Fuzzy Logic Practical Report on Coronary Heart Disease Risk Assessment

Dinis Marques Firmino (P13240786)

January 12, 2018

1 Introduction

The objective of this practical assignment was to design, develop and test a fuzzy inference system (FIS) for an advisory task of personal choice. The system domain chosen for this assignment was the evaluation of patient risk of coronary heart disease (CHD). All efforts during each stage of the project will be put towards making sure the FIS assesses patients correctly, as if the system were to be used in a real environment.

The interest in pursuing this problem stems from various literature sources that have also attempted to tackle this problem [4, 5, 6, 7]. These systems are formally called expert systems, typically because they require an expert with extensive knowledge of the problem domain (medicine in this case) to construct the rule-bases and aid in the design of the fuzzy sets. However, the abundance of real medical data on CHD [2], opens up doors for researchers in AI with no previous domain knowledge to tackle this problem and many others.

Two main tools were used during the design and development of the FIS. QTFuzzyLite [1] will be the main tool in the creation of the FIS. This choice was because it is a very lightweight software package when compared to Matlab, and offers a much faster work flow to create and modify various properties of the system. The tool also allows for exporting of the created FIS into *.fis* files which can be imported into Matlab s Fuzzy Toolbox. Matlab will be used in the evaluation phase to test the performance of the FIS on real data, and for the observation of the surface views produced by the FIS. Also, Python will be used to conduct some basic data analysis of the dataset by generating plots, and for applying pre-processing techniques to the original dataset. Excel was also used as a convenient tool to manipulate datasets quickly.

The structure of the report is as follows. Section 2 provides an introduction to the basis of the work by describing and justifying the choice of data used for this study. Section 3 follows on by detailing the analysis conducted and pre-processing techniques used on the dataset to extract key information to be used during the design and development of the FIS. Section 4 details the structure of the FIS and all its components. Section 5 explains the testing methodology used and similarly, Section 6 the tuning methodology. Section 7 focuses on analysing the best performing FIS and Section 8 provides an in depth discussion on the results of the development and aims to provide justification for the design decisions made and suggest better approaches to avoid the various pitfalls of the original approach. Finally, Section 9 wraps up the report by summarising the work done at each stage of the project.

2 Data

The prediction of heart disease risk is considered a critical topic. Meaning that results outputted from the system could impact someone's life directly. While this work is not aimed at creating a production ready system, the process and results should be grounded by reality. For this to happen, a dataset containing a generous amount of samples of real labelled medical data on patients that have undergone CHD examinations needs to be analysed to provide a solid basis for creating the FIS.

The dataset used in this project was obtained from [2], and consists of a total of 304 samples of anonymized patient data which includes heart disease related attributes along with a predicted target column. Out of the 304 samples, 164 patients were diagnosed with no heart disease (0), the other 139 were diagnosed with some kind of CHD (1-4). The table in Appendix A contains the 14 attributes present in the dataset along with their descriptions.

The main attribute which is of high relevance to the design of our FIS output is the predicted column which classifies the patient based on the severity of the CHD. Classes 1-4 signify the presence of heart disease, in order of increasing severity such that a patient classified as 4 has been diagnosed with the most severe type of CHD, whereas 1 is the least worse. The severity is measured by the narrowing of the coronary artery [**?**]

3 Analysis

The sheer number of attributes that are used to define the class of heart disease in the patient would be a significant burden on the fuzzy inference system if all were to be used as input variables. This is primarily because a very large rule-base would be needed to assess the level of heart disease risk, which would make the process very time consuming and impractical. Factors such as system performance would also be affected. In order to identify attributes that relate most to the target variable, feature selection techniques were employed to reduce the dimensionality of the problem. Once the variables were selected, a brief analysis of these variables was performed to extract key information.

3.1 Feature Selection

Firstly, to reduce the dimensionality of the problem i.e the number of input attributes into the system, a popular univariate feature selection technique called Chi2 was employed. In summary, the technique excludes variables that have low correlation to the target variable, keeping only a user specified k number of values. k was set to 4 for this selection, which after some consideration was thought to be a suitable number of inputs to the system as it largely decreases the total number of rules needed while still providing enough attributes for an accurate risk assessment. In addition, a decision to remove categorical values from this selection was made. This is because these attributes, while potentially valuable for risk assessment, don't contain any "fuzzyness" to model with a fuzzy set, and crisp sets would suffice. The table in Appendix B shows the selected attributes along with their scores assigned by the algorithm.

3.2 Statistics

With the input attributes selected, a basic analysis of how the data is distributed was the final step before proceeding to the development of the FIS. This was necessary to obtain some insights on possible suitable membership functions and their characteristic values to aid the design phase. Two types of plots were produced for this investigation, histograms and box plots. These plots sufficed to obtain knowledge on how the sample attributes differed between target classes. This was important to gauge the difference between patients at lower and higher risks of CHD.

First, some brief analysis of the histograms was conducted. The age attribute showed a subtle increasing pattern as the condition worsened, with most cases of heart condition occurring at 55 years and above. Cholesterol showed mean levels of above equal to or above 250 for at risk patients. The maximum heart rate achieved (thalach) was also characterized by a decreasing trend in mean values as the heart disease cases worsen, with the exception of severe cases (prediction = 4). Oldpeak distributions shows the biggest difference between classes, with an increasing mean value as the heart disease worsens. The histograms are attached in Appendix C.

Lastly, the box plots were examined following the initial insights obtained from the histograms. Instead of writing the about these analysis, the results were collected and used to construct a table of approximate boundaries for each class such deducted from variables such as the first and third quartile ranges, the means obtained from the histograms and minimum and maximum values of each attribute. This knowledge is of high value for the development of the FIS in the following section. The table in Appendix E shows the final analysis results, and Appendix D shows the box plots used to obtain the ranges.

4 Development of the Mandami FIS

After the initial analysis and feature selection was complete, and the key information obtained from the medical data, the development of a knowledge grounded FIS was possible. The inference system will be a Mandami type to ensure the rules can be constructed linguistically which both increases the understandability of the system, and aids in the development and testing of the rule-base. This section is split into the main stages of development such as input and output variable creation, constructing the rule-base and the selection choices for the key FIS parameters such as defuzzification technique and aggregation, conjunction, disjunction and implication operators.

4.1 Inputs

The input variables of the FIS were the starting point of the development. The inputs were obtained directly from the feature selection in the previous section, and their design was entirely guided by the statistical information obtained from the histograms and box-plots such as the ranges of each attribute, and suitable domain ranges that the fuzzy set membership functions should cover.

The FIS inputs are attached in Appendix F.

4.1.1 Age

The input variable "Age" consists of 4 fuzzy sets that cover the domain of discourse. These are "young", "middle", "old" and "very old". The membership functions for the sets "young" and "very old" are trapezoidal, whereas "middle" and "old" are triangular. The range of the domain is 0-100 years of age.

4.1.2 Cholesterol

The input variable "Cholesterol" represents the serum cholesterol levels of a patient. Serum cholesterol is composed by LDL, HDL and triglyceride lipids [10], and is therefore more accurate than using measurements from a single lipid. The domain of discourse consists of 4 fuzzy sets, "low", "normal", "elevated" and "very high". The membership functions for the sets "low" and "very high" are trapezoidal, whereas "normal" and "elevated" are triangular. The range of the domain is 100-400.

4.1.3 Old Peak

The input variable "OldPeak" represents the patients ST depression induced by exercise relative to rest [?]. The input domain contains 3 fuzzy sets, "good", "bad" and "very bad". The membership functions for the sets "good" and "very bad" are trapezoidal, and "bad" is triangular. The input domain range is 0-7.

4.1.4 Thalach (Maximum Heart Rate)

The last input variable is "Maximum heart rate" also known as "Thalach" in the dataset. This input domain consists of 3 fuzzy sets, "low", "normal" and "high". The sets "low" and "high" use trapezoidal membership functions, and "normal" uses triangular. The input domain range is 50-200.

4.2 Output

The FIS has a single output variable called "Risk". This domain contains 5 fuzzy sets, "very low", "low", "medium", "high" and "very high". **The fuzzy sets all use the triangular membership function!** The range is from 0-1. 1 being the highest risk a patient will be at of having CHD and should seek immediate medical assistance. The output was designed in a way that matches the target attribute in the dataset, such that patients with attributes matching those of positively diagnosed patients, then the possibility of being in increasingly higher risk is covered by the various sets.

The FIS output is attached in Appendix G.

4.3 Rule-base

To build an effective rule-base, one must ensure the rules are able to describe the domain well enough to predict varying degrees of risk with good accuracy. The first consideration to make to achieve a suitable performance is to decide on a suitable number of rules that accommodate most combinations of the antecedents along with the resultant consequent. The performance of the FIS rule-base should then be checked by using findings from medical literature and the initial dataset analysis findings to decide whether the output results are inline with reality. This will be discussed in depth in the testing section.

Initially, the construction of the rule-base was guided mainly by common sense and glances at the data analysis findings. Primarily, this was to quickly obtain quantitative test results on the initial performance to base as the starting point for the tuning phase which requires more time and focus in order to achieve the best possible end result.

After the testing and tuning phases which are described in detail in the following sections, the final rule-base consists of a total of **144** rules. The linguistic rules can be seen in appendix H. The number of rules was tested to ensure it covered all possible combinations of antecedents, therefore, making the FIS rule-base complete. Complete rule-bases are usually a problem in time critical systems because they significantly hinder performance, but the difference of milliseconds/seconds isn't really a hindrance for a medical risk assessment tool such as this [13].

4.4 Operators and Defuzzification

The aggregation operator used was maximum. The conjunction (AND) operator used was the union also known as minimum. The disjunction (OR) operator was the intersection also known as maximum. The implication operator used was also the minimum. The defuzzification method used was centroid.

The selection of operators and defuzzification technique were guided by literature that states that for the majority of fuzzy inference systems, the parameters chosen are sufficient to achieve desirable results [3]. However, in the testing and tuning phases, there is a possibility that experimentation with different settings will be undertaken.

5 Testing

Testing is the most crucial stage in the FIS development. The goal of testing is to ensure the outputs from the FIS are accurate and in accordance with real medical data so that the system could be trusted by a person or expert using it to assess risk of CHD.

The first step in testing is to obtain testing data. Conveniently, the original dataset with the excluded unnecessary attributes suits the testing needs of the FIS perfectly. This is because it provides 304 samples of data which can be used to evaluate the FIS. In addition, the target attribute/class for each sample is available to be used for comparison with the FIS risk assessment output.

The FIS was tested as soon as the initial rule-base was constructed. Using Matlab and the fuzzy logic toolbox, the **evalfis** function was used to evaluate the performance of the FIS on the test data samples. The process returns a single column matrix of the evaluated risks for each sample. This column was then merged with the testing dataset in a spreadsheet to aid the analysis of the FIS performance. The merged table can be seen in Appendix I. Some tables are cropped because they contain too many test samples.

This data was then sorted so that the predicted class was in ascending order (0 to 4), followed by the calculation of the mean, min and max values for the evaluated risk output for each class. The mean was used in order to obtain a single value that represents all the evaluated samples for each individual class, thus, making it easier to assess the overall FIS performance for that class of heart disease (seriousness). Min and Max values are also useful to analyse the range of the risk assessments for each class. Appendix J shows the performance results of the initial FIS. As expected, the results were very mixed and featured a large number of incorrect assessments. This was clearly seen by observing the mean and median values for each class. For the samples containing diagnosed patients (1-4), the central tendencies of the evaluated risks were in the ranges 0.59-0.65 and showed an increasing trend with respect to the severity of the heart disease. This range of values shows that the FIS mainly predicted patients in these classes to be members of the moderate and high risk fuzzy sets. This indicated that the system was (for the most part) evaluating a fair amount of diagnosed patients correctly. However, looking at the min values for each of these classes, patients with attributes indicating heart disease, were assessed with risks as low as 0.31. The same problem is present in the samples of patients with no signs of heart disease, albeit worse performance is exhibited. The mean risk was 0.53 which indicates that most patients with no signs of heart disease were considered to be in moderate risk. The mean was skewed due to a significant amount of patients being evaluated with risks as high as 0.85, when in reality the correct evaluation should be as close to 0 as possible. The vice-versa should apply for patients with severe signs of heart disease.

The test shed light on several problems with the initial system. One of them being the rule-base which may be favouring certain variables more than others and biasing the output of the system. The other could be that the fuzzy set membership functions and their characteristics aren't totally precise. The latter seems less probable to be the major setback of the FIS because the characteristics were chosen based on real data and looking at previous research approaches [5, 6]. The goal was then to eliminate the errors in the rule-base, so that the means for the evaluated risks of patients who don't have heart disease are minimized closer to 0, and those with signs of heart disease are maximized closer to 1 depending on the severity of heart disease associated by the input attributes.

6 Tuning

Now that the testing procedure has been set-up, incremental tuning of the FIS was the last step to ensure the best possible performance before the final system evaluation.

The rule-base was the component of the FIS to undergo most changes. The initial test showed problems such as missing rules, and biased consequents. The missing rules were identified because certain combinations of inputs would cause the system to output erroneous values due to no rules firing. After careful manual inspection, the missing rules that were needed to make the system complete were added to the rule-base.

