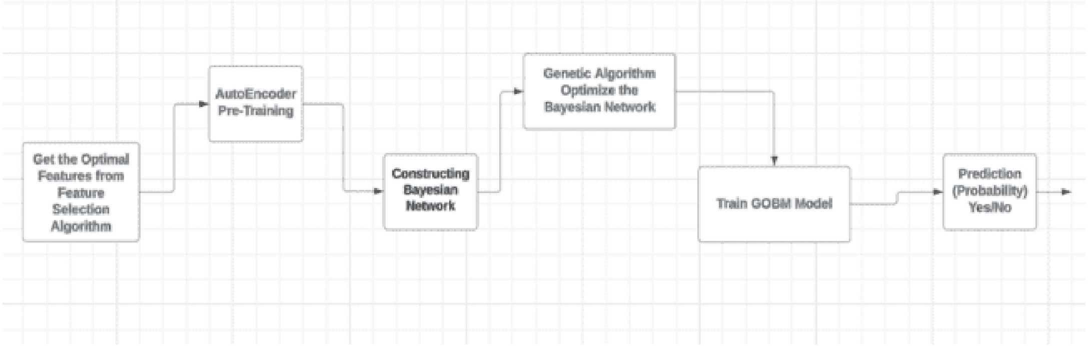

Get the Optimal Features from Feature Selection Algorithm
AutoEncoder Pre-Training
Constructing Bayesian Network
Genetic Algorithm Optimize the Bayesian Network
Train GOBM Model
Prediction (Probability) Yes/No

**Fig 3.3:** Architecture of Prediction of Mets

Fig 3.3 represents phase 2 architecture where prediction of the patient has metabolic syndrome through Genetically optimized Bayesian Network.

## Phase -3:

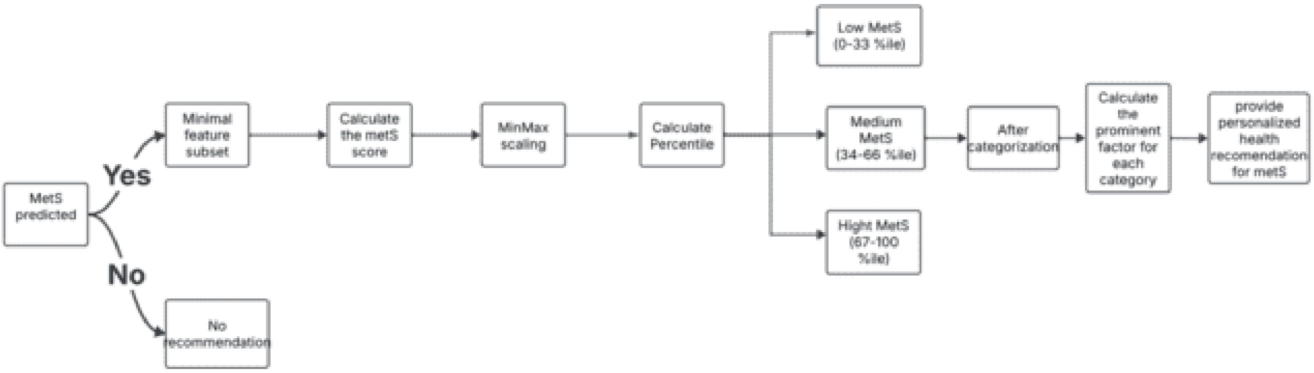


**Fig 3.4:** Architecture of Severity Score calculation and Classification

In this Phase (fig 3.4) we will calculate severity score and classify the patients who have metabolic syndrome into Low, Medium, High categories.

## Phase -4:

Diet Plan{Protiens , carbs,...etc}
Exercise Plan

**Fig 3.5:** Architecture of Personalized Healthcare and Dietary Plan

In this Phase / Final Phase after the patient classified as they have metabolic syndrome and severity category (Low, Medium, High), personalized healthcare plan and dietary plan (fig 3.5) will be provided based on gender, age, blood rate.

## 3.2 Algorithm Design

In our proposed architecture (Fig 3.1) we have four phases of solving our problem statement

1. Feature Selection Phase
2. Prediction of Metabolic Syndrome Phase
3. Mets_Score calculation and Severity Classification Phase
4. Personalized Healthcare and Dietary Plan Recommendation Phase

### 3.2.1 Modules Description

### Module – 1: Feature Selection (Fig 3.2)

### Filter Method:

Correlation coefficient: Choosing the correlation coefficient as a filter method for feature selection in the context of metabolic syndrome detection has several advantages compared to other filter methods. Many features associated with metabolic syndrome (e.g., BMI, blood pressure, glucose levels) are continuous numerical variables.

### Why Correlation?

The correlation coefficient effectively quantifies the linear relationships between these continuous features and the target variable (e.g., presence or absence of metabolic syndrome), making it a suitable choice for this type of data.

### Wrapper Method:

Genetic Algorithm: Using a wrapper method with a Genetic Algorithm (GA) for feature selection in early detection of metabolic syndrome can be an optimal approach.

### Why Genetic Algorithm?

While there are evolutionary algorithms that may outperform GA in specific scenarios or datasets related to metabolic syndrome, GA remains a strong choice due to its robustness and diversity.

GA is robust in handling noisy data and uncertainties, which is common in medical datasets.
GA's mechanisms for maintaining diversity through mutation and crossover allow it to explore the feature space effectively.

## Module 2 – Prediction of Mets (Fig 3.3)

**Autoencoders Pre-Training**: Autoencoders compress the input features to capture essential patterns while reducing dimensionality. This process enhances the feature representation for better downstream performance.

**Building Bayesian Network**: A Bayesian Network is constructed using the learned feature representations to model probabilistic relationships. This network allows for inference regarding the presence of Metabolic Syndrome.

**Genetic Optimization**: Genetic algorithms are used to optimize the structure and parameters of the Bayesian Network. This optimization aims to improve the model's predictive accuracy and efficiency.

**Train the Model:** The optimized Bayesian Network is trained on the dataset to learn the associations with Metabolic Syndrome. Training adjusts the model parameters to minimize prediction errors.

**Predict Metabolic Syndrome:** The trained model predicts the likelihood of Metabolic Syndrome in new patients. It outputs a probability score those aids in early diagnosis and intervention.

## Module 3 – MetS Score Calculation and Classification Severity (Fig 3.4)

### Step -1 :

After Proven the patient have metabolic Syndrome, Calculate the MetS Score (Severity Score)

### Step – 2:

For this we require additional Features other than initial ones.

### STEP 3:

Severity score (MetS) =A+B

A = Bayesian network Probabaility (value between 0 and 1)

B = cMetS-S score derived using following equations (Fig.6)

### STEP 4:

Normalize the computer MetS score using Minmax scaler.

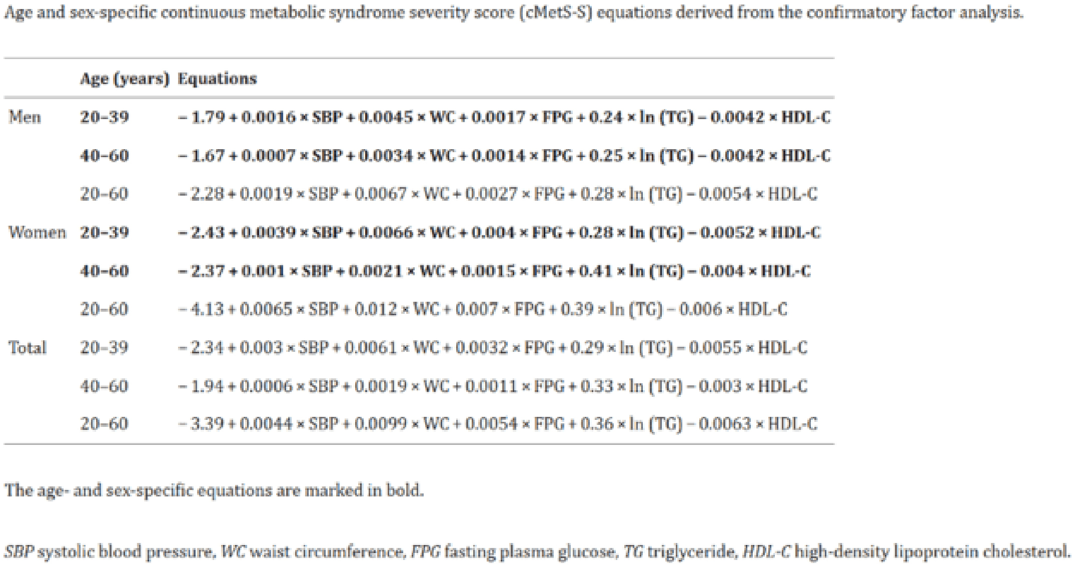
### STEP 5:

Calculate the percentile for calculated score.

### STEP 6: Severity Classification

Based on percentile classify the severity.

- 0-33%ile – **Low Severity**
- 34-66%ile – **Medium Severity**
- 67-100%ile – **High Severity**



Age and sex-specific continuous metabolic syndrome severity score (cMetS-S) equations derived from the confirmatory factor analysis.

|  | Age (years) | Equations |
|---|---|---|
| Men | **20–39** | **$-1.79 + 0.0016 \times \mathrm{SBP} + 0.0045 \times \mathrm{WC} + 0.0017 \times \mathrm{FPG} + 0.24 \times \ln(\mathrm{TG}) - 0.0042 \times \mathrm{HDL\text{-}C}$** |
|  | **40–60** | **$-1.67 + 0.0007 \times \mathrm{SBP} + 0.0034 \times \mathrm{WC} + 0.0014 \times \mathrm{FPG} + 0.25 \times \ln(\mathrm{TG}) - 0.0042 \times \mathrm{HDL\text{-}C}$** |
|  | 20–60 | $-2.28 + 0.0019 \times \mathrm{SBP} + 0.0067 \times \mathrm{WC} + 0.0027 \times \mathrm{FPG} + 0.28 \times \ln(\mathrm{TG}) - 0.0054 \times \mathrm{HDL\text{-}C}$ |
| Women | **20–39** | **$-2.43 + 0.0039 \times \mathrm{SBP} + 0.0066 \times \mathrm{WC} + 0.004 \times \mathrm{FPG} + 0.28 \times \ln(\mathrm{TG}) - 0.0052 \times \mathrm{HDL\text{-}C}$** |
|  | **40–60** | **$-2.37 + 0.001 \times \mathrm{SBP} + 0.0021 \times \mathrm{WC} + 0.0015 \times \mathrm{FPG} + 0.41 \times \ln(\mathrm{TG}) - 0.004 \times \mathrm{HDL\text{-}C}$** |
|  | 20–60 | $-4.13 + 0.0065 \times \mathrm{SBP} + 0.012 \times \mathrm{WC} + 0.007 \times \mathrm{FPG} + 0.39 \times \ln(\mathrm{TG}) - 0.006 \times \mathrm{HDL\text{-}C}$ |
| Total | 20–39 | $-2.34 + 0.003 \times \mathrm{SBP} + 0.0061 \times \mathrm{WC} + 0.0032 \times \mathrm{FPG} + 0.29 \times \ln(\mathrm{TG}) - 0.0055 \times \mathrm{HDL\text{-}C}$ |
|  | 40–60 | $-1.94 + 0.0006 \times \mathrm{SBP} + 0.0019 \times \mathrm{WC} + 0.0011 \times \mathrm{FPG} + 0.33 \times \ln(\mathrm{TG}) - 0.003 \times \mathrm{HDL\text{-}C}$ |
|  | 20–60 | $-3.39 + 0.0044 \times \mathrm{SBP} + 0.0099 \times \mathrm{WC} + 0.0054 \times \mathrm{FPG} + 0.36 \times \ln(\mathrm{TG}) - 0.0063 \times \mathrm{HDL\text{-}C}$ |

The age- and sex-specific equations are marked in bold.

*SBP* systolic blood pressure, *WC* waist circumference, *FPG* fasting plasma glucose, *TG* triglyceride, *HDL-C* high-density lipoprotein cholesterol.

**Fig 3.6:** MetS Score Formula

## Module 4: Personalized Healthcare Recommendation (Fig 3.5)

1.After classifying the Patient into LOW, MEDIUM, HIGH we now move into giving a Personalized Healthcare and dietary plan

2. Based on Age, Gender, and Severity, and other factors through which we can give more personalized healthcare to an individual.

### Lifestyle Interventions for MetS

Dietary Approaches

Physical Activity

Weight Management

### Specific Nutritional Components

Fats

Whole Grains and Legumes

Fruits and Vegetables

Dairy and Nuts

Sugars and Sweetened Beverages

### Role of Antioxidants

Antioxidants, such as carotenoids, vitamins, and selenium, play a significant role in reducing oxidative stress and its associated metabolic complications. Polyphenols, especially flavonoids, demonstrate potent antioxidant effects, improving cardiometabolic outcomes.

### Age and Socioeconomic Stratification

Age and socioeconomic factors modulate the relationship between lifestyle changes and MetS risk. Younger adults (40–49 years) exhibit greater responsiveness to lifestyle interventions, possibly due to faster metabolism and proactive health-seeking behaviors.

### More Personalized Healthcare and Diet Recommender

Recent advancements in AI have enabled the development of intelligent systems for personalized healthcare. For T2DM patients, genetic algorithms (GA) have been applied to create optimized exercise schedules by considering user-specific constraints, such as heart rate, body condition, and activity preferences. Systems like RUNNER and SHADE have demonstrated the potential of leveraging contextual data for generating tailored fitness recommendations.

## 3.3 Datasets for the Study and Platform

### 3.3.1 Dataset

**Cited**

Dataset posted on 2022-08-27, 14:56 authored by Yan Zhang, Xiaoxu Zhang, Jaina Razbek, Deyang Li, Wenjun Xia, Liangliang Bao, Hongkai Mao, Mayisha Daken, Mingqin Cao

### Dataset Description

The dataset utilized in this study is sourced from the research titled "Opening the black box: interpretable machine learning for predictor finding of metabolic syndrome", published on Springer Nature's Fig share platform. The dataset provides clinical and biochemical parameters that are essential for the prediction and analysis of metabolic syndrome using machine learning techniques.

The dataset consists of anonymized clinical and biochemical health records from multiple individuals, aimed at identifying risk factors associated with metabolic syndrome. It includes demographic, hematological, and metabolic parameters, such as age, gender, blood pressure, glucose levels, cholesterol levels, and various blood cell counts.

The target variable is **Metabolic Syndrome** (0 = no, 1 = yes), which helps in building predictive models for early identification. The data is well-structured and ideal for machine learning applications due to its numerical consistency and medically relevant features.

## Summary:

This dataset consists of 54 features collected from individuals to assess various health metrics, including indicators of metabolic syndrome (MetS). The primary focus is on blood parameters, lifestyle factors, and medical history, which are crucial for understanding metabolic health and related conditions.This comprehensive dataset is essential for healthcare research, particularly in understanding and managing metabolic syndrome and related health issues.

## List of Features and Their Descriptions

**Gender (0 = female, 1 = male):** Biological sex of the individual

**Age (years):** Age of the individual in years

**Neutrophil percentage (%):** Proportion of neutrophils among white blood cells

**Lymphocyte percentage (%):** Proportion of lymphocytes among white blood cells

**Neutrophil count ($10^9$/L):** Absolute count of neutrophils in the blood

**Mean red blood cell volume (FL):** Average size of red blood cells

**Mean haemoglobin concentration (g/L):** Average concentration of haemoglobin in RBCs

**RBC distribution width SD:** Variation in red blood cell sizes

**Mean platelet volume (FL):** Average size of platelets

Platelet distribution width standard deviation: Variability in platelet size

**BMI (kg/$m^2$):** Body Mass Index, a measure of body fat

**Waist circumference (cm):** Abdominal measurement used to assess fat distribution

**Systolic blood pressure (mmHg):** Pressure in arteries during heartbeats

**Diastolic blood pressure (mmHg):** Pressure in arteries between heartbeats

**Fasting blood glucose (mmol/L):** Blood sugar level after fasting

**Low-density lipoprotein cholesterol (mmol/L):** "Bad" cholesterol level

**Total cholesterol (mmol/L):** Combined measure of all cholesterol types

**Triglycerides (mmol/L):** Type of fat (lipid) found in blood

**High-density lipoprotein cholesterol (mmol/L):** "Good" cholesterol level

**Metabolic Syndrome (0 = no, 1 = yes):** Target variable indicating presence of metabolic syndrome

## Justification for selection

This dataset was selected for our project on Metabolic Syndrome Prediction and Personalized Healthcare Recommendation due to its comprehensive inclusion of medically significant parameters directly linked to the diagnostic criteria of metabolic syndrome. It offers a balanced combination of demographic, hematologic, and metabolic indicators that are essential for building robust predictive models.

Moreover, the presence of a clearly defined target variable and high-quality numerical data allows for effective training, evaluation, and interpretability of machine learning algorithms. Its relevance to real-world clinical conditions ensures that any insights or models developed from this dataset can be practically applied to improve preventive healthcare and personalized treatment plans.

## 3.3.2 Platform and Specifications

The experimental work and data analysis were conducted using the following platform and system configuration:

- Platform: Jupyter Notebook (via Anaconda Distribution)
- Programming Language: Python 3.9
- Key Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, shap, xgboost
- Operating System: Windows 10 (64-bit)

These specifications ensured efficient data preprocessing, modeling, and result visualization.

## Referred Datasets:

1.The National Health and Nutrition Examination Survey (NHANES) :

https://www.kaggle.com/datasets/nguyenvy/nhanes-19882018

2.Metabolic Syndrome A Comprehensive Dataset on Risk Factors and Health Indicators

https://www.kaggle.com/datasets/antimoni/metabolic-syndrome

3. Early-Stage Diabetes Risk Prediction Dataset

https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset

4. PIMA Indian Diabetes (PID) Source: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)

https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

# CHAPTER 4

# RESULTS AND INFERENCES

In this section, we present the outcomes obtained at each stage of our modular experiment pipeline and provide a detailed analysis of the results from each module. The entire process was designed to predict Metabolic Syndrome (MetS) with optimal accuracy and support the delivery of personalized healthcare interventions based on prediction severity.

Based on the experimental setup, we obtained results at each stage of our modular pipeline. The first step involved Feature Selection using a hybrid model (filter + wrapper methods) to extract the most relevant features. This reduced the dimensionality and ensured that only the most impactful variables were used for further modeling.

Next, we used autoencoders for pre-training to learn compressed representations of the selected features. These representations were then used to construct a Bayesian Network using BIC Score and the Hill Climb Search algorithm. This helped model the relationships among variables effectively, and after further optimization, we built a final structure suitable for accurate prediction of Metabolic Syndrome (MetS).

Based on the prediction outcomes, individuals were classified into three categories according to their risk severity. Depending on the severity level, we provided personalized healthcare and dietary plans, ensuring tailored recommendations for better health management.

## 4.1 Metrics for evaluation

### Feature Selection Module

Filter Method -Accuracy, Count of Features.

Wrapper (GA) - Fitness Score, Loss, Accuracy

### Prediction Module

Bic Score, Accuracy

### Analysis of Feature Selection Module

Random forest Classifier, Accuracy, Count of Features.

## 4.2 Parameter Settings

**Pre-Processng Techniques** – {Missing Values, Numeric Type Conversion...}

**Filter Method** (Correlation Coefficient) – Threshold = 0.08

**Genetic Algorithms (GA) -** Generations = 20, Population = 100

**Bayesian Network Construction** – Hill Climb Search, Bic Score, Threshold = 0.65 or 65%

**Autoencoders** – Encoding Dimension =6

### Severity Classification and MetS Score Calculation

### Thresholds:

0 to 0.3 -> Low Category

0.31 to 0.60 -> Medium Category

0.61 to 0.99 -> High Category

## 4.3 Results and Discussion

**Pre-Processng Techniques** – {Missing Values, Numeric Type Conversion...}

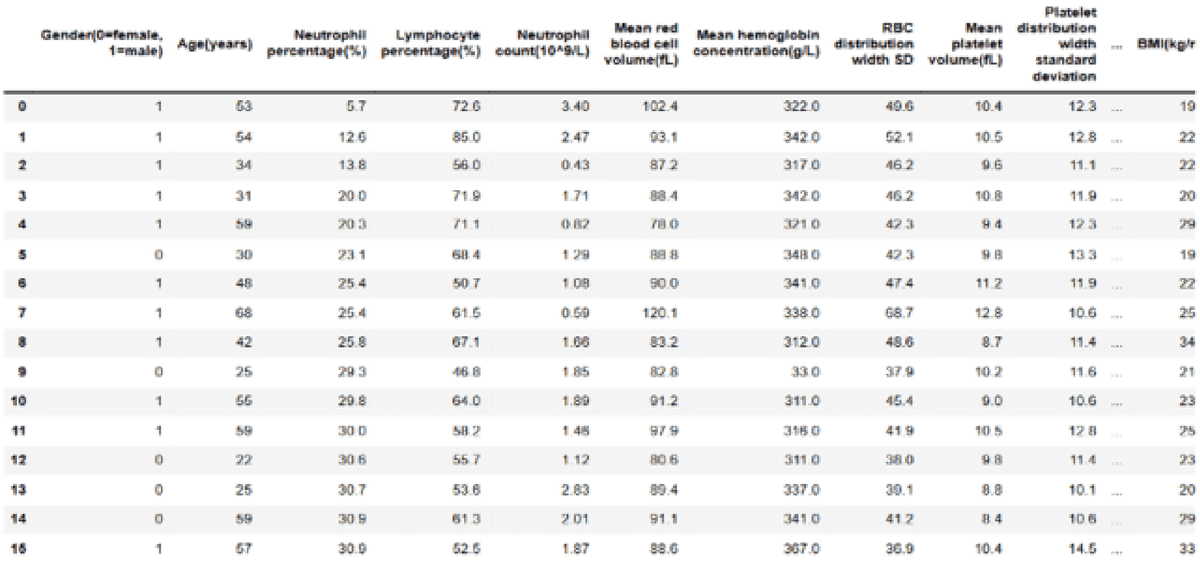
**Filter Method** (Correlation Coefficient) – Threshold = 0.08

**Genetic Algorithms (GA)** - Generations = 20, Population = 100

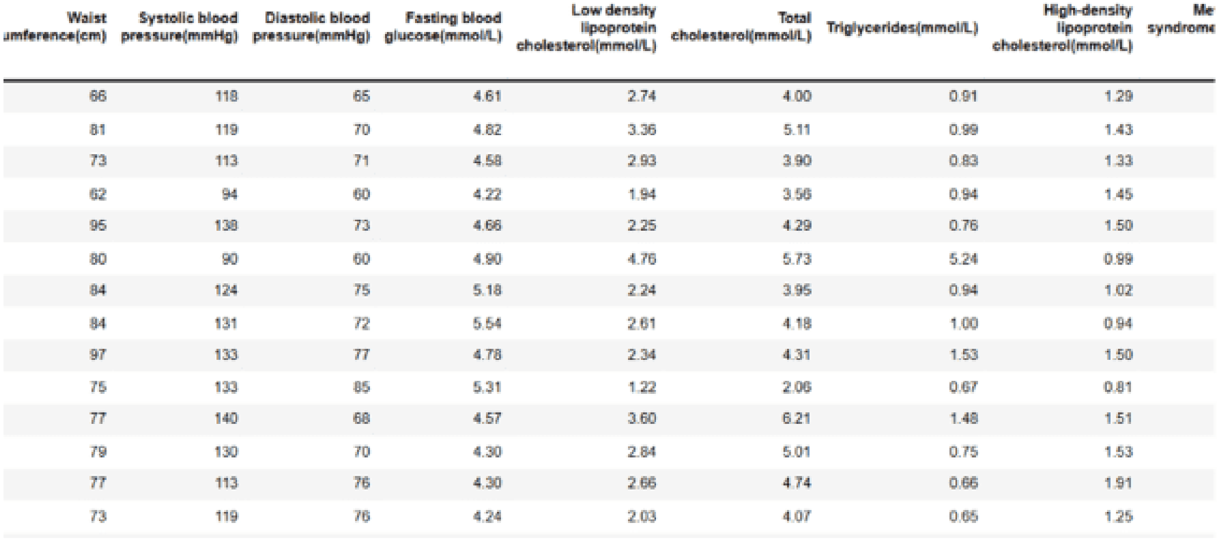
**Bayesian Network Construction** –Hill Climb Search, Bic Score, Threshold = 0.65 or 65%

**Autoencoders** – Encoding Dimension =6

## Initial Dataset:



| | Gender(0=female, 1=male) | Age(years) | Neutrophil percentage(%) | Lymphocyte percentage(%) | Neutrophil count(10^9/L) | Mean red blood cell volume(fL) | Mean hemoglobin concentration(g/L) | RBC distribution width SD | Mean platelet volume(fL) | Platelet distribution width standard deviation | ... | BMI(kg/r |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 53 | 5.7 | 72.6 | 3.40 | 102.4 | 322.0 | 49.6 | 10.4 | 12.3 | ... | 19 |
| **1** | 1 | 54 | 12.6 | 85.0 | 2.47 | 93.1 | 342.0 | 52.1 | 10.5 | 12.8 | ... | 22 |
| **2** | 1 | 34 | 13.8 | 56.0 | 0.43 | 87.2 | 317.0 | 46.2 | 9.6 | 11.1 | ... | 22 |
| **3** | 1 | 31 | 20.0 | 71.9 | 1.71 | 88.4 | 342.0 | 46.2 | 10.8 | 11.9 | ... | 20 |
| **4** | 1 | 59 | 20.3 | 71.1 | 0.82 | 78.0 | 321.0 | 42.3 | 9.4 | 12.3 | ... | 29 |
| **5** | 0 | 30 | 23.1 | 68.4 | 1.29 | 88.8 | 348.0 | 42.3 | 9.8 | 13.3 | ... | 19 |
| **6** | 1 | 48 | 25.4 | 50.7 | 1.08 | 90.0 | 341.0 | 47.4 | 11.2 | 11.9 | ... | 22 |
| **7** | 1 | 68 | 25.4 | 61.5 | 0.59 | 120.1 | 338.0 | 68.7 | 12.8 | 10.6 | ... | 25 |
| **8** | 1 | 42 | 25.8 | 67.1 | 1.66 | 83.2 | 312.0 | 48.6 | 8.7 | 11.4 | ... | 34 |
| **9** | 0 | 25 | 29.3 | 46.8 | 1.85 | 82.8 | 33.0 | 37.9 | 10.2 | 11.6 | ... | 21 |
| **10** | 1 | 55 | 29.8 | 64.0 | 1.89 | 91.2 | 311.0 | 45.4 | 9.0 | 10.6 | ... | 23 |
| **11** | 1 | 59 | 30.0 | 58.2 | 1.48 | 97.9 | 316.0 | 41.9 | 10.5 | 12.8 | ... | 25 |
| **12** | 0 | 22 | 30.6 | 55.7 | 1.12 | 80.6 | 311.0 | 38.0 | 9.8 | 11.4 | ... | 23 |
| **13** | 0 | 25 | 30.7 | 53.6 | 2.83 | 89.4 | 337.0 | 39.1 | 8.8 | 10.1 | ... | 20 |
| **14** | 0 | 59 | 30.9 | 61.3 | 2.01 | 91.1 | 341.0 | 41.2 | 8.4 | 10.6 | ... | 29 |
| **15** | 1 | 57 | 30.9 | 52.5 | 1.87 | 88.6 | 367.0 | 36.9 | 10.4 | 14.5 | ... | 33 |



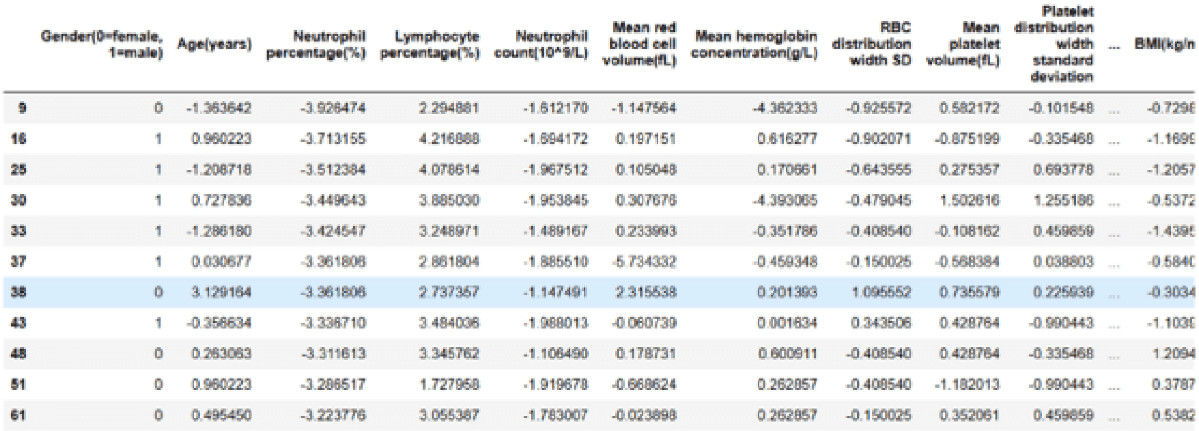
| Waist umference(cm) | Systolic blood pressure(mmHg) | Diastolic blood pressure(mmHg) | Fasting blood glucose(mmol/L) | Low density lipoprotein cholesterol(mmol/L) | Total cholesterol(mmol/L) | Triglycerides(mmol/L) | High-density lipoprotein cholesterol(mmol/L) | Me syndrome |
|---|---|---|---|---|---|---|---|---|
| 66 | 118 | 65 | 4.61 | 2.74 | 4.00 | 0.91 | 1.29 | |
| 81 | 119 | 70 | 4.82 | 3.36 | 5.11 | 0.99 | 1.43 | |
| 73 | 113 | 71 | 4.58 | 2.93 | 3.90 | 0.83 | 1.33 | |
| 62 | 94 | 60 | 4.22 | 1.94 | 3.56 | 0.94 | 1.45 | |
| 95 | 138 | 73 | 4.66 | 2.25 | 4.29 | 0.76 | 1.50 | |
| 80 | 90 | 60 | 4.90 | 4.76 | 5.73 | 5.24 | 0.99 | |
| 84 | 124 | 75 | 5.18 | 2.24 | 3.95 | 0.94 | 1.02 | |
| 84 | 131 | 72 | 5.54 | 2.61 | 4.18 | 1.00 | 0.94 | |
| 97 | 133 | 77 | 4.78 | 2.34 | 4.31 | 1.53 | 1.50 | |
| 75 | 133 | 85 | 5.31 | 1.22 | 2.06 | 0.67 | 0.81 | |
| 77 | 140 | 68 | 4.57 | 3.60 | 6.21 | 1.48 | 1.51 | |
| 79 | 130 | 70 | 4.30 | 2.84 | 5.01 | 0.75 | 1.53 | |
| 77 | 113 | 76 | 4.30 | 2.66 | 4.74 | 0.66 | 1.91 | |
| 73 | 119 | 76 | 4.24 | 2.03 | 4.07 | 0.65 | 1.25 | |

**Fig 4.1:** Initial Dataset

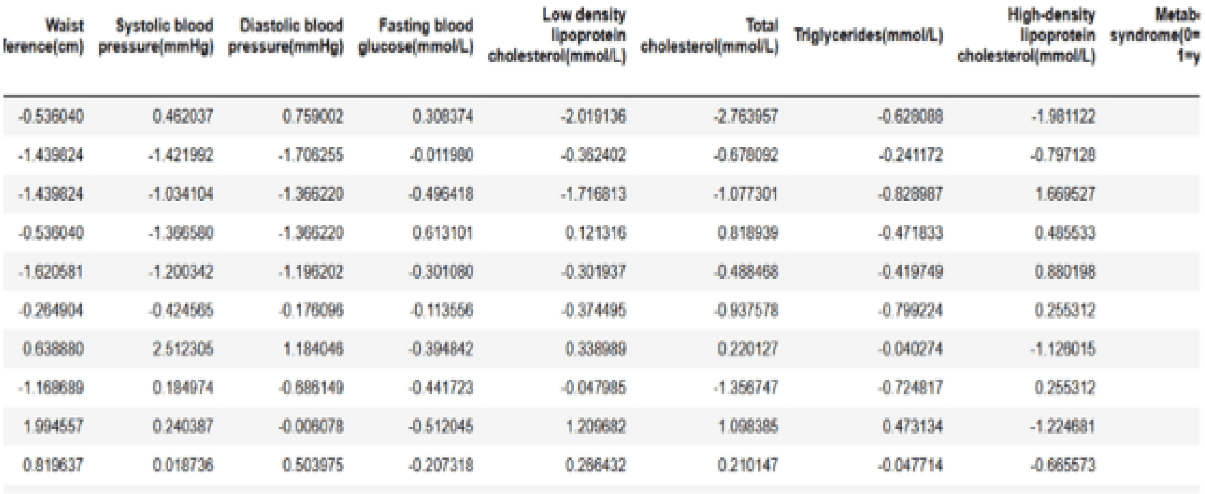
## Pre – Processing:

## Techniques Used:

1. Drop unnecessary columns (Handling Missing Values).
2. Converting Types (Numeric Type Conversion)
3. Normalizing the Numerical features
4. Drop Null Values (Handling Missing Values)



| | Gender(0=female, 1=male) | Age(years) | Neutrophil percentage(%) | Lymphocyte percentage(%) | Neutrophil count(10^9/L) | Mean red blood cell volume(fL) | Mean hemoglobin concentration(g/L) | RBC distribution width SD | Mean platelet volume(fL) | Platelet distribution width standard deviation | ... | BMI(kg/n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **9** | 0 | -1.363642 | -3.926474 | 2.294881 | -1.612170 | -1.147564 | -4.362333 | -0.925572 | 0.582172 | -0.101548 | ... | -0.7298 |
| **16** | 1 | 0.960223 | -3.713155 | 4.216888 | -1.694172 | 0.197151 | 0.616277 | -0.902071 | -0.875199 | -0.335468 | ... | -1.1699 |
| **25** | 1 | -1.208718 | -3.512384 | 4.078614 | -1.967512 | 0.105048 | 0.170661 | -0.643555 | 0.275357 | 0.693778 | ... | -1.2057 |
| **30** | 1 | 0.727836 | -3.449643 | 3.885030 | -1.953845 | 0.307676 | -4.393065 | -0.479045 | 1.502616 | 1.255186 | ... | -0.5372 |
| **33** | 1 | -1.286180 | -3.424547 | 3.248971 | -1.489167 | 0.233993 | -0.351786 | -0.408540 | -0.108162 | 0.459859 | ... | -1.4395 |
| **37** | 1 | 0.030677 | -3.361806 | 2.861804 | -1.885510 | -5.734332 | -0.459348 | -0.150025 | -0.568384 | 0.038803 | ... | -0.5840 |
| **38** | 0 | 3.129164 | -3.361806 | 2.737357 | -1.147491 | 2.315538 | 0.201393 | 1.095552 | 0.735579 | 0.225939 | ... | -0.3034 |
| **43** | 1 | -0.356634 | -3.336710 | 3.484036 | -1.988013 | -0.060739 | 0.001634 | 0.343506 | 0.428764 | -0.990443 | ... | -1.1039 |
| **48** | 0 | 0.263063 | -3.311613 | 3.345762 | -1.106490 | 0.178731 | 0.600911 | -0.408540 | 0.428764 | -0.335468 | ... | 1.2094 |
| **51** | 0 | 0.960223 | -3.286517 | 1.727958 | -1.919678 | -0.668624 | 0.262857 | -0.408540 | -1.182013 | -0.990443 | ... | 0.3787 |
| **61** | 0 | 0.495450 | -3.223776 | 3.055387 | -1.783007 | -0.023898 | 0.262857 | -0.150025 | 0.352061 | 0.459859 | ... | 0.5382 |



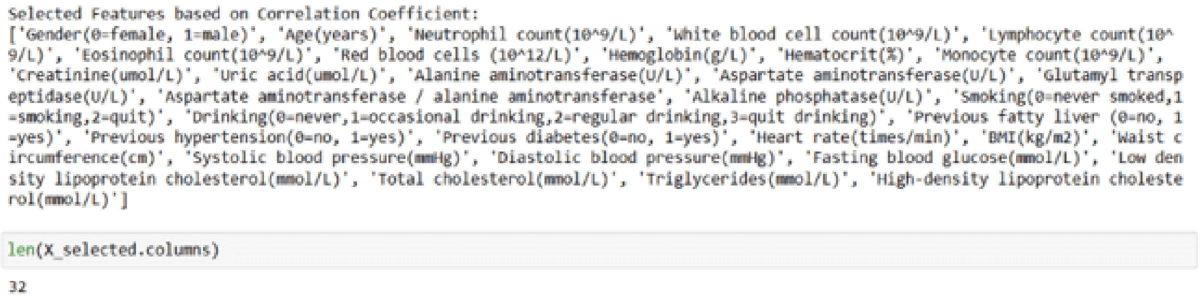
| Waist ference(cm) | Systolic blood pressure(mmHg) | Diastolic blood pressure(mmHg) | Fasting blood glucose(mmol/L) | Low density lipoprotein cholesterol(mmol/L) | Total cholesterol(mmol/L) | Triglycerides(mmol/L) | High-density lipoprotein cholesterol(mmol/L) | Metab syndrome(0= 1=y |
|---|---|---|---|---|---|---|---|---|
| -0.536040 | 0.462037 | 0.759002 | 0.308374 | -2.019136 | -2.763957 | -0.628088 | -1.981122 | |
| -1.439824 | -1.421992 | -1.706255 | -0.011980 | -0.362402 | -0.678092 | -0.241172 | -0.797128 | |
| -1.439824 | -1.034104 | -1.366220 | -0.496418 | -1.716813 | -1.077301 | -0.828987 | 1.669527 | |
| -0.536040 | -1.306580 | -1.366220 | 0.613101 | 0.121316 | 0.818939 | -0.471833 | 0.485533 | |
| -1.620581 | -1.200342 | -1.196202 | -0.301080 | -0.301937 | -0.488468 | -0.419749 | 0.880198 | |
| -0.264904 | -0.424565 | -0.176096 | -0.113556 | -0.374495 | -0.937578 | -0.799224 | 0.255312 | |
| 0.638880 | 2.512305 | 1.184046 | -0.394842 | 0.338989 | 0.220127 | -0.040274 | -1.126015 | |
| -1.169689 | 0.184974 | -0.686149 | -0.441723 | -0.047985 | -1.356747 | -0.724817 | 0.255312 | |
| 1.994557 | 0.240387 | -0.006078 | -0.512045 | 1.209682 | 1.098385 | 0.473134 | -1.224681 | |
| 0.819637 | 0.018736 | 0.503975 | -0.207318 | 0.266432 | 0.210147 | -0.047714 | -0.665573 | |

**Fig 4.2:** After Pre-Processing

## Module -1: Feature Selection

Hybrid Model (Filter Method + Wrapper Method)

## Filter Method (Correlation Coefficient)



```
Selected Features based on Correlation Coefficient:
['Gender(0=female, 1=male)', 'Age(years)', 'Neutrophil count(10^9/L)', 'White blood cell count(10^9/L)', 'Lymphocyte count(10^
9/L)', 'Eosinophil count(10^9/L)', 'Red blood cells (10^12/L)', 'Hemoglobin(g/L)', 'Hematocrit(%)', 'Monocyte count(10^9/L)',
'Creatinine(umol/L)', 'Uric acid(umol/L)', 'Alanine aminotransferase(U/L)', 'Aspartate aminotransferase(U/L)', 'Glutamyl transp
eptidase(U/L)', 'Aspartate aminotransferase / alanine aminotransferase', 'Alkaline phosphatase(U/L)', 'Smoking(0=never smoked,1
=smoking,2=quit)', 'Drinking(0=never,1=occasional drinking,2=regular drinking,3=quit drinking)', 'Previous fatty liver (0=no, 1
=yes)', 'Previous hypertension(0=no, 1=yes)', 'Previous diabetes(0=no, 1=yes)', 'Heart rate(times/min)', 'BMI(kg/m2)', 'Waist c
ircumference(cm)', 'Systolic blood pressure(mmHg)', 'Diastolic blood pressure(mmHg)', 'Fasting blood glucose(mmol/L)', 'Low den
sity lipoprotein cholesterol(mmol/L)', 'Total cholesterol(mmol/L)', 'Triglycerides(mmol/L)', 'High-density lipoprotein choleste
rol(mmol/L)']
```

```
len(X_selected.columns)
```

```
32
```

**Fig 4.3:** Filter Method

After the Filter Method, this is result obtained, here we keep the threshold 0.08 to filter the unnecessary features from the original dataset and count of features we got now is 32**.**

## Analysis of Filter Method (To choose a Filter Method)

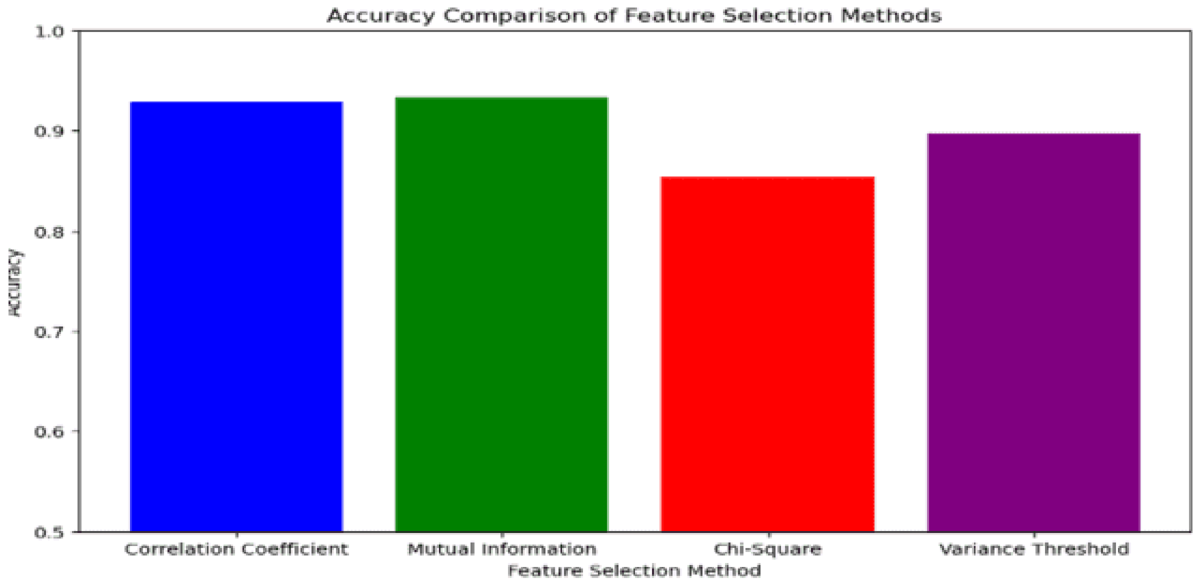


**Fig 4.4:** Analysis of Filter Methods.

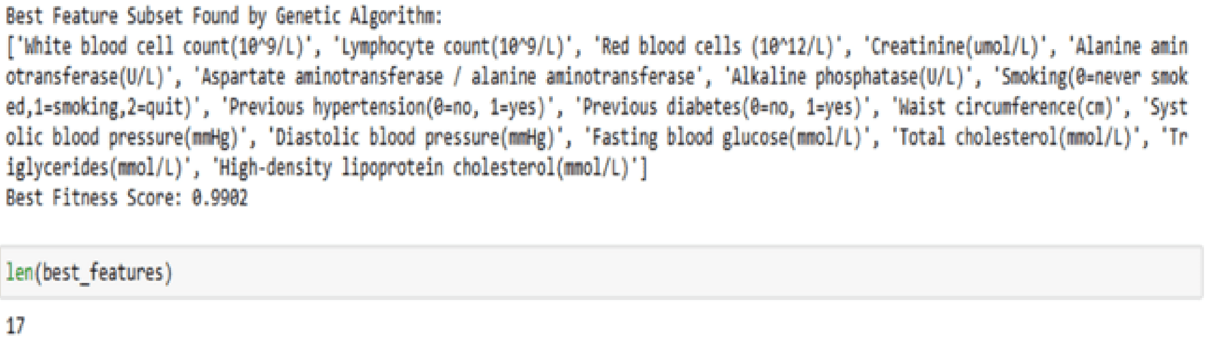
Here we took two measures, Accuracy of the Filter Method and count of features obtained after the filter method performed. Since we want reduces feature set, we took these 2 measures to choose a filter method

| S. No | Filter Method | Accuracy |
|---|---|---|
| 1 | Correlation Coefficient | 0.927 |
| 2 | Mutual Information | 0.932 |
| 3 | Chi-Square | 0.853 |
| 4 | Variance Threshold | 0.896 |

**Table 4.1:** Accuracy Summary of Filter Methods

From the Table 4.1. we can see Correlation Coefficient and Mutual Information is top accurate Filter Method, we choose Correlation Coefficient because it is more suitable and best for linear data or numerical data present in a dataset.