An even bigger problem which aids the incorrect evaluations was caused by the current definition of the consequents to rules i.e the risk. Careful consideration of the results from the initial analysis of the medical data helped shed light on the problems that lead to bad consequent decisions. Namely, by looking at the scores from the feature selection in Appendix B, one can see that the variables *Thalach (Max heart rate)* and *Oldpeak* were the most significant in determining the presence of heart disease. Whereas, *Age* was the least significant, followed by *Cholesterol*. However, in the initial design of the rule-base, age was a predominant factor in determining the risk and as such, members of the *old* and *very old* fuzzy sets were usually considered to be at higher risks. But looking at the box plots and histograms in Appendix D and C, the central tendencies of age are practically equal for all classes of heart disease and no heart disease. The same can be said for cholesterol. This lead to the first modification of the rule-base which was to disregard *age* and *cholesterol* as priority variables, and increase risk to *high* and *very high* only when *oldpeak* and *thalach* values were considered dangerous. The vice-versa is applied for determining lower risks.

An alternative approach to the selection of membership functions was also incorporated into the system tuning to check if there was any improvement in performance. A FIS using Gaussian membership functions was developed, and the results can be seen in Appendix K under system name . The performance was worse using these types of membership functions because the membership onset and offset with these functions uses two much input domain space. The meaning of this is that certain classes have distinct range values, where a member should be fully in the set and not partially for most of the set. Trapezoidal MF's offer this potential unlike Gaussians, especially as the "shoulders/ends" of the input domains as seen in Appendix F.

The various tweaks to the rule-base were split into various iterations and the performance of each iteration is displayed in the table in Appendix K. The performances will be compared and analysed further in the next section.

7 Results

The tuning results table in Appendix K shows 4 tuning iterations. Each with a small but significant improvement. However, picking the best performing iteration was still complicated because there was a small trade-off between a significant improvement in correctly assessing patients with no signs of heart disease, and a slight decrease in performance when assessing patients diagnosed with heart disease. The trade-off is clear by comparing the initial results in Appendix J with those in K. At each iteration there was a significant decrease in the mean/median risk assessment for all classes due to the trade-off. The acceptance of this trade-off lead to decision that the best iteration and therefore, the final FIS, was the 4th iteration. A better way to visually compare the initial FIS rule-base performance with the tuned FIS was to create box plots based on the results for of the assessments of each system of the patients in each class of heart disease. The plots can be seen in Appendix L. At a quick glance, there doesn't seem to be much difference. However, by looking at the box plots for patients with no heart disease (class 0), it is clear that the mean risk is now below 5 which means most patients are being correctly assessed as "moderate" risk is the highest fuzzy set they can be a member of, as opposed to moderate-high for the initial FIS. Patients with the least severe form of heart disease (class 1), now have a stronger similarity in risk assessments compared to the inital FIS. This was due to the trade-off mentioned above, and due to uncertainty in the data which is discussed in depth in the next section. The remaining patients in more severe classes of heart disease (2-4) all suffered from a slight drop in mean risk but not significant enough for the system to be considered worse performing as most patients are still assessed as being in "high-veryhigh" risk.

A look at the surface viewer generated by Matlab offered a new perspective into the performance of the system. Appendix O shows 4 different perspectives of input variables and how the risk is evaluated by the FIS. The first surface (age vs cholesterol) shows a steady increase in risk at the cholesterol and age increases. Also, when the cholesterol levels are considered "normal" i.e below 200, the risk drops massively as expected. This is also backed by [10] which confirms that the normal levels of serum cholesterol should be below 200. The second surface (age vs oldpeak) shows a similar trend with surface risk steadily increasing as age and oldpeak increase. The third surface (age vs thalach/max heart rate) shows that risk is highest when age is above 60 and heart rate is low. In contrast, risk is very low when age is young and max heart rate is considered high. This is because younger people, particularly babies and children have much higher heart rates than adults [9]. Finally, the last surface (Oldpeak vs Thalach), as expected, shows an increased risk as the max heart rate drops and old peak increase. Strangely, when Oldpeak is at its highest (7) there is a sudden drop of risk to 0.5. This could be caused by a incorrect rule.

Overall, the results are mostly in-line with what was expected at the start of the project as we based the system on real data. However, even after an extensive testing and tuning procedure, the system still has many flaws which will be discussed in greater depth in the next section.

An overview of the final FIS when imported into Matlab can be seen in Appendix M. The inputs, output, operators, defuzzification and number of rules can be observed in this window. Appendix N shows a snapshot of the QT FuzzyLite workflow with the final FIS loaded. And finally, Appendix P shows the Matlab code (.fis file) for the final FIS as exported from QT FuzzyLite.

8 Discussion

This section aims to discuss the various pitfalls encountered throughout the design and development of the FIS, and ultimately describe what could be done to achieve better overall performance in the evaluation of the risk of CHD.

Firstly, the rule-base tuning procedure turned out to be very difficult to get right for a number of reasons. Firstly, the lack of expert knowledge in CHD was an obstacle in making the right decision of the correct risks (low, moderate, high...) for the various set of input attributes and rules. This knowledge barrier could have been partially resolved if there was time to perform an extensive review of the literature associated with CHD. Secondly, even with 4 input variables (out of the original 14), the total number of rules (144) required for a complete FIS made the construction of the rule-base a tedious procedure. This procedure was made worse due to the lack of expert knowledge mentioned above. The sheer number of rules also made the manual optimization of the rule-base complicated, meaning that even by following a strict quantitative testing and tuning methodology, the impact of the changes to the rule-base on the overall system performance were minimal after 4 iterations. Therefore, suggesting that many more labour intensive iterations would be necessary to observe a real impact. The tuning of the rule-base at each iteration could have been sped up by applying some kind of optimization algorithm that only selects the rules that are most important to the system, and discarding irrelevant rules. A final incomplete FIS would be left after this process, that performs just as well or better than its complete counterpart, and with much less rules to manually optimize. Thirdly, the data in the dataset used in the project featured a high degree of uncertainty, which in turn affected the design of the FIS. The uncertainty was clearly observable in the testing and tuning phases, where the attributes of patients with no heart disease (0), and those with the least severe form of heart disease (1) were quite similar, and in cases the same. This was also the case with patients with more severe types of heart disease (2-4), but less so. This uncertainty affected the risk assessments of the system gravely. One of the main reasons (along with the others mentioned above) behind these incorrect assessments was due to the choice of membership function parameters possibly being affected by this uncertainty, thus requiring a more sophisticated approach to avoid these problems.

As mentioned above, there are many problems caused by the current approach to this problem that largely affect the performance of the system. However, most of these problems could have been resolved by simply adopting a different inference system architecture. Given the problems faced, only two other suitable approaches come to mind. The Adaptive-Neuro Fuzzy Inference System (ANFIS) and Type-2 fuzzy logic.

The lack of expert knowledge that lead to the tedious manual construction of the FIS rule-base and lengthy tuning stage was definitely a large shortcoming of the project. In hindsight, the best method to overcome this would have been to adopt the ANFIS approach.

ANFIS is a type of supervised learning and FIS hybrid architecture. As such, all supervised learning systems require a dataset of input attributes and a target attribute/prediction. The CHD dataset used in the study conforms to this criteria, and offers plenty of samples in each respective class. This means that the dataset can be split into training and testing datasets while still maintain class balance and enough samples for training. The ANFIS could then be trained to learn the mappings between the input attributes and target variable and using optimization algorithms based on gradient descent, determine the suitable membership function characteristics and rules for the system automatically.

There is reason to believe the ANFIS approach would have achieved much better results as more attributes could have been used to train the system, and be much more time efficient due to the large decrease in manual labour due to the optimization algorithms eliminating the bulk of the work and simultaneously the need to reference medical knowledge.

The other large shortcoming of the approach was the systems inability to handle large degrees of uncertainty. This deficiency essentially makes the system very prone to incorrect assessments, which in turn means its unsuitable to give reliable medical advice of any kind and as it stands, should be used with caution. Uncertainty prevails in many parts of the system, and mostly stems from the medical data which guided the design of the membership functions and rules. This uncertainty emerged through the similarities in the attributes of patients in different classes of heart disease, which suggests that some patients with attributes indicating no signs of heart disease, could actually have severe heart disease, and vice-versa. The characteristic values of the membership functions are very prone to this uncertainty as they are essentially fixed values and are not themselves "fuzzy". This relates to the experience during the design phase where there was a large indecision of where the boundaries of the fuzzy sets should be. There were also difficulties in determining the consequents to rules, this is typical of problems with large uncertainty, and even experts may have varying opinions on what the correct consequents to rules for this system should be.

Type-2 fuzzy sets, proposed by Prof. Zadeh [11], have been proven to achieve state of art performance when compared to type-1 fuzzy sets in problem domains that feature large degrees of uncertainty[12, 13]. The key differences between the two techniques is that type-2 sets comprise membership function characteristics that are themselves "fuzzy" or vague. This allows them to model noisy data and therefore uncertainty much better than traditional type-1 sets. The performance benefits would be very significant if applied to the problem of CHD risk assessment. This is shown to be the case in [13] where various examples of approaches using type-2 fuzzy sets to solve substantially more complex problems in the medicine sector were reviewed. The problems include the diagnosis of diseases, modelling compatibility of symptoms to disease, assessing health of new born babies and many more. Due to nature of medical risk assessment tools, the increased computational complexity of general type-2 sets would not be a hindrance as much as it would be for time critical systems, in which interval type-2 sets are preferred. Therefore, Type-2 sets offer the best possibility of increasing the performance of the system while maintaining the interpretability of the Mandami linguistic antecedents and consequents which is important to be able to understand the conclusions that the system arrives at.

The Type-2 approach could also be augmented by incorporating some form of intelligent optimization algorithm such as evolutionary algorithms. Research in [14] demonstrates a interval Type-2 FIS that uses Hierarchical Genetic Algorithms to fine tune the type-2 membership functions and optimize the rule-base. A similar approach could be developed for the problem of CHD risk assessment to with the ultimate goal of optimizing the mean risk assessments for each class of CHD.

9 Conclusion

The initial aim of the project was to create a FIS that could correctly assess the risk of CHD in patients. It is important to note, that even from the beginning of the project, there was recognition that the lack of domain knowledge would pose some difficulties in the success of the project. However, choosing to use real medical data collected by experts of patients diagnosed for the presence of CHD, somewhat masked the need for expert knowledge as the design of the system was guided entirely by the data analysis findings at the start of the project. The labelled classes of CHD severity in the data were used to aid many aspects of the project such as differentiating between the attributes of patients diagnosed with CHD and those without CHD. The testing and tuning of the system also depended on these class labels to isolate the risk evaluation results for every patient in each class and produce a mean/median value of risk for each class, as opposed to an overall risk which wouldn't be a good metric to arrive at a suitable final performance conclusion.

The final system produced fairly accurate assessments of CHD risk most of the time as discussed in the results section. The limitations of the system were analysed and alternative approaches were considered in the discussion section in which a clear point was made that if the project was to be undertaken again the most likely approaches to achieve better results would be either an ANFIS based approach, or an Optimized Type-2 FIS. These approaches would eliminate problems such as the lack of expert knowledge, and the tedious construction and manual optimization of rule-base, membership functions and their characteristics, operators and defuzzification techniques. Type-2 FIS would also deal with the problem of uncertainty in the data.

Overall, the system performance indicates that it would be adequate to be used by people with access to their medical records, but only if used with caution, as the risk assessments could be very inaccurate in some cases. With the extra work suggested in the discussion section, it could possibly be tested by experts to further assess its performance in the real world. However, the adoption of expert systems in the medical sector has always been plagued by the ethical and moral consequences related to computer made decisions on suitable patient care. This is because all it would take would be 1 false positive/negative assessment to cause possible confusion or harm to a human being, which is why the decisions of these systems need to be carefully reviewed by an expert for decades to come.

References

[1] Juan Rada-Vilela, Ph.D. *Fuzzy Lite Libraries for Fuzzy Logic Control.* https://www.fuzzylite.com/

[2] Robert Detrano, M.D., Ph.D. *UCI Heart Disease Data Set* . V.A. Medical Center, Long Beach and Cleveland Clinic Foundation

[3] Jaume Polo (2001) *Development and Testing of a Number of Matlab Based Fuzzy System Applications.*

[4] Ziasabounchi, Negar; Askerzade, Iman(2014) *ANFIS Based Classification Model for Heart Disease Prediction .*

[5] Adeli, Ali; Neshat, Mehdi(2010) *A Fuzzy Expert System for Heart Disease Diagnosis.* Proceedings of the International MultiConference of Engineeers and Computer Scientists

[6] Kumar, Sanjeev; Kaur, Gursimranjeet (2013) *Detection of Heart Diseases using Fuzzy Logic.* International Journal of Engineering Trends and Technology

[7] Chen, a H; Huang, S Y; Hong, P S; Cheng, C H; Lin, E J (2011) *HDPS: Heart disease prediction system.* Computing in Cardiology (CinC)

[8] *Can your cholesterol be too low?* https://www.webmd.com/men/features/can-your-cholesterol-be-too-low-feature

[9] *Monitoring heart rate normality* https://www.webmd.com/heart-disease/heart-failure/watching-rate-monitor

[10] *What Is Serum Cholesterol and Why Is It Important?* https://www.healthline.com/health/serum-cholesterol

[11] L. A. Zadeh (1975). *The Concept of a Linguistic Variable and Its Application to Approximate Reasoning1* Information Sciences, vol. 8, pp. 199249.