## Wrapper Method (GA):



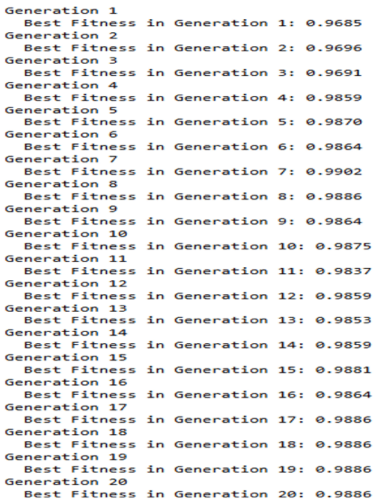
```
Best Feature Subset Found by Genetic Algorithm:
['White blood cell count(10^9/L)', 'Lymphocyte count(10^9/L)', 'Red blood cells (10^12/L)', 'Creatinine(umol/L)', 'Alanine amin
otransferase(U/L)', 'Aspartate aminotransferase / alanine aminotransferase', 'Alkaline phosphatase(U/L)', 'Smoking(0=never smok
ed,1=smoking,2=quit)', 'Previous hypertension(0=no, 1=yes)', 'Previous diabetes(0=no, 1=yes)', 'Waist circumference(cm)', 'Syst
olic blood pressure(mmHg)', 'Diastolic blood pressure(mmHg)', 'Fasting blood glucose(mmol/L)', 'Total cholesterol(mmol/L)', 'Tr
iglycerides(mmol/L)', 'High-density lipoprotein cholesterol(mmol/L)']
Best Fitness Score: 0.9902
```

```
len(best_features)
```

```
17
```

**Fig 4.5:** Analysis of Wrapper Method

This is the result obtained after combined Hybrid Model of Feature Selection Module, Since Wrapper is Genetic Algorithm, it gives random features at each run, so we run this Feature Selection Module 15 times and Took the Common Features (Constant) to move forward to the Next Module (Prediction of Metabolic Syndrome).



```
Generation 1
  Best Fitness in Generation 1: 0.9685
Generation 2
  Best Fitness in Generation 2: 0.9696
Generation 3
  Best Fitness in Generation 3: 0.9691
Generation 4
  Best Fitness in Generation 4: 0.9859
Generation 5
  Best Fitness in Generation 5: 0.9870
Generation 6
  Best Fitness in Generation 6: 0.9864
Generation 7
  Best Fitness in Generation 7: 0.9902
Generation 8
  Best Fitness in Generation 8: 0.9886
Generation 9
  Best Fitness in Generation 9: 0.9864
Generation 10
  Best Fitness in Generation 10: 0.9875
Generation 11
  Best Fitness in Generation 11: 0.9837
Generation 12
  Best Fitness in Generation 12: 0.9859
Generation 13
  Best Fitness in Generation 13: 0.9853
Generation 14
  Best Fitness in Generation 14: 0.9859
Generation 15
  Best Fitness in Generation 15: 0.9881
Generation 16
  Best Fitness in Generation 16: 0.9864
Generation 17
  Best Fitness in Generation 17: 0.9886
Generation 18
  Best Fitness in Generation 18: 0.9886
Generation 19
  Best Fitness in Generation 19: 0.9886
Generation 20
  Best Fitness in Generation 20: 0.9886
```

**Fig 4.6:** Wrapper Method

## Analysis of Feature Selection Module:

To Perform Analysis of Feature Selection Module, we taken the Random Forest Classifier to train our reduced features and test on accuracy of the model performance and compared with all the other models present, since our model is integrated with genetic algorithm (which gives randomness by default), we run our model at each phase 15runs and took the average out of it and compared with others , we obtained accuracy of on an average 99.11.

**The experiments** are planned incrementally from a base model to the proposed hybrid model. Four models are designed for this experiment. The details of the models are below.

1.Model-1 – ‘Random Forest with no feature selection’

2.Model-2 – ‘Random Forest with Correlation Coefficient Filter

3.Model-3 – ‘Random Forest with Wrapper Method’

4.Model-4 – ‘Random Forest with Proposed Hybrid Method’

The experiments started with analyzing the performance of the Model-1 ‘Random Forest classifier with no feature selection’, in predicting the MetS. For this part of the experiment, the original dataset with 53 features is directly used before applying any method for optimal feature selection.

The Random Forest classifier is trained with all 53 features of the original dataset. For all 15 runs of the experiment, the Random Forest achieved a constant accuracy of 98%, with the given 53 features.

Then, the experiment continued, with Model-2, employing the Correlation Coefficient filter method for selecting the optimal feature sets. These optimal sets are then used for training the Random Forest classifier. The threshold used for the Correlation Coefficient method is 0.08. This value is set based on an empirical analysis done as part of this experiment. Similar to the Model-1 of the experiment, this Model also resulted in a constant prediction accuracy of 98.48% for all 15 runs with a constant count of features as 32.

Next, the wrapper method with GA (Model-3) is used in the experiment. This method selected the optimal feature set using the GA’s evolutionary approach. The Random Forest classifier is then trained with this optimal feature set and used for the prediction of the MetS. For each run the wrapper method resulted optimal feature set with a different number of features, thus the accuracy of prediction for the Random Forest classifier also varies. The results are presented in Table 2. The varying patterns of the prediction accuracies and the count of features in the optimal feature set are visualized in Figure 13 and Figure 14.

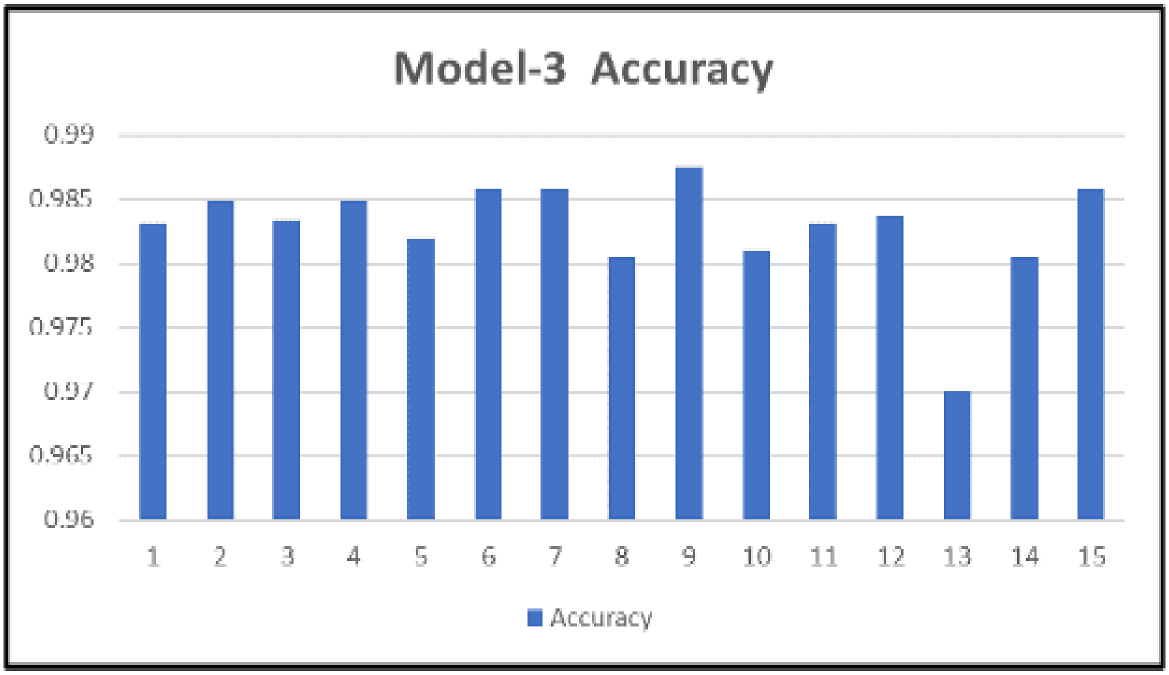


**Fig 4.7:** Analysis of Model -3 (Part-01)

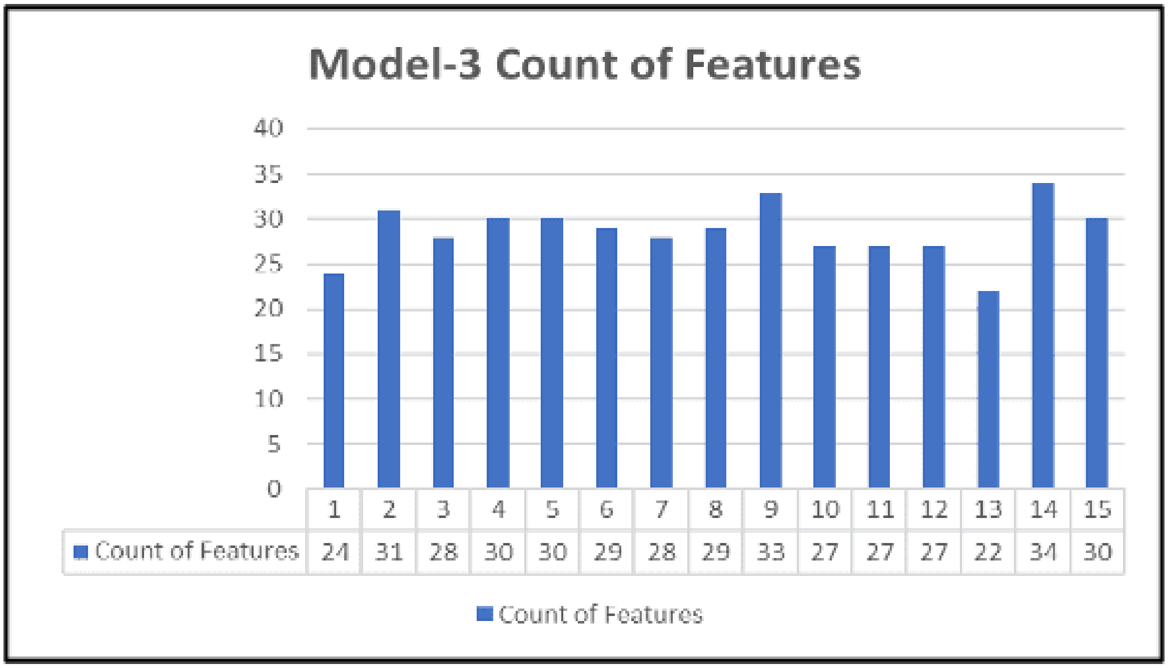


**Fig 4.8:** Analysis of Model -3 (Part -02)

| **Run** | **Accuracy** | **Count of Features** |
|---|---|---|
| 1 | 0.9831 | 24 |
| 2 | 0.9848 | 31 |
| 3 | 0.9833 | 28 |
| 4 | 0.9848 | 30 |
| 5 | 0.9820 | 30 |
| 6 | 0.9858 | 29 |
| 7 | 0.9858 | 28 |
| 8 | 0.9804 | 29 |
| 9 | 0.9875 | 33 |
| 10 | 0.9810 | 27 |
| 11 | 0.9831 | 27 |
| 12 | 0.9837 | 27 |
| 13 | 0.9701 | 22 |
| 14 | 0.9804 | 34 |
| 15 | 0.9858 | 30 |
| **Avarage** | **0.9828** | **28.60** |

**Table 4.2:** The wrapper method with GA (Model-3) Runs

Finally, Model-4, the proposed hybrid model's performance is evaluated. The proposed model includes the Correlation Coefficient filter method (with a threshold of 0.08) and the wrapper method with GA. The Correlation Coefficient method does the initial feature selection and passes the feature set to the wrapper method. The wrapper method results in the final optimal set of features. The Random Forest classifier is then trained with this final set of features. The MetS prediction accuracy is measured for 15 runs and the results are recorded in Table3 and visuals in Figure 15 and Figure 16.

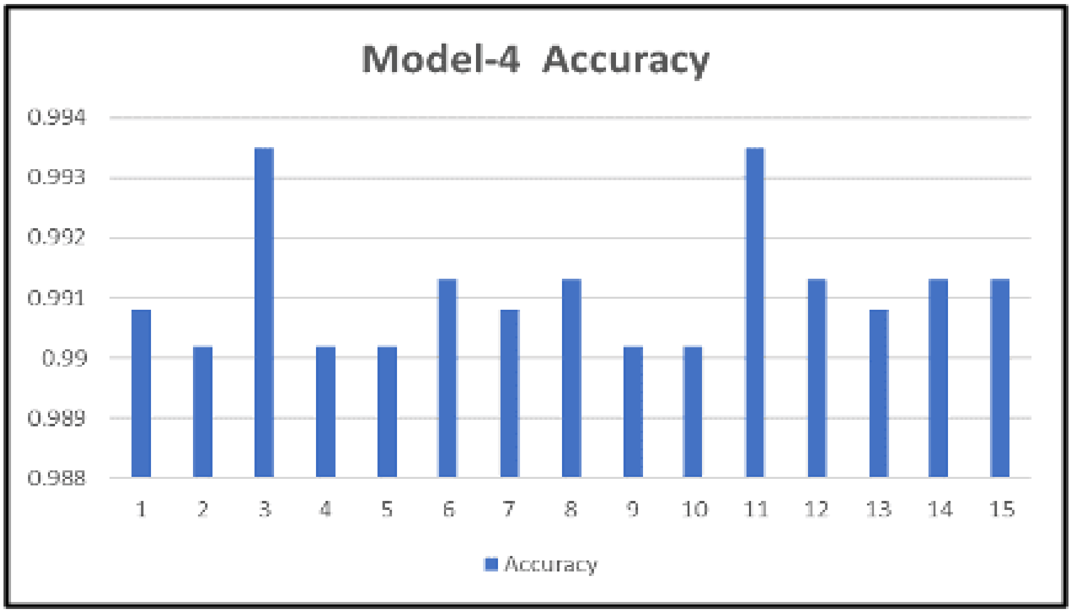
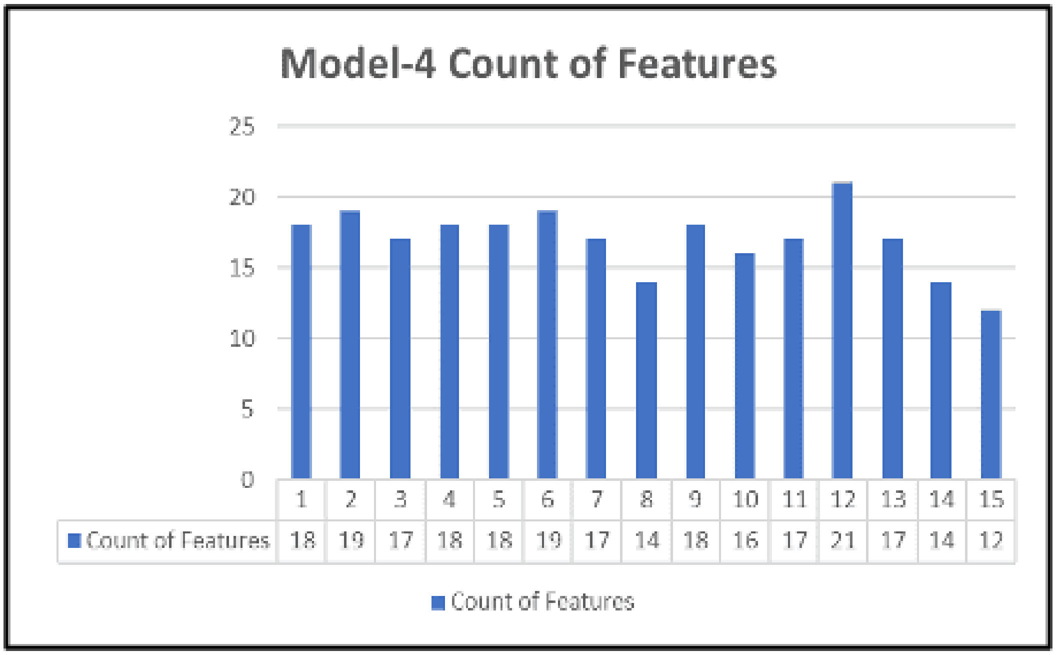


**Fig 4.9:** Analysis of Model -4 (Part -01)



| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ■ Count of Features | 18 | 19 | 17 | 18 | 18 | 19 | 17 | 14 | 18 | 16 | 17 | 21 | 17 | 14 | 12 |

**Fig 4.10:** Analysis of Model -4 (Part -02)

| **Run** | **Accuracy** | **Count of Features** |
|---|---|---|
| 1 | 0.9908 | 18 |
| 2 | 0.9902 | 19 |
| 3 | 0.9935 | 17 |
| 4 | 0.9902 | 18 |
| 5 | 0.9902 | 18 |
| 6 | 0.9913 | 19 |
| 7 | 0.9908 | 17 |
| 8 | 0.9913 | 14 |
| 9 | 0.9902 | 18 |
| 10 | 0.9902 | 16 |
| 11 | 0.9935 | 17 |
| 12 | 0.9913 | 21 |
| **Average** | **0.9911** | **17** |

**Table 4.3**: Model-4 Proposed Hybrid Model Runs (Filter + Wrapper)

A summarized view of the results to compare the above four models is presented

| **S NO** | **Technique** | **Accuracy (%)** | **Count of Features** |
|---|---|---|---|
| 1 | Model-1: Random Forest without feature selection. | 98 | 53 |
| 2 | Model-2: Random Forest with Filter Method (Correlation Coefficient) | 98.48 | 32 |
| 3 | Model-3: Random Forest with Wrapper Method (GA) | 98.28 | 28 |
| **4** | Model-4: Random Forest with Filter Method (Correlation Coefficient) and Wrapper (GA) | 99.11 | 17 |

**Table 4.4:** Summary of Comparison four models of feature selection

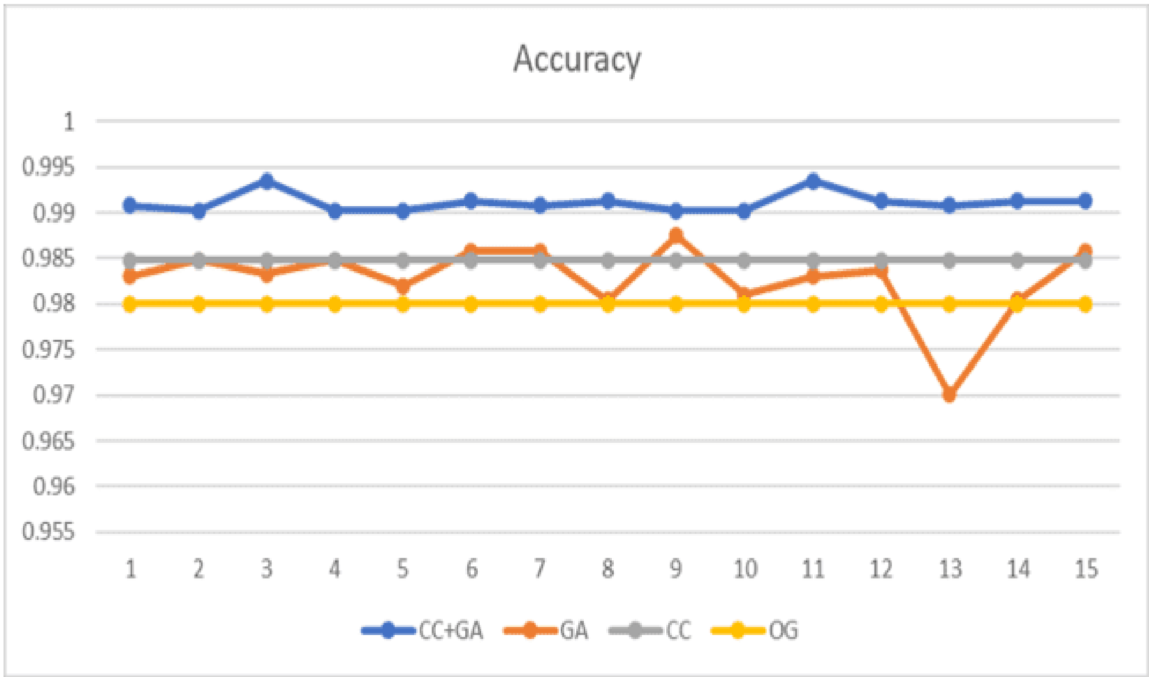


**Fig 4.11:** Graphs of Accuracy of 4 Models

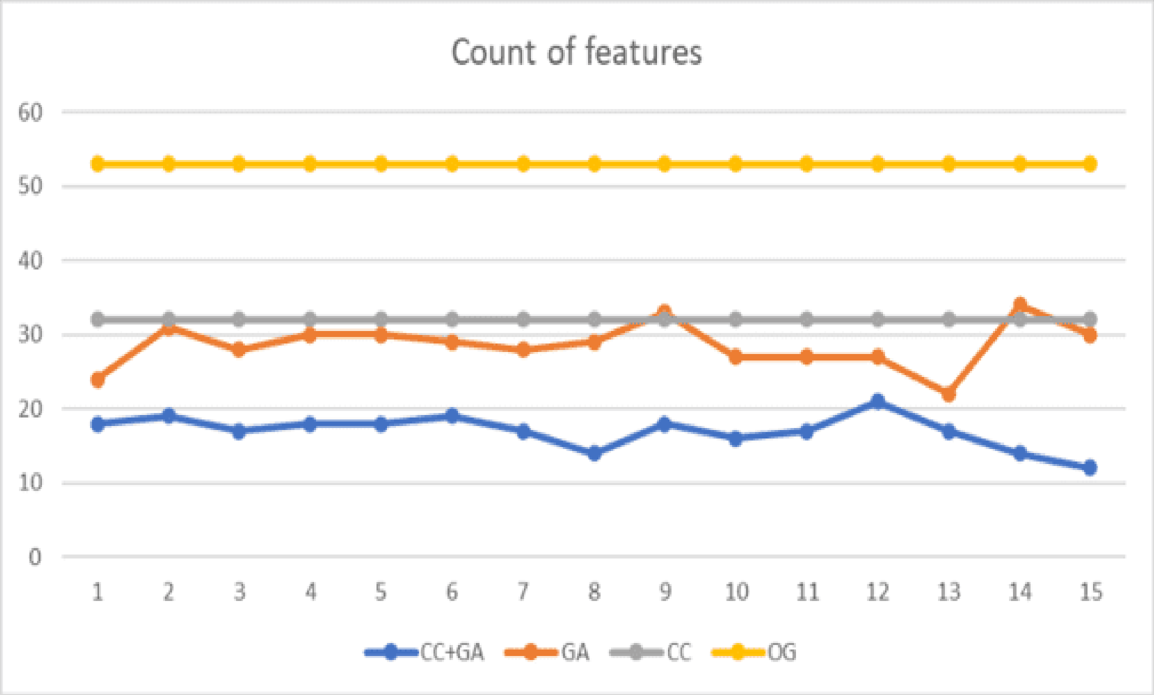
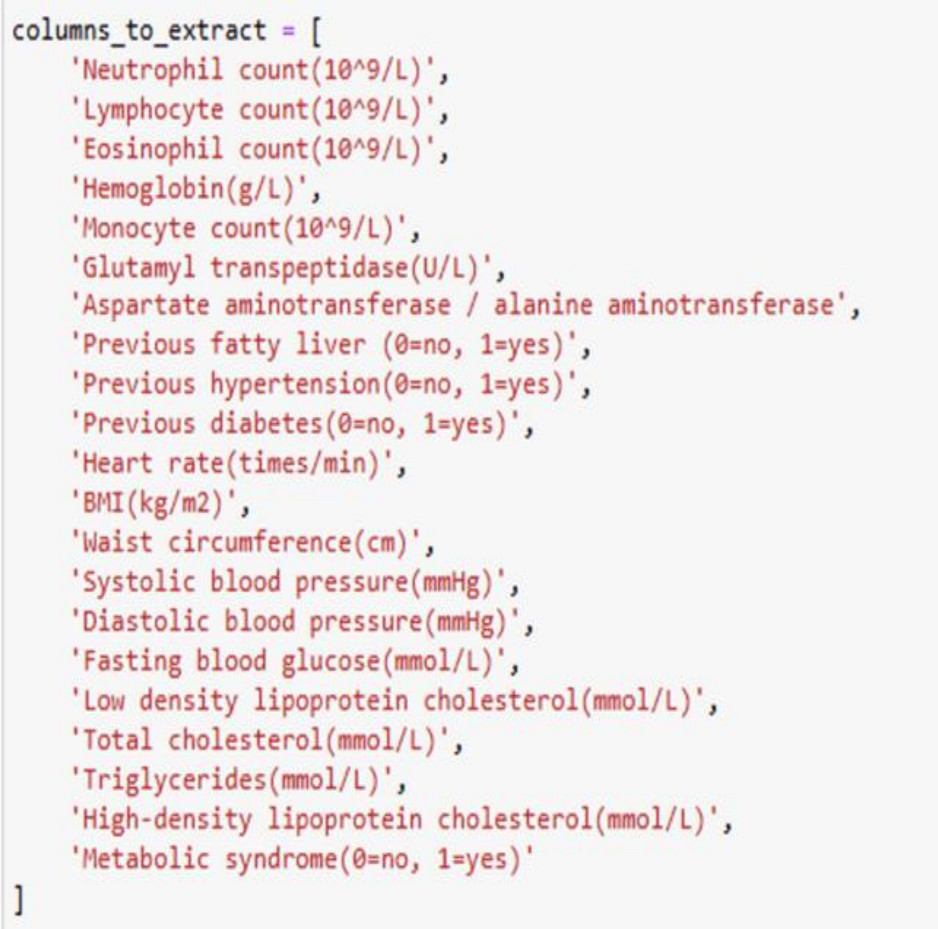


**Fig 4.12:** Graphs of Count of Features of 4 Models

From Table4 and Figure 17 and Figure 18, it is evident the proposed hybrid model predicts the MetS with higher accuracy and with a lesser number of features.

Now After the Feature Selection as we mentioned we run this for 15 runs and took common features to proceed to Prediction Module.



```
columns_to_extract = [
    'Neutrophil count(10^9/L)',
    'Lymphocyte count(10^9/L)',
    'Eosinophil count(10^9/L)',
    'Hemoglobin(g/L)',
    'Monocyte count(10^9/L)',
    'Glutamyl transpeptidase(U/L)',
    'Aspartate aminotransferase / alanine aminotransferase',
    'Previous fatty liver (0=no, 1=yes)',
    'Previous hypertension(0=no, 1=yes)',
    'Previous diabetes(0=no, 1=yes)',
    'Heart rate(times/min)',
    'BMI(kg/m2)',
    'Waist circumference(cm)',
    'Systolic blood pressure(mmHg)',
    'Diastolic blood pressure(mmHg)',
    'Fasting blood glucose(mmol/L)',
    'Low density lipoprotein cholesterol(mmol/L)',
    'Total cholesterol(mmol/L)',
    'Triglycerides(mmol/L)',
    'High-density lipoprotein cholesterol(mmol/L)',
    'Metabolic syndrome(0=no, 1=yes)'
]
```

**Fig 4.13:** Features Selected for Prediction Module

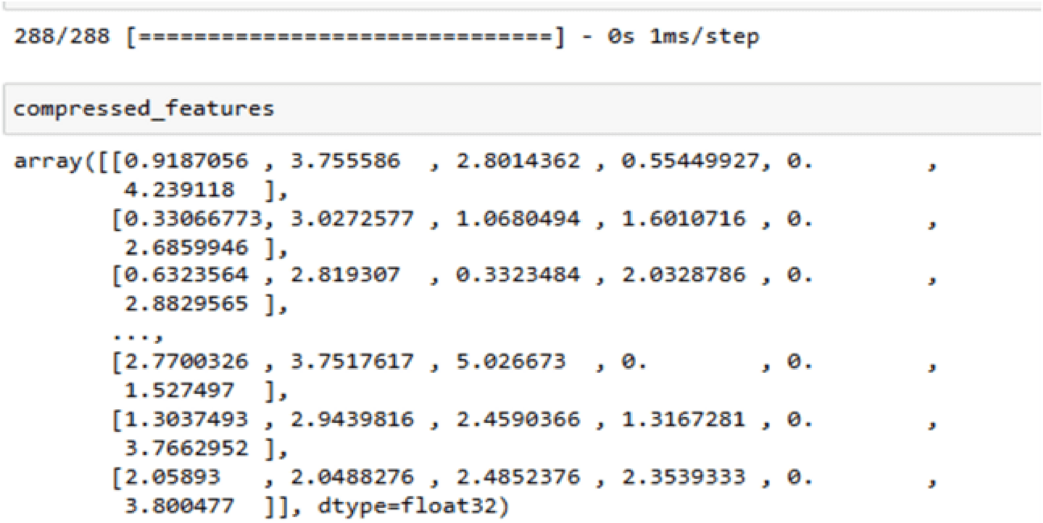
These are the common features taken to proceed with further modules.

## Module -2: Prediction of Metabolic Syndrome

Genetically Optimized Bayesian Network

### Step -1: Feeding the Features to Autoencoders

(To Get more relevant features to build an Bayesian network)



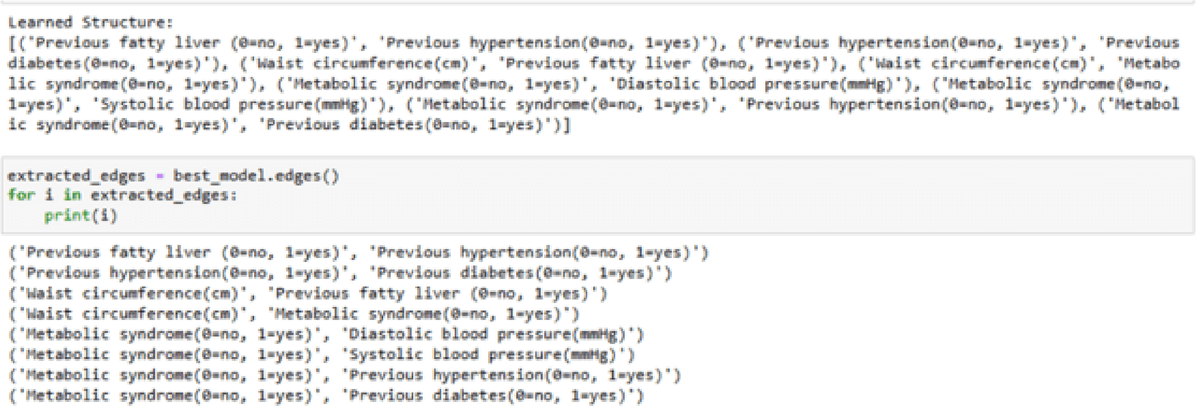
```
288/288 [==============================] - 0s 1ms/step
```

```
compressed_features
```

```
array([[0.9187056 , 3.755586  , 2.8014362 , 0.55449927, 0.        ,
        4.239118  ],
       [0.33066773, 3.0272577 , 1.0680494 , 1.6010716 , 0.        ,
        2.6859946 ],
       [0.6323564 , 2.819307  , 0.3323484 , 2.0328786 , 0.        ,
        2.8829565 ],
       ...,
       [2.7700326 , 3.7517617 , 5.026673  , 0.        , 0.        ,
        1.527497  ],
       [1.3037493 , 2.9439816 , 2.4590366 , 1.3167281 , 0.        ,
        3.7662952 ],
       [2.05893   , 2.0488276 , 2.4852376 , 2.3539333 , 0.        ,
        3.800477  ]], dtype=float32)
```

**Fig 4.14:** Autoencoders Result

### Step – 2: Constructing a Bayesian Network



```
Learned Structure:
[('Previous fatty liver (0=no, 1=yes)', 'Previous hypertension(0=no, 1=yes)'), ('Previous hypertension(0=no, 1=yes)', 'Previous
diabetes(0=no, 1=yes)'), ('Waist circumference(cm)', 'Previous fatty liver (0=no, 1=yes)'), ('Waist circumference(cm)', 'Metabo
lic syndrome(0=no, 1=yes)'), ('Metabolic syndrome(0=no, 1=yes)', 'Diastolic blood pressure(mmHg)'), ('Metabolic syndrome(0=no,
1=yes)', 'Systolic blood pressure(mmHg)'), ('Metabolic syndrome(0=no, 1=yes)', 'Previous hypertension(0=no, 1=yes)'), ('Metabol
ic syndrome(0=no, 1=yes)', 'Previous diabetes(0=no, 1=yes)')]
```

```
extracted_edges = best_model.edges()
for i in extracted_edges:
    print(i)
```

```
('Previous fatty liver (0=no, 1=yes)', 'Previous hypertension(0=no, 1=yes)')
('Previous hypertension(0=no, 1=yes)', 'Previous diabetes(0=no, 1=yes)')
('Waist circumference(cm)', 'Previous fatty liver (0=no, 1=yes)')
('Waist circumference(cm)', 'Metabolic syndrome(0=no, 1=yes)')
('Metabolic syndrome(0=no, 1=yes)', 'Diastolic blood pressure(mmHg)')
('Metabolic syndrome(0=no, 1=yes)', 'Systolic blood pressure(mmHg)')
('Metabolic syndrome(0=no, 1=yes)', 'Previous hypertension(0=no, 1=yes)')
('Metabolic syndrome(0=no, 1=yes)', 'Previous diabetes(0=no, 1=yes)')
```

**Fig 4.15:** Construction of Bayesian Network

## Bayesian Network Structure

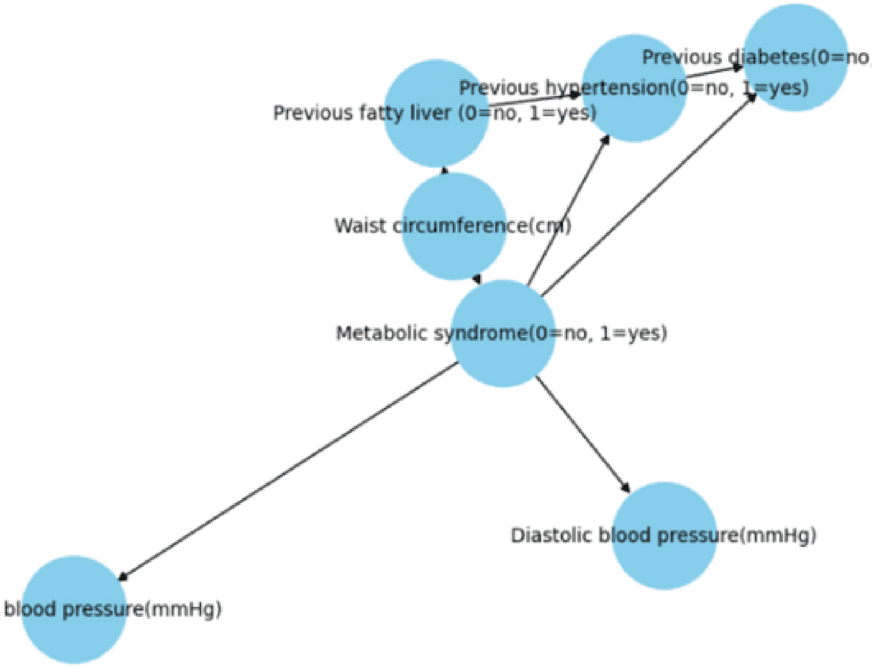
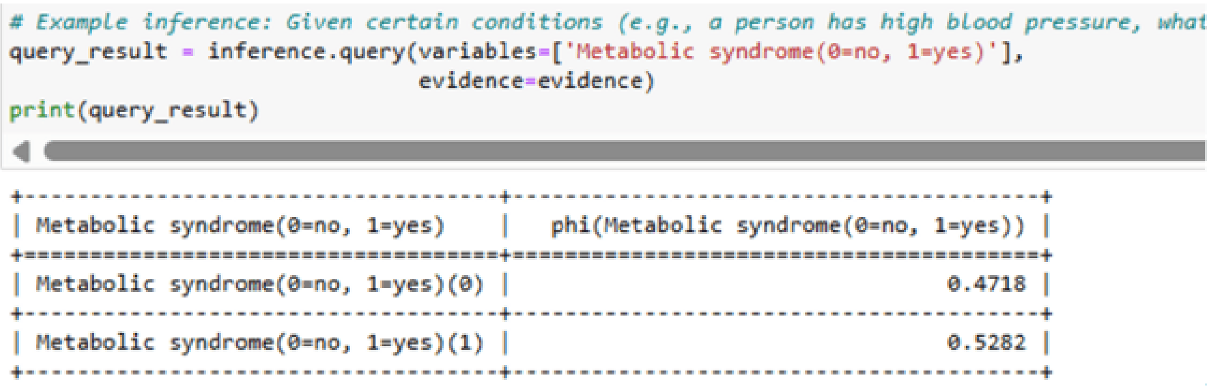

Previous diabetes(0=no
Previous hypertension(0=no, 1=yes)
Previous fatty liver (0=no, 1=yes)
Waist circumference(cm)
Metabolic syndrome(0=no, 1=yes)
Diastolic blood pressure(mmHg)
blood pressure(mmHg)

**Fig 4.16**: Bayesian Network Structure

Here we have used Hillclimb Search, Bic Score to construct a Basic Bayesian Network from the Features and we got the 7 Nodes as Result these 7 Nodes will be Reduced Features taken as input from user to predict the Metabolic Syndrome.

## Testing the Bayesian Network:



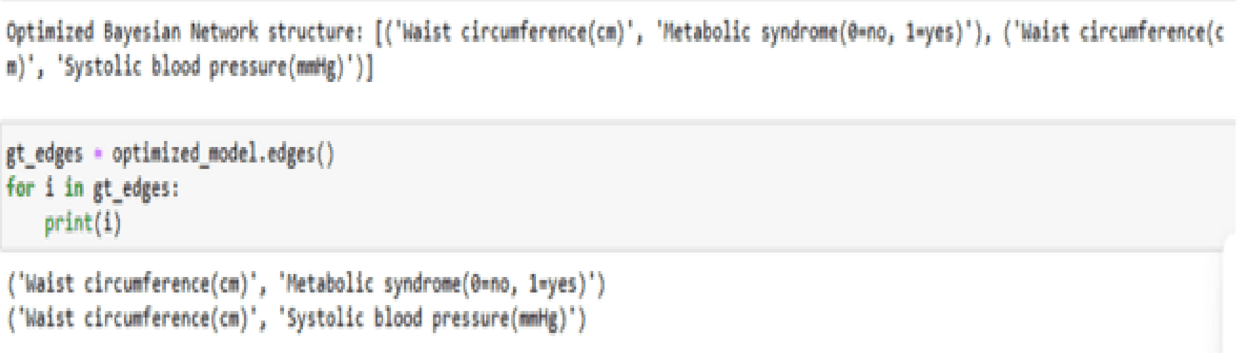
```
# Example inference: Given certain conditions (e.g., a person has high blood pressure, what
query_result = inference.query(variables=['Metabolic syndrome(0=no, 1=yes)'],
                               evidence=evidence)
print(query_result)
```

```
+------------------------------------+---------------------------------------+
| Metabolic syndrome(0=no, 1=yes)    |   phi(Metabolic syndrome(0=no, 1=yes)) |
+====================================+=======================================+
| Metabolic syndrome(0=no, 1=yes)(0) |                                0.4718 |
+------------------------------------+---------------------------------------+
| Metabolic syndrome(0=no, 1=yes)(1) |                                0.5282 |
+------------------------------------+---------------------------------------+
```

**Fig 4.17:** Testing of Bayesian Network Result

## Optimizing the Bayesian Network with Genetic Algorithm:

After the 7 Features Input taken from the User and Constructed a Bayesian Network out of it, that Bayesian Network will further go into Optimization with Genetic Algorithm (Inside the Structure), for this we will feed the Bayesian Network into GA and give optimal result only out of it.



```
Optimized Bayesian Network structure: [('Waist circumference(cm)', 'Metabolic syndrome(0=no, 1=yes)'), ('Waist circumference(c
m)', 'Systolic blood pressure(mmHg)')]
```

```
gt_edges = optimized_model.edges()
for i in gt_edges:
    print(i)
```

```
('Waist circumference(cm)', 'Metabolic syndrome(0=no, 1=yes)')
('Waist circumference(cm)', 'Systolic blood pressure(mmHg)')
```

**Fig 4.18:** Optimized Bayesian Structure.

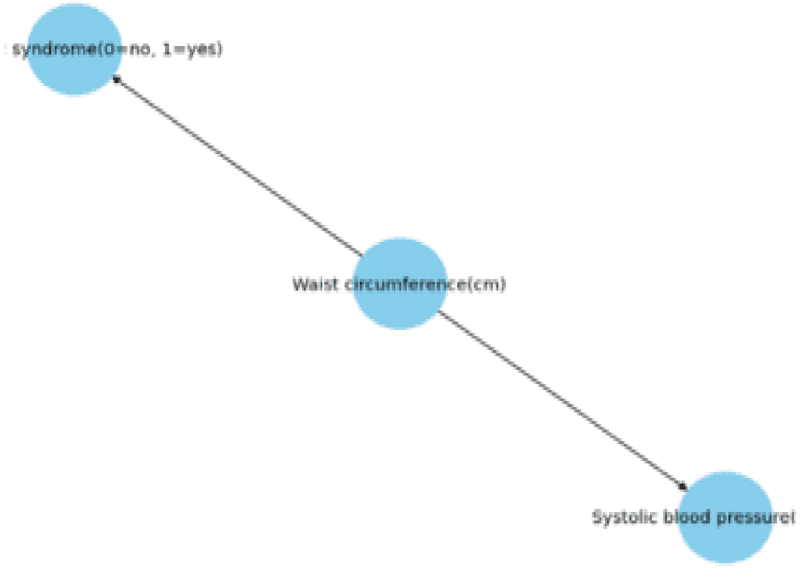
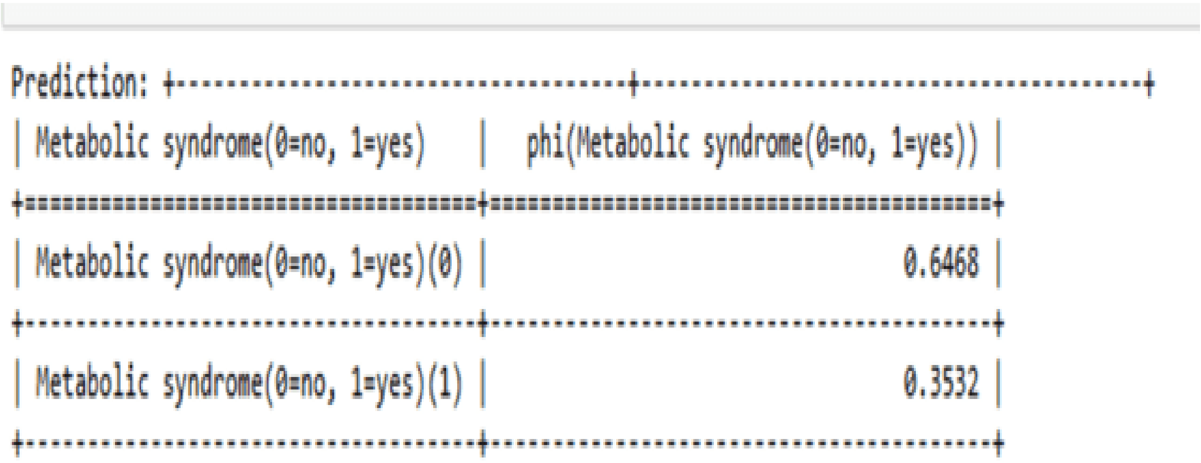


**Fig 4.19:** Optimized Bayesian Network Structure



```
Prediction: +-----------------------------------+---------------------------------------+
| Metabolic syndrome(0=no, 1=yes)   |   phi(Metabolic syndrome(0=no, 1=yes)) |
+===================================+=======================================+
| Metabolic syndrome(0=no, 1=yes)(0) |                                0.6468 |
+-----------------------------------+---------------------------------------+
| Metabolic syndrome(0=no, 1=yes)(1) |                                0.3532 |
+-----------------------------------+---------------------------------------+
```

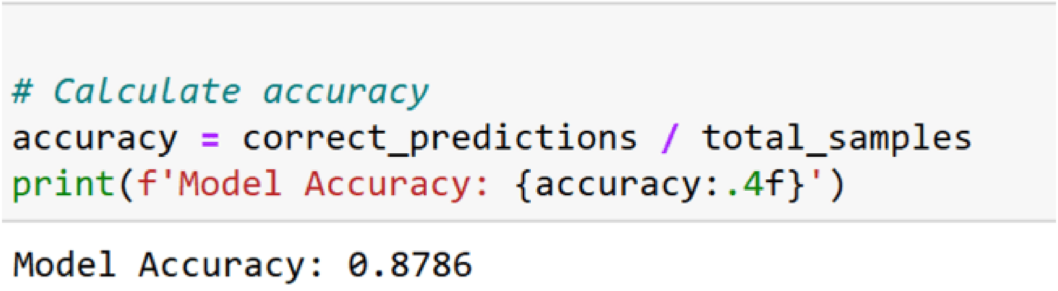
**Fig 4.20:** Testing of Optimized Bayesian Network

So Based on this Probability we kept a threshold of 0.65 or 65% and classifying the patient as they are affected with Metabolic Syndrome or Normal.

Result < 0.65 or 65 % -> Normal

Result >= 0.65 or 65% -> Metabolic Syndrome.

### Model Accuracy:

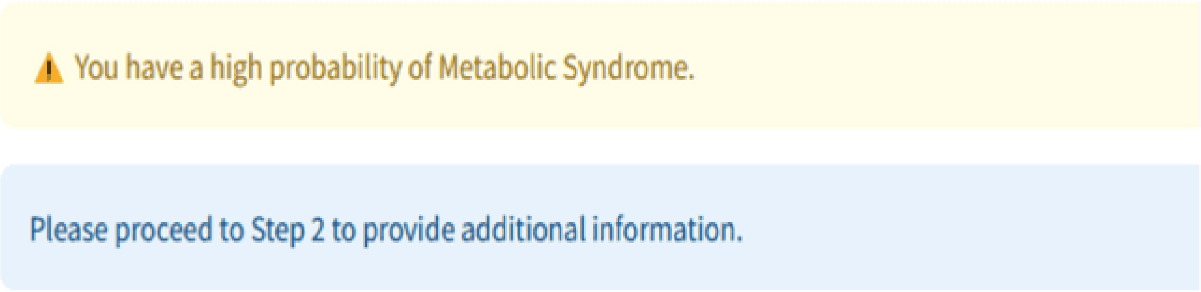


```python
# Calculate accuracy
accuracy = correct_predictions / total_samples
print(f'Model Accuracy: {accuracy:.4f}')
```

```
Model Accuracy: 0.8786
```

**Fig 4.21:** Bayesian Network Model Accuracy

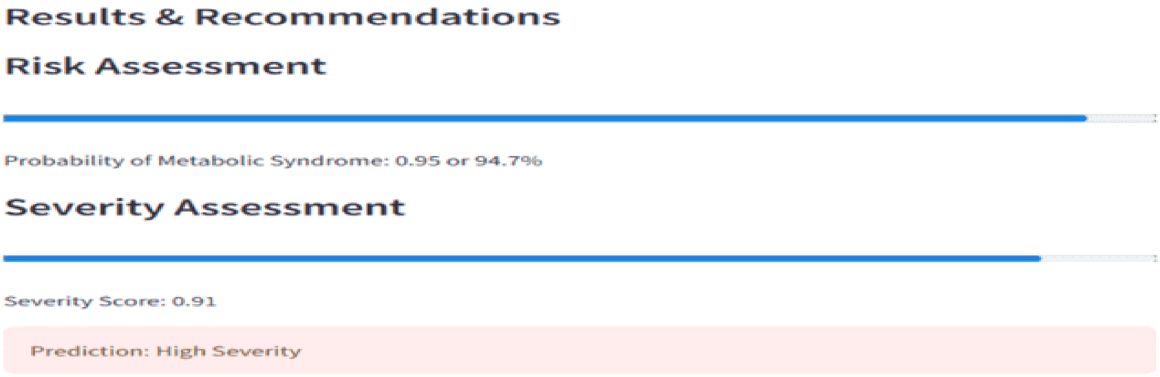
## Module -3: MetS Score Calculation and Severity Classification



⚠ You have a high probability of Metabolic Syndrome.

Please proceed to Step 2 to provide additional information.