[12] Mendel, J.M., and John, R.I.B. (2002). *Type-2 fuzzy sets made simple.* Fuzzy Systems, IEEE Transactions On 10, 117127.

[13] Robert John and Simon Coupland (). *Extensions to Type-1 Fuzzy Logic: Type-2 Fuzzy Logic and Uncertainty*

[14] Castillo, O., and Melin, P. (2012). *A review on the design and optimization of interval type-2 fuzzy controllers.* Applied Soft Computing Journal 12, 12671278.

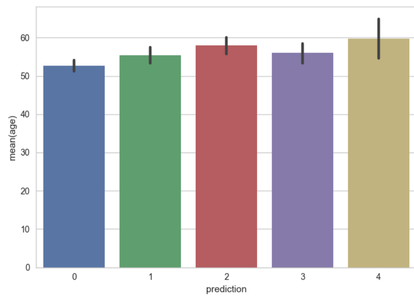
A Dataset Attributes

Attribute	Description
age	age in years
sex	sex (1 = male; 0 = female)
cp	chest pain type
trestbps	resting blood pressure (in mm Hg on admission to the hospital)
chol	serum cholestoral in mg/dl
fbs	(fasting blood sugar ¿ 120 mg/dl) (1 = true; 0 = false)
restecg	resting electrocardiographic results
thalach	maximum heart rate achieved
exang	exercise induced angina (1 = yes; 0 = no)
oldpeak	ST depression induced by exercise relative to rest
slope	the slope of the peak exercise ST segment
ca	number of major vessels (0-3) colored by flourosopy
thal	3 = normal; 6 = fixed defect; 7 = reversable defect
prediction	diagnosis of heart disease (angiographic disease status)

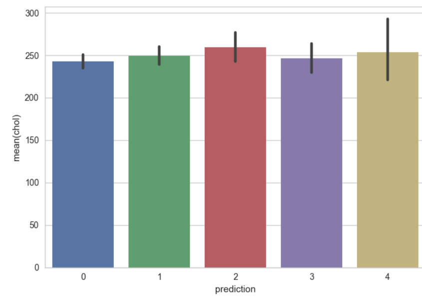
B Selected Attributes

Attribute	Description	Score
age	age in years	27.97523271
chol	serum cholestoral in mg/dl	38.16745159
thalach	maximum heart rate achieved	219.88355659
oldpeak	ST depression induced by exercise relative to rest	98.50552313

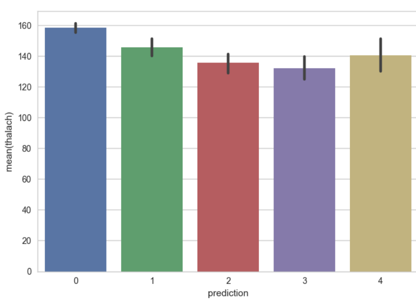
C Histograms



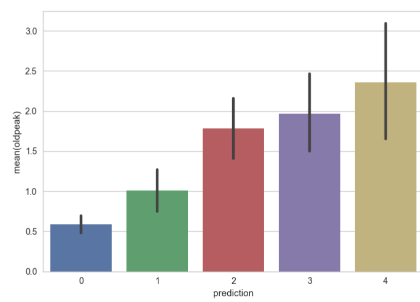
(a) Age



(b) Cholesterol



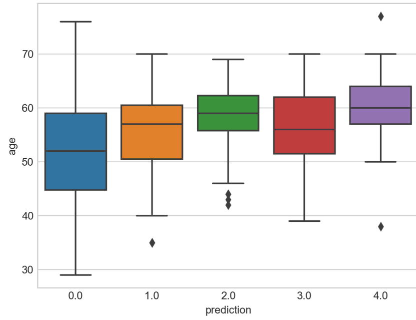
(c) Thalach



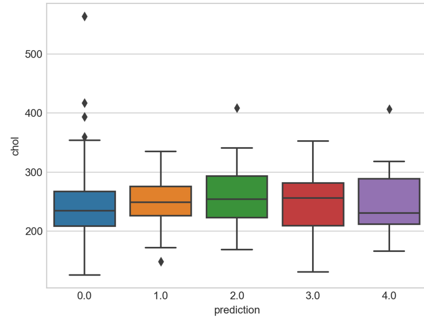
(d) Oldpeak

Figure 1: caption

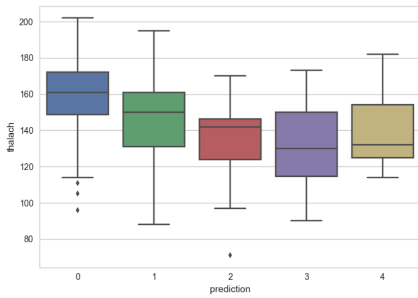
D Box Plots



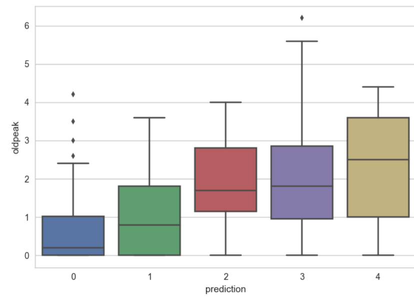
(a) Age



(b) Cholesterol



(c) Thalach



(d) Oldpeak

E Ranges and Means

Attribute	Class	First Quartile	Third Quartile	Mean	Min	Max
Oldpeak	0	0	1.025	0.586585	0	6.2
	1	0	1.8	1.005455	-	-
	2	1.15	2.8	1.780556	-	-
	3	0.95	2.85	1.962857	-	-
	4	1	3.6	2.361538	-	-
Age	0	44.75	59	52.585366	29	76
	1	50.5	60.5	55.381818	-	-
	2	55.75	62.25	58.027778	-	-
	3	51.5	62	56	-	-
	4	57	64	59.692308	-	-
Chol	0	208.75	267.25	242.640244	126	409
	1	226	275.5	249.109091	-	-
	2	223.25	293.25	259.277778	-	-
	3	209	281.5	246.457143	-	-
	4	212	289	253.384615	-	-
Thalach	0	148.75	172	158.378049	71	202
	1	131	161	145.927273	-	-
	2	123.75	146.25	135.583333	-	-
	3	114.5	150	132.057143	-	-
	4	125	154	140.615385	-	-

F FIS Inputs

Figure 2: Age Input

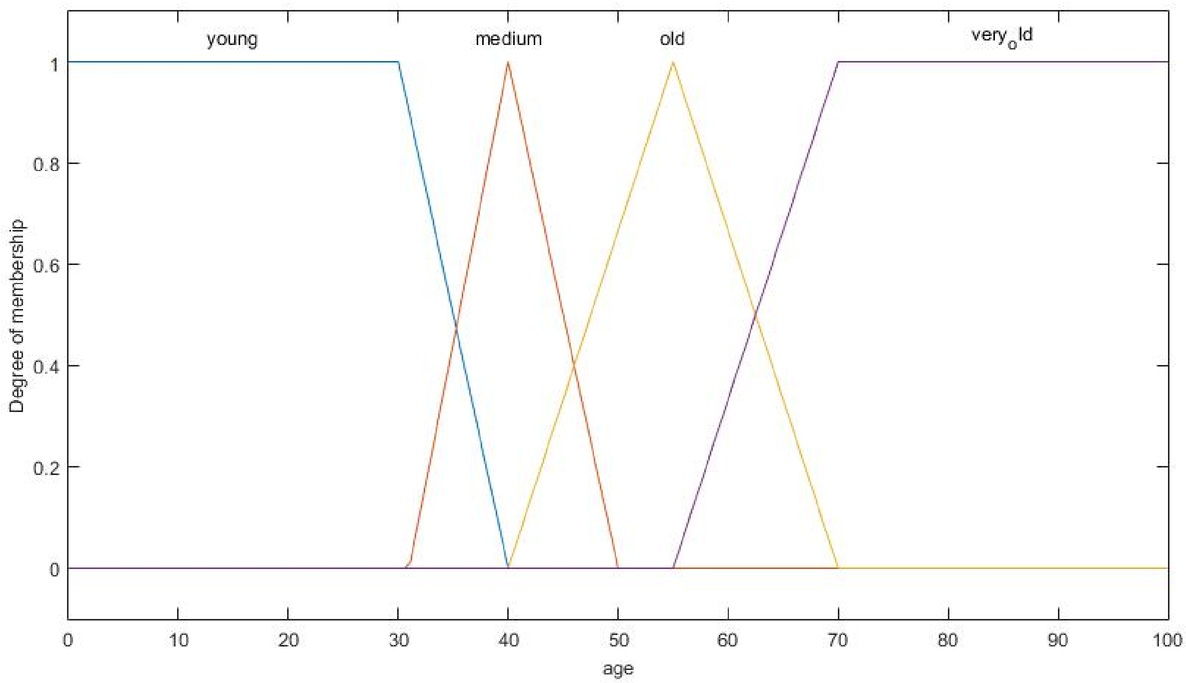


Figure 3: Cholesterol Input

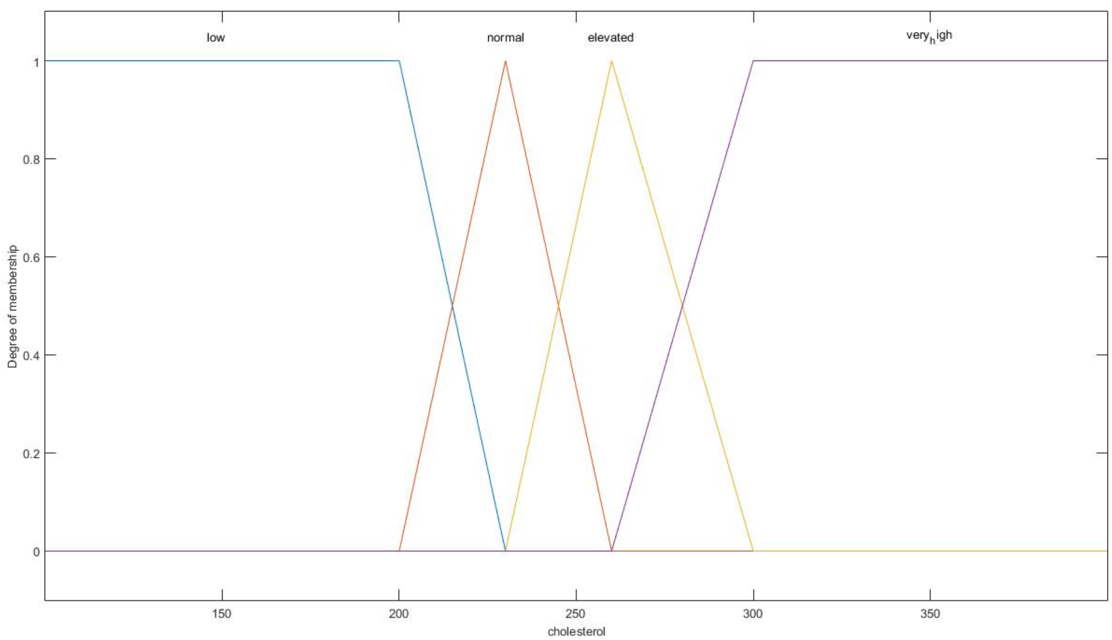


Figure 4: Oldpeak Input

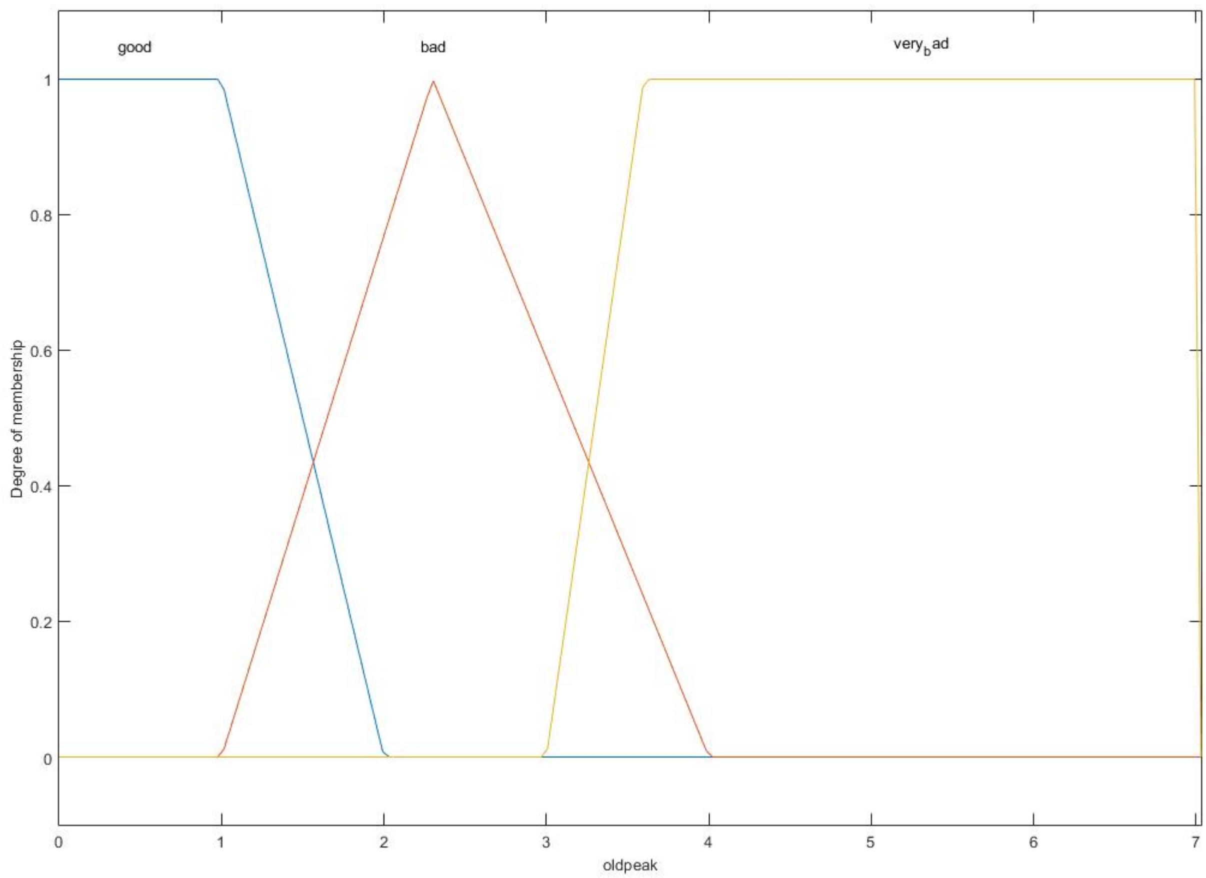
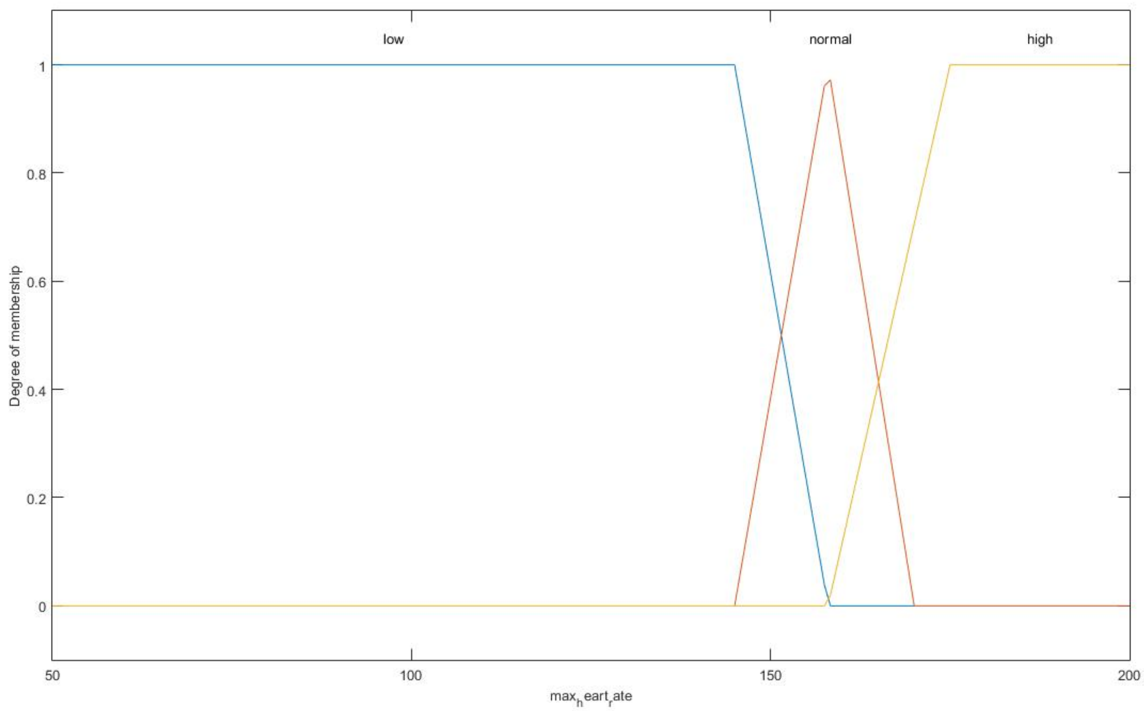
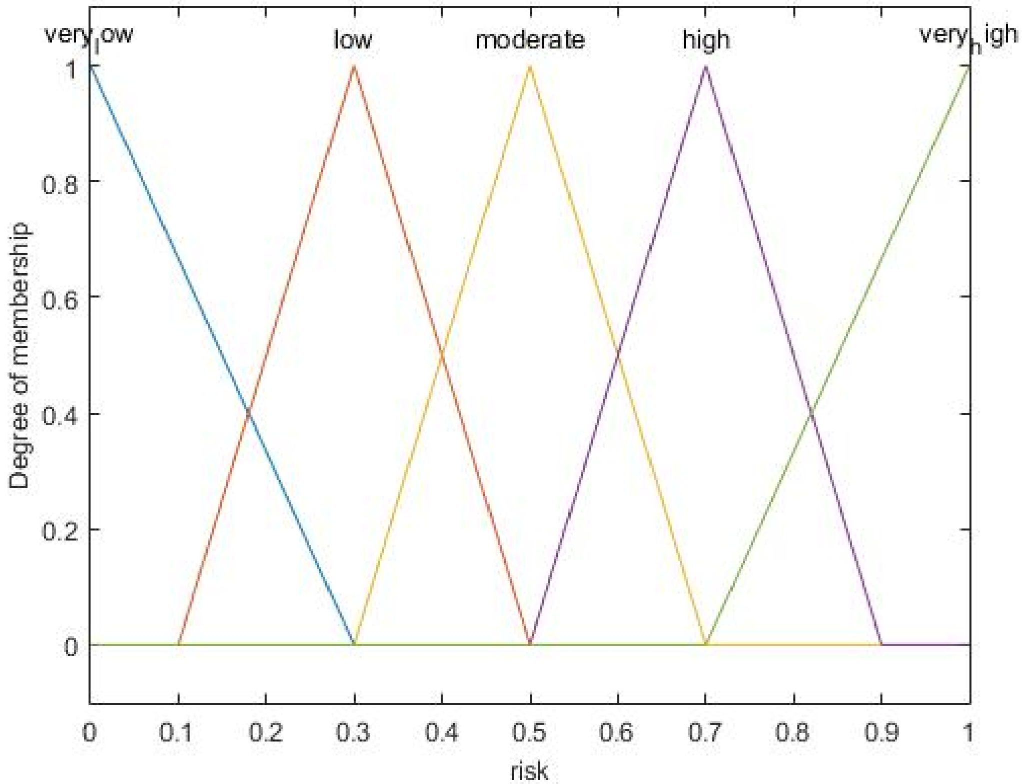


Figure 5: Thalach Input



G FIS Output



very_low
low
moderate
high
very_high
1
0.8
0.6
0.4
0.2
0
Degree of membership
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
risk

H Rule-base

```
if age is young and max_heart_rate is low and cholesterol is low and oldpeak is good then risk is very_low
if age is young and max_heart_rate is low and cholesterol is low and oldpeak is bad then risk is moderate
if age is young and max_heart_rate is low and cholesterol is low and oldpeak is very_bad then risk is high
if age is young and max_heart_rate is low and cholesterol is normal and oldpeak is good then risk is very_low
if age is young and max_heart_rate is low and cholesterol is normal and oldpeak is bad then risk is low
if age is young and max_heart_rate is low and cholesterol is normal and oldpeak is very_bad then risk is moderate
if age is young and max_heart_rate is low and cholesterol is elevated and oldpeak is good then risk is low
if age is young and max_heart_rate is low and cholesterol is elevated and oldpeak is bad then risk is moderate
if age is young and max_heart_rate is low and cholesterol is elevated and oldpeak is very_bad then risk is high
if age is young and max_heart_rate is low and cholesterol is very_high and oldpeak is good then risk is moderate
if age is young and max_heart_rate is low and cholesterol is very_high and oldpeak is bad then risk is high
if age is young and max_heart_rate is low and cholesterol is very_high and oldpeak is very_bad then risk is very_high
if age is young and max_heart_rate is normal and cholesterol is low and oldpeak is good then risk is very_low
if age is young and max_heart_rate is normal and cholesterol is low and oldpeak is bad then risk is low
if age is young and max_heart_rate is normal and cholesterol is low and oldpeak is very_bad then risk is moderate
if age is young and max_heart_rate is normal and cholesterol is normal and oldpeak is good then risk is very_low
if age is young and max_heart_rate is normal and cholesterol is normal and oldpeak is bad then risk is very_low
if age is young and max_heart_rate is normal and cholesterol is normal and oldpeak is very_bad then risk is low
if age is young and max_heart_rate is normal and cholesterol is elevated and oldpeak is good then risk is low
if age is young and max_heart_rate is normal and cholesterol is elevated and oldpeak is bad then risk is moderate
if age is young and max_heart_rate is normal and cholesterol is elevated and oldpeak is very_bad then risk is moderate
if age is young and max_heart_rate is normal and cholesterol is very_high and oldpeak is good then risk is low
if age is young and max_heart_rate is normal and cholesterol is very_high and oldpeak is bad then risk is moderate
if age is young and max_heart_rate is normal and cholesterol is very_high and oldpeak is very_bad then risk is moderat
if age is young and max_heart_rate is high and cholesterol is low and oldpeak is good then risk is very_low
if age is young and max_heart_rate is high and cholesterol is low and oldpeak is bad then risk is low
if age is young and max_heart_rate is high and cholesterol is low and oldpeak is very_bad then risk is moderate
if age is young and max_heart_rate is high and cholesterol is normal and oldpeak is good then risk is very_low
if age is young and max_heart_rate is high and cholesterol is normal and oldpeak is bad then risk is low
if age is young and max_heart_rate is high and cholesterol is normal and oldpeak is very_bad then risk is low
if age is young and max_heart_rate is high and cholesterol is elevated and oldpeak is good then risk is low
if age is young and max_heart_rate is high and cholesterol is elevated and oldpeak is very_bad then risk is moderate
if age is young and max_heart_rate is high and cholesterol is elevated and oldpeak is bad then risk is moderate
if age is young and max_heart_rate is high and cholesterol is very_high and oldpeak is good then risk is low
if age is young and max_heart_rate is high and cholesterol is very_high and oldpeak is bad then risk is moderate
if age is young and max_heart_rate is high and cholesterol is very_high and oldpeak is very_bad then risk is moderate
if age is medium and max_heart_rate is low and cholesterol is low and oldpeak is good then risk is low
if age is medium and max_heart_rate is low and cholesterol is low and oldpeak is bad then risk is moderate
if age is medium and max_heart_rate is low and cholesterol is low and oldpeak is very_bad then risk is high
if age is medium and max_heart_rate is low and cholesterol is normal and oldpeak is good then risk is low
if age is medium and max_heart_rate is low and cholesterol is normal and oldpeak is bad then risk is moderate
if age is medium and max_heart_rate is low and cholesterol is normal and oldpeak is very_bad then risk is high
if age is medium and max_heart_rate is low and cholesterol is elevated and oldpeak is good then risk is moderate
if age is medium and max_heart_rate is low and cholesterol is elevated and oldpeak is bad then risk is high
if age is medium and max_heart_rate is low and cholesterol is elevated and oldpeak is very_bad then risk is very_high
if age is medium and max_heart_rate is low and cholesterol is very_high and oldpeak is good then risk is moderate
if age is medium and max_heart_rate is low and cholesterol is very_high and oldpeak is bad then risk is very_high
if age is medium and max_heart_rate is low and cholesterol is very_high and oldpeak is very_bad then risk is very_high
if age is medium and max_heart_rate is normal and cholesterol is low and oldpeak is good then risk is very_low
if age is medium and max_heart_rate is normal and cholesterol is low and oldpeak is bad then risk is low
if age is medium and max_heart_rate is normal and cholesterol is low and oldpeak is very_bad then risk is moderate
if age is medium and max_heart_rate is normal and cholesterol is normal and oldpeak is good  then risk is very_low
if age is medium and max_heart_rate is normal and cholesterol is normal and oldpeak is bad then risk is low
if age is medium and max_heart_rate is normal and cholesterol is normal and oldpeak is very_bad then risk is moderate
if age is medium and max_heart_rate is normal and cholesterol is elevated and oldpeak is good then risk is low
if age is medium and max_heart_rate is normal and cholesterol is elevated and oldpeak is bad then risk is moderate
if age is medium and max_heart_rate is normal and cholesterol is elevated and oldpeak is very_bad then risk is high
if age is medium and max_heart_rate is normal and cholesterol is very_high and oldpeak is good then risk is low
if age is medium and max_heart_rate is normal and cholesterol is very_high and oldpeak is bad then risk is high
if age is medium and max_heart_rate is normal and cholesterol is very_high and oldpeak is very_bad then risk is high
if age is medium and max_heart_rate is high and cholesterol is low and oldpeak is good then risk is low
if age is medium and max_heart_rate is high and cholesterol is low and oldpeak is bad then risk is low
if age is medium and max_heart_rate is high and cholesterol is low and oldpeak is very_bad then risk is moderate
if age is medium and max_heart_rate is high and cholesterol is normal and oldpeak is good then risk is low
if age is medium and max_heart_rate is high and cholesterol is normal and oldpeak is bad then risk is moderate
if age is medium and max_heart_rate is high and cholesterol is normal and oldpeak is very_bad then risk is moderate
if age is medium and max_heart_rate is high and cholesterol is elevated and oldpeak is good then risk is low
if age is medium and max_heart_rate is high and cholesterol is elevated and oldpeak is bad then risk is moderate
if age is medium and max_heart_rate is high and cholesterol is elevated and oldpeak is very_bad then risk is moderate
if age is medium and max_heart_rate is high and cholesterol is very_high and oldpeak is good  then risk is moderate
if age is medium and max_heart_rate is high and cholesterol is very_high and oldpeak is bad then risk is high
if age is medium and max_heart_rate is high and cholesterol is very_high and oldpeak is very_bad then risk is very_high
```

```
if age is old and max_heart_rate is low and cholesterol is low and oldpeak is good then risk is moderate
if age is old and max_heart_rate is low and cholesterol is low and oldpeak is bad then risk is high
if age is old and max_heart_rate is low and cholesterol is low and oldpeak is very_bad then risk is very_high
if age is old and max_heart_rate is low and cholesterol is normal and oldpeak is good then risk is low
if age is old and max_heart_rate is low and cholesterol is normal and oldpeak is bad then risk is moderate
if age is old and max_heart_rate is low and cholesterol is normal and oldpeak is very_bad then risk is high
if age is old and max_heart_rate is low and cholesterol is elevated and oldpeak is good then risk is moderate
if age is old and max_heart_rate is low and cholesterol is elevated and oldpeak is bad then risk is high
if age is old and max_heart_rate is low and cholesterol is elevated and oldpeak is very_bad then risk is high
if age is old and max_heart_rate is low and cholesterol is very_high and oldpeak is good then risk is moderate
if age is old and max_heart_rate is low and cholesterol is very_high and oldpeak is bad then risk is high
if age is old and max_heart_rate is low and cholesterol is very_high and oldpeak is very_bad then risk is very_high
if age is old and max_heart_rate is normal and cholesterol is low and oldpeak is good then risk is moderate
if age is old and max_heart_rate is normal and cholesterol is low and oldpeak is bad then risk is moderate
if age is old and max_heart_rate is normal and cholesterol is low and oldpeak is very_bad then risk is high
if age is old and max_heart_rate is normal and cholesterol is normal and oldpeak is good then risk is low
if age is old and max_heart_rate is normal and cholesterol is normal and oldpeak is bad then risk is low
if age is old and max_heart_rate is normal and cholesterol is normal and oldpeak is very_bad then risk is moderate
if age is old and max_heart_rate is normal and cholesterol is elevated and oldpeak is good then risk is moderate
if age is old and max_heart_rate is normal and cholesterol is elevated and oldpeak is bad then risk is high
if age is old and max_heart_rate is normal and cholesterol is elevated and oldpeak is very_bad then risk is high
if age is old and max_heart_rate is normal and cholesterol is very_high and oldpeak is good then risk is moderate
if age is old and max_heart_rate is normal and cholesterol is very_high and oldpeak is bad then risk is high
if age is old and max_heart_rate is normal and cholesterol is very_high and oldpeak is very_bad then risk is very_high
if age is old and max_heart_rate is high and cholesterol is low and oldpeak is good then risk is moderate
if age is old and max_heart_rate is high and cholesterol is low and oldpeak is bad then risk is moderate
if age is old and max_heart_rate is high and cholesterol is low and oldpeak is very_bad then risk is high
if age is old and max_heart_rate is high and cholesterol is normal and oldpeak is good then risk is moderate
if age is old and max_heart_rate is high and cholesterol is normal and oldpeak is bad then risk is moderate
if age is old and max_heart_rate is high and cholesterol is normal and oldpeak is very_bad then risk is moderate
if age is old and max_heart_rate is high and cholesterol is elevated and oldpeak is good then risk is moderate
if age is old and max_heart_rate is high and cholesterol is elevated and oldpeak is bad then risk is high
if age is old and max_heart_rate is high and cholesterol is elevated and oldpeak is very_bad then risk is high
if age is old and max_heart_rate is high and cholesterol is very_high and oldpeak is good then risk is high
if age is old and max_heart_rate is high and cholesterol is very_high and oldpeak is bad then risk is high
if age is old and max_heart_rate is high and cholesterol is very_high and oldpeak is very_bad then risk is very_high
```

```
if age is very_old and max_heart_rate is low and cholesterol is low and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is low and cholesterol is low and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is low and cholesterol is low and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is low and cholesterol is normal and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is low and cholesterol is normal and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is low and cholesterol is normal and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is low and cholesterol is elevated and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is low and cholesterol is elevated and oldpeak is bad then risk is very_high
if age is very_old and max_heart_rate is low and cholesterol is elevated and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is low and cholesterol is very_high and oldpeak is good then risk is high
if age is very_old and max_heart_rate is low and cholesterol is very_high and oldpeak is bad then risk is very_high
if age is very_old and max_heart_rate is low and cholesterol is very_high and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is normal and cholesterol is low and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is normal and cholesterol is low and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is normal and cholesterol is low and oldpeak is very_bad then risk is high
if age is very_old and max_heart_rate is normal and cholesterol is normal and oldpeak is good then risk is low
if age is very_old and max_heart_rate is normal and cholesterol is normal and oldpeak is bad then risk is moderate
if age is very_old and max_heart_rate is normal and cholesterol is normal and oldpeak is very_bad then risk is high
if age is very_old and max_heart_rate is normal and cholesterol is elevated and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is normal and cholesterol is elevated and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is normal and cholesterol is elevated and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is normal and cholesterol is very_high and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is normal and cholesterol is very_high and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is normal and cholesterol is very_high and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is high and cholesterol is low and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is high and cholesterol is low and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is high and cholesterol is low and oldpeak is very_bad then risk is high
if age is very_old and max_heart_rate is high and cholesterol is normal and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is high and cholesterol is normal and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is high and cholesterol is normal and oldpeak is very_bad then risk is high
if age is very_old and max_heart_rate is high and cholesterol is elevated and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is high and cholesterol is elevated and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is high and cholesterol is elevated and oldpeak is very_bad then risk is very_high
if age is very_old and max_heart_rate is high and cholesterol is very_high and oldpeak is good then risk is moderate
if age is very_old and max_heart_rate is high and cholesterol is very_high and oldpeak is bad then risk is high
if age is very_old and max_heart_rate is high and cholesterol is very_high and oldpeak is very_bad then risk is very_high
```