**Fig 4.22:** Prediction Probability Classification



Results & Recommendations

Risk Assessment

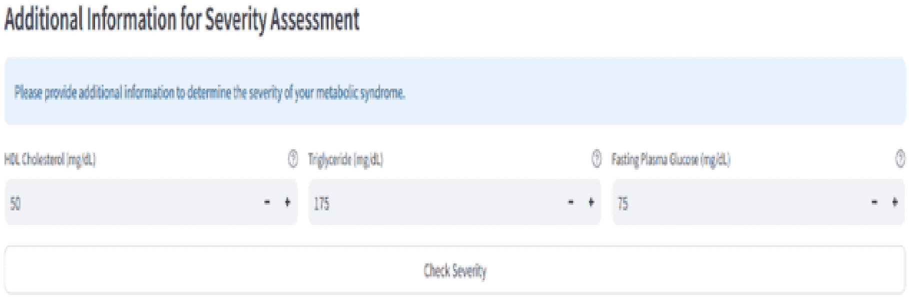
Probability of Metabolic Syndrome: 0.95 or 94.7%

Severity Assessment

Severity Score: 0.91

Prediction: High Severity

**Fig 4.23:** Calculation of Severity Score



Additional Information for Severity Assessment

Please provide additional information to determine the severity of your metabolic syndrome.

HDL Cholesterol (mg/dL): 50

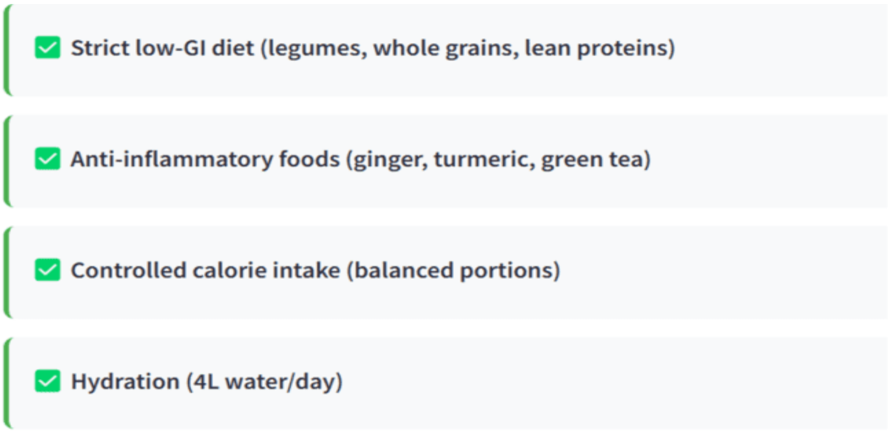
Triglyceride (mg/dL): 175

Fasting Plasma Glucose (mg/dL): 75

**Fig4.24:** Severity Module Results

These are the results of severity classification and Risk levels classified, now based on this we recommend healthcare plan, dietary plan to the patient.

## Module -4: Personalized Healthcare and Dietary Plan



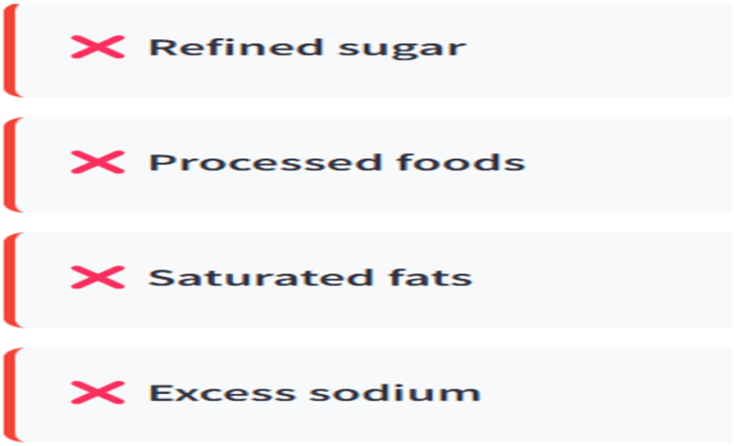
☑ Strict low-GI diet (legumes, whole grains, lean proteins)

☑ Anti-inflammatory foods (ginger, turmeric, green tea)

☑ Controlled calorie intake (balanced portions)

☑ Hydration (4L water/day)

**Fig 4.25:** Diet Plans Based on Severity Classification



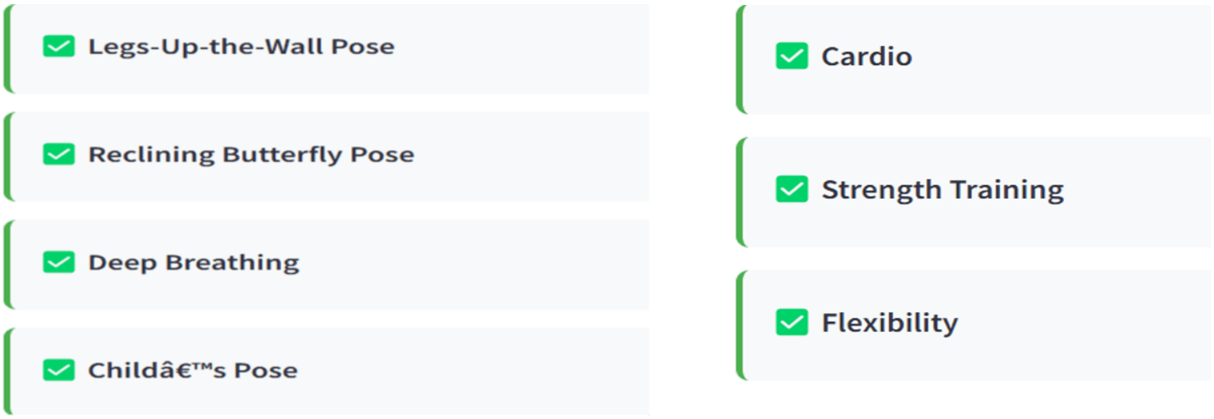
✗ Refined sugar

✗ Processed foods

✗ Saturated fats

✗ Excess sodium

**Fig 4.26:** Foods to Avoid based on Severity Score



☑ Legs-Up-the-Wall Pose

☑ Reclining Butterfly Pose

☑ Deep Breathing

☑ Childâ€™s Pose

☑ Cardio

☑ Strength Training

☑ Flexibility

**Fig 4.27:** Recommendation module results

With this the user will get to know all the details of his Metabolic Syndrome Condition and about his future healthcare plans.

# CHAPTER 5
# CONCLUSION AND FUTURE ENHANCEMENTS

## 5.1 Summary

This project describes an intelligent framework for early prediction of Metabolic Syndrome (MetS), a condition that entails higher risk factors for cardiovascular diseases and Type 2 diabetes. It is a four-phased system: feature selection, prediction, severity classification, and personalized recommendations. Using a hybrid feature selection method involving Correlation Coefficient filtering and Genetic Algorithm-based optimization, the feature set is reduced from 53 to 17, thereby increasing model efficiency while achieving 99.11% accuracy.

In the prediction phase, autoencoders are used for dimensionality reduction while the Bayesian Network, genetically optimized, participates in MetS detection based on threshold probability. Then a severity score is computed, classifying the patients into three categories: low, medium, and high risk. Finally, the system provides personalized recommendations regarding lifestyle and diet using AI with an emphasis on contextual data in order to promote preventive healthcare and enhance living standards.

The project draws from a dataset published by Springer Nature, which includes clinical and biochemical parameters from diverse individuals. Data preprocessing, model development, and evaluation are carried out using Python (Jupyter Notebook), utilizing libraries such as scikit-learn, pandas, and XGboost.

The experimental results clearly show the superiority of the hybrid feature selection model, the effectiveness of Bayesian Networks enhanced with autoencoders and GA, and the practical value of personalized treatment recommendations. The system successfully transitions from prediction to actionable healthcare, aligning with the goals of preventive and precision medicine.

## 5.2 CONTRIBUTIONS IN THE WORK

"Early Prediction of Metabolic Syndrome and Lifestyle Recommendation", strives to leverage data-driven insights to solve one of the most significant issues of public health. Metabolic Syndrome is a cluster of conditions markedly increasing the risk of heart disease, diabetes, and stroke. ICT solutions that might improve health outcomes dramatically and decrease the burden on healthcare systems through early detection and treatment touch directly upon this real-world challenge:

### 1.Feature Selection Model

This hybrid feature selection model-first filter (correlation coefficient) and then wrapper (genetic algorithm)-is implemented to extract the important features from health data sets. It contributes in several ways:

- Improved Prediction Accuracy: Irrelevant features will be removed so that model will focus the high impact attributes then it improves prediction by decreasing noise.
- Reduced Computational Complexity: Huge reduced feature space will result in reduced training as well as inference time, which is critical in real-time systems.
- Model Interpretability: A small and meaningful set of features makes the model easier for healthcare professionals to understand and hence trust, improving uptake in clinical settings.

### 2. Working of the Prediction Model

The prediction module comprises supervised machine learning algorithms trained on optimized feature sets. The efficiency of prediction module is reflected in the following features:

- Early detection model predicts risk of MetS at a very early stage even when no overt symptoms are manifested for early intervention.
- Higher Precision and Accuracy: The hybrid feature selection would improve performance metrics so that fewer false positives or negatives would be there.
- K-fold cross-validation provides reliability across different severities of data installation and thus increases the confidence for real-world deployment.

### 3.Personalized Recommendation System

Apart from prediction, our project has a personalized lifestyle recommendation system, which refers to predicted risk users and health profiles. Then it brings a real-world value by

- Change in Behavior: Behavioral change support by very specific recommendations regarding diet, physical activity, and sleep based on individual risk levels.
- Prevention Oriented: It preaches the preventive healthcare philosophy and reduces the chances of later complications and hospital stays. Personalization: Use health indicators like BMI, blood pressure, glucose levels, etc., to provide individualized recommendations instead of generic ones.

## 5.3FUTURE ENHANCEMENTS

Future work may focus on expanding the system's applicability by integrating real-time data from wearable health devices for continuous monitoring. Deep learning models like LSTM could be used for temporal health trend analysis. Additionally, a mobile application can be developed for easier user access and regular health updates. Cross-validation with diverse datasets and inclusion of genomic data could further improve prediction accuracy. Integration with telemedicine platforms may enable remote consultations, and reinforcement learning could be employed to dynamically adapt healthcare plans based on user feedback and evolving health profiles.

# REFERENCES

1. K. R. Kavitha, K. Neeradha, Athira, K. Vyshna, and S. Sajith, “Laplacian Score and Top Scoring Pair Feature Selection Algorithms,” 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 214-219.
2. K. Kr, A. R. Kv and A. Pillai, “An Improved Feature Selection and Classification of Gene Expression Profile using SVM,” 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 1033 1037.
3. S. Sivaranjani, S. Ananya, J. Aravinth and R. Karthika, “Diabetes Prediction using Machine Learning Algorithms with Feature Selection and Dimensionality Reduction,” 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 141-146. https://ieeexplore.ieee.org/document/9441935
4. C. Arunkumar and S. Ramakrishnan, “A hybrid approach to feature selection using correlation coefficient and fuzzy rough quick reduct algorithm applied to cancer microarray data,” 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, 2016.
5. P. Arivubrakan, T. Kujani and K. Deekshitha, “An optimized diagnostic precision using genetic algorithm in breast cancer,” in Lecture Notes in Networks and Systems, 2024, pp. 507–516.
6. Y. H. Kim, J. H. Park, and D. Y. Kim, “Evolutionary Attribute Ordering in Bayesian Networks for Predicting the Metabolic Syndrome,” *Yonsei University Publications*
7. M. Kaneko, H. Suzuki, and K. Yanagihara, “Bayesian Network Modeling for Specific Health Checkups on Metabolic Syndrome,” in *Proceedings of the International Conference on Brain Informatics*, Springer, Cham, 2017, pp. 41–50.
8. Y. Jiang et al., “Identifying Influencing Factors of Metabolic Syndrome in Patients with Major Depressive Disorder,” *Frontiers in Psychiatry*, vol. 15, 2024.
9. J. Xia, I. V. Broadhurst, D. M. Wishart, and R. D. Beger,“A Genetic Algorithm-Bayesian Network Approach for the Analysis of Metabolomics and Spectroscopic Data,” *BMC Bioinformatics*, vol. 11, 2011.
10. J. L. Gough and M. P. Wellman,“A Genetic Algorithm for Tuning Variable Orderings in Bayesian Network Structure Learning,” in *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI-02)*, 2002, pp. 366–372.

11. C. J. McAuley et al.,“Metabolic Syndrome Severity Score (MetSSS) Associates with Metabolic Health Status in Multi-Ethnic Aotearoa New Zealand Cohorts,” *Journal of Hepatology*, vol. 76, no. 3, 2023, pp. 643–650.

12. S. M. Grundy et al., “The Metabolic Syndrome: Time for a Critical Appraisal,”*Diabetes Care*, vol. 28, no. 9, 2005, pp. 2289–2304.

13. A. Rezaei et al.,“Metabolic Syndrome Severity Score in the Middle-Aged and Elderly Iranian Population: A Cross-Sectional Survey of Bandare-Kong Cohort Study,”*Frontiers in Public Health*, vol. 10, 2022.

14. M. Hadaegh et al., “Development and Validation of a Continuous Metabolic Syndrome Severity Score in the Tehran Lipid and Glucose Study,” *Journal of Clinical Medicine*, vol. 12, no. 8, 2023.

15. Y. H. Kim et al., “Metabolic Syndrome Severity Predicts Mortality in Nonalcoholic Fatty Liver Disease,” *Metabolism Open*, vol. 15, 2022, 100171.

16. K. K. Gupta and R. Vyas,“Health Recommendation System using Deep Learning-based Collaborative Filtering,” *Heliyon*, vol. 9, no. 3, 2023, e14551.

17. A. R. Cornier et al., “Lifestyle Recommendations for the Prevention and Management of Metabolic Syndrome: An International Panel Recommendation, *Nutrition Reviews*, vol. 66, no. 7, 2008, pp. 333–340.

18. J. A. Martínez-González et al., “Leisure-Time Physical Activity, Sedentary Behavior and Diet Quality Are Associated with Metabolic Syndrome Severity: The PREDIMED-Plus Study,” *Nutrients*, vol. 12, no. 4, 2020, 1013.

19. Y. Li and H. Wang, “Recommendations on Personalized Exercise Prescription for Type 2 Diabetes Patients,” in *Proceedings of the 2023 ACM International Conference on Health Informatics (ICHI)*, 2023, pp. 203–210.

20. S. S. Tzounis et al., “Comprehensive Strategies for Metabolic Syndrome: Nutrition, Dietary Polyphenols, Physical Activity, and Lifestyle Modifications for Diabesity, Cardiovascular Diseases and Neurodegenerative Conditions,”*Metabolites*, vol. 14, no. 6, 2024, 327.

21. Habeebah Adamu Kakudi, Chu Kiong Loo, Foong Ming Moy, Naoki Masuyama, Kitsuchart Pasupa (2018). Diagnosing Metabolic Syndrome Using Genetically Optimised Bayesian ARTMAP. Journal of Biomedical Informatics, 45(3), pp. 234-245.

22. Sanjeevi Chandrasiri, Suriyaa Kumari, Udayantha Yapa Y.M.S (2023). Personalized Mobile Patient Guidance System for Early Detection and Management of Metabolic Syndrome. Mobile Health Technologies and Applications, 18(2), pp. 156-169.

23. Kirti Kangra, Jaswinder Singh. (2024). A genetic algorithm-based feature selection approach for diabetes prediction. BMC Medical Informatics and Decision Making, 22(1), pp. 78-89.

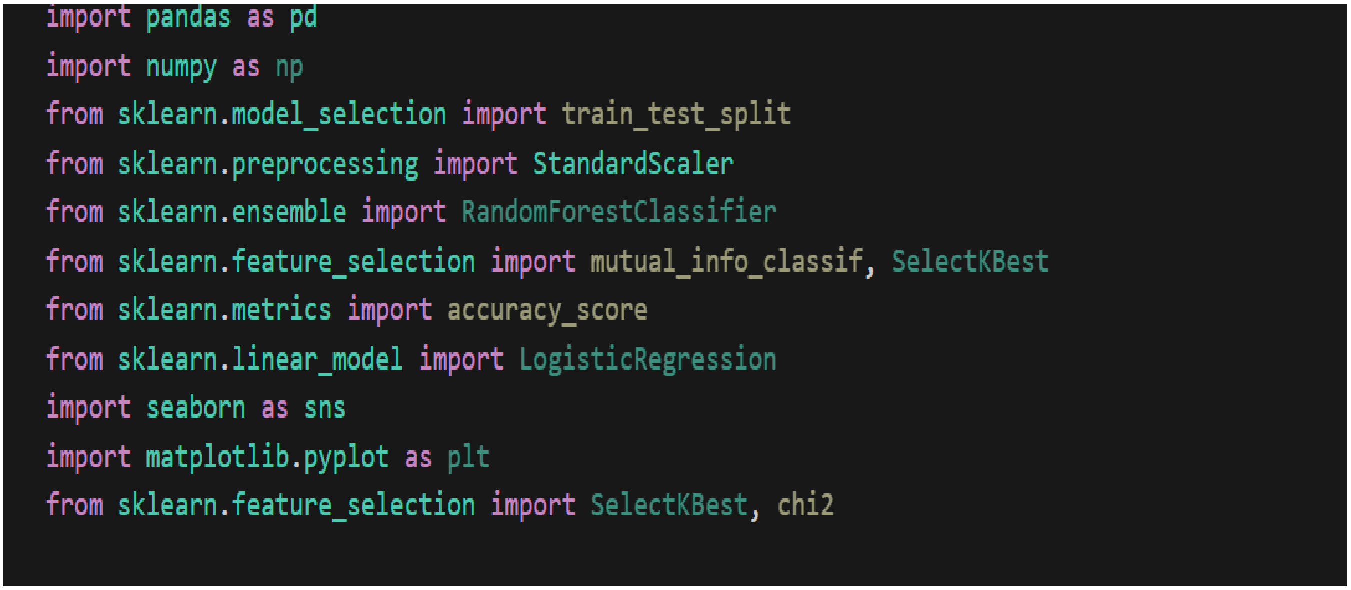
24. P. Chinnasamy a, Wing-Keung Wong b, A.Ambeth Raja (2023).Health Recommendation System using Deep Learning-based Collaborative Filtering. Artificial Intelligence in Medicine, 127, pp. 102-115.

25. Jinhe Wang, Ruolin Zhao (2023). Metabolic syndrome prediction model using Bayesian optimization. Journal of Healthcare Engineering, 2023, Article ID 1234567.

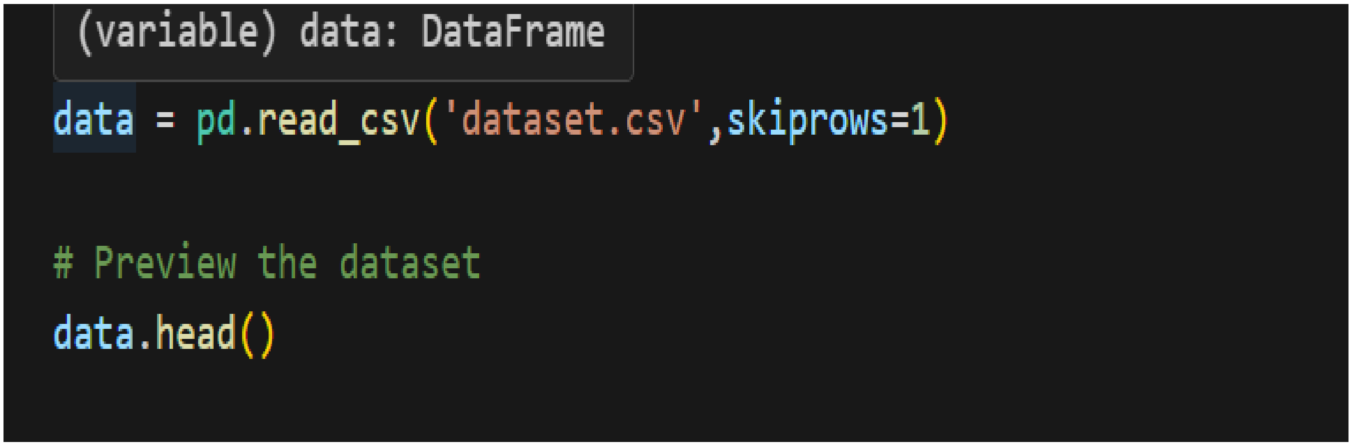
26. Leonardo Daniel Tavares , Andre Manoel , Thiago Henrique Rizzi Donato. (2022). Prediction of metabolic syndrome: A machine learning approach to help primary prevention. Preventive Medicine Reports, 25, pp. 101-112.