I Full FIS Performance Results

Age	Thalach	Chol	Oldpeak	Class	Initial FIS	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration (FINAL)
34	174	182	0	0	0.204	0.201	0.197	0.197	0.198
34	192	210	0.7	0	0.204	0.201	0.197	0.197	0.198
39	152	220	0	0	0.222	0.221	0.217	0.217	0.217
35	174	192	0	0	0.225	0.224	0.219	0.219	0.220
41	153	235	0	0	0.257	0.258	0.273	0.273	0.254
37	170	215	0	0	0.259	0.259	0.250	0.250	0.250
43	165	213	0.2	0	0.275	0.275	0.273	0.273	0.288
35	182	183	1.4	0	0.280	0.281	0.219	0.219	0.220
38	173	175	0	0	0.292	0.292	0.272	0.272	0.272
43	161	211	0	0	0.293	0.293	0.284	0.284	0.247
39	179	199	0	0	0.319	0.319	0.286	0.286	0.286
45	152	236	0.2	0	0.326	0.327	0.356	0.344	0.277
41	132	203	0	0	0.364	0.364	0.313	0.313	0.318
41	182	157	0	0	0.364	0.364	0.313	0.313	0.318
41	168	198	0	0	0.366	0.366	0.316	0.316	0.295
44	169	226	0	0	0.371	0.371	0.337	0.337	0.353
42	178	226	0	0	0.376	0.376	0.325	0.325	0.335
42	173	209	0	0	0.378	0.378	0.327	0.327	0.337
42	150	180	0	0	0.379	0.379	0.328	0.328	0.280
44	149	242	0.3	0	0.380	0.380	0.403	0.403	0.338
46	152	243	0	0	0.383	0.383	0.428	0.428	0.328
44	188	219	0	0	0.399	0.399	0.350	0.350	0.366
44	175	141	0.6	0	0.399	0.399	0.350	0.350	0.366
44	170	220	0	0	0.399	0.399	0.350	0.350	0.366
51	154	227	0	0	0.401	0.401	0.352	0.321	0.329
40	178	199	1.4	0	0.404	0.404	0.300	0.300	0.300
53	160	234	0	0	0.409	0.376	0.325	0.325	0.335
45	138	160	0	0	0.412	0.412	0.364	0.364	0.382
41	172	204	1.4	0	0.421	0.421	0.354	0.354	0.339
56	163	221	0	0	0.422	0.422	0.385	0.385	0.372
46	156	197	0	0	0.425	0.425	0.380	0.380	0.334
46	172	204	0	0	0.425	0.425	0.380	0.380	0.400
44	179	233	0.4	0	0.426	0.399	0.350	0.350	0.366
42	194	240	0.8	0	0.435	0.405	0.356	0.356	0.337
54	160	239	1.2	0	0.435	0.435	0.409	0.409	0.414
47	156	257	0	0	0.436	0.436	0.442	0.442	0.400
59	161	234	0.5	0	0.436	0.436	0.405	0.405	0.345
47	143	204	0.1	0	0.440	0.440	0.399	0.399	0.420
44	180	235	0	0	0.440	0.399	0.350	0.350	0.366
54	165	232	1.6	0	0.443	0.443	0.409	0.409	0.423
45	175	234	0.6	0	0.445	0.412	0.364	0.364	0.382
42	178	244	0.8	0	0.449	0.421	0.376	0.376	0.342
48	186	222	0	0	0.457	0.457	0.424	0.424	0.443
41	179	250	0	0	0.460	0.445	0.407	0.407	0.320
50	162	244	1.1	0	0.466	0.438	0.406	0.406	0.418
43	171	247	1.5	0	0.470	0.470	0.455	0.455	0.447
59	164	221	0	0	0.470	0.470	0.458	0.458	0.385
50	159	254	0	0	0.475	0.461	0.431	0.431	0.449
58	165	211	0	0	0.475	0.475	0.463	0.463	0.406
64	155	227	0.6	0	0.479	0.479	0.434	0.431	0.360
52	169	223	0	0	0.481	0.481	0.464	0.464	0.475
46	160	177	1.4	0	0.488	0.488	0.483	0.483	0.331
52	147	233	0.1	0	0.491	0.491	0.484	0.367	0.327
54	163	201	0	0	0.491	0.491	0.483	0.483	0.489

Age	Thalach	Chol	Oldpeak	Class	Initial FIS	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration (FINAL)
35	130	198	1.6	1	0.315	0.315	0.315	0.315	0.336
40	181	223	0	1	0.350	0.350	0.300	0.300	0.300
41	158	172	0	1	0.364	0.364	0.313	0.313	0.156
52	160	230	0	1	0.374	0.374	0.323	0.323	0.331
48	168	229	1	1	0.387	0.387	0.384	0.384	0.407
44	177	197	0	1	0.399	0.399	0.350	0.350	0.366
47	152	243	0	1	0.411	0.411	0.447	0.446	0.347
35	156	282	0	1	0.421	0.421	0.376	0.376	0.348
50	163	233	0.6	1	0.429	0.403	0.354	0.354	0.371
57	164	232	0	1	0.446	0.446	0.421	0.421	0.385
57	150	229	0.4	1	0.466	0.466	0.446	0.390	0.339
43	143	247	0.1	1	0.471	0.471	0.456	0.456	0.411
49	126	149	0.8	1	0.477	0.477	0.456	0.456	0.469
52	161	255	0	1	0.502	0.468	0.441	0.441	0.457
46	144	249	0.8	1	0.504	0.504	0.506	0.506	0.404
59	162	204	0.8	1	0.520	0.520	0.533	0.533	0.463
65	158	248	0.6	1	0.526	0.526	0.525	0.525	0.416
57	174	236	0	1	0.536	0.526	0.549	0.549	0.500
67	150	212	0.8	1	0.547	0.547	0.529	0.529	0.418
45	132	264	1.2	1	0.562	0.562	0.603	0.603	0.572
63	136	197	0	1	0.579	0.579	0.624	0.624	0.500
62	154	244	1.4	1	0.579	0.494	0.492	0.492	0.524
47	118	275	1	1	0.590	0.590	0.638	0.638	0.500
57	123	241	0.2	1	0.604	0.559	0.599	0.464	0.378
56	144	249	1.2	1	0.613	0.591	0.639	0.536	0.475
59	159	288	0.2	1	0.619	0.619	0.657	0.657	0.518
61	138	207	1.9	1	0.627	0.627	0.655	0.609	0.603
67	129	229	2.6	1	0.650	0.650	0.515	0.515	0.652
53	155	203	3.1	1	0.650	0.650	0.620	0.620	0.566
64	158	335	0	1	0.650	0.650	0.700	0.700	0.500
54	109	266	2.2	1	0.650	0.650	0.700	0.700	0.700
58	131	216	2.2	1	0.650	0.650	0.614	0.614	0.595
60	155	185	3	1	0.650	0.650	0.700	0.700	0.576
51	173	299	1.6	1	0.650	0.643	0.693	0.693	0.690
56	105	184	2.1	1	0.650	0.650	0.700	0.700	0.700
54	195	283	0	1	0.650	0.584	0.630	0.630	0.612
58	105	218	2	1	0.650	0.650	0.604	0.604	0.584
59	143	249	0	1	0.665	0.591	0.639	0.536	0.422
46	147	231	3.6	1	0.680	0.575	0.424	0.534	0.571
57	141	261	0.3	1	0.690	0.659	0.703	0.703	0.507
57	88	274	1.2	1	0.694	0.694	0.720	0.720	0.571
57	112	276	0.6	1	0.696	0.696	0.721	0.721	0.539
60	144	253	1.4	1	0.697	0.605	0.655	0.566	0.569
63	169	269	1.8	1	0.703	0.680	0.695	0.695	0.643
60	161	305	0	1	0.705	0.705	0.726	0.726	0.547
58	171	300	0	1	0.708	0.708	0.728	0.728	0.651
59	125	273	0	1	0.725	0.725	0.739	0.739	0.563
60	141	258	2.8	1	0.737	0.650	0.685	0.685	0.729
61	169	330	0	1	0.751	0.751	0.759	0.759	0.616
61	146	307	1	1	0.751	0.751	0.759	0.759	0.584
58	160	284	1.8	1	0.751	0.751	0.752	0.752	0.646
62	99	267	1.8	1	0.764	0.710	0.729	0.729	0.718
65	174	282	1.4	1	0.780	0.723	0.723	0.723	0.579
64	131	309	1.8	1	0.789	0.791	0.795	0.795	0.736
70	109	322	2.4	1	0.860	0.877	0.886	0.886	0.903

Age	Thalach	Chol	Oldpeak	Class	Initial FIS	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration (FINAL)
60	160	230	1.4	2	0.435	0.435	0.404	0.404	0.411
52	156	204	1	2	0.474	0.474	0.451	0.451	0.465
58	156	234	0.1	2	0.485	0.424	0.393	0.393	0.339
56	150	409	1.9	2	0.500	0.500	0.500	0.500	0.500
44	153	290	0	2	0.502	0.502	0.504	0.504	0.382
42	125	315	1.8	2	0.529	0.529	0.554	0.554	0.752
43	136	341	3	2	0.541	0.541	0.573	0.573	0.825
61	161	203	0	2	0.542	0.542	0.567	0.567	0.469
59	162	177	0	2	0.547	0.547	0.582	0.582	0.500
44	144	169	2.8	2	0.549	0.549	0.586	0.586	0.566
54	113	188	1.4	2	0.554	0.554	0.592	0.592	0.572
46	120	311	1.8	2	0.575	0.575	0.620	0.620	0.710
66	132	212	0.1	2	0.601	0.601	0.650	0.555	0.434
59	142	239	1.2	2	0.635	0.550	0.587	0.445	0.448
56	142	256	0.6	2	0.646	0.624	0.675	0.619	0.466
51	142	305	1.2	2	0.650	0.650	0.700	0.700	0.540
59	134	218	2.2	2	0.650	0.650	0.604	0.604	0.584
58	146	225	2.8	2	0.650	0.650	0.528	0.528	0.528
67	71	237	1	2	0.678	0.613	0.664	0.590	0.451
56	103	283	1.6	2	0.679	0.679	0.712	0.712	0.617
57	143	335	3	2	0.696	0.696	0.721	0.721	0.721
63	147	254	1.4	2	0.703	0.564	0.587	0.571	0.579
66	120	246	0	2	0.729	0.598	0.647	0.551	0.431
60	170	293	1.2	2	0.737	0.704	0.715	0.715	0.627
60	142	282	2.8	2	0.744	0.744	0.753	0.753	0.753
61	145	234	2.6	2	0.757	0.650	0.554	0.554	0.606
62	97	263	1.2	2	0.764	0.680	0.713	0.713	0.583
62	106	294	1.9	2	0.764	0.764	0.771	0.771	0.741
63	144	187	4	2	0.777	0.650	0.700	0.700	0.888
59	140	326	3.4	2	0.800	0.725	0.739	0.739	0.805
55	117	327	3.4	2	0.800	0.650	0.700	0.700	0.805
68	141	193	3.4	2	0.800	0.650	0.700	0.700	0.805
65	127	254	2.8	2	0.802	0.650	0.662	0.662	0.745
67	108	286	1.5	2	0.805	0.786	0.790	0.790	0.679
61	140	260	3.6	2	0.808	0.650	0.700	0.700	0.759
69	146	254	2	2	0.841	0.650	0.634	0.634	0.768

Age	Thalach	Chol	Oldpeak	Class	Initial FIS	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration (FINAL)
39	140	219	1.2	3	0.348	0.348	0.316	0.316	0.327
52	168	212	1	3	0.465	0.465	0.436	0.436	0.453
48	150	256	0	3	0.488	0.488	0.532	0.532	0.414
40	114	167	2	3	0.500	0.500	0.500	0.500	0.500
50	126	200	0.9	3	0.500	0.500	0.500	0.500	0.500
54	108	206	0	3	0.500	0.500	0.500	0.435	0.452
66	165	228	1	3	0.519	0.519	0.499	0.499	0.399
57	115	131	1.2	3	0.529	0.529	0.554	0.554	0.539
43	120	177	2.5	3	0.538	0.538	0.568	0.568	0.550
48	166	274	0.5	3	0.539	0.522	0.533	0.533	0.529
59	90	176	1	3	0.546	0.546	0.580	0.580	0.500
67	163	254	0.2	3	0.557	0.557	0.576	0.576	0.445
45	147	309	0	3	0.562	0.562	0.603	0.603	0.452
49	139	188	2	3	0.627	0.627	0.678	0.678	0.669
58	173	224	3.2	3	0.650	0.650	0.444	0.444	0.561
55	132	353	1.2	3	0.650	0.650	0.700	0.700	0.539
55	111	217	5.6	3	0.650	0.650	0.609	0.700	0.765
54	126	239	2.8	3	0.650	0.650	0.587	0.587	0.567
53	95	282	2	3	0.650	0.650	0.700	0.700	0.700
55	130	205	2	3	0.650	0.650	0.669	0.669	0.658
60	157	258	2.6	3	0.677	0.650	0.649	0.649	0.661
58	111	270	0.8	3	0.708	0.708	0.728	0.728	0.549
58	130	259	3	3	0.714	0.650	0.692	0.692	0.721
58	152	319	0	3	0.717	0.717	0.733	0.733	0.557
68	150	274	1.6	3	0.746	0.723	0.729	0.729	0.681
54	116	286	3.2	3	0.750	0.650	0.700	0.700	0.757
62	160	268	3.6	3	0.764	0.764	0.771	0.771	0.770
62	145	164	6.2	3	0.764	0.650	0.700	0.700	0.888
62	103	281	1.4	3	0.765	0.765	0.771	0.771	0.644
63	132	330	1.8	3	0.777	0.778	0.783	0.783	0.727
64	96	246	2.2	3	0.783	0.650	0.624	0.624	0.692
67	125	299	0.9	3	0.826	0.834	0.837	0.837	0.652
70	112	269	2.9	3	0.851	0.717	0.734	0.734	0.893
56	133	288	4	3	0.852	0.673	0.710	0.710	0.815
51	122	298	4.2	3	0.853	0.650	0.700	0.700	0.878

Age	Thalach	Chol	Oldpeak	Class	Initial FIS	1st Iteration	2nd Iteration	3rd Iteration	4th Iteration (FINAL)
58	165	230	2.5	4	0.500	0.500	0.349	0.349	0.459
63	154	407	4	4	0.500	0.500	0.500	0.500	0.500
38	182	231	3.8	4	0.500	0.500	0.434	0.434	0.451
65	114	225	1	4	0.595	0.595	0.644	0.545	0.427
60	132	206	2.4	4	0.650	0.650	0.662	0.662	0.649
50	128	243	2.6	4	0.650	0.650	0.609	0.609	0.589
70	125	174	2.6	4	0.650	0.650	0.700	0.700	0.700
55	145	289	0.8	4	0.650	0.650	0.700	0.700	0.500
64	132	212	2	4	0.650	0.650	0.634	0.634	0.616
57	124	289	1	4	0.692	0.692	0.719	0.719	0.536
77	162	304	0	4	0.719	0.719	0.735	0.735	0.500
61	125	166	3.6	4	0.751	0.650	0.700	0.700	0.819
58	140	318	4.4	4	0.855	0.707	0.728	0.728	0.900

J Initial FIS Performance Results

System	Class	Mean Risk	Median Risk	Min	Max	Standard deviation
Initial FIS	0	0.536033978	0.529330148	0.203813559	0.857857143	0.136072277
	1	0.597592874	0.65	0.315455798	0.859897436	0.130365234
	2	0.659735244	0.664202683	0.434963548	0.841400811	0.117218923
	3	0.647564717	0.65	0.347809692	0.853333333	0.128565597
	4	0.643327239	0.65	0.5	0.855480226	0.10363406

K Tuning Iterations Performance

Iteration	Class	Mean Risk	Median Risk	Min	Max	Standard deviation
1st Iteration	0	0.52	0.51	0.20	0.79	-
	1	0.58	0.59	0.32	0.88	-
	2	0.61	0.64	0.42	0.79	-
	3	0.62	0.65	0.35	0.83	-
	4	0.62	0.65	0.50	0.72	-
2nd Iteration	0	0.52	0.50	0.20	0.79	-
	1	0.59	0.62	0.30	0.89	-
	2	0.63	0.64	0.39	0.79	-
	3	0.63	0.65	0.32	0.84	-
	4	0.62	0.66	0.35	0.73	-
3rd Iteration	0	0.51	0.50	0.20	0.79	-
	1	0.58	0.60	0.30	0.89	-
	2	0.61	0.60	0.39	0.79	-
	3	0.63	0.67	0.32	0.84	-
	4	0.62	0.66	0.35	0.73	-
4th Iteration (Final)	0	0.46	0.48	0.20	0.79	-
	1	0.52	0.52	0.16	0.90	-
	2	0.61	0.58	0.34	0.89	-
	3	0.61	0.57	0.33	0.89	-
	4	0.59	0.54	0.43	0.90	-

L Initial vs Tuned FIS Box Plots

Figure 6: Initial FIS Box Plot

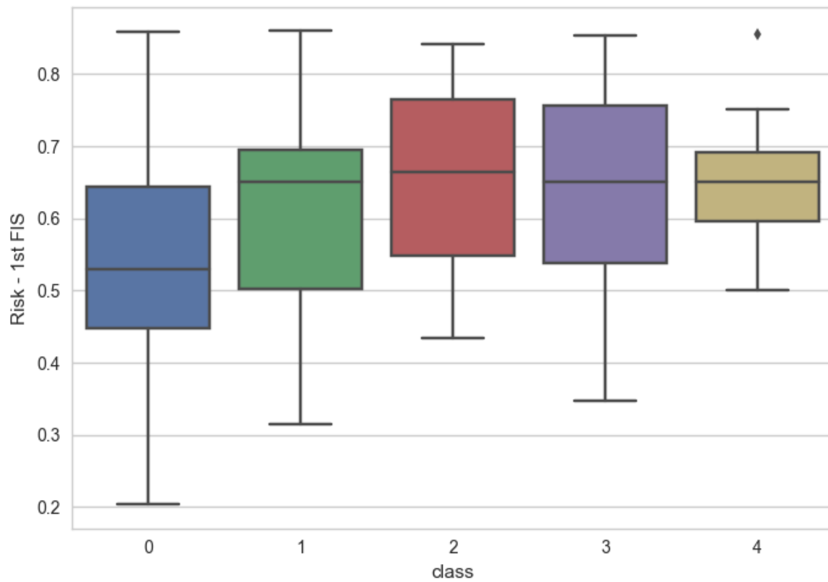
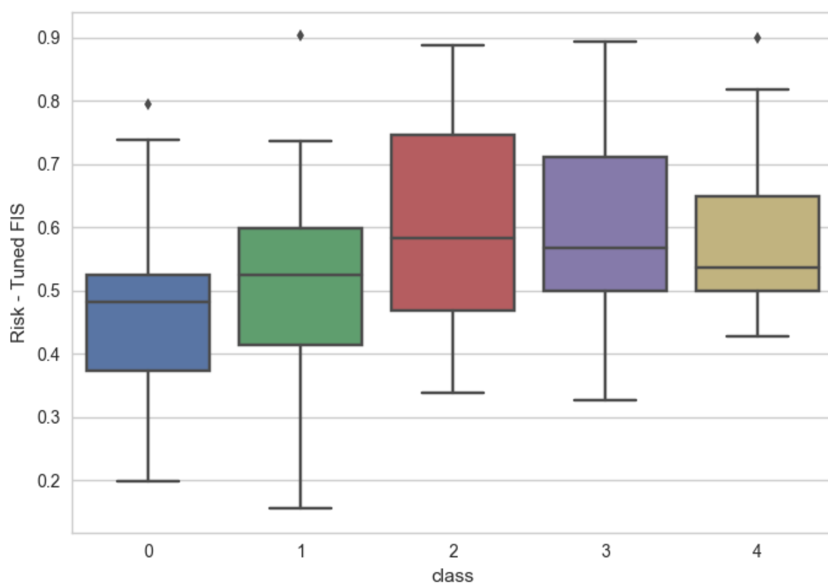
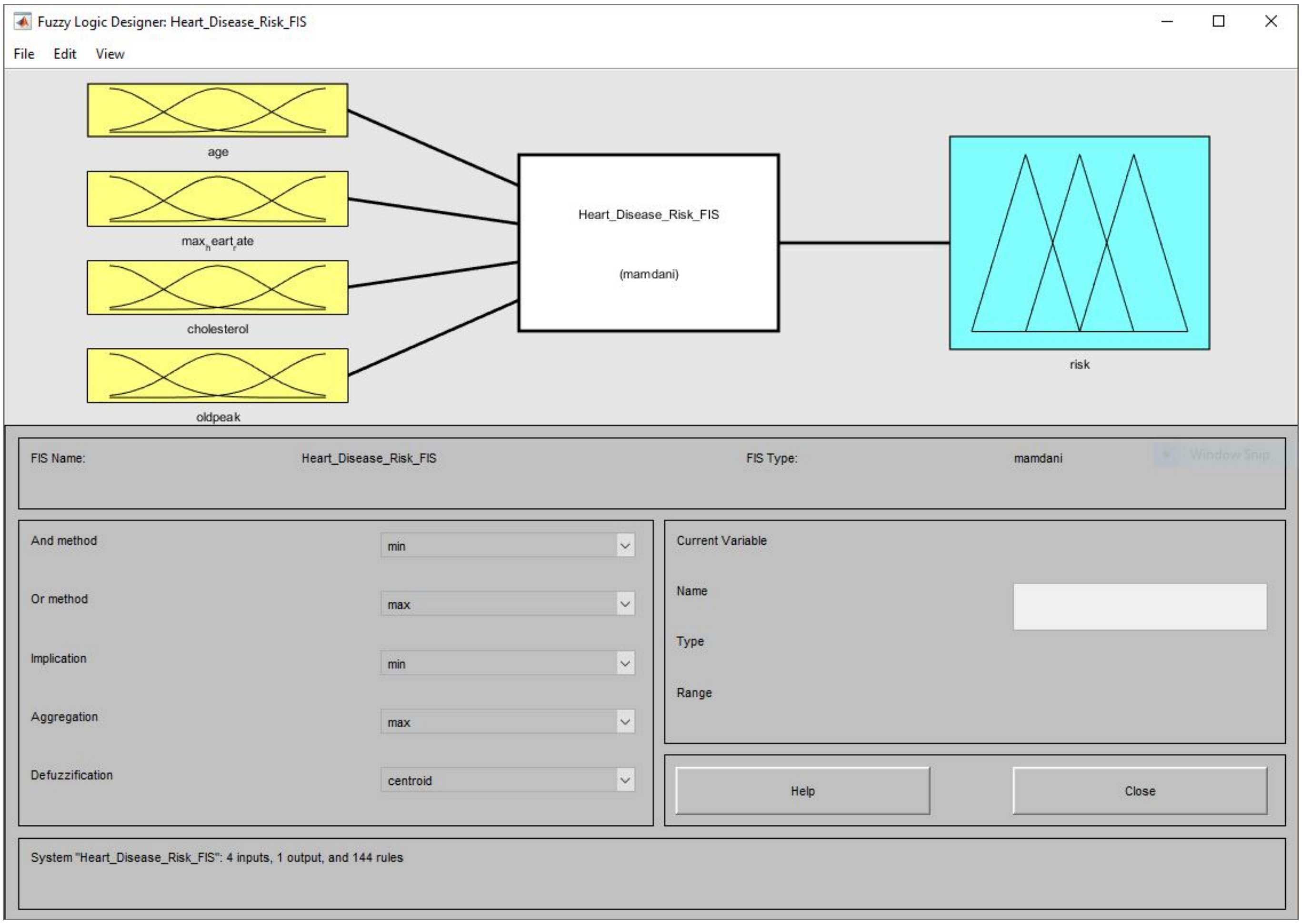


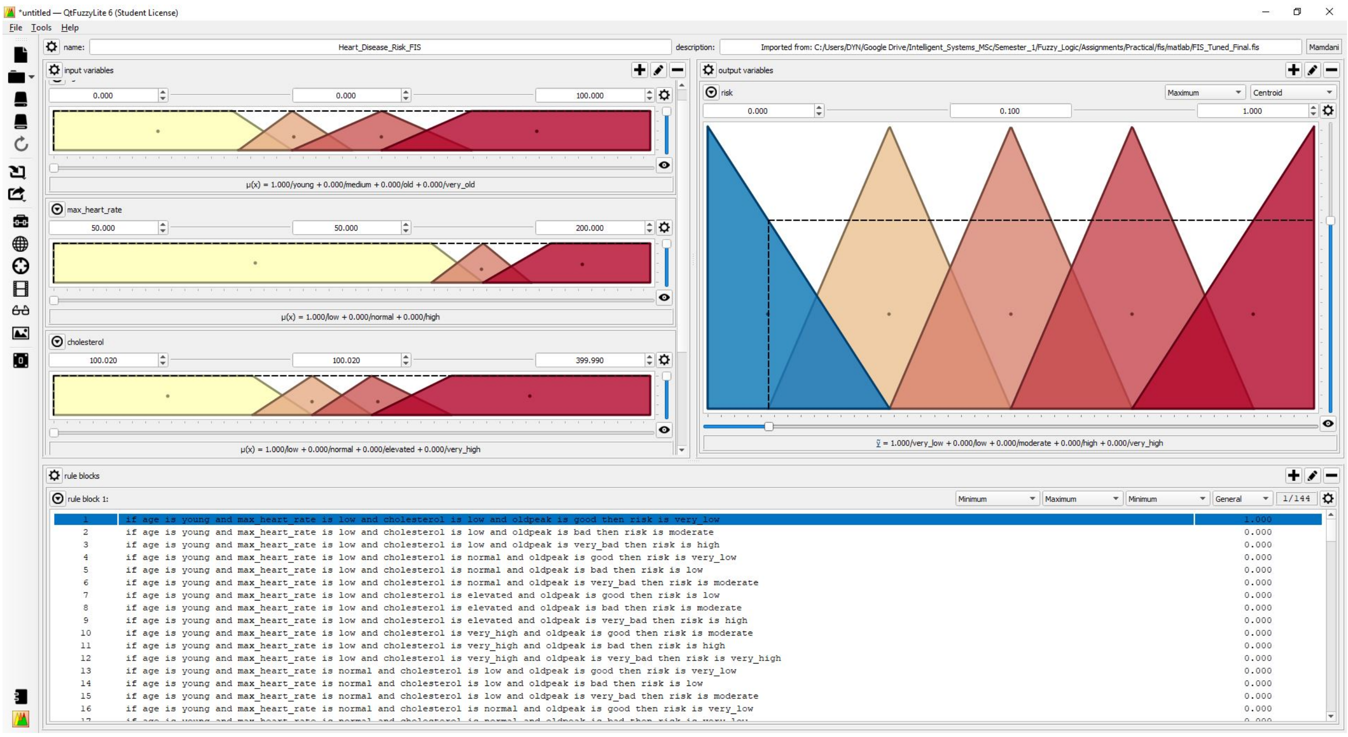
Figure 7: Tuned FIS Box Plot



M FIS Matlab Overview



N FIS QT FuzzyLite Overview



O FIS Surface View

Figure 8: Age vs Cholesterol

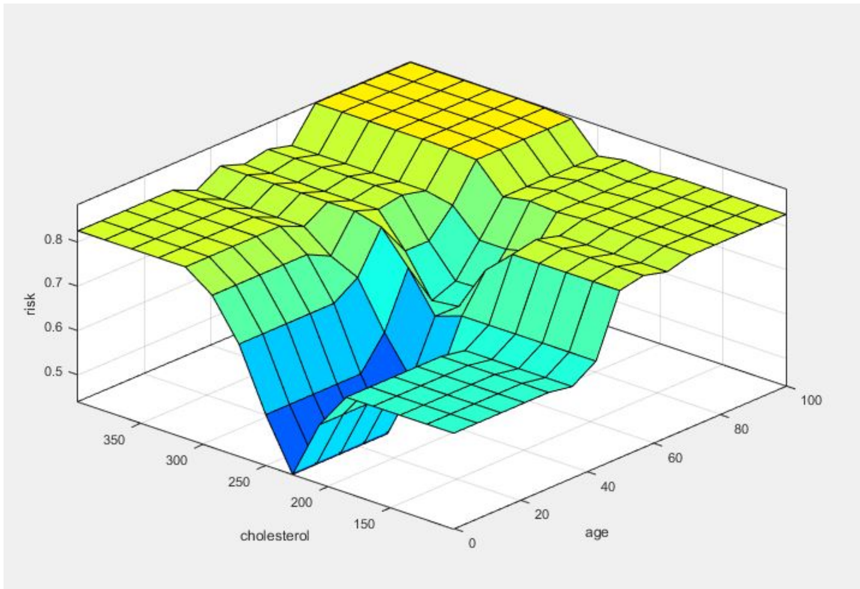


Figure 9: Age vs OldPeak

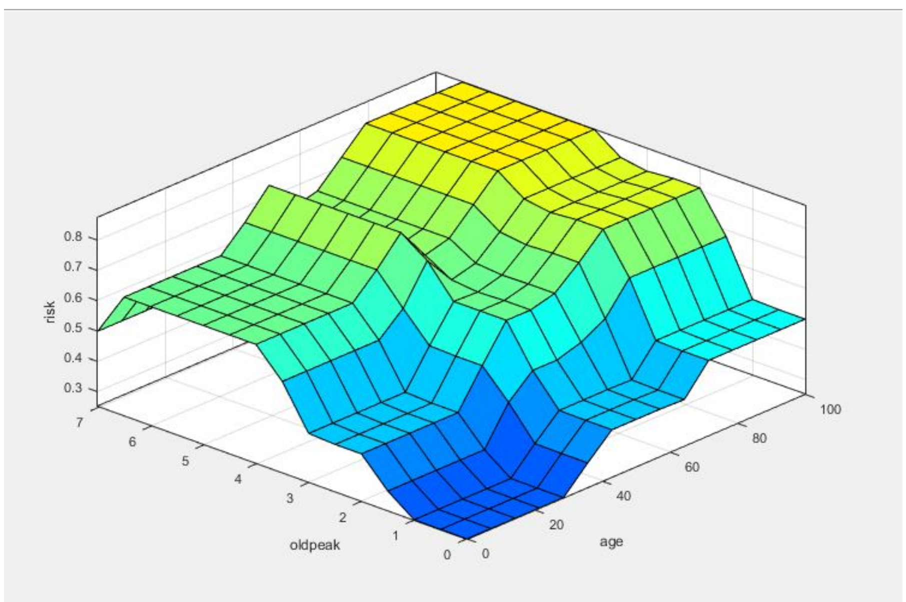


Figure 10: Age vs Thalach

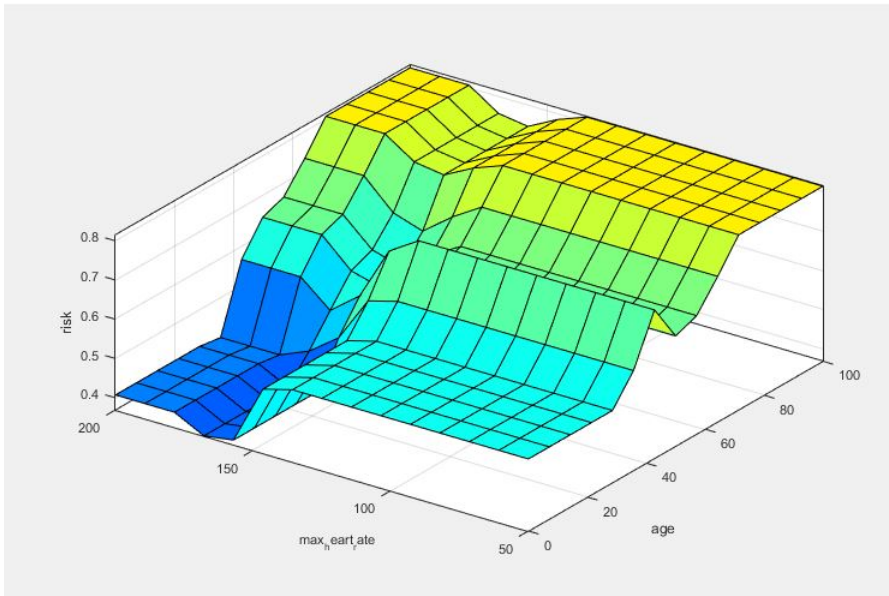
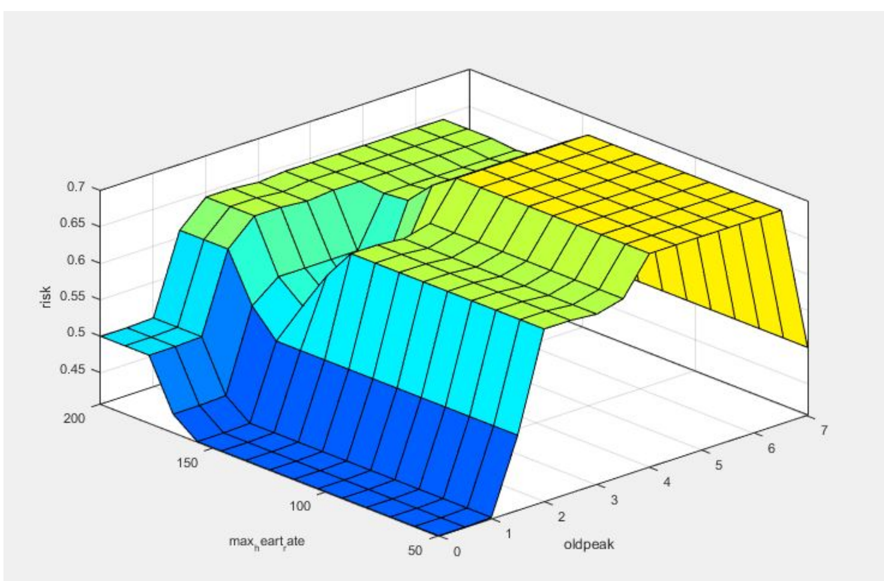


Figure 11: Oldpeak vs Thalach



P FIS Matlab Code

```
#Code automatically generated with fuzzylite 6.0.

[System]
Name='Heart_Disease_Risk_FIS'
Type='mamdani'
Version=6.0
NumInputs=4
NumOutputs=1
NumRules=144
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'

[Input1]
Name='age'
Range=[0.000 100.000]
NumMFs=4
MF1='young':'trapmf',[0.000 0.000 30.000 40.000]
MF2='medium':'trimf',[31.000 40.000 50.000]
MF3='old':'trimf',[40.000 55.000 70.000]
MF4='very_old':'trapmf',[55.000 70.000 100.000 100.000]

[Input2]
Name='max_heart_rate'
Range=[50.000 200.000]
NumMFs=3
MF1='low':'trapmf',[0.000 0.000 145.000 158.000]
MF2='normal':'trimf',[145.000 158.000 170.000]
MF3='high':'trapmf',[158.000 175.000 200.000 200.000]

[Input3]
Name='cholesterol'
Range=[100.020 399.990]
NumMFs=4
MF1='low':'trapmf',[100.000 100.000 200.000 230.000]
MF2='normal':'trimf',[200.000 230.000 260.000]
MF3='elevated':'trimf',[230.000 260.000 300.000]
MF4='very_high':'trapmf',[260.000 300.000 400.000 400.000]

[Input4]
Name='oldpeak'
Range=[0.000 7.030]
NumMFs=3
MF1='good':'trapmf',[0.000 0.000 1.000 2.000]
MF2='bad':'trimf',[1.000 2.300 4.000]
MF3='very_bad':'trapmf',[3.000 3.600 7.000 7.000]
```

```
[Output1]
Name='risk'
Range=[0.000 1.000]
NumMFs=5
MF1='very_low':'trimf',[0.000 0.000 0.300]
MF2='low':'trimf',[0.100 0.300 0.500]
MF3='moderate':'trimf',[0.300 0.500 0.700]
MF4='high':'trimf',[0.500 0.700 0.900]
MF5='very_high':'trimf',[0.700 1.000 1.000]
```

```
[Rules]
1.000 1.000 1.000 1.000 , 1.000 (1.000) : 1
1.000 1.000 1.000 2.000 , 3.000 (1.000) : 1
1.000 1.000 1.000 3.000 , 4.000 (1.000) : 1
1.000 1.000 2.000 1.000 , 1.000 (1.000) : 1
1.000 1.000 2.000 2.000 , 2.000 (1.000) : 1
1.000 1.000 2.000 3.000 , 3.000 (1.000) : 1
1.000 1.000 3.000 1.000 , 2.000 (1.000) : 1
1.000 1.000 3.000 2.000 , 3.000 (1.000) : 1
1.000 1.000 3.000 3.000 , 4.000 (1.000) : 1
1.000 1.000 4.000 1.000 , 3.000 (1.000) : 1
1.000 1.000 4.000 2.000 , 4.000 (1.000) : 1
1.000 1.000 4.000 3.000 , 5.000 (1.000) : 1
1.000 2.000 1.000 1.000 , 1.000 (1.000) : 1
1.000 2.000 1.000 2.000 , 2.000 (1.000) : 1
1.000 2.000 1.000 3.000 , 3.000 (1.000) : 1
1.000 2.000 2.000 1.000 , 1.000 (1.000) : 1
1.000 2.000 2.000 2.000 , 1.000 (1.000) : 1
1.000 2.000 2.000 3.000 , 2.000 (1.000) : 1
1.000 2.000 3.000 1.000 , 2.000 (1.000) : 1
1.000 2.000 3.000 2.000 , 3.000 (1.000) : 1
1.000 2.000 3.000 3.000 , 3.000 (1.000) : 1
1.000 2.000 4.000 1.000 , 2.000 (1.000) : 1
1.000 2.000 4.000 2.000 , 3.000 (1.000) : 1
1.000 2.000 4.000 3.000 , 3.000 (1.000) : 1
1.000 3.000 1.000 1.000 , 1.000 (1.000) : 1
1.000 3.000 1.000 2.000 , 2.000 (1.000) : 1
1.000 3.000 1.000 3.000 , 3.000 (1.000) : 1
1.000 3.000 2.000 1.000 , 1.000 (1.000) : 1
1.000 3.000 2.000 2.000 , 2.000 (1.000) : 1
1.000 3.000 2.000 3.000 , 2.000 (1.000) : 1
1.000 3.000 3.000 1.000 , 2.000 (1.000) : 1
1.000 3.000 3.000 3.000 , 3.000 (1.000) : 1
1.000 3.000 3.000 2.000 , 3.000 (1.000) : 1
1.000 3.000 4.000 1.000 , 2.000 (1.000) : 1
1.000 3.000 4.000 2.000 , 3.000 (1.000) : 1
1.000 3.000 4.000 3.000 , 3.000 (1.000) : 1
```

```
2.000 1.000 1.000 1.000 , 2.000 (1.000) : 1
2.000 1.000 1.000 2.000 , 3.000 (1.000) : 1
2.000 1.000 1.000 3.000 , 4.000 (1.000) : 1
2.000 1.000 2.000 1.000 , 2.000 (1.000) : 1
2.000 1.000 2.000 2.000 , 3.000 (1.000) : 1
2.000 1.000 2.000 3.000 , 4.000 (1.000) : 1
2.000 1.000 3.000 1.000 , 3.000 (1.000) : 1
2.000 1.000 3.000 2.000 , 4.000 (1.000) : 1
2.000 1.000 3.000 3.000 , 5.000 (1.000) : 1
2.000 1.000 4.000 1.000 , 3.000 (1.000) : 1
2.000 1.000 4.000 2.000 , 5.000 (1.000) : 1
2.000 1.000 4.000 3.000 , 5.000 (1.000) : 1
2.000 2.000 1.000 1.000 , 1.000 (1.000) : 1
2.000 2.000 1.000 2.000 , 2.000 (1.000) : 1
2.000 2.000 1.000 3.000 , 3.000 (1.000) : 1
2.000 2.000 2.000 1.000 , 1.000 (1.000) : 1
2.000 2.000 2.000 2.000 , 2.000 (1.000) : 1
2.000 2.000 2.000 3.000 , 3.000 (1.000) : 1
2.000 2.000 3.000 1.000 , 2.000 (1.000) : 1
2.000 2.000 3.000 2.000 , 3.000 (1.000) : 1
2.000 2.000 3.000 3.000 , 4.000 (1.000) : 1
2.000 2.000 4.000 1.000 , 2.000 (1.000) : 1
2.000 2.000 4.000 2.000 , 4.000 (1.000) : 1
2.000 2.000 4.000 3.000 , 4.000 (1.000) : 1
2.000 3.000 1.000 1.000 , 2.000 (1.000) : 1
2.000 3.000 1.000 2.000 , 2.000 (1.000) : 1
2.000 3.000 1.000 3.000 , 3.000 (1.000) : 1
2.000 3.000 2.000 1.000 , 2.000 (1.000) : 1
2.000 3.000 2.000 2.000 , 3.000 (1.000) : 1
2.000 3.000 2.000 3.000 , 3.000 (1.000) : 1
2.000 3.000 3.000 1.000 , 2.000 (1.000) : 1
2.000 3.000 3.000 2.000 , 3.000 (1.000) : 1
2.000 3.000 3.000 3.000 , 3.000 (1.000) : 1
2.000 3.000 4.000 1.000 , 3.000 (1.000) : 1
2.000 3.000 4.000 2.000 , 4.000 (1.000) : 1
2.000 3.000 4.000 3.000 , 5.000 (1.000) : 1
```

```
3.000 1.000 1.000 1.000 , 3.000 (1.000) : 1
3.000 1.000 1.000 2.000 , 4.000 (1.000) : 1
3.000 1.000 1.000 3.000 , 5.000 (1.000) : 1
3.000 1.000 2.000 1.000 , 2.000 (1.000) : 1
3.000 1.000 2.000 2.000 , 3.000 (1.000) : 1
3.000 1.000 2.000 3.000 , 4.000 (1.000) : 1
3.000 1.000 3.000 1.000 , 3.000 (1.000) : 1
3.000 1.000 3.000 2.000 , 4.000 (1.000) : 1
3.000 1.000 3.000 3.000 , 4.000 (1.000) : 1
3.000 1.000 4.000 1.000 , 3.000 (1.000) : 1
3.000 1.000 4.000 2.000 , 4.000 (1.000) : 1
3.000 1.000 4.000 3.000 , 5.000 (1.000) : 1
3.000 2.000 1.000 1.000 , 3.000 (1.000) : 1
3.000 2.000 1.000 2.000 , 3.000 (1.000) : 1
3.000 2.000 1.000 3.000 , 4.000 (1.000) : 1
3.000 2.000 2.000 1.000 , 2.000 (1.000) : 1
3.000 2.000 2.000 2.000 , 2.000 (1.000) : 1
3.000 2.000 2.000 3.000 , 3.000 (1.000) : 1
3.000 2.000 3.000 1.000 , 3.000 (1.000) : 1
3.000 2.000 3.000 2.000 , 4.000 (1.000) : 1
3.000 2.000 3.000 3.000 , 4.000 (1.000) : 1
3.000 2.000 4.000 1.000 , 3.000 (1.000) : 1
3.000 2.000 4.000 2.000 , 4.000 (1.000) : 1
3.000 2.000 4.000 3.000 , 5.000 (1.000) : 1
3.000 3.000 1.000 1.000 , 3.000 (1.000) : 1
3.000 3.000 1.000 2.000 , 3.000 (1.000) : 1
3.000 3.000 1.000 3.000 , 4.000 (1.000) : 1
3.000 3.000 2.000 1.000 , 3.000 (1.000) : 1
3.000 3.000 2.000 2.000 , 3.000 (1.000) : 1
3.000 3.000 2.000 3.000 , 3.000 (1.000) : 1
3.000 3.000 3.000 1.000 , 3.000 (1.000) : 1
3.000 3.000 3.000 2.000 , 4.000 (1.000) : 1
3.000 3.000 3.000 3.000 , 4.000 (1.000) : 1
3.000 3.000 4.000 1.000 , 4.000 (1.000) : 1
3.000 3.000 4.000 2.000 , 4.000 (1.000) : 1
3.000 3.000 4.000 3.000 , 5.000 (1.000) : 1
```

```
4.000 1.000 1.000 1.000 , 3.000 (1.000) : 1
4.000 1.000 1.000 2.000 , 4.000 (1.000) : 1
4.000 1.000 1.000 3.000 , 5.000 (1.000) : 1
4.000 1.000 2.000 1.000 , 3.000 (1.000) : 1
4.000 1.000 2.000 2.000 , 4.000 (1.000) : 1
4.000 1.000 2.000 3.000 , 5.000 (1.000) : 1
4.000 1.000 3.000 1.000 , 3.000 (1.000) : 1
4.000 1.000 3.000 2.000 , 5.000 (1.000) : 1
4.000 1.000 3.000 3.000 , 5.000 (1.000) : 1
4.000 1.000 4.000 1.000 , 4.000 (1.000) : 1
4.000 1.000 4.000 2.000 , 5.000 (1.000) : 1
4.000 1.000 4.000 3.000 , 5.000 (1.000) : 1
4.000 2.000 1.000 1.000 , 3.000 (1.000) : 1
4.000 2.000 1.000 2.000 , 4.000 (1.000) : 1
4.000 2.000 1.000 3.000 , 4.000 (1.000) : 1
4.000 2.000 2.000 1.000 , 2.000 (1.000) : 1
4.000 2.000 2.000 2.000 , 3.000 (1.000) : 1
4.000 2.000 2.000 3.000 , 4.000 (1.000) : 1
4.000 2.000 3.000 1.000 , 3.000 (1.000) : 1
4.000 2.000 3.000 2.000 , 4.000 (1.000) : 1
4.000 2.000 3.000 3.000 , 5.000 (1.000) : 1
4.000 2.000 4.000 1.000 , 3.000 (1.000) : 1
4.000 2.000 4.000 2.000 , 4.000 (1.000) : 1
4.000 2.000 4.000 3.000 , 5.000 (1.000) : 1
4.000 3.000 1.000 1.000 , 3.000 (1.000) : 1
4.000 3.000 1.000 2.000 , 4.000 (1.000) : 1
4.000 3.000 1.000 3.000 , 4.000 (1.000) : 1
4.000 3.000 2.000 1.000 , 3.000 (1.000) : 1
4.000 3.000 2.000 2.000 , 4.000 (1.000) : 1
4.000 3.000 2.000 3.000 , 4.000 (1.000) : 1
4.000 3.000 3.000 1.000 , 3.000 (1.000) : 1
4.000 3.000 3.000 2.000 , 4.000 (1.000) : 1
4.000 3.000 3.000 3.000 , 5.000 (1.000) : 1
4.000 3.000 4.000 1.000 , 3.000 (1.000) : 1
4.000 3.000 4.000 2.000 , 4.000 (1.000) : 1
4.000 3.000 4.000 3.000 , 5.000 (1.000) : 1
```