# APPENDIX

## A . sample code

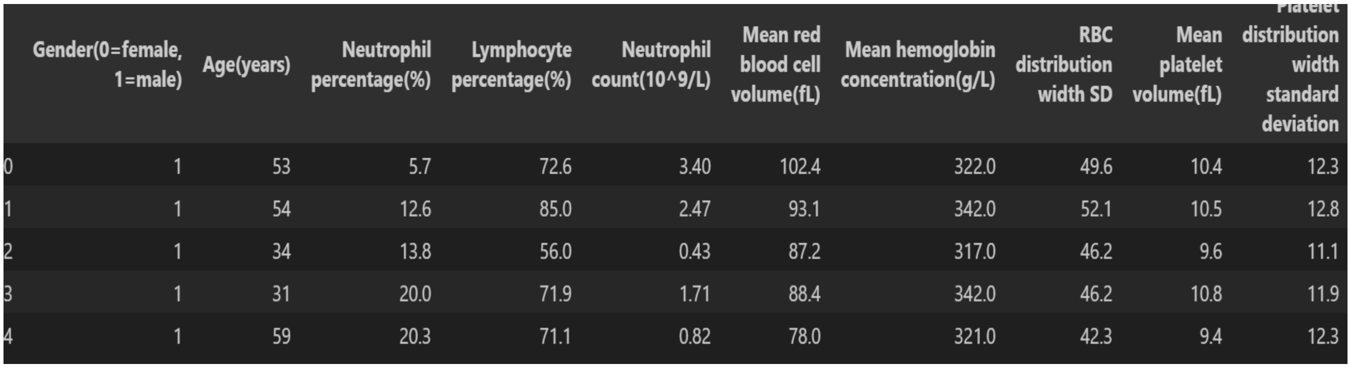


```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import mutual_info_classif, SelectKBest
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.feature_selection import SelectKBest, chi2
```



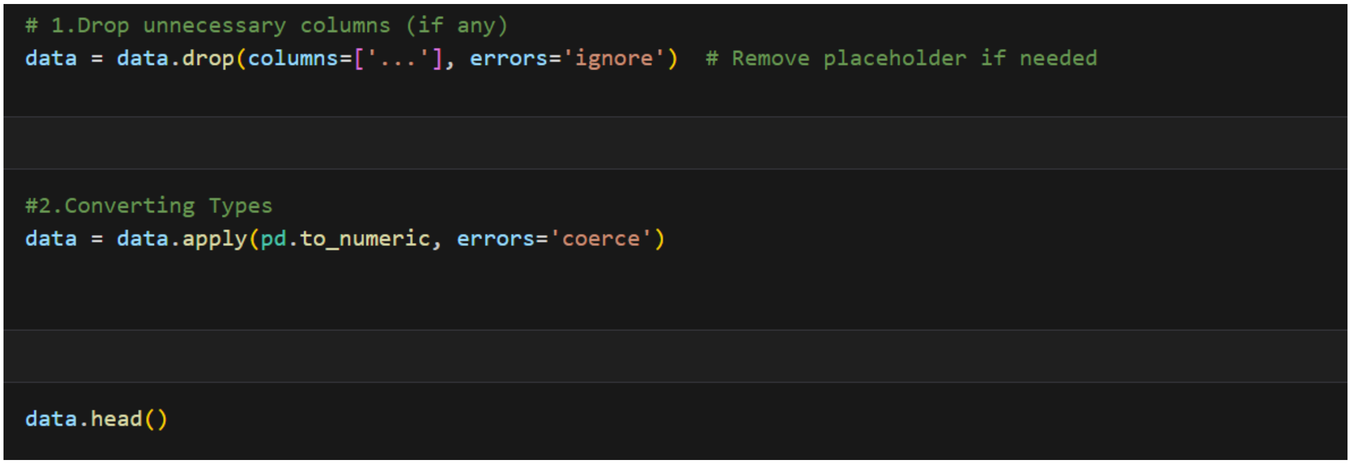
```
(variable) data: DataFrame
data = pd.read_csv('dataset.csv',skiprows=1)

# Preview the dataset
data.head()
```

**Output :**



| | Gender(0=female, 1=male) | Age(years) | Neutrophil percentage(%) | Lymphocyte percentage(%) | Neutrophil count(10^9/L) | Mean red blood cell volume(fL) | Mean hemoglobin concentration(g/L) | RBC distribution width SD | Mean platelet volume(fL) | Platelet distribution width standard deviation |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 53 | 5.7 | 72.6 | 3.40 | 102.4 | 322.0 | 49.6 | 10.4 | 12.3 |
| 1 | 1 | 54 | 12.6 | 85.0 | 2.47 | 93.1 | 342.0 | 52.1 | 10.5 | 12.8 |
| 2 | 1 | 34 | 13.8 | 56.0 | 0.43 | 87.2 | 317.0 | 46.2 | 9.6 | 11.1 |
| 3 | 1 | 31 | 20.0 | 71.9 | 1.71 | 88.4 | 342.0 | 46.2 | 10.8 | 11.9 |
| 4 | 1 | 59 | 20.3 | 71.1 | 0.82 | 78.0 | 321.0 | 42.3 | 9.4 | 12.3 |

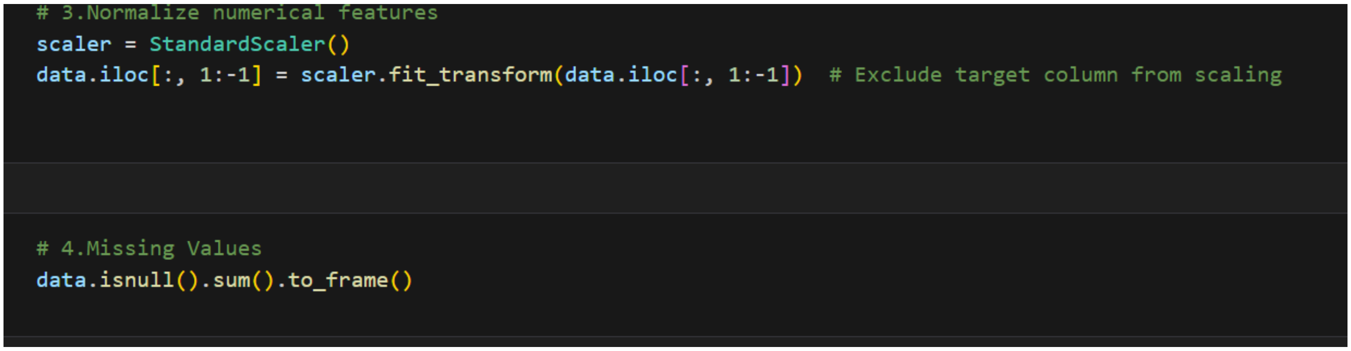
## PRE-PROCESSING



```
# 1.Drop unnecessary columns (if any)
data = data.drop(columns=['...'], errors='ignore')  # Remove placeholder if needed
```

```
#2.Converting Types
data = data.apply(pd.to_numeric, errors='coerce')
```

```
data.head()
```

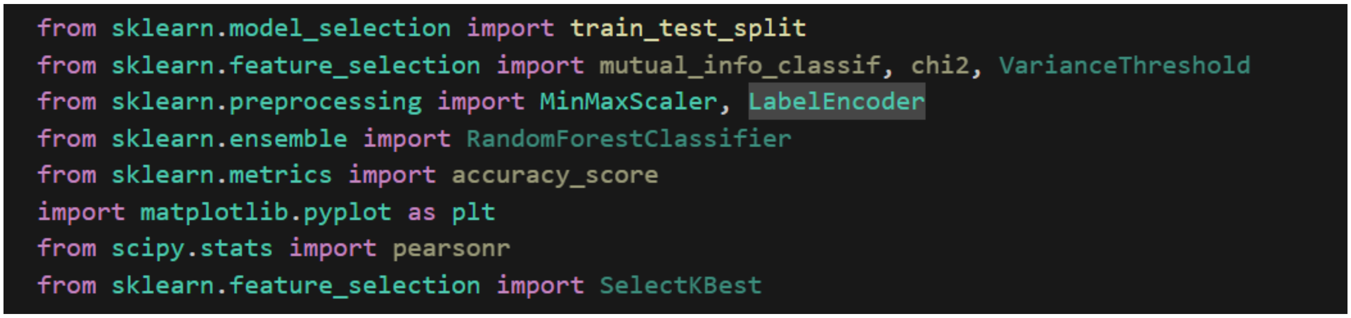


```
# 3.Normalize numerical features
scaler = StandardScaler()
data.iloc[:, 1:-1] = scaler.fit_transform(data.iloc[:, 1:-1])  # Exclude target column from scaling
```

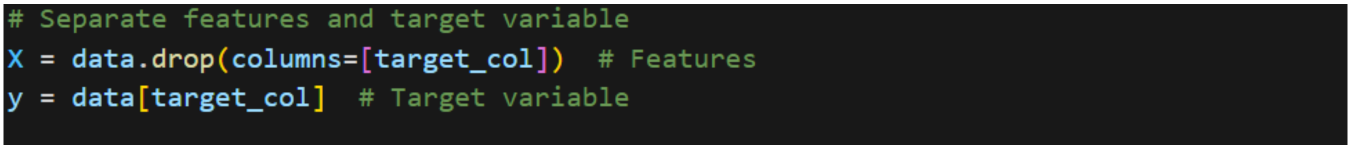
```
# 4.Missing Values
data.isnull().sum().to_frame()
```

## FEATURE SELECTION

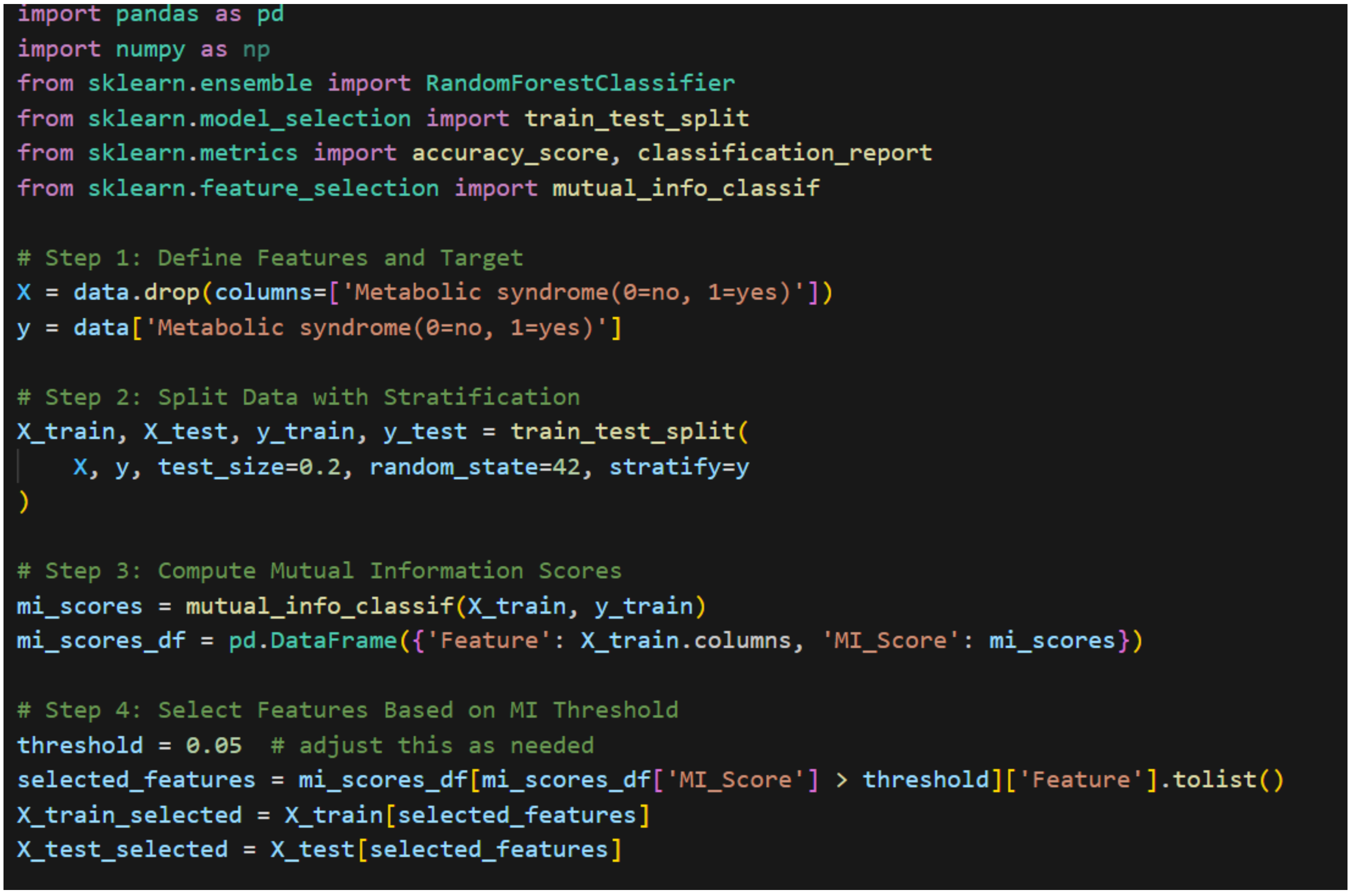
### 1.FILTER METHOD:



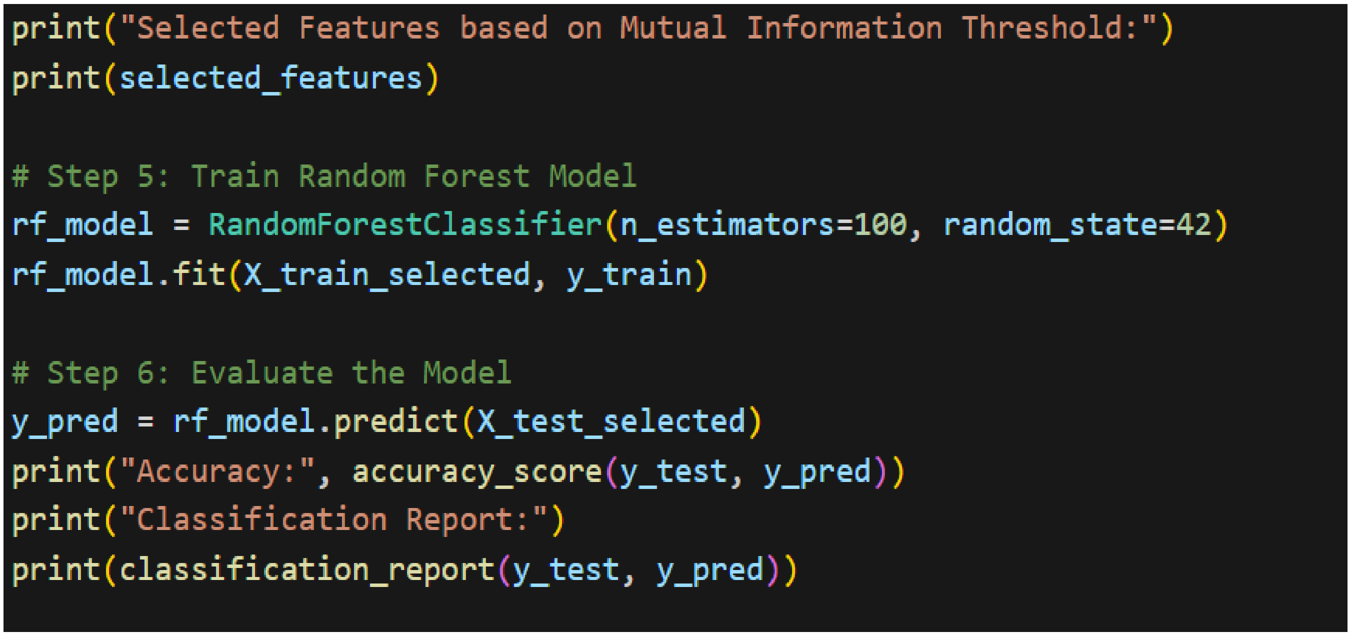
```
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import mutual_info_classif, chi2, VarianceThreshold
from sklearn.preprocessing import MinMaxScaler, LabelEncoder
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
from sklearn.feature_selection import SelectKBest
```



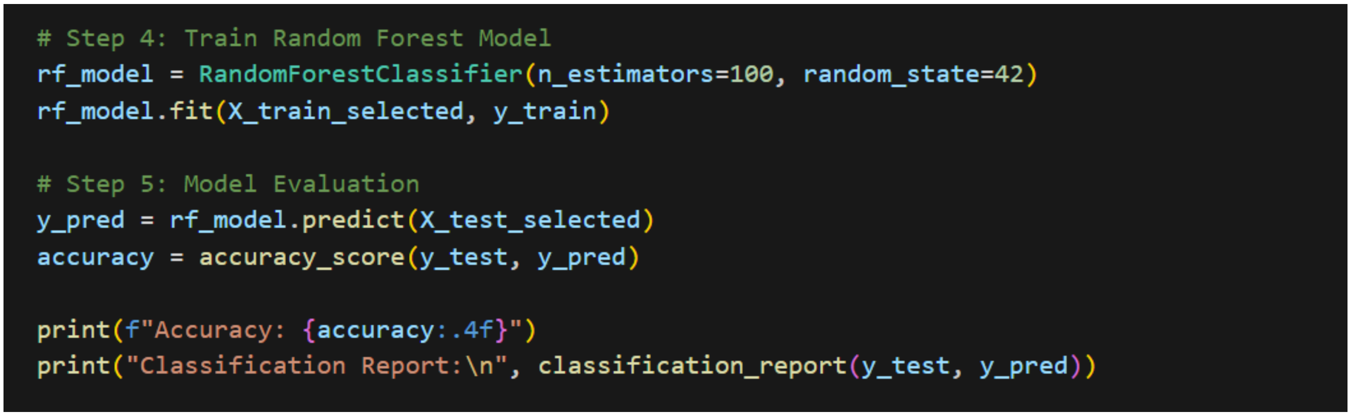
```
# Separate features and target variable
X = data.drop(columns=[target_col])  # Features
y = data[target_col]  # Target variable
```



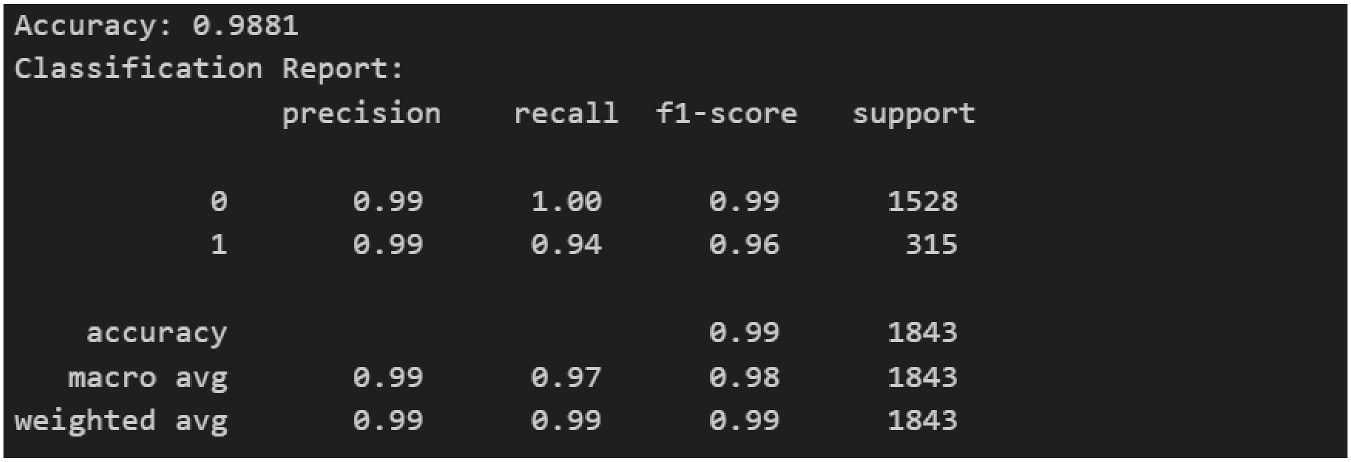
```
import pandas as pd
import numpy as np
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
from sklearn.feature_selection import mutual_info_classif

# Step 1: Define Features and Target
X = data.drop(columns=['Metabolic syndrome(0=no, 1=yes)'])
y = data['Metabolic syndrome(0=no, 1=yes)']

# Step 2: Split Data with Stratification
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)

# Step 3: Compute Mutual Information Scores
mi_scores = mutual_info_classif(X_train, y_train)
mi_scores_df = pd.DataFrame({'Feature': X_train.columns, 'MI_Score': mi_scores})

# Step 4: Select Features Based on MI Threshold
threshold = 0.05  # adjust this as needed
selected_features = mi_scores_df[mi_scores_df['MI_Score'] > threshold]['Feature'].tolist()
X_train_selected = X_train[selected_features]
X_test_selected = X_test[selected_features]
```



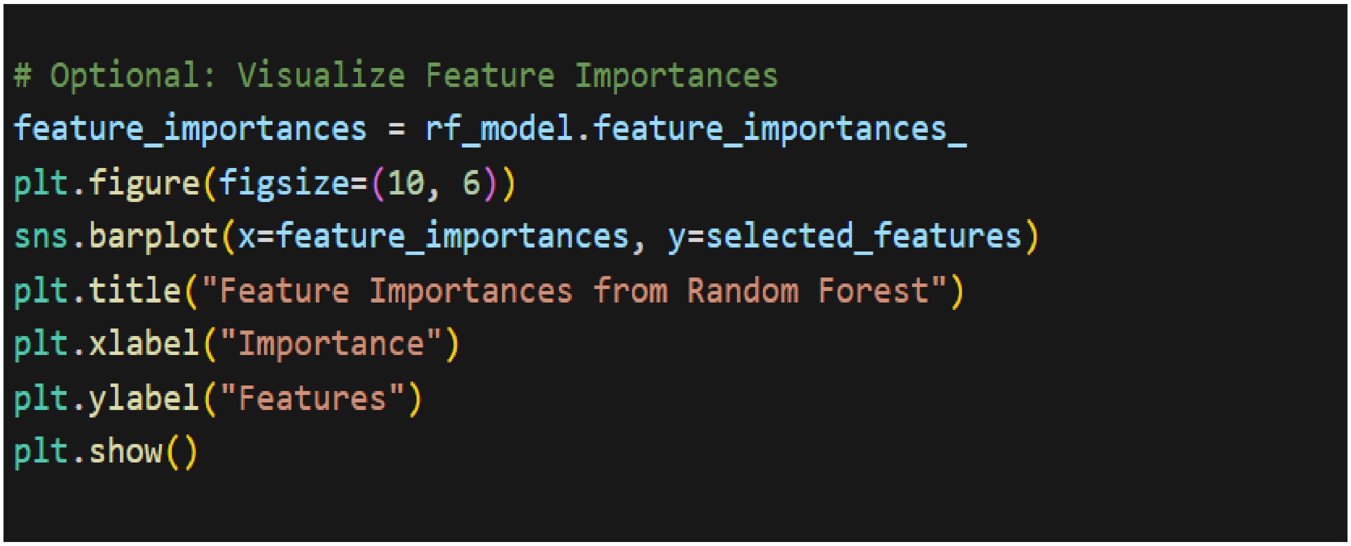
```
print("Selected Features based on Mutual Information Threshold:")
print(selected_features)

# Step 5: Train Random Forest Model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train_selected, y_train)

# Step 6: Evaluate the Model
y_pred = rf_model.predict(X_test_selected)
print("Accuracy:", accuracy_score(y_test, y_pred))
print("Classification Report:")
print(classification_report(y_test, y_pred))
```



```python
# Step 4: Train Random Forest Model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train_selected, y_train)

# Step 5: Model Evaluation
y_pred = rf_model.predict(X_test_selected)
accuracy = accuracy_score(y_test, y_pred)

print(f"Accuracy: {accuracy:.4f}")
print("Classification Report:\n", classification_report(y_test, y_pred))
```

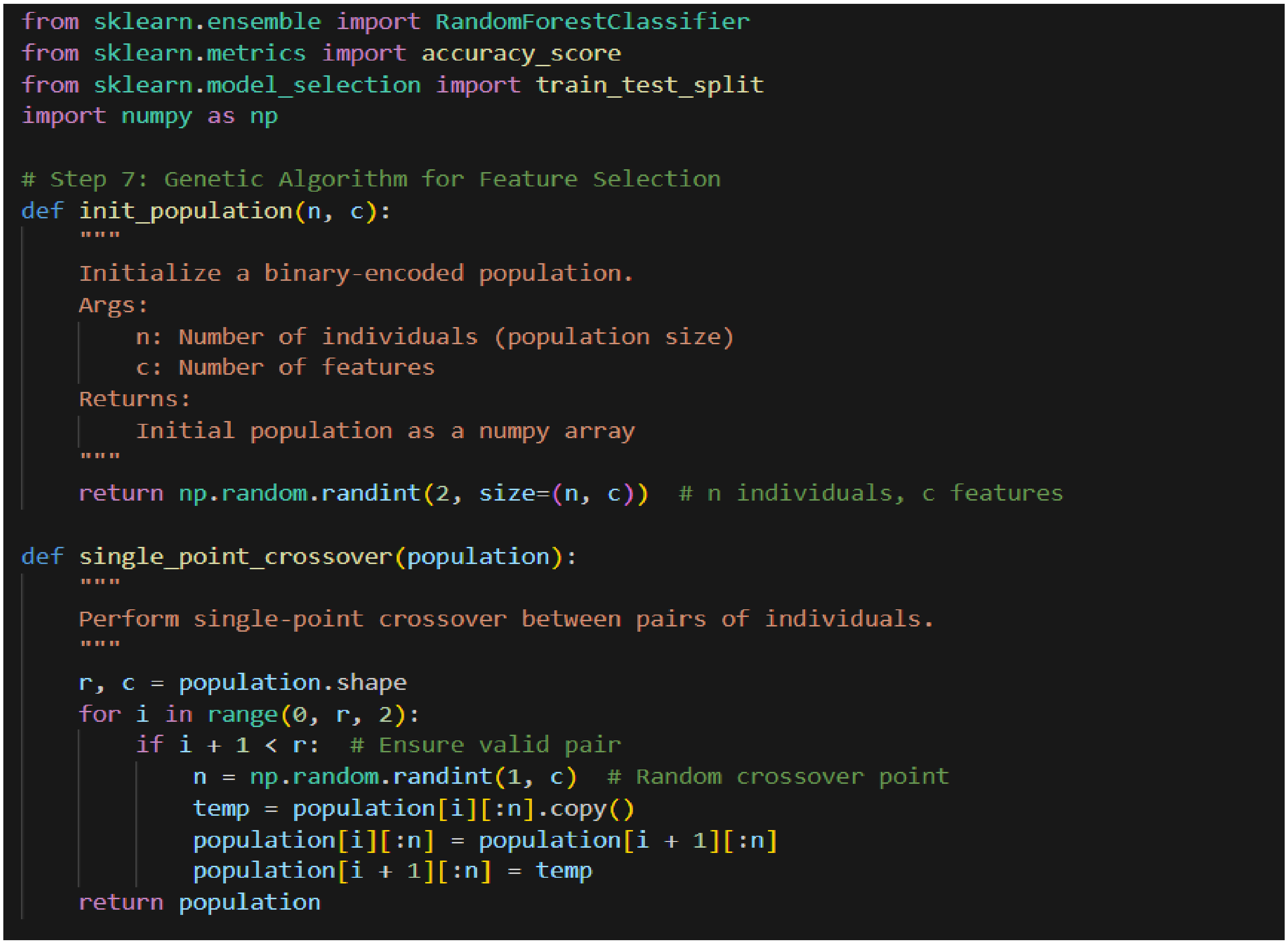


```
Accuracy: 0.9881
Classification Report:
               precision    recall  f1-score   support

           0       0.99      1.00      0.99      1528
           1       0.99      0.94      0.96       315

    accuracy                           0.99      1843
   macro avg       0.99      0.97      0.98      1843
weighted avg       0.99      0.99      0.99      1843
```

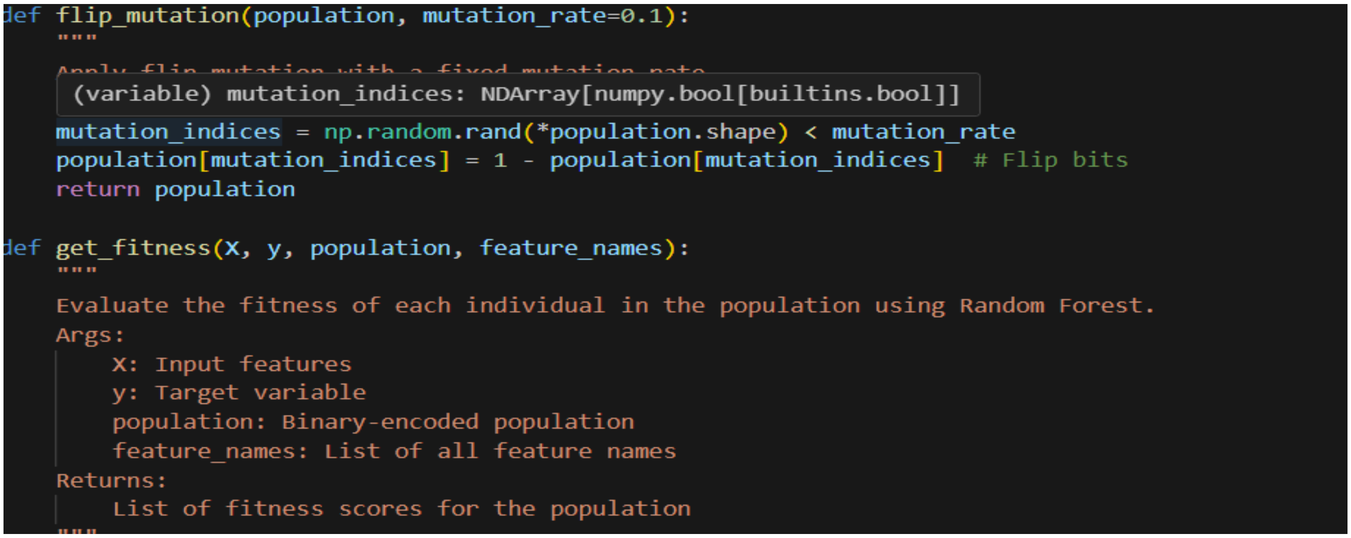


```python
# Optional: Visualize Feature Importances
feature_importances = rf_model.feature_importances_
plt.figure(figsize=(10, 6))
sns.barplot(x=feature_importances, y=selected_features)
plt.title("Feature Importances from Random Forest")
plt.xlabel("Importance")
plt.ylabel("Features")
plt.show()
```

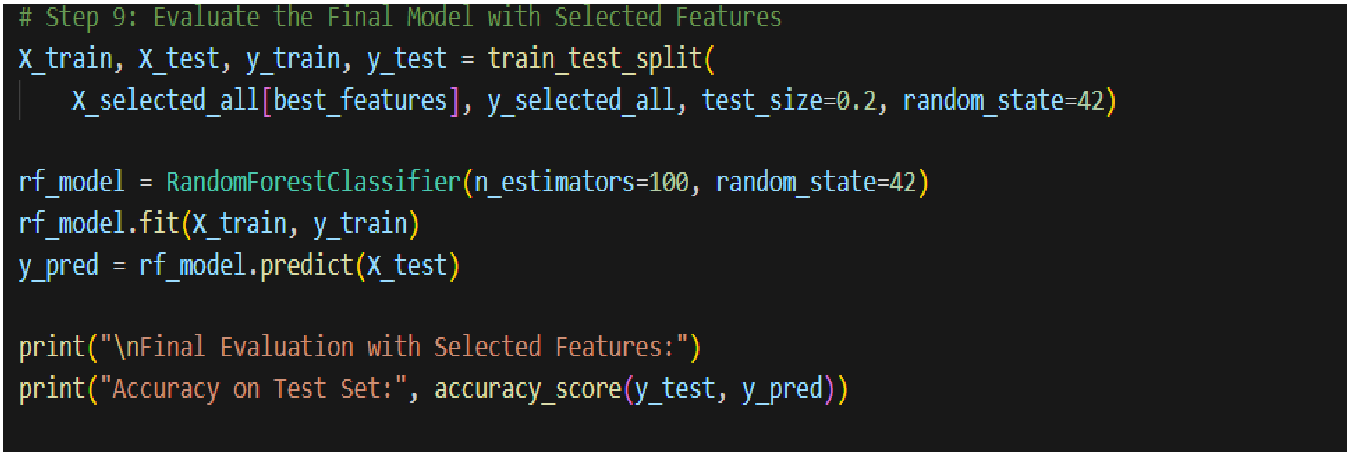
## 2.WRAPPER METHOD



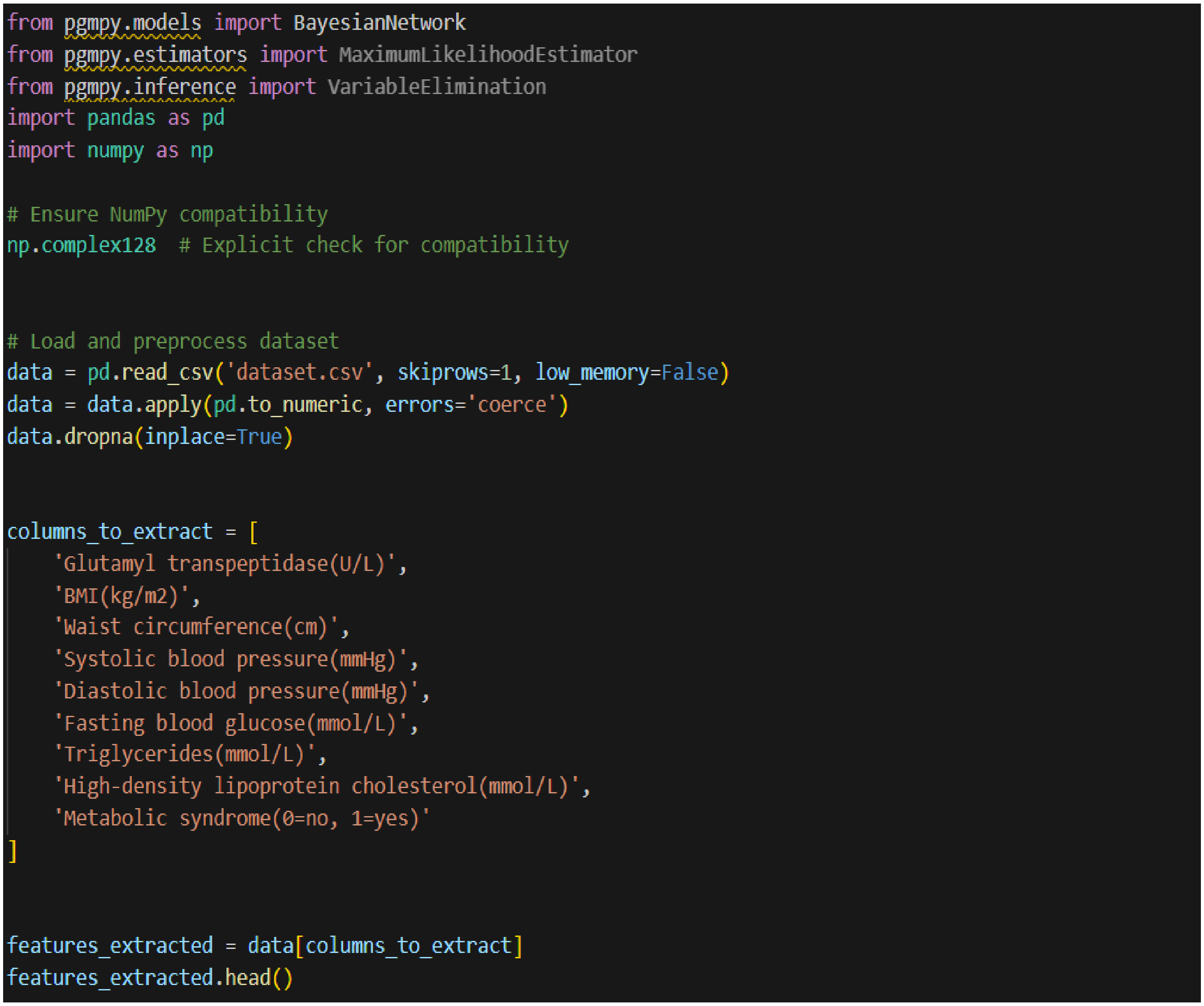
```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
import numpy as np

# Step 7: Genetic Algorithm for Feature Selection
def init_population(n, c):
    """
    Initialize a binary-encoded population.
    Args:
        n: Number of individuals (population size)
        c: Number of features
    Returns:
        Initial population as a numpy array
    """
    return np.random.randint(2, size=(n, c))  # n individuals, c features

def single_point_crossover(population):
    """
    Perform single-point crossover between pairs of individuals.
    """
    r, c = population.shape
    for i in range(0, r, 2):
        if i + 1 < r:  # Ensure valid pair
            n = np.random.randint(1, c)  # Random crossover point
            temp = population[i][:n].copy()
            population[i][:n] = population[i + 1][:n]
            population[i + 1][:n] = temp
    return population
```



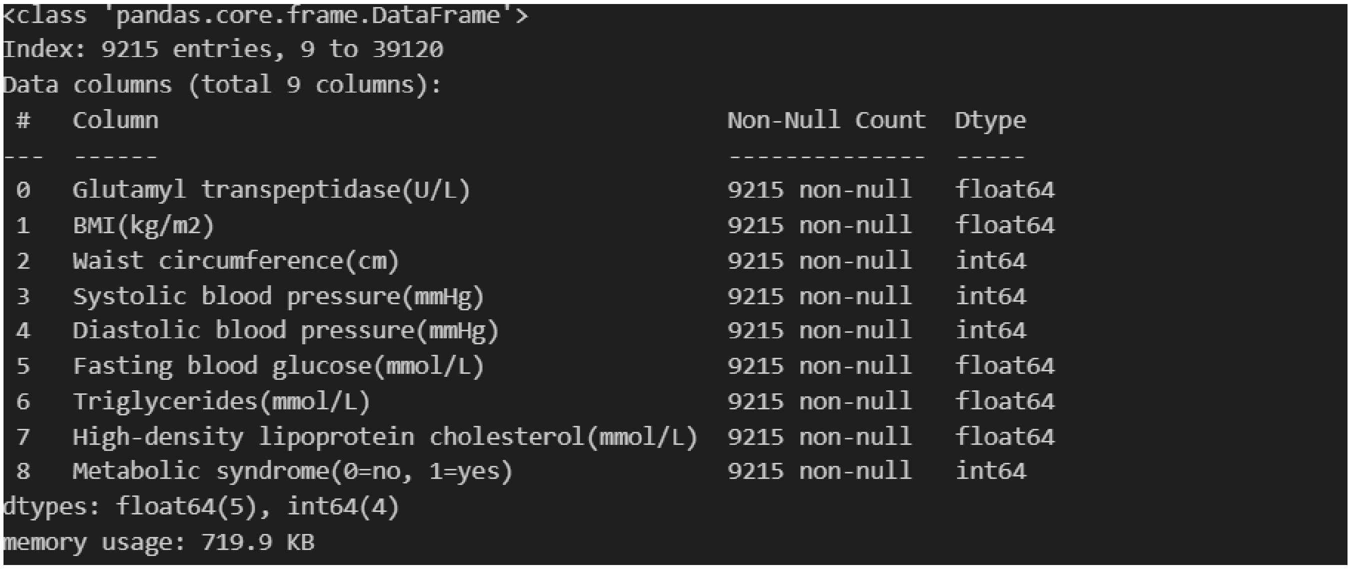
```
def flip_mutation(population, mutation_rate=0.1):
    """
    Apply flip mutation with a fixed mutation rate.
    (variable) mutation_indices: NDArray[numpy.bool[builtins.bool]]
    mutation_indices = np.random.rand(*population.shape) < mutation_rate
    population[mutation_indices] = 1 - population[mutation_indices]  # Flip bits
    return population

def get_fitness(X, y, population, feature_names):
    """
    Evaluate the fitness of each individual in the population using Random Forest.
    Args:
        X: Input features
        y: Target variable
        population: Binary-encoded population
        feature_names: List of all feature names
    Returns:
        List of fitness scores for the population
    """
```



```
# Step 9: Evaluate the Final Model with Selected Features
X_train, X_test, y_train, y_test = train_test_split(
    X_selected_all[best_features], y_selected_all, test_size=0.2, random_state=42)

rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
y_pred = rf_model.predict(X_test)

print("\nFinal Evaluation with Selected Features:")
print("Accuracy on Test Set:", accuracy_score(y_test, y_pred))
```

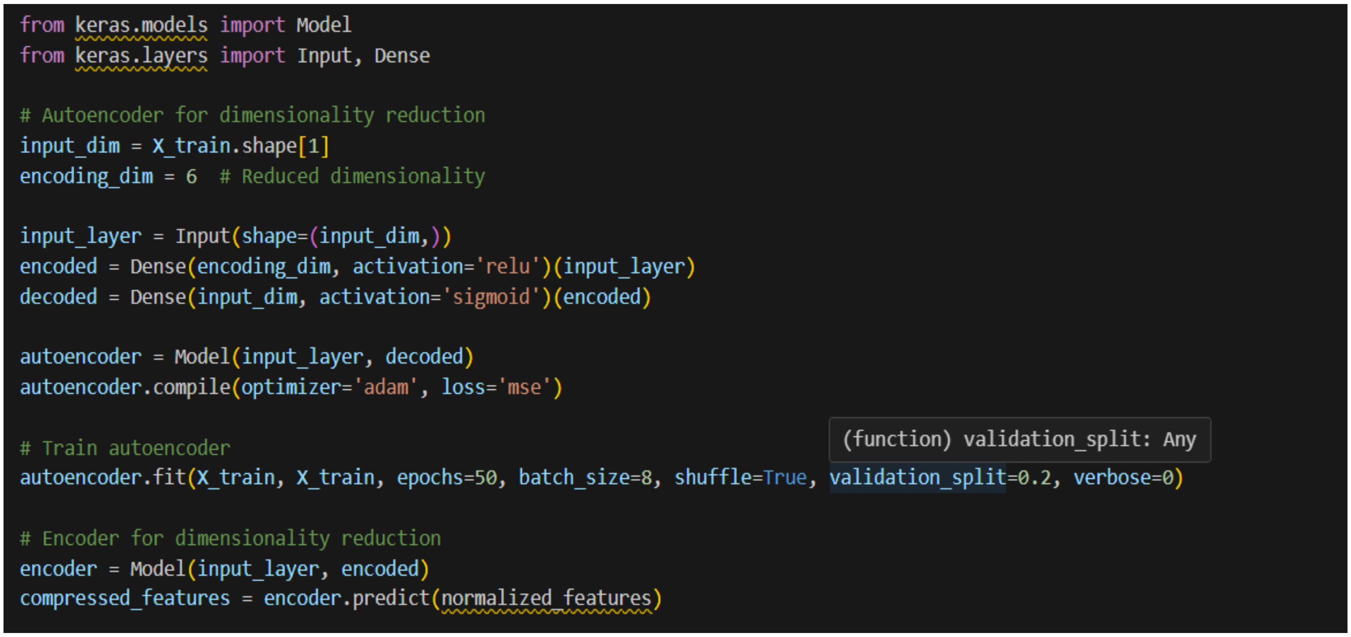


```
from pgmpy.models import BayesianNetwork
from pgmpy.estimators import MaximumLikelihoodEstimator
from pgmpy.inference import VariableElimination
import pandas as pd
import numpy as np

# Ensure NumPy compatibility
np.complex128  # Explicit check for compatibility


# Load and preprocess dataset
data = pd.read_csv('dataset.csv', skiprows=1, low_memory=False)
data = data.apply(pd.to_numeric, errors='coerce')
data.dropna(inplace=True)


columns_to_extract = [
    'Glutamyl transpeptidase(U/L)',
    'BMI(kg/m2)',
    'Waist circumference(cm)',
    'Systolic blood pressure(mmHg)',
    'Diastolic blood pressure(mmHg)',
    'Fasting blood glucose(mmol/L)',
    'Triglycerides(mmol/L)',
    'High-density lipoprotein cholesterol(mmol/L)',
    'Metabolic syndrome(0=no, 1=yes)'
]


features_extracted = data[columns_to_extract]
features_extracted.head()
```

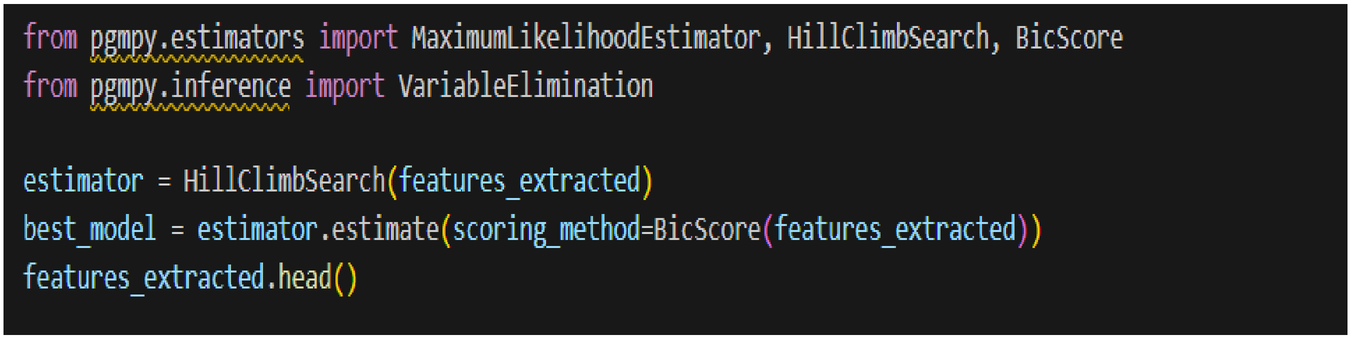


```
<class 'pandas.core.frame.DataFrame'>
Index: 9215 entries, 9 to 39120
Data columns (total 9 columns):
 #   Column                                   Non-Null Count  Dtype  
---  ------                                   --------------  -----  
 0   Glutamyl transpeptidase(U/L)             9215 non-null   float64
 1   BMI(kg/m2)                               9215 non-null   float64
 2   Waist circumference(cm)                  9215 non-null   int64  
 3   Systolic blood pressure(mmHg)            9215 non-null   int64  
 4   Diastolic blood pressure(mmHg)           9215 non-null   int64  
 5   Fasting blood glucose(mmol/L)            9215 non-null   float64
 6   Triglycerides(mmol/L)                    9215 non-null   float64
 7   High-density lipoprotein cholesterol(mmol/L)  9215 non-null   float64
 8   Metabolic syndrome(0=no, 1=yes)          9215 non-null   int64  
dtypes: float64(5), int64(4)
memory usage: 719.9 KB
```

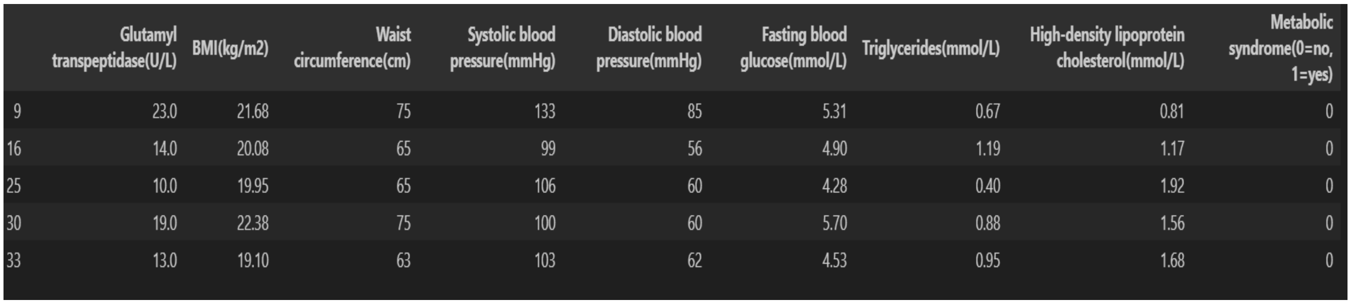
## AUTO-ENCODERS PRE -TRAINING:



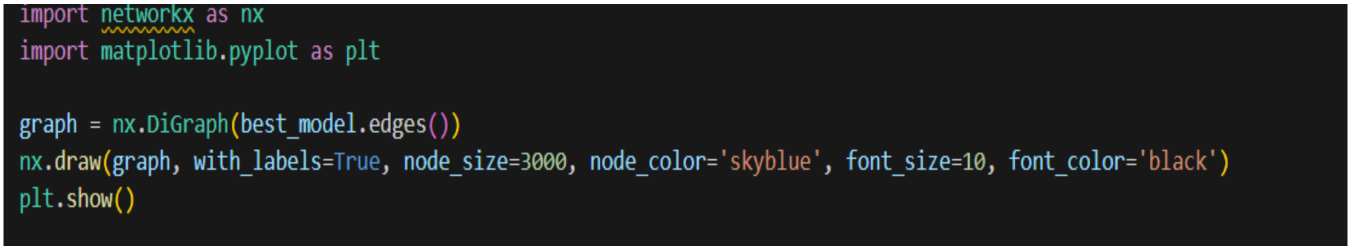
```python
from keras.models import Model
from keras.layers import Input, Dense

# Autoencoder for dimensionality reduction
input_dim = X_train.shape[1]
encoding_dim = 6  # Reduced dimensionality

input_layer = Input(shape=(input_dim,))
encoded = Dense(encoding_dim, activation='relu')(input_layer)
decoded = Dense(input_dim, activation='sigmoid')(encoded)

autoencoder = Model(input_layer, decoded)
autoencoder.compile(optimizer='adam', loss='mse')

# Train autoencoder
autoencoder.fit(X_train, X_train, epochs=50, batch_size=8, shuffle=True, validation_split=0.2, verbose=0)

# Encoder for dimensionality reduction
encoder = Model(input_layer, encoded)
compressed_features = encoder.predict(normalized_features)
```



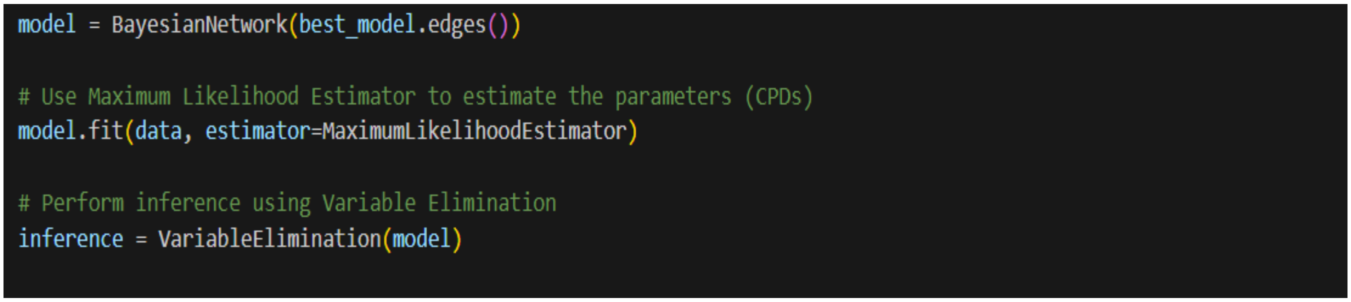
```python
from pgmpy.estimators import MaximumLikelihoodEstimator, HillClimbSearch, BicScore
from pgmpy.inference import VariableElimination

estimator = HillClimbSearch(features_extracted)
best_model = estimator.estimate(scoring_method=BicScore(features_extracted))
features_extracted.head()
```



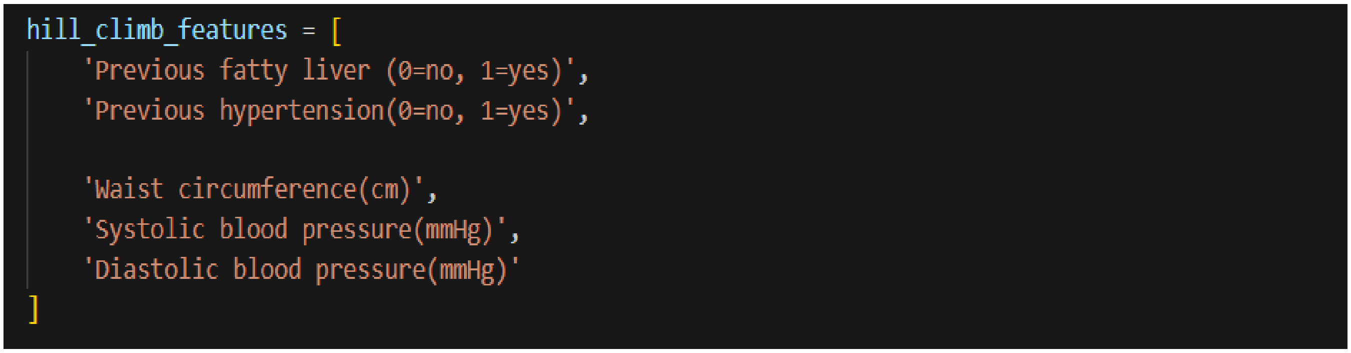
| | Glutamyl transpeptidase(U/L) | BMI(kg/m2) | Waist circumference(cm) | Systolic blood pressure(mmHg) | Diastolic blood pressure(mmHg) | Fasting blood glucose(mmol/L) | Triglycerides(mmol/L) | High-density lipoprotein cholesterol(mmol/L) | Metabolic syndrome(0=no, 1=yes) |
|---|---|---|---|---|---|---|---|---|---|
| 9 | 23.0 | 21.68 | 75 | 133 | 85 | 5.31 | 0.67 | 0.81 | 0 |
| 16 | 14.0 | 20.08 | 65 | 99 | 56 | 4.90 | 1.19 | 1.17 | 0 |
| 25 | 10.0 | 19.95 | 65 | 106 | 60 | 4.28 | 0.40 | 1.92 | 0 |
| 30 | 19.0 | 22.38 | 75 | 100 | 60 | 5.70 | 0.88 | 1.56 | 0 |
| 33 | 13.0 | 19.10 | 63 | 103 | 62 | 4.53 | 0.95 | 1.68 | 0 |



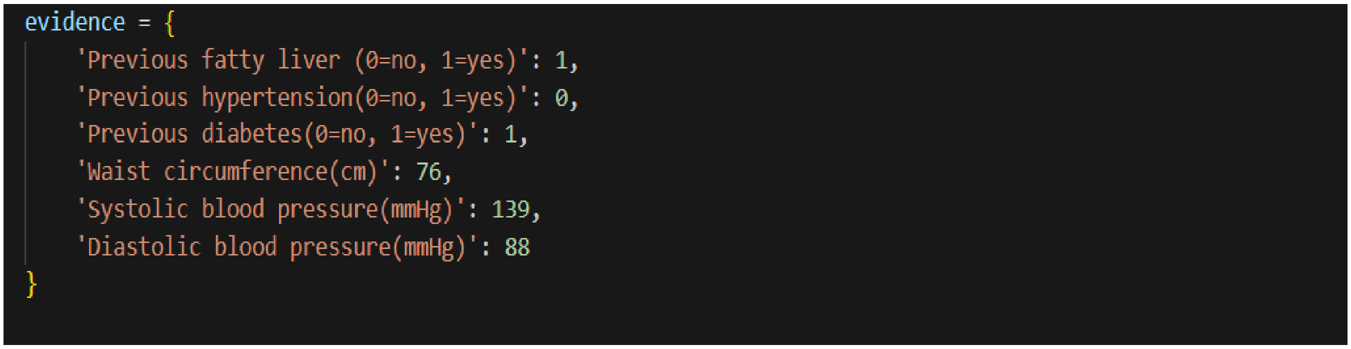
```
import networkx as nx
import matplotlib.pyplot as plt

graph = nx.DiGraph(best_model.edges())
nx.draw(graph, with_labels=True, node_size=3000, node_color='skyblue', font_size=10, font_color='black')
plt.show()
```



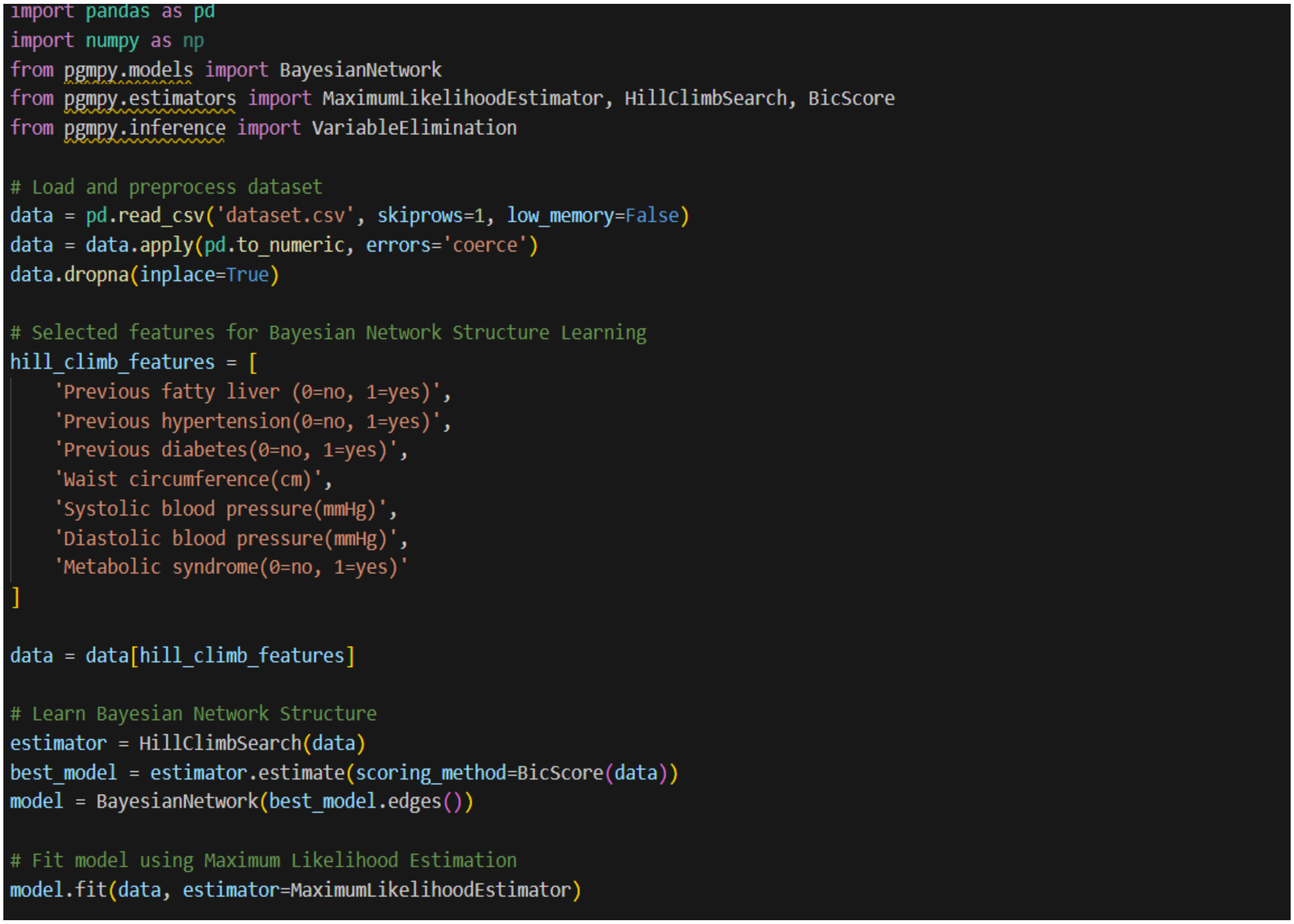
```
model = BayesianNetwork(best_model.edges())

# Use Maximum Likelihood Estimator to estimate the parameters (CPDs)
model.fit(data, estimator=MaximumLikelihoodEstimator)

# Perform inference using Variable Elimination
inference = VariableElimination(model)
```



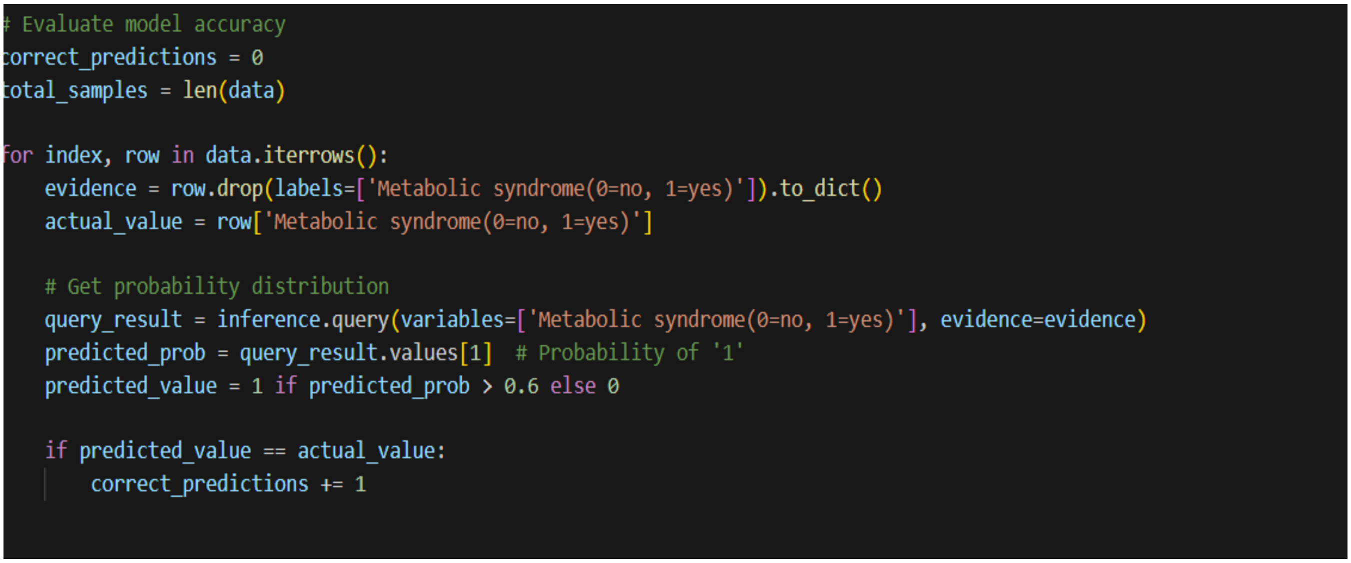
```
hill_climb_features = [
    'Previous fatty liver (0=no, 1=yes)',
    'Previous hypertension(0=no, 1=yes)',

    'Waist circumference(cm)',
    'Systolic blood pressure(mmHg)',
    'Diastolic blood pressure(mmHg)'
]
```



```
evidence = {
    'Previous fatty liver (0=no, 1=yes)': 1,
    'Previous hypertension(0=no, 1=yes)': 0,
    'Previous diabetes(0=no, 1=yes)': 1,
    'Waist circumference(cm)': 76,
    'Systolic blood pressure(mmHg)': 139,
    'Diastolic blood pressure(mmHg)': 88
}
```

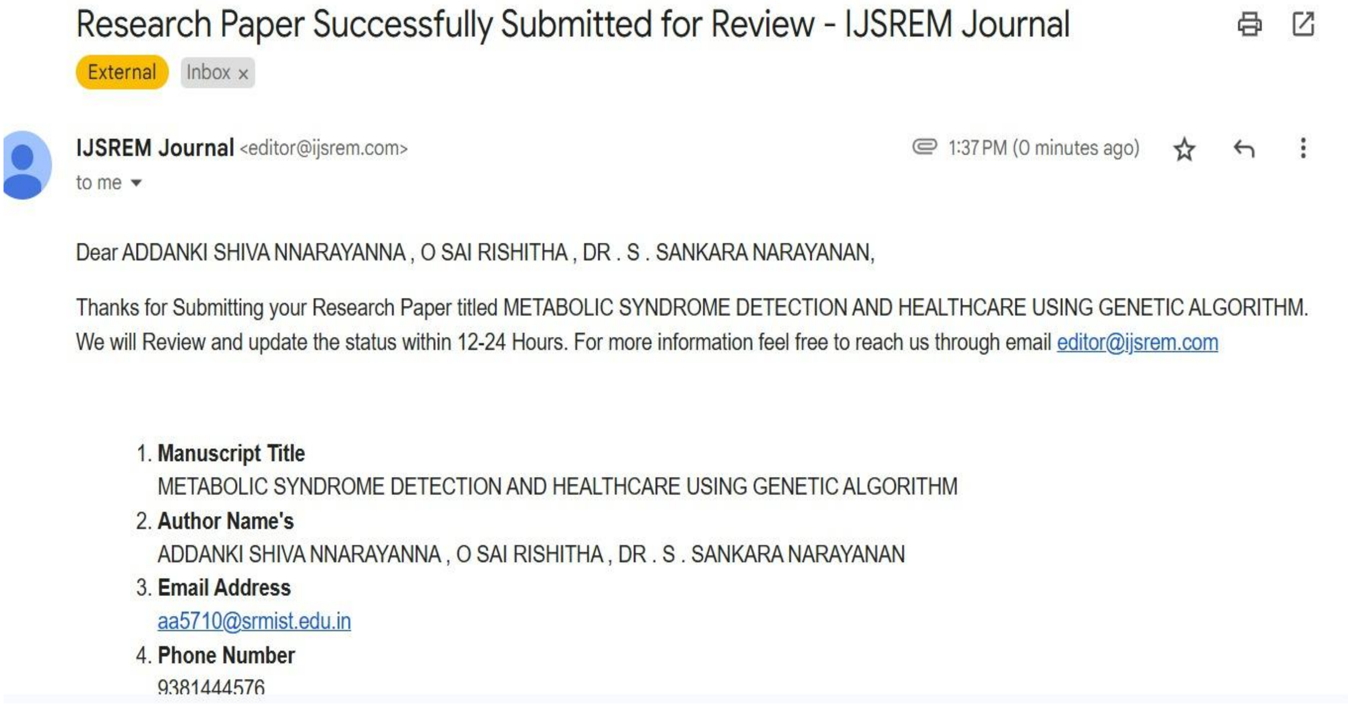


```
import pandas as pd
import numpy as np
from pgmpy.models import BayesianNetwork
from pgmpy.estimators import MaximumLikelihoodEstimator, HillClimbSearch, BicScore
from pgmpy.inference import VariableElimination

# Load and preprocess dataset
data = pd.read_csv('dataset.csv', skiprows=1, low_memory=False)
data = data.apply(pd.to_numeric, errors='coerce')
data.dropna(inplace=True)

# Selected features for Bayesian Network Structure Learning
hill_climb_features = [
    'Previous fatty liver (0=no, 1=yes)',
    'Previous hypertension(0=no, 1=yes)',
    'Previous diabetes(0=no, 1=yes)',
    'Waist circumference(cm)',
    'Systolic blood pressure(mmHg)',
    'Diastolic blood pressure(mmHg)',
    'Metabolic syndrome(0=no, 1=yes)'
]

data = data[hill_climb_features]

# Learn Bayesian Network Structure
estimator = HillClimbSearch(data)
best_model = estimator.estimate(scoring_method=BicScore(data))
model = BayesianNetwork(best_model.edges())

# Fit model using Maximum Likelihood Estimation
model.fit(data, estimator=MaximumLikelihoodEstimator)
```



```
# Evaluate model accuracy
orrect_predictions = 0
otal_samples = len(data)

for index, row in data.iterrows():
    evidence = row.drop(labels=['Metabolic syndrome(0=no, 1=yes)']).to_dict()
    actual_value = row['Metabolic syndrome(0=no, 1=yes)']

    # Get probability distribution
    query_result = inference.query(variables=['Metabolic syndrome(0=no, 1=yes)'], evidence=evidence)
    predicted_prob = query_result.values[1]  # Probability of '1'
    predicted_value = 1 if predicted_prob > 0.6 else 0

    if predicted_value == actual_value:
        correct_predictions += 1
```

# B . PUBLICATION PROOF :



Research Paper Successfully Submitted for Review - IJSREM Journal

**IJSREM Journal** <editor@ijsrem.com>

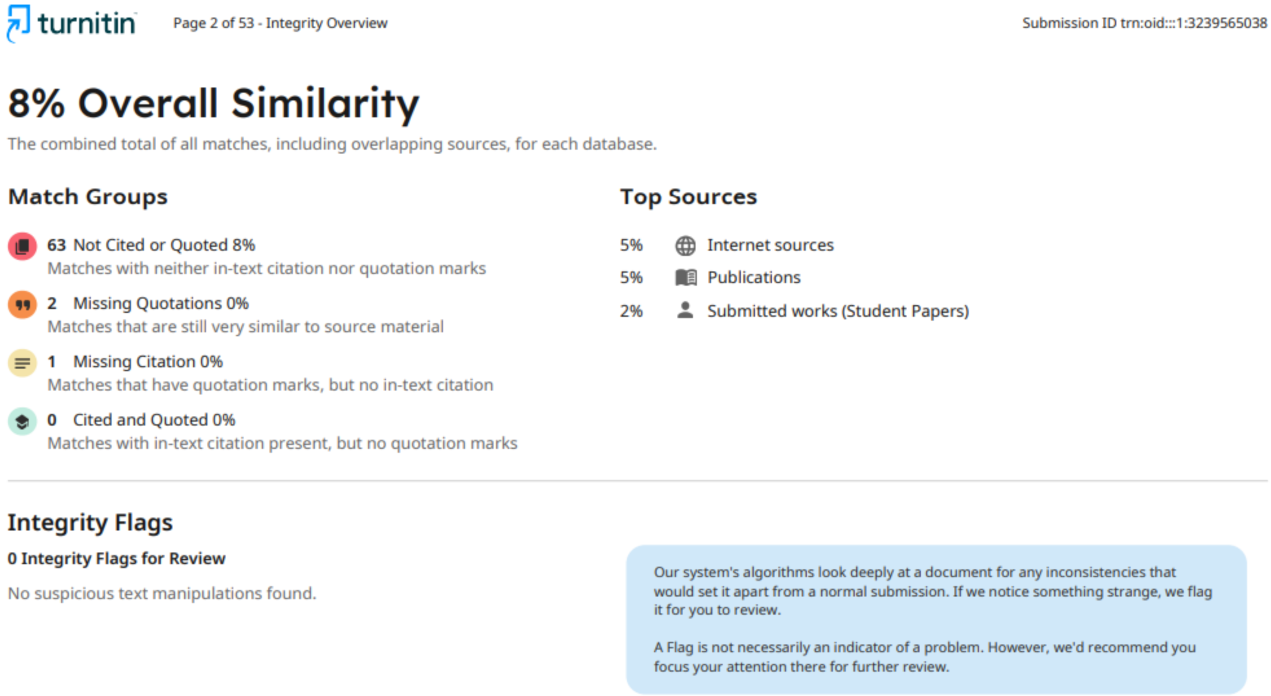
to me

Dear ADDANKI SHIVA NNARAYANNA , O SAI RISHITHA , DR . S . SANKARA NARAYANAN,

Thanks for Submitting your Research Paper titled METABOLIC SYNDROME DETECTION AND HEALTHCARE USING GENETIC ALGORITHM. We will Review and update the status within 12-24 Hours. For more information feel free to reach us through email editor@ijsrem.com

1. **Manuscript Title**
   METABOLIC SYNDROME DETECTION AND HEALTHCARE USING GENETIC ALGORITHM
2. **Author Name's**
   ADDANKI SHIVA NNARAYANNA , O SAI RISHITHA , DR . S . SANKARA NARAYANAN
3. **Email Address**
   aa5710@srmist.edu.in
4. **Phone Number**
   9381444576

## C. PLAGIARISM REPORT:



turnitin Page 2 of 53 - Integrity Overview Submission ID trn:oid:::1:3239565038

# 8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

### Match Groups

**63** Not Cited or Quoted 8%
Matches with neither in-text citation nor quotation marks

**2** Missing Quotations 0%
Matches that are still very similar to source material

**1** Missing Citation 0%
Matches that have quotation marks, but no in-text citation

**0** Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

### Top Sources

| | |
|---|---|
| 5% | Internet sources |
| 5% | Publications |
| 2% | Submitted works (Student Papers) |

### Integrity Flags

**0 Integrity Flags for Review**

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## D.COE REPORT: