**A comparative Analysis of Machine learning Models to Determine**

**Obesity Levels in Humans**

Ananthu Karukampilly Surendran

Aleena Alby

Faculty of Engineering Environment

And Computing

Coventry University

Coventry, United Kingdom

karukampia@uni.coventry.ac.uk

albya@uni.coventry.ac.uk

***Abstract – Obesity is a complicated condition characterized by an excess of body fat. Obesity must be addressed because, in addition to generating apparent physical changes, it can lead to a variety of severe and potentially life-threatening ailments. This paper presents a comparative study of different machine learning algorithms such as Logistic Regression, Support Vector Classification, Decision Tree, XGBoost and Gaussian Process Classification to identify the obesity level in humans. Before Modelling the data is subjected to some pre-processing techniques such as data format conversion, scaling etc. All the models were trained on the same dataset and XGBoost and Decision Tree Classifier achieved an accuracy of 99% and 98% respectively.***

***Keywords – Logistic Regression, Support Vector Classifier (SVC), Decision Tree, XGBoost, Gaussian Process classification (GPC).***

# Introduction

Obesity is a chronic condition that affects both children and adults equally. It is prevalent in both developing and developed countries. Obesity was also associated with a higher death rate as per a research conducted in the twentieth century (WHO, 2000). Understanding, predicting, and controlling obesity therefore has the potential to save people's lives and billions of dollars (Abdullah, Hussain, & Hussain, 2017).

Obesity is described as a condition in which an abnormal or excessive quantity of fat accumulates in the adipose tissue as a result of a negative energy balance and weight growth (WHO, 2000). The WHO (World Health Organization) cautioned the public in 1998 by classifying obesity as a public health hazard, but in 2004, the WHO proclaimed obesity to be a worldwide epidemic (Río, et al., 2019). Nowadays, childhood obesity is a worldwide problem. Occasionally, this condition has a related issue like type II diabetes. Children who are obese are at risk for developing a variety of short-, medium-, and long-term health issues, including an increase in risk factors for cardiovascular and metabolic diseases, psychological and social issues brought on by peer taunting and comments, and long-term obesity persistence, an increase in disabilities, or early death. Childhood obesity is alarming not just because it is becoming more prevalent in populations around the world (Río, et al., 2019).

According to figures presented by NHS, 1 in 4 adults and 1 in 5 children are presumed to be obese in the UK (NHS, n.d.). Body Mass Index (BMI) gives us a good way to gauge the extent of obesity, but it does not take into consideration the body's varying levels of fat.

This paper's major goal is to identify persons who are obese and educate them about the risk factors for obesity. The procedure is divided into two phases, the first of which involves reading the data, applying certain pre-processing procedures, and then visualizing the data to look for any underlying relationships between the attributes. Later in phase two, we will apply machine learning algorithms to the data in order to identify the most effective model for forecasting an individual's level of obesity.

# Dataset

The dataset for the project was procured from the website called UCI Machine learning repository. The website link to same is shown below.

<https://archive.ics.uci.edu/ml/datasets/Estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition+>

Based on their dietary habits and physical health, people in Mexico, Peru, and Colombia were used to collect the data.

The dataset contains 2111 rows of data across 17 characteristics. There are seven different categories for obesity levels. 23% of the data was directly obtained from users via a web platform, while 77% of the data were created synthetically using the Weka tool and the SMOTE filter, more information about the attributes is shown in appendix a.

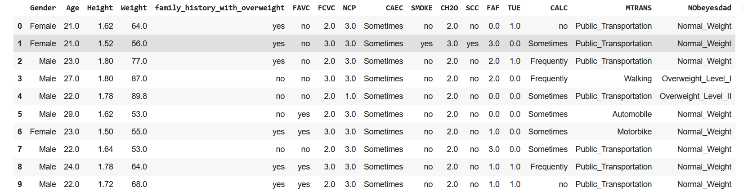
# Data understanding and exploration

The modifications and preparatory processes that were made prior to modelling are described in this section of the report.

The procedures involved in ore processing include scaling, cleaning, and encoding of the data. We also make use of other techniques such as dimensionality reduction and feature engineering to generate new features.

## Data Understanding

The dataset had 17 attributes of different data types. The first 10 rows of data from our obesity dataset are depicted in the picture beneath. Where 8 of the variables were numerical and the rest were categorical. We also checked whether our dataset have any null values and found none.



*Figure 1 First 10 rows of Dataset*

## Data Pre-processing

Data pre-processing is done to transform the raw data into something meaningful.

### Feature Construction : We used pandas dataframe to construct a new feature called BMI (Body Mass Index).The following equation is used to calculate body mas index. BMI is determined using a patient's height and weight, which are available in our core dataset.

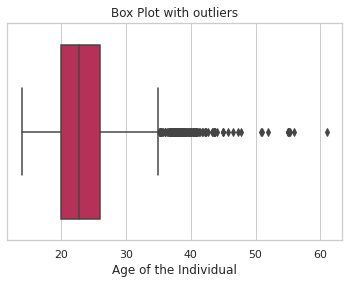
*Equation 1 BMI Equation.*

### Encoding: Categorical encoding is required since machine learning algorithms only understand numbers; Label Encoding and One-Hot Encoding can be used for this. The process of translating categories into numbers is known as categorical encoding. One-Hot Encoding generates extra features dependent on the number of distinct values in the category column. We employed Label Encoding for encoding in our project; with this method, each value or label in a category column is allocated a unique integer based on alphabetical ordering (Sethi, 2020).

### Handling Outliers: An outlier is a dataset observation that deviates dramatically from the rest of the data. It might happen owing to measurement variability or misunderstanding when filling out data points. The most obvious causes of outliers include data input errors, measurement errors, and so on (Suresh, 2020). Outlier identification and treatment is one of the most significant aspects of data preparation since it can have a negative influence on statistical analysis and the training process of a machine learning system, resulting in lower accuracy (Bonthu, 2021). For detecting outliers we have made of the IQR (Inter Quartile Range) method. The data points which are beyond -1.5\*IQR to 1.5\*IQR are treated as outliers.

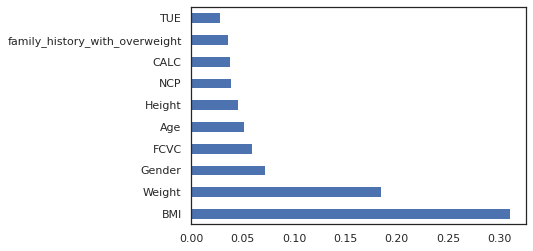
*Equation 2 IQR Equation*

*Q1 and Q3 denote the first and third quartiles, respectively. There were 168 outlier instances in the Age column, and all of these data points were removed.*



*Figure 2 Plotting Outliers in Age Column using Boxplot*

### Feature Selection: The process of extracting crucial traits from data such that the output of the model is precise and satisfies the criteria is referred to as feature selection. One of the key steps in the feature engineering technique is feature selection, which is done to cut down on the number of dependent variables. In order to boost the models' accuracy, feature selection is done. To identify the best characteristics, we employed a model-based strategy called Extra Tree Classifier. This approach tree-based approach to select the import features (Manoj, 2021).



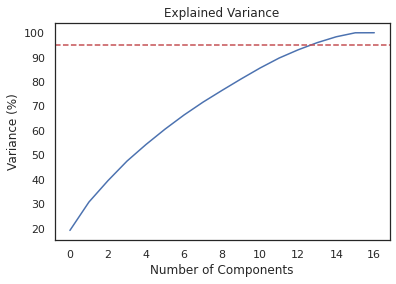
*Figure 3 Features with their Scores*

### Feature Scaling: The feature scaling stage of data preparation is very important. It is carried out to convert the dataset's characteristics into a limited range. Normalization and standardisation are two of the most used feature scaling techniques in machine learning. Data are shifted and rescaled using the normalisation scaling method such that they fall between 0 and 1. Additionally called Min-Max scaling. standardisation, when the data are centred with one standard deviation from the mean. The distribution that results has a unit standard deviation and the attribute's mean changes to zero as a result (aniruddha, 2020).

*Equation 3 Min-Max Scaler Equation*

*Equation 4 StandardScaler Equation*

### Dimentionality Reduction: The practise of reducing the number of features (columns) while keeping the best quality is known as "dimensionality reduction". The quantity of features (columns) is simply decreased but the most information is maintained by dimensionality reduction since it is difficult to analyse every variable at a microscopic level. Dimensionality reduction techniques include Kernel PCA, Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA). PCA - We must identify k distinct directions of data variation, often known as key main components, in an N-dimensional space. A number of essential procedures make up PCA. First, we standardise the data. Next, the covariance matrix is computed and utilised to identify the directions of largest variance in order to study the link between the attributes. These are computed using the eigenvalues and eigenvectors of the covariance matrix. Data may be reduced to two or three dimensions with the use of dimensionality reduction. The below figure depicts that we retain 95% important features when components is 12.



*Figure 4 PCA*

# Methods

This section will describe about the machine learning algorithms used to determine the obesity level in humans.

## Logistic Regression

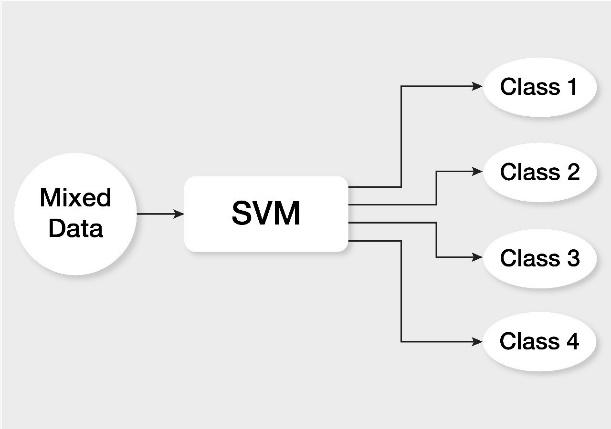
Logistic regression is a classification technique used in machine learning. Logistic regression is a statistical method for analysing a dataset where one or more independent factors influence the outcome. Finding the model that best reflects the relationship between the dependent and independent variables is the goal of logistic regression. In order to represent the dependent variable, a logistic function is used. There are only two viable classes since the dependent variable is dichotomous (Raj, 2020). (For instance, whether the person is overweight or not).

Even though binary target variables are commonly predicted using logistic regression, it may be broadened and further separated into three types: Multinomial, Ordinal, and Binomial.

## Support Vector Classifier

A linear model called Support Vector Machine, or SVM, is employed to address classification and regression issues.

It can resolve both linear and non-linear problems and is effective for a variety of real-world challenges. Using the data as input, the SVM algorithm, assuming it is able, generates a line that separates the classes into distinct segments. SVM applies the input data in a larger nonlinear space, called the characteristic space, to provide the optimal separation boundaries between variables (Ray, 2017).



*Figure 5 SVM Classifier (Source: towardsdatascience)*

## Decision Tree

A decision tree (DT) may be used to solve both classification and regression issues. The name implies that it use a flowchart-like tree structure to display the predictions resulting from a sequence of feature-based splits. It begins with a root node and concludes with a choice made by the leaves (Dang, Sangshin Kwak, & Seungdeog Choi, 2022). Decision trees are, to put it simply, a collection of if-else expressions. It determines if the condition is met and, if so, advances to the following node in the decision chain.

Once we reach the leaf node, we will get our output. Then it will choose another node according on the criteria (Ugale, Nimeesha Venkatavelu, Pranay Patil, & Suraj Rane, 2022).



*Figure 6 Decision Tree (Source: towardsdatascience)*

## XGBoost

A decision tree-based machine learning method called Extreme Gradient Boosting, or XGBoost, employs boosting to improve performance (NVIDIA, n.d.). The process of gradient boosting entails building new models that predict the outputs of older models, which are then integrated to produce the final result (Abbas, Hilal Tayara, & Kil To Chong, 2022). Classification and regression issues can be solved with XGBoost. To reduce loss, the method also uses a gradient descent technique.

## Gaussian Process Classification (GPC)

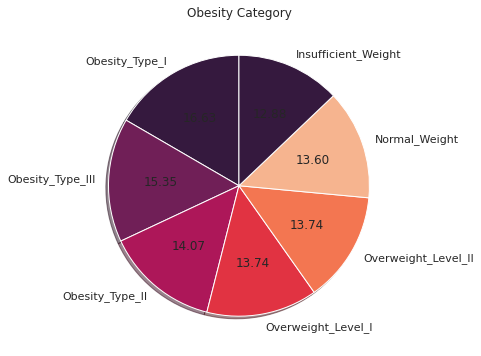
A machine learning method called the Gaussian process may be used for both classification and regression. A stochastic process with a Gaussian normal distribution as its kernel is known as a Gaussian process. Non-parametric in nature, GP holds that the data is evenly and uniformly distributed. A Gaussian distribution is used by the Gaussian Process model to produce predictions (Yi, 2019). The mean of a model reveals the typical or anticipated value of a prediction, whereas the variance reveals the degree of uncertainty associated with that prediction (Yi, 2019). In the case of Gaussian Process Regression (GPR), they can be obtained by generalizing the linear regression, but in the case of Gaussian Process Classification the likelihood is non-Gaussian. The Laplace approximation for GP can be used for both binary and multi-classification (Rasmussen & Williams, 2006).

# EXPERIMENTAL RESULTS AND ANALYSIS

As we mentioned we have done a comparative analysis of machine learning algorithms which we have mentioned in the Methods section. The Programming language for this entire purpose is Python 3 and we have used Jupyter Notebook as our IDE. The Jupyter Notebook App is a server-client application that allows editing and running notebook documents via a web browser. The steps involve EDA (Exploratory Data Analysis), Pre-processing, preparing different machine learning models and compare them on the basis of accuracy.

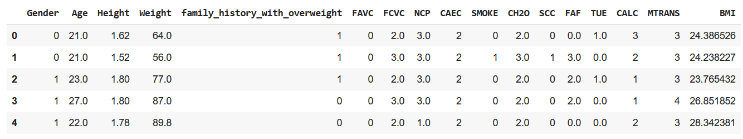
## EDA – (Exploratory Data Analysis)

The Figure below shows the class distribution in our dataset. We have created a pie-chart to show the distribution of data in all classes, and we can see that the data is evenly distributed between each class.



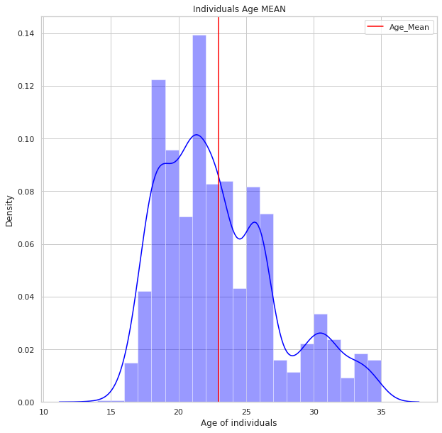
*Figure 7 Class Distribution in Dataset*

The figure 8, shows the top 5 rows of the dataset after applying feature construction. We can see that the last column BMI was custom made by us as BMI an important feature for calculating one’s obesity level. In Figure 2 we have shown that our dataset contains outliers and that we have eliminated them using the IQR method.



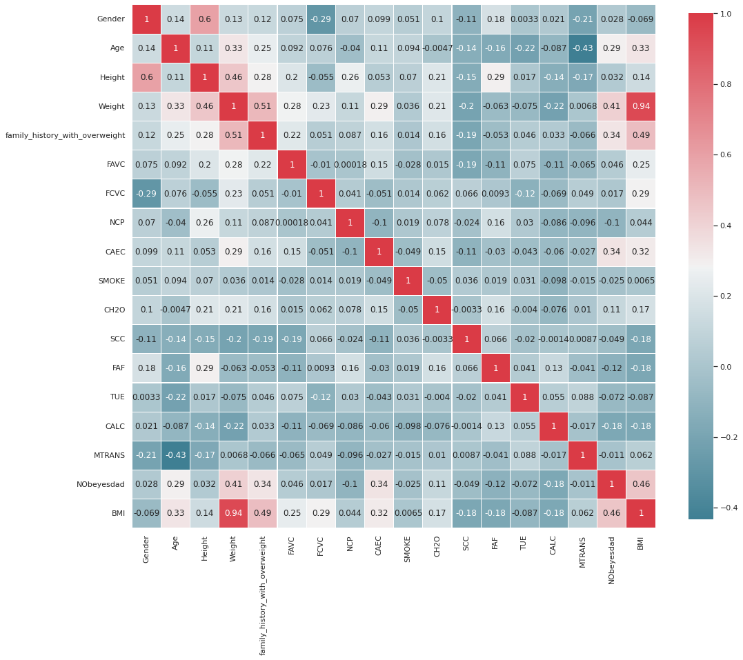
*Figure 8 Dataset after Feature Construction*

The below figure shows the distribution of Age column after we have removed the outliers from the data. A log transformation is applied to data and we can notice that the data tends to follow normal Gaussian distribution.



*Figure 9 Distplot of Age*

A correlation heatmap between the variables was generated using the help of seaborn library. We can see some variables have positive relations while others have an extreme negative correlation between them. As we have created the feature ‘BMI’ from weight we can see a positive correlation of 0.96 between them. Some features which have a positive correlation with our target variable are Height, Weight, BMI, history for overweight within family etc.



*Figure 10 Heatmap*

## Results

The findings of our models are described in this section.

The modified dataset comprised 1943 rows and 18 columns, while the original dataset had 2111 rows and 17 features.

The training data makes up 67% of the dataset.

The same dataset is used to train all of the models, which are then assessed using the F1-score, recall, precision, and accuracy.

The correct number of classifications are made for the overall number of classifications is called accuracy. Recall illustrates how many actual positive cases we were able to correctly predict using our model, while Precision shows how many of the accurately expected cases were positive.

F1 Score is the harmonic mean of precision and recall.

*5, Equations for Evaluating the Model*

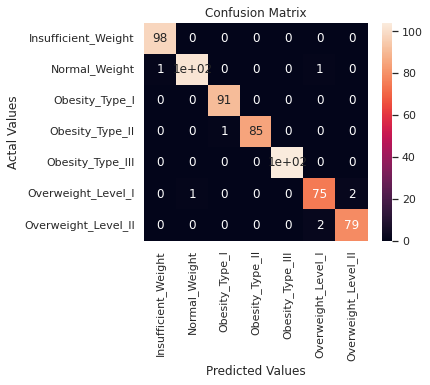
XGBoost model outperformed all the other models and achieved an accuracy of 99% in test data. Table 1 shows the performance metrics for each model we have implemented. We have also implemented hyperparameter tuning in each individual model with the help of GridSearchCV. It will return the best possible parameters for our model from different parameters passed into the parameter grid.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Models** | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| LG | 84% | 84% | 84% | 84% |
| SVC | 95% | 95% | 95% | 95% |
| DT | 98% | 98% | 98% | 98% |
| XGB | 99% | 99% | 99% | 99% |
| GPC | 91% | 91% | 91% | 91% |

Accuracy, recall, precision, and F1-Score scored by the logistic regression model is 84%. The model is instructed to give more weight to the training data by the hyperparameter provided, which was 'C'. The Support vector classifier received tolerance and kernel as hyperparameters, and the model scored 95% in terms of accuracy, recall, precision, and F1-Score. In terms of accuracy, recall, precision, and F1-Score, the decision tree model scored 98%.

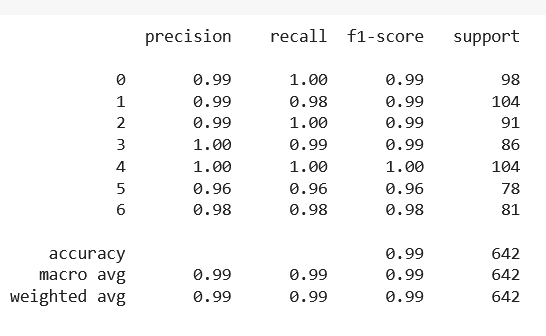
The minimal sample leaf was the hyperparameter provided.

Gaussian Process Classification was also used for the classification purpose and the model achieved a score of 91% in accuracy, recall, precision and F1-score. The hyperparameter was the different types of kernel. We tried to improve the score by using PCA. From figure 4, we can understand that when the no of components of 12, it can retrieve 95% of the important features. The new model achieved an accuracy of 97%.



*Figure 11 XGBoost Model ConfusionMatrix*

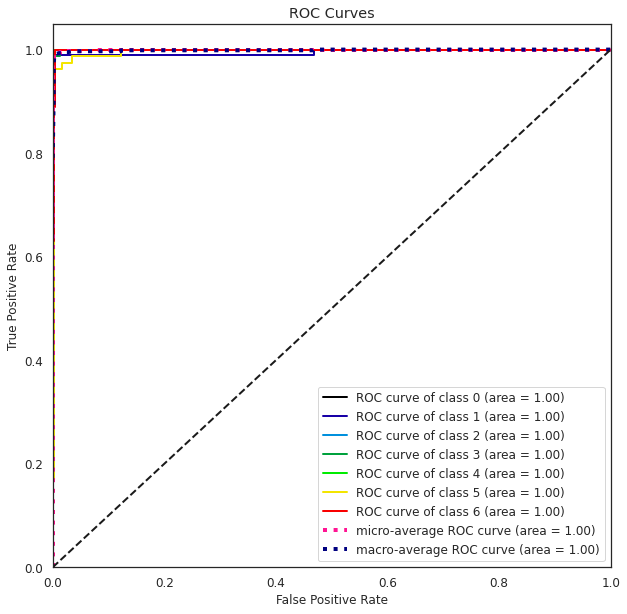
The XGBoost model was the superior model among all the other to find the obesity level in humans. The model achieved a score of 99% in terms of recall, precision, F1-Score and accuracy.



*Figure 12 Classification Report of XGBoost*

The hyperparameters "n\_estimator," "learning rate," and "max depth" were fed to the XGBoost model.

GridSearchCV returned a learning rate of 0.15 and a max depth of 3 with 70 n estimators as the optimal parameters for our model. The ROC-curve for the model is displayed in the figure below. Plotting the TPR (True Positive Rate) against the FPR yields the ROC curve (False positive rate).

The ROC curves in the image move up and to the left, indicating a strong model. Even though all the curves are closely packed, class 6 does show a little improvement over the other classes. 

*Figure 13 ROC – Curve*

# Social and ethical consideration

Overweight and obesity prevention programs are required to promote individual and public health, enable informed choice and reduce the social expenses associated with being overweight (Have, 2014). New proposals and policies regarding SSB (Sugar-sweetened beverages) consumption could be made (Kass, Kenneth Hecht, Amy Paul, & Kerry Birnbach, 2014). Healthy eating habits might be promoted through campaigns.

The necessity of appropriate eating and food should be presented in academics so that children may comprehend the value of public health advantages at a young age.

Obesity also has a significant social impact; research has shown that persons of comparable weight are usually connected. As a result, these individuals exhibit similar behavioral traits or tendencies. (Smith, Paul N Zivich, & Leah Frerichs, 2020).

# Discussion and conclusion

Obesity and overweight are considered epidemiological issues because they are associated with a variety of ailments, including type 2 diabetes, osteoarthritis, stroke, various forms of cancer, and, in some cases, mortality.

Obesity prevention is a complex task that requires the participation of the government, corporations, and professionals in the field. As a result, lowering obesity and its consequences necessitates considerable lifestyle changes among the general public, notably in terms of food and physical activity. In this regard, owing to technological breakthroughs, methods such as artificial intelligence and machine learning may now be used to support the prevention and diagnosis of overweight and obese people.

In this day and age, computer-aided diagnostics with the use of deep learning and machine learning can have a substantial impact. The conventional approach is time-consuming and has several drawbacks. A succeeding model might be created using ANN (Artificial Neural Networks) to evaluate if it outperforms the existing model.

In order to develop an intelligent model for the identification of people with obesity or overweight, six machine learning models, including Decision Trees, Support Vector Classifier, Logistic Regression, Gaussian Process Classification, and XGBoost, were tested in this study. XGBoost model was found to be the best optimal for classifying obesity levels in humans, this model will support decision-making for experts in the field.

# References

Abbas, Z., Hilal Tayara, & Kil To Chong. (2022). Alzheimer's disease prediction based on continuous feature representation. *Elsevier*.

Abdullah, A., Hussain, A., & Hussain, I. (2017). Predicting Obesity Using Longitudinal Near Infra-Red Spectroscopy (NIRS) Data. (pp. 123-128). ICCDA.

aniruddha. (2020, April 3). *analyticsvidhya*. Retrieved from analyticsvidhya: https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machine-learning-normalization-standardization/

Bonthu, H. (2021, May 21). *Analytics Vidhya*. Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2021/05/detecting-and-treating-outliers-treating-the-odd-one-out/?cv=1

Dang, H.-L., Sangshin Kwak, & Seungdeog Choi. (2022). Parallel DC Arc Failure Detecting Methods Based on Artificial Intelligent Techniques. *IEE*, 26058 - 26067.

Have, M. T. (2014). Ethical aspects of obesity prevention . *Elsevier*.

Kass, N., Kenneth Hecht, Amy Paul, & Kerry Birnbach. (2014). Ethics and Obesity Prevention: Ethical Considerations in 3 Approaches to Reducing Consumption of Sugar-Sweetened Beverages. *NIH*.

Manoj, S. (2021, April 23). *Analytics Vidhya*. Retrieved from https://www.analyticsvidhya.com/blog/2021/04/discovering-the-shades-of-feature-selection-methods/

NHS. (n.d.). *NHS.UK*. Retrieved from NHS: https://www.nhs.uk/conditions/obesity/

NVIDIA. (n.d.). *NVIDIA*. Retrieved from NVIDIA: https://www.nvidia.com/en-us/glossary/data-science/xgboost/

Raj, A. (2020, November 7). *towardsdatascience*. Retrieved from towardsdatascience: https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning.* MIT Press.

Ray, S. (2017, September 13). *AnalyticsVidhya*. Retrieved from AnalyticsVidhya: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

Río, N. G., Carina S. González-González, Francisco Gacía-Peñalvo, Honorio Armas-Ramos, Yeray del Cristo Barrios-Fleitas, Norberto Marrero-Gordillo, . . . Vicente Navarro-Adelantado. (2019). Treatment of children obesity and diabetes through. *TEEM'19: Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, (pp. 452-458).

Sethi, A. (2020, March 6). *AnalyticsVidhya*. Retrieved from https://www.analyticsvidhya.com/blog/2020/03/one-hot-encoding-vs-label-encoding-using-scikit-learn/

Smith, N. R., Paul N Zivich, & Leah Frerichs. (2020). Social influences on obesity: current knowledge, emerging methods, and directions for future research and practice. *NIH*.

Suresh, A. (2020, November 30). *Medium*. Retrieved from https://medium.com/analytics-vidhya/how-to-remove-outliers-for-machine-learning-24620c4657e8

Ugale, M., Nimeesha Venkatavelu, Pranay Patil, & Suraj Rane. (2022). Review of Machine Learning Techniques for Crop Recommendation System. *IJRASET*.

WHO. (2000). *World Health Organization (2000) Obesity: preventing and managing the global epidemic. .* Geneva: WHO.

Yi, W. (2019, December 1). *towardsdatascience*. Retrieved from https://towardsdatascience.com/understanding-gaussian-process-the-socratic-way-ba02369d804

# **Appendix**

|  |  |
| --- | --- |
| **Attributes** | **Answers** |
| Gender | * Female * Male |
| Age | Numeric value |
| Height | In meters |
| Weight | In KG |
| Family History for Obesity | Yes / No |
| FAVC (Caloric Food consumption frequency) | Yes / No |
| FCVC ( Vegetable Consumption Frequency ) | Never /Sometimes /Always |
| NCP (No of Main Meals) | Between 1 y 2 / Three / More than three |
| CAEC ( Food consumption in between meals ) | No / Sometimes / Frequently /Always |
| SMOKE | Yes / No |
| CH2O (Water use on a daily basis) | Less than a litre / Between 1 and 2 L / More than 2 L |
| SCC ( squamous cell carcinoma ) | Yes / No |
| FAF ( Physical Activity Frequency ) | I do not have/1 or 2 days/2 or 4 days/4 or 5 days |
| TUE ( Therapeutic Use Exceptions ) | Numeric value |
| CALC ( Consumption of alcohol ) | I do not drink/Sometimes/Frequently/Always |
| MTRANS ( Which transportation usually use ) | Automobile/Motorbike/Bike/Public Transportation/Walking |
| NObeyesdad (Obesity level) | * Underweight * Normal Weight * Overweight I * Overweight II * Obesity I * Obesity II * Obesity III |

Github link for code: https://github.com/AleenaAlby/Modelling-and-Optimisation-Under-Uncertainty

**Fuzzy Logic Controller in Residencies for Disabled**

***Abstract*— This paper uses fuzzy logic controllers to demonstrate an assistance system designed for people with disabilities. The Fuzzy Logic Controller analyses inputs like temperature, heart rate, and other factors to produce the desired output. With the aid of the MATLAB Fuzzy Logic Toolbox, FLC was implemented. Optimization of the membership function was done using the help of Genetic Algorithm.**

***Keywords— Fuzzy Logic Controller (FLC), Fuzzy Inference System (FLS), Genetic Algorithm.***

# I. Introduction

Technology has given us a plethora of facilities and made our daily lives so much easier than we would have ever imagined. It has bought revolutionary changes in many field especially in medical field (Verma, 2021). The use of intelligent applications and assistive mechanisms to meet the needs of old aged individuals and patients who live solely is becoming ever more popular today (Galliakis, Skourlas, Galiotou, & Voyiatiz, 2018). The use of smart homes and their number of users are increasing exponentially. It was anticipated that the market for smart home technology alone would increase from $40 million in 2012 to $26 billion in 2019. The crucial component of smart homes is the distribution of information among devices (heating, lighting, and other sensors) and rooms, which is then employed to regulate the environment (Wilson, 2014).

Fuzzy systems are being used more than ever before.

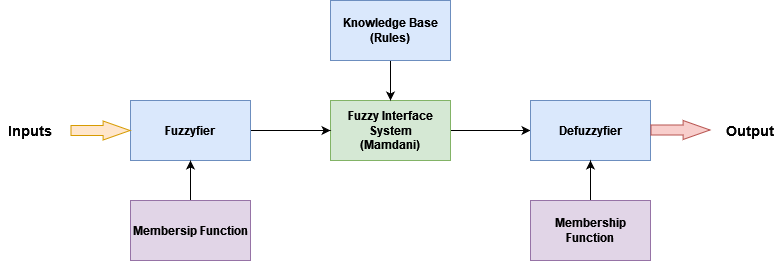
Because real-world manifestations are intrinsically hazy and ambiguous, fuzzy set theory was developed to model them.

The system's main feature is knowledge presentation in the form of if-else statements (Leski, Michal Jezewski & Robert Czabanski, 2017). Fuzzy Logic controllers could be used to automate smart residences for the disabled. There are numerous types of fuzzy inference systems available, but we chose the Mamdani fuzzy system for our project.

# FUZZY LOGIC CONTROLLER DESIGN

The Fuzzy Logic Toolbox from MATLAB is used to implement the Fuzzy Logic Controller (FLC).

The FLC passes inputs like heart rate, temperature, smoke level, volume intensity, and so forth to the FIS (fuzzy inference system) for processing, from which we receive the desired output. Design for smart residencies for disabled is shown in the figure below.



*Figure 1 Fuzzy Design*

To determine the output or surface graph, the Input and Output Membership Functions are sent to the FIS.

Each rule generates a fuzzy set as its output, which is derived from the output function and FIS. The fuzzy set we obtained as an output is given to the defuzzyfier in order to obtain the final crisp output (MathWorks).

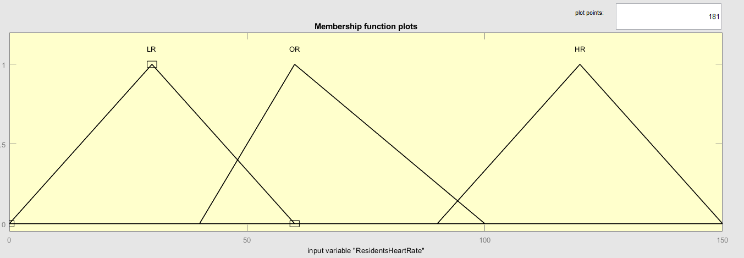
## Input and Membership Functions

The FLC which we have developed takes twelve inputs which are ResidentsHeartRate, ResidentsBloodPressureSystolic, ResidentsBloodPressureDiastolic, Time, Gestures, ResidentBodyTemperature, FlatAmbientTemp, SmokeAnalyzer, VolumeIntensity, PIRSensor, LDRSensor and WaterTemperature. A relevant Membership functions for each input is declared. The membership functions are of type triangle or trapezoidal. The section below describes each of the input variable.

### ResidentsHeartRate

We must constantly keep an eye on the resident's heart rate.

People may experience stress if their BPM is not optimal (BeatsPerMinute). This may also be the cause of numerous other issues. So to maintain an optimal level always is important. We have divided the ResidentsHeartRate variable into LR (Low Rate), OR (Optimal Rate), HR (High Rate).



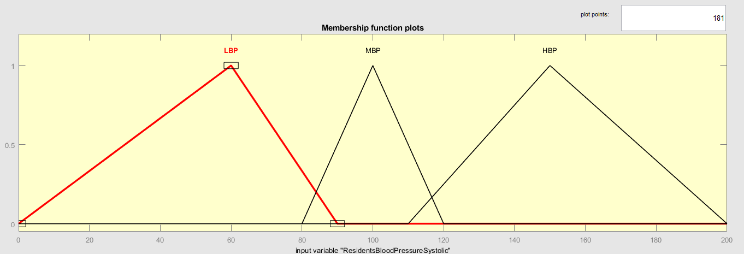
*Figure 2 ResidentsHeartRate Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (BPM)** | **Type** |
| LR | 0 - 60 | Triangle |
| OR | 40 - 100 | Triangle |
| HR | 90 - 150 | Triangle |

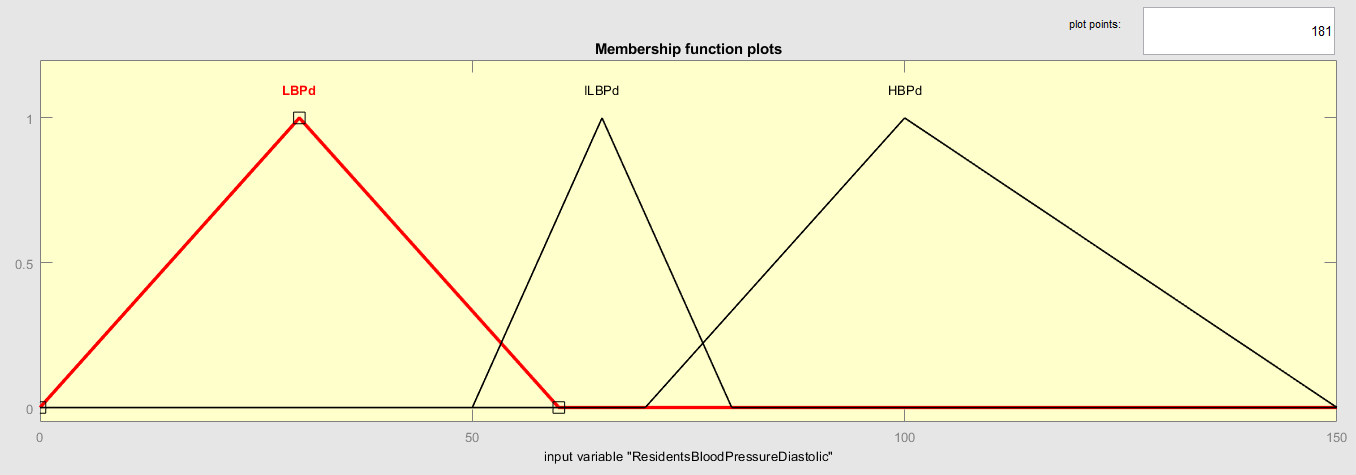
*Table 1 ResidentsHeartRate Membership Function*

### ResidentsBloodPressureSystolic & Diastolic

Blood pressure is the force with which our body's blood circulates. This facilitates the delivery of oxygen and nutrients to our tissues and organs. Monitoring blood pressure levels is therefore essential. ResidentsBloodPressureDiastolic is divided into LBPd (Low Blood Pressure diastolic), ILBPd (Ideal Level Blood Pressure diastolic), and HBPd. ResidentsBloodPressureSystolic is divided into LBP (Low Blood Pressure), MBP (Moderate Blood Pressure), and HBP (High Blood Pressure) (High Blood Pressure diastolic).



*Figure 3 ResidentsBloodPressureSystolic Membership Function*



*Figure 4 ResidentsBloodPressureDiastolic Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (mm Hg)** | **Type** |
| LBP | 0 - 90 | Triangle |
| MBP | 80 - 120 | Triangle |
| HBP | 110 - 200 | Triangle |

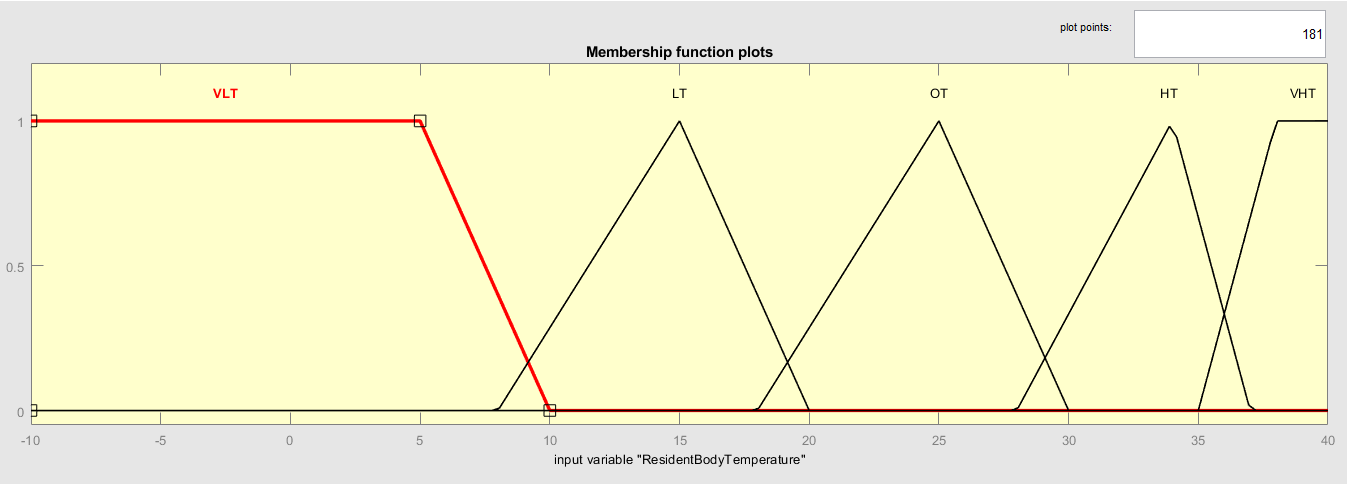
*Table 2 ResidentsBloodPressureSystolic Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (mm Hg)** | **Type** |
| LBPd | 0 - 60 | Triangle |
| ILBPd | 50 - 80 | Triangle |
| HBPd | 70 - 150 | Triangle |

*Table 3 ResidentsBloodPressureDiastolic Membership Function*

### ResidentBodyTemperature

Individuals suffering from illness may have varying body temperature all time. Low and extreme body temperature may also harm the body is divided into VLT (Very Low Temperature), LT (Low Temperature), NT (Normal Temperature), HT (High Temperature), VHT (Very High Temperature)



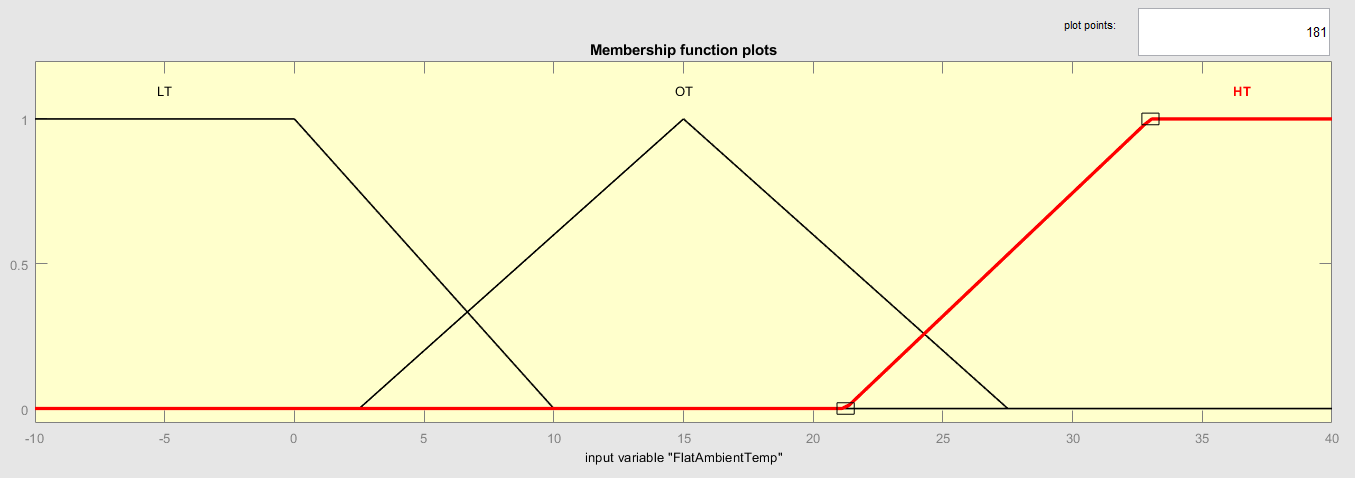
*Figure 5 ResidentBodyTemperature Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (Celsius)** | **Type** |
| VLT | -10 - 10 | Trapezoid |
| LT | 8 - 20 | Triangle |
| NT | 18 - 35 | Triangle |
| HT | 28 - 37 | Triangle |
| VHT | 35-40 | Trapezoid |

*Table 4 ResidentBodyTemperature Membership Function*

### FlatAmbientTemp

Ambient Temperature also plays a vital role for the disabled ones. Sometimes if they are exposed to low and high temperature, they have a probability of getting sick. The Temperature could be controlled using Temperature Controller. FlatAmbientTemp is divided into LT (Low Temperature), NT (Normal Temperature), HT (High Temperature).

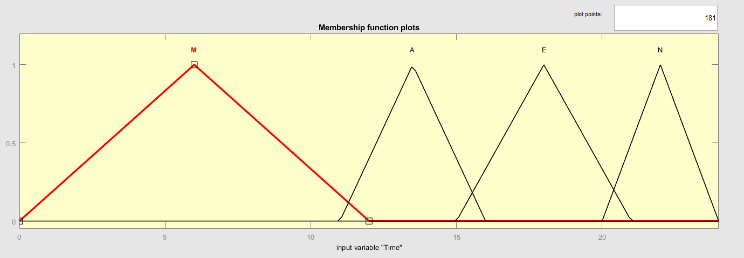


*Figure 6 FlatAmbientTemp Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (Celsius)** | **Type** |
| LT | -10 - 10 | Trapezoid |
| NT | 2.5 – 27.5 | Triangle |
| HT | 21.25 - 40 | Trapezoid |

### Time

Time input was introduced so other sensors could work in a timely manner. A 24hr format is used in the membership function and same. Time is divided into M (Morning), A (Afternoon), E (Evening), N (Night).



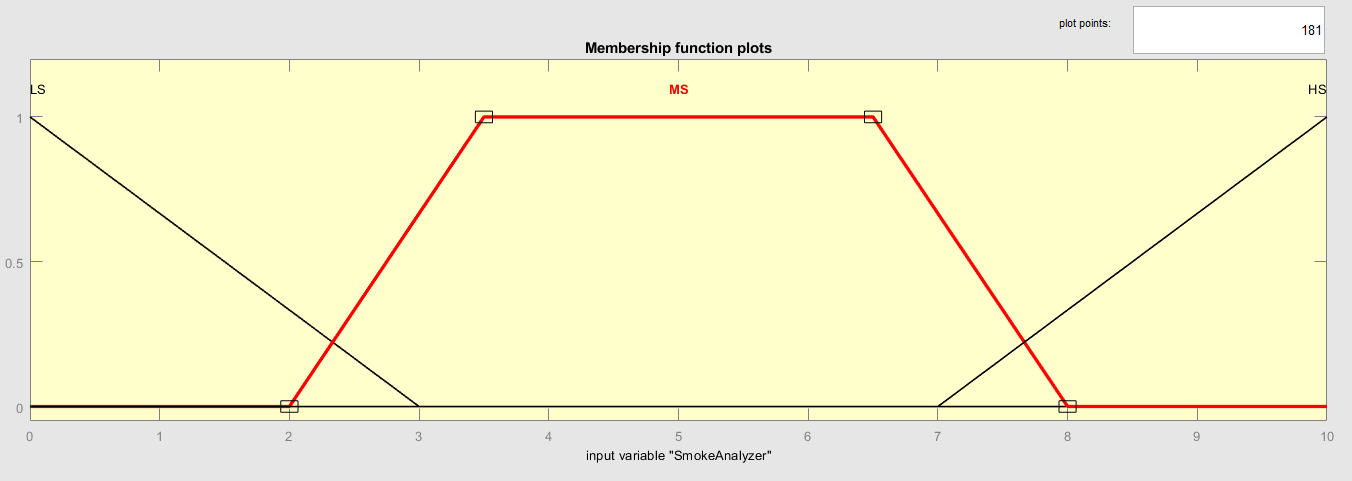
*Figure 7 Time Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (24hr)** | **Type** |
| M | 0 - 12 | Triangle |
| A | 11 - 16 | Triangle |
| E | 15 - 21 | Triangle |
| N | 20 - 24 | Triangle |

*Table 5 Time Membership Function*

### SmokeAnalyzer

This is used to identify if there is any fire breakout in the flat. It can be used to identify the intensity of the fire and do the needful. SmokeAnalyzer is divided into LS (Low Smoke), MS (Moderate Smoke), and HS (High Smoke).



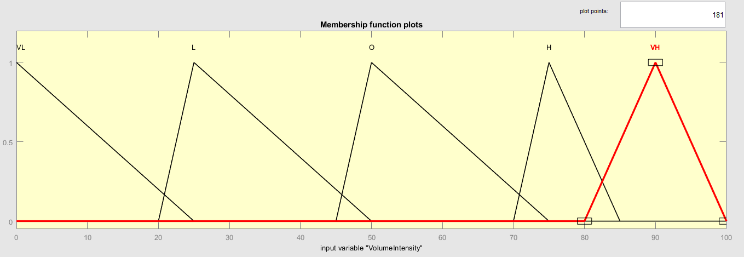
*Figure 8 SmokeAnalyzer Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| LS | 0 - 3 | Triangle |
| MS | 2 – 8 | Trapezoid |
| HS | 7 - 10 | Triangle |

*Table 6 SmokeAnalyzer Membership Function*

### VolumeIntensity

This describes the volume level in its surroundings from the appliances in the flat or room. VolumeIntensity is divided into VL (Very Low), L (Low), O (Optimum), H (High), and VH (Very High).



*Figure 9 VolumeIntensity Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| VL | 0 - 25 | Triangle |
| L | 20 - 50 | Triangle |
| O | 45 - 75 | Triangle |
| H | 70 - 85 | Triangle |
| VH | 80 - 100 | Triangle |

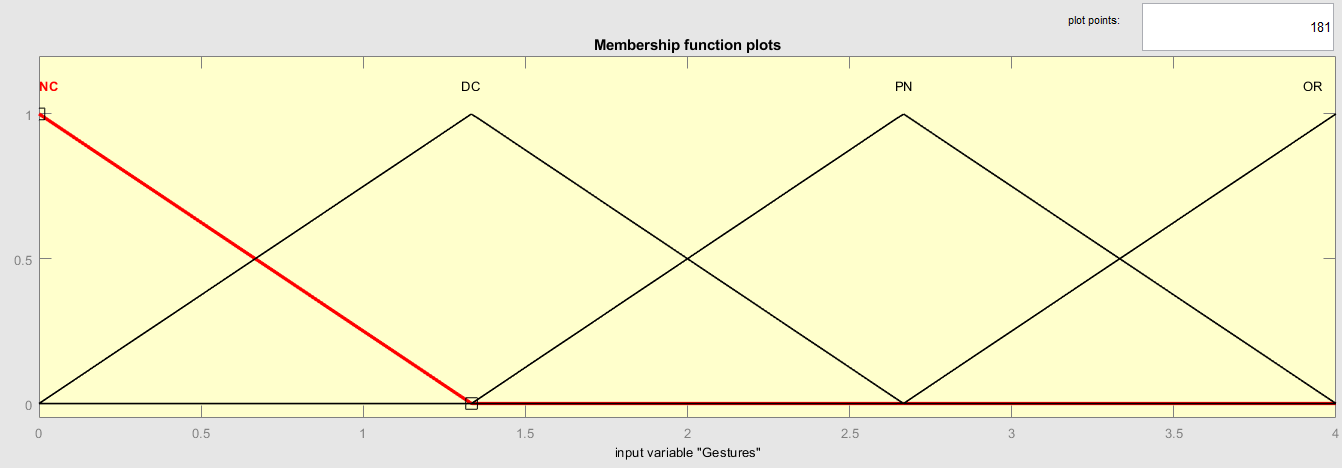
*Table 7 VolumeIntensity Membership Function*

### Gestures

Gestures are used for individuals for those who are differently abled. They can use this to communicate in case of emergencies. Gestures are divided into NA (Nurse Attention), DA (Doctor Attention), PA (Personal Attention), and OA (Other Attention).

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| NA | 0 - 1 | Triangle |
| DA | 11 - 16 | Triangle |
| PA | 15 - 21 | Triangle |
| OA | 20 - 24 | Triangle |

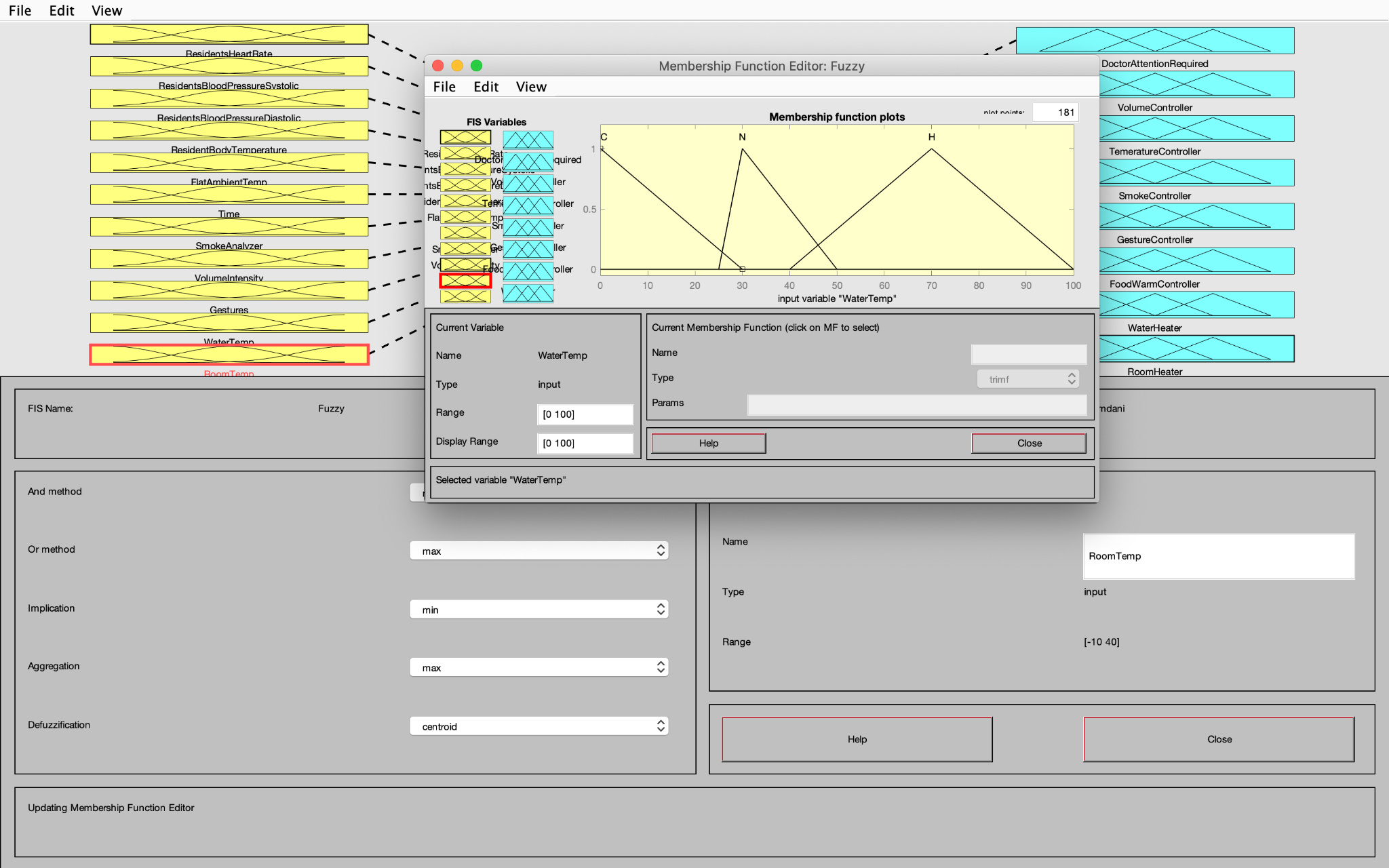
*Table 8 Gestures Membership Function*



*Figure 10 Gestures Membership Function*

### WaterTemperature

Water temperature is used to measure the current water temperature. It can be used to identify the water heater temperature and adjust the temperature of the water heater accordingly. Water temperature is divided into C (Cold water), N (Normal water) and H (Hot water)



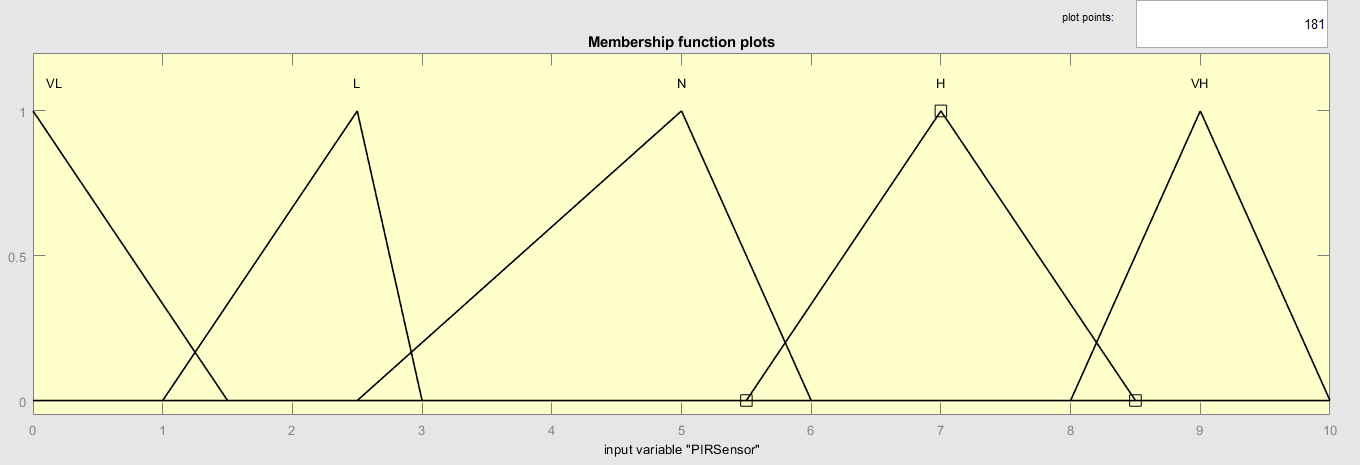
*Figure 11 WaterTemperature Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| C | 0 - 30 | Triangle |
| N | 25 – 50 | Trapezoid |
| H | 40 - 100 | Triangle |

*Table 9 WaterTemperature Membership Function*

### PIRSensor

An electronic sensor that measures the infrared (IR) light emitted by objects in its field of vision is known as a passive infrared sensor (PIR sensor). Most frequently, they are utilized in PIR-based motion detectors. PIR sensors are frequently utilized in autonomous lighting and security alarm systems. Linguistic variables for PIRSensor VL (Very Low), L (Low), M (Medium), H (High), and VH (Very High).



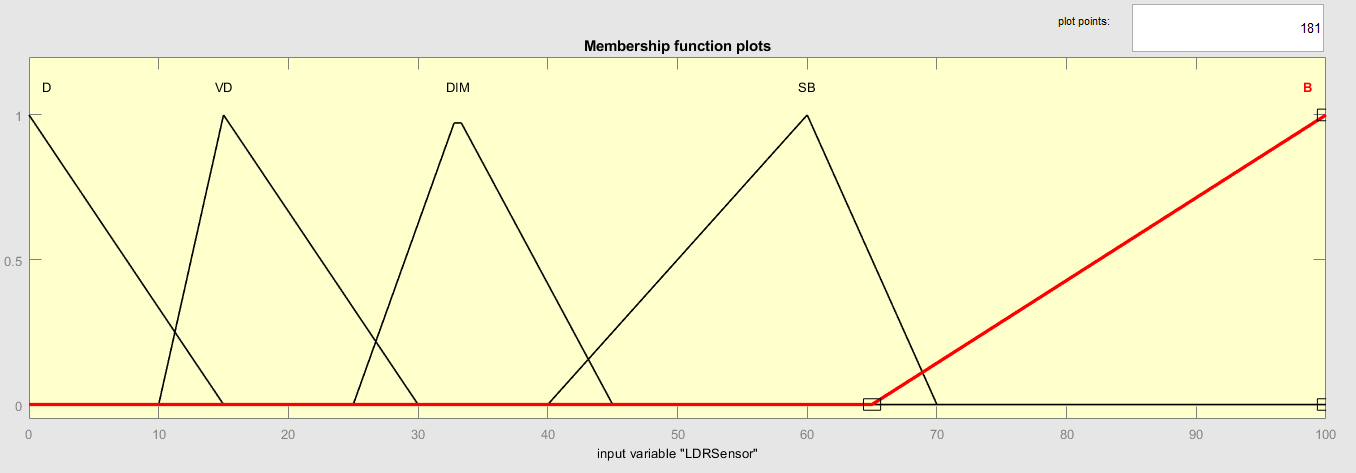
*Figure 12 PIRSensor Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| VL | 0 – 1.5 | Triangle |
| L | 1 - 3 | Triangle |
| M | 2.5 - 6 | Triangle |
| H | 5.5 – 8.5 | Triangle |
| VH | 8 - 10 | Triangle |

*Table 10 PIRSensor Membership Function*

### LDRSensor

LDR are utilised to detect the presence and absence of light. Alarm clocks, lights, and light intensity meters are devices that use LDR as light sensors. Linguistic variables for PIRSensor D (Dark), VD (Very Dark), DIM, SB (somewhat bright), and B (Bright).

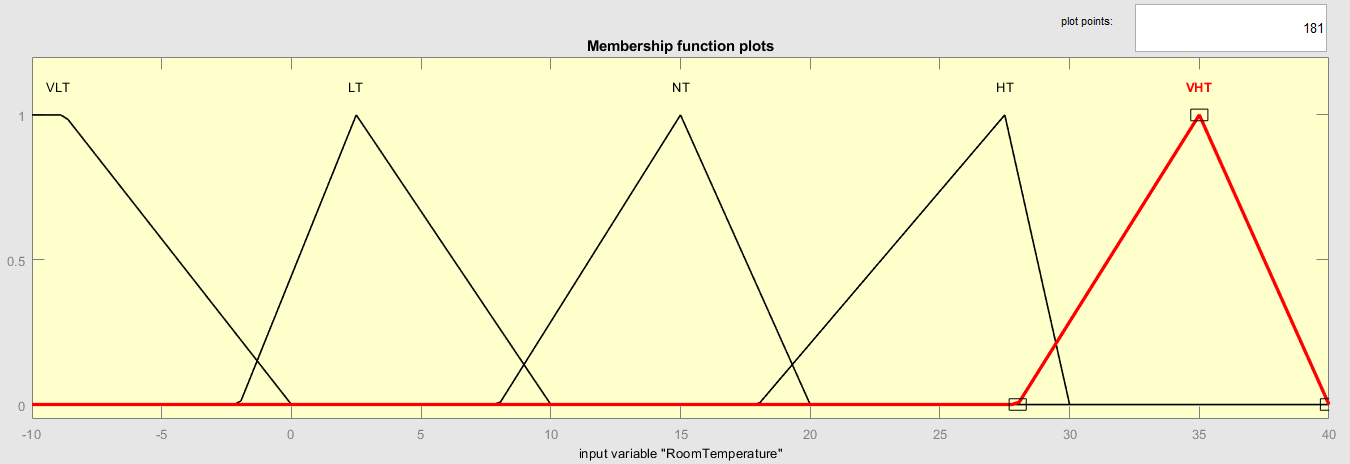
   
 *Figure 13 LDRSensor Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| D | 0 - 15 | Triangle |
| VD | 10 - 30 | Triangle |
| DIM | 25 - 45 | Triangle |
| SB | 40 - 70 | Triangle |
| B | 65 - 100 | Triangle |

*Table 11 LDRSensor Membership Function*

### RoomTemperature

The air temps most people prefer for indoor settings are referred to as room temperature. It can be used to adjust the room temperature accordingly. Room temperature is divided into VLT (Very Low Temperature), LT (Low Temperature), NT (Normal Temperature), HT (High Temperature), and VHT (Very High Temperature).



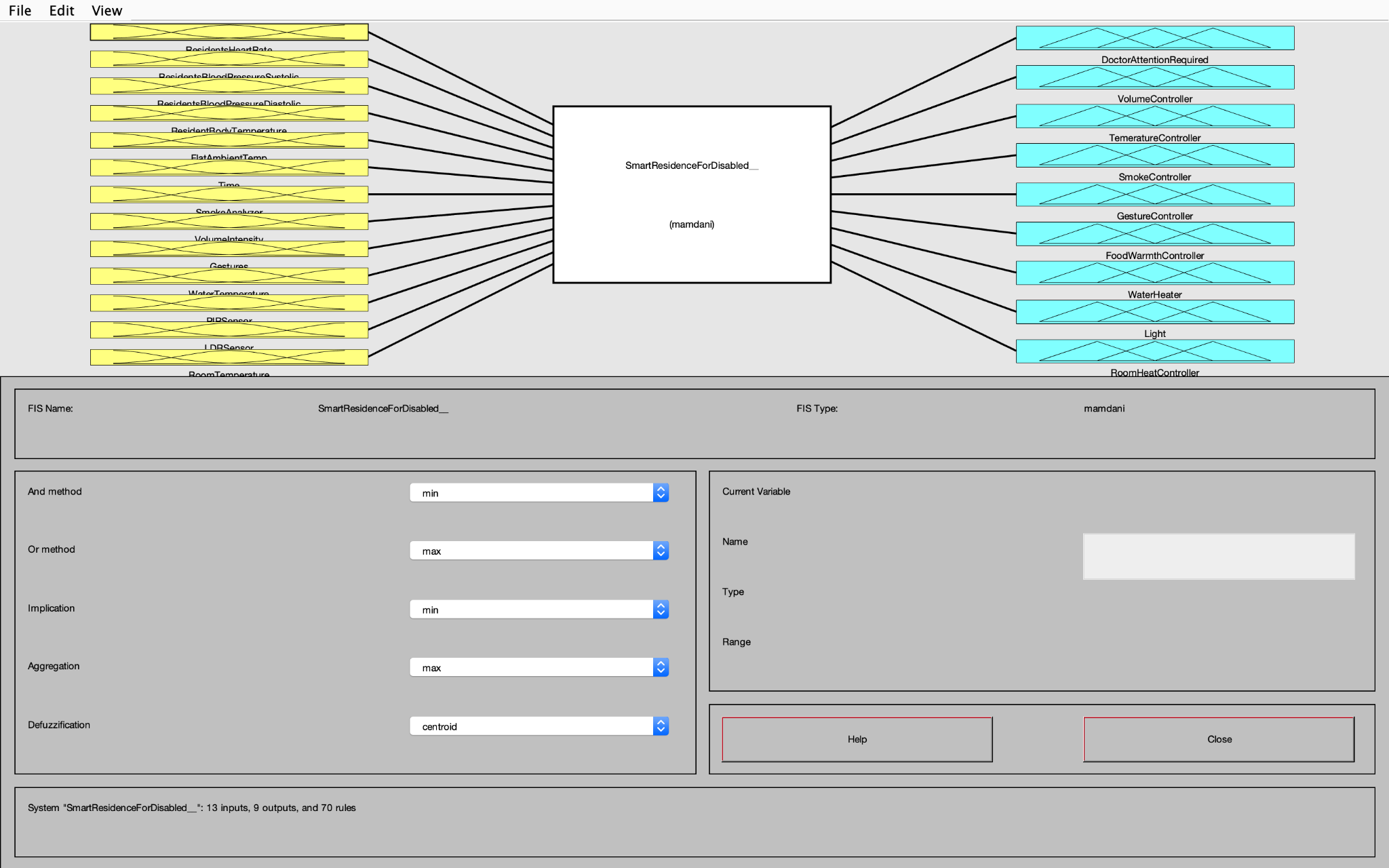
*Figure 14 RoomTemperature Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| VL | -10 - -5 | Triangle |
| L | -5 - 14 | Triangle |
| O | 15 - 26 | Triangle |
| H | 26 - 32 | Triangle |
| VH | 32 - 40 | Triangle |

*Table 12 RoomTemperature Membership Function*

## Fuzzy Inference System

The inputs are connected with the outputs using the fuzzy sets available in the Fuzzy Inference System. Mainly there are two types of inference system which are Mamdani and Sugeno Fuzzy inference system. For this project as mentioned earlier we have implemented the Mamdani FIS. This system was proposed in 1975 by Ebhasim Mamdani. Essentially, it was planned to control a steam engine and boiler combination by synthesizing a set of fuzzy rules obtained from system developers (tutoialspoint, n.d.). Mamdani systems are well suited to expert system applications where the rules are created from human expert knowledge, such as medical diagnostics, because their rule bases are more logical and simpler to understand (E.H.Mamdani & S.Assilian, 1975).



*Figure 15 Mamdani FIS*

The main steps involved in FIS are:

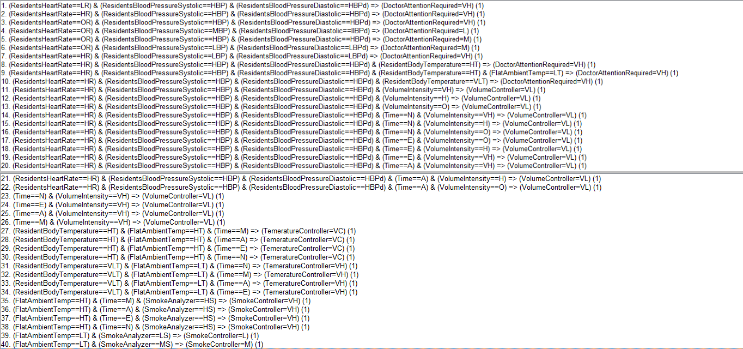
### Fuzzification

Fuzzification refers to the technique of converting clean input to fuzzy. The linguistic variable is another term for the fuzzy variable. For fuzziness to be expressed, a membership function is necessary.

### Rules

To obtain the desired output after passing the input to the FIS, we use the help of Fuzzy Rules, these are also known as linguistic representation. For our Fuzzy logic Controller to work with maximum performance these rules should we return by someone who have good domain knowledge. The system makes its decisions on the basis of rules which we have set in the knowledge base. The rules are basically if-then condition with OR and AND operator. For example if ResidentsHeartRate and ResidentsBloodPressure is high they may be a need to call the doctor for further assistance.

We have created a total of 68 rules and some of them are shown in the figure below.



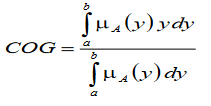
*Figure 14 Fuzzy Rules (snippet)*

### Aggregation

This is the process where we combine all the rules output into a single output. All the rules declared will be analyzed and reviewed to get this final output.

### Defuzzification

Defuzzification is the method in which we will convert the fuzzy input into crisp output. While we developed the Fuzzy Inference System we have used the centroid method for this purpose. The centroid method works on the basis average value obtained from the membership function. The Centre of Area (COA) is calculated using the equation shown below.



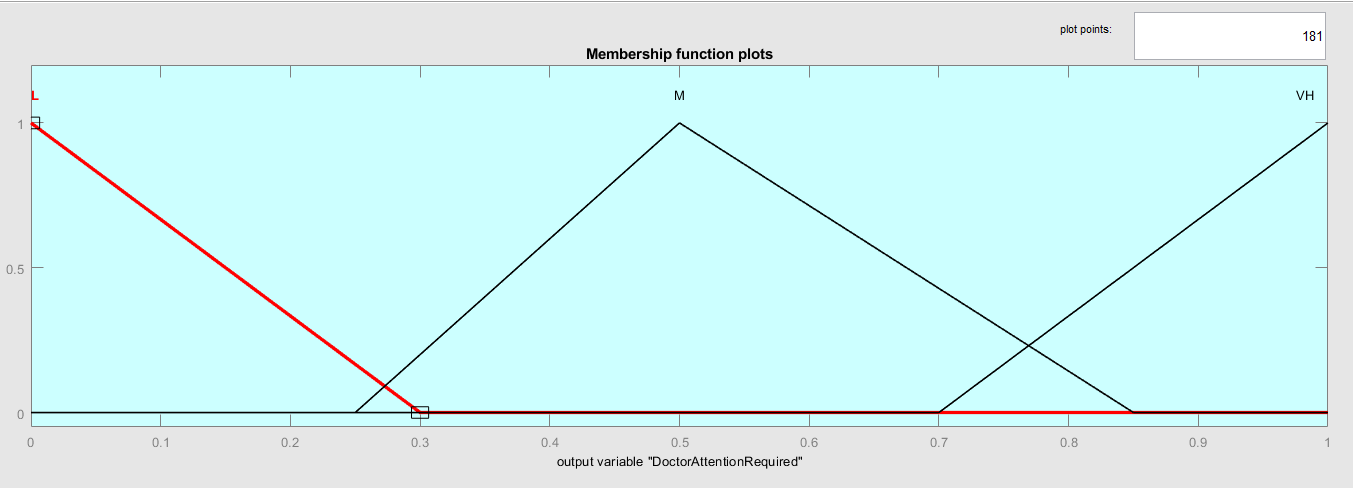
*Equation 1 COA Equation*

## Output and Membership Function

The FLC which we have taken eight outputs which are Doctor Attention Required, Volume Controller, Temerature Controller, Smoke Controller, Gesture Controller, Food Warmth Controller, Light Controller, RoomHeat Controller and WaterHeater. A relevant Membership function for each output is declared. The membership functions are of type triangle or trapezoidal. The section below describes each of the output variables

### DoctorAttentionRequired

Doctor attention means the level of medical attention the person needs. This output depends on the Resident Heart Rate, Blood Pressure and Body Temperature. *DoctorAttentionRequired* is divided into L (Low), M (Medium), and VH (Very High).



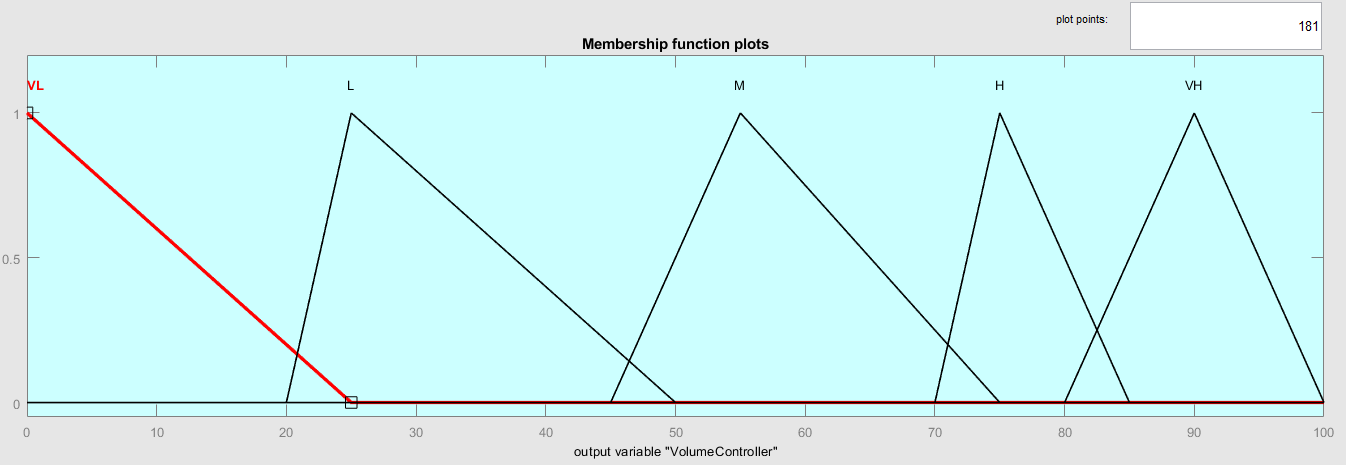
*Figure 16 DoctorAttentionRequired Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| L | 0 – 0.3 | Triangle |
| M | 0.25 – 0.85 | Triangle |
| VH | 0.7 - 1 | Triangle |

*Table 13 DoctorAttentionRequired Membership Function*

### VolumeController

To control the audio system sound volume controller is used. It depends on the volume intensity input. Volume Controller is divided into VL (Very Low), L (Low), M (Medium), H (High), and VH (Very High).



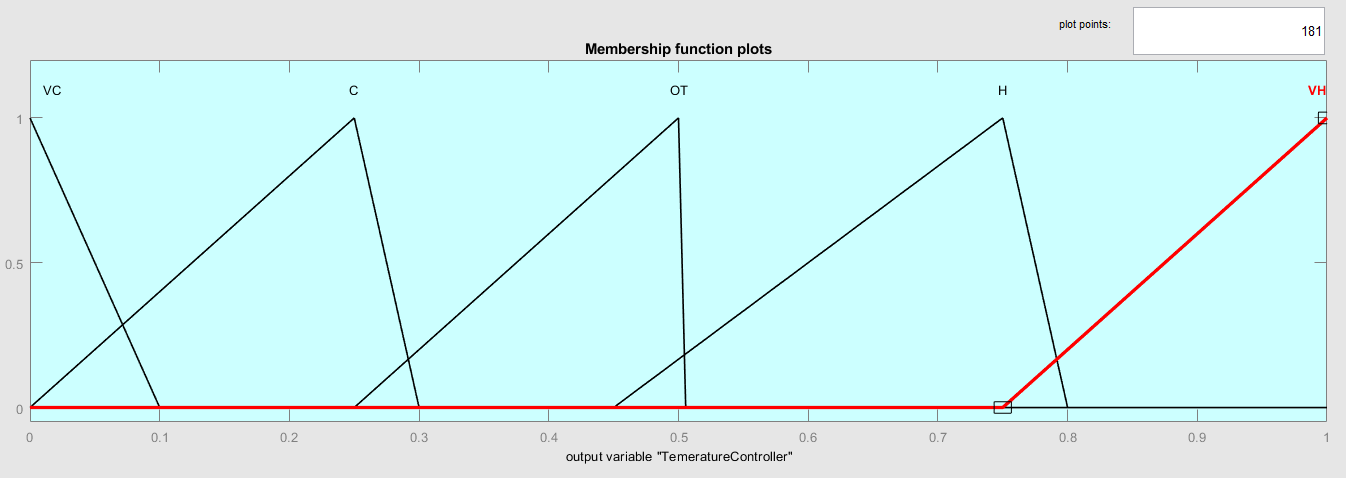
*Figure 17 VolumeController Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| VL | 0 - 25 | Triangle |
| L | 20 - 50 | Triangle |
| O | 45 - 75 | Triangle |
| H | 70 - 85 | Triangle |
| VH | 80 - 100 | Triangle |

*Table 14 VolumeController Membership Function*

### TemperatureController

This controller can be used to control the main cooling and heating system installed in the flat. Temperature Controller is divided into VC (Very Cold), C (Cold), OT (Optimum Temperature), H (Hot), VH (Very Hot).



*Figure 18 TemperatureController Membership Function*

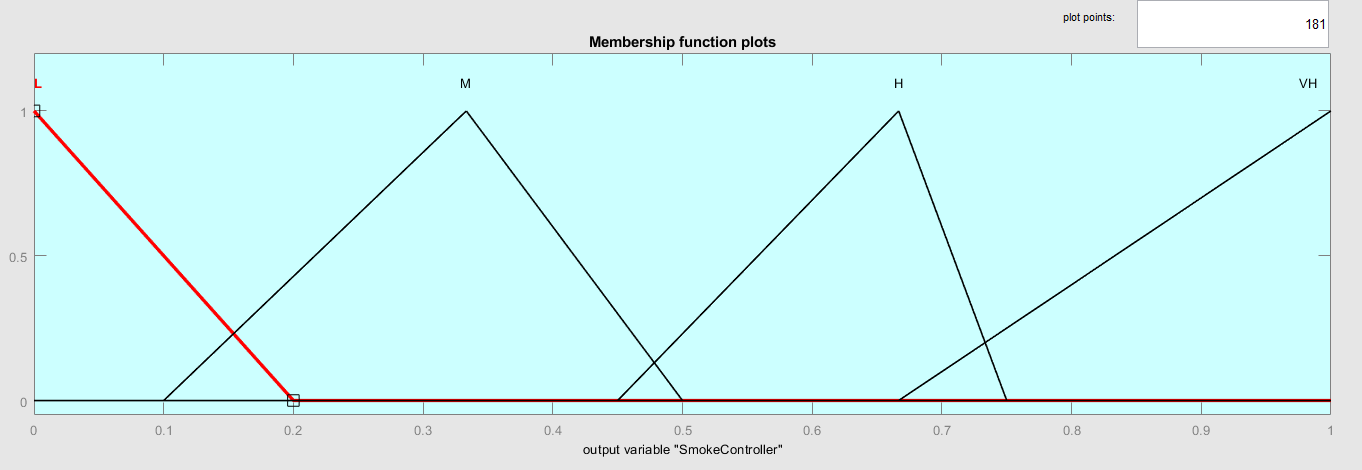
|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range (Celsius)** | **Type** |
| VC | 0 – 0.1 | Triangle |
| C | 0 – 0.3 | Triangle |
| OT | 0.25 – 0.5 | Triangle |
| H | 0.45 – 0.8 | Triangle |
| VH | 0.75 – 1 | Triangle |

*Table 15 TemperatureController Membership Function*

### SmokeController

To control the smoke inside the flat smoke controller is used. It can sense minute airborne particles to identify fires.

it depends on the SmokeAnalyzer input. The Smoke Controller is divided into L (Low), M (Medium), H (High), and VH (Very High).



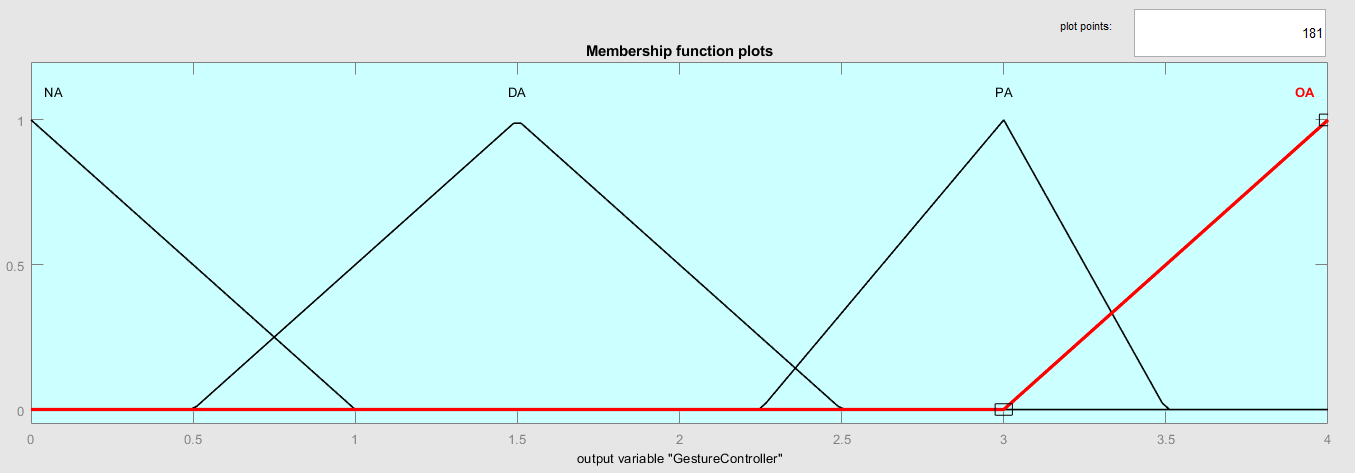
*Figure 19 SmokeController Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| L | 0 – 0.33 | Triangle |
| M | 0.1 – 0.5 | Triangle |
| H | 0.45 – 0.75 | Triangle |
| VH | 0.7 – 1 | Triangle |

*Table 16 SmokeController Membership Function*

### GestureController

This is used to identify what gesture is the user intending. The gestures are categorized as NA (Nurse Assistance), DA (Doctor Assistance), PA (Personal Assistance), and OA (Other Assistance)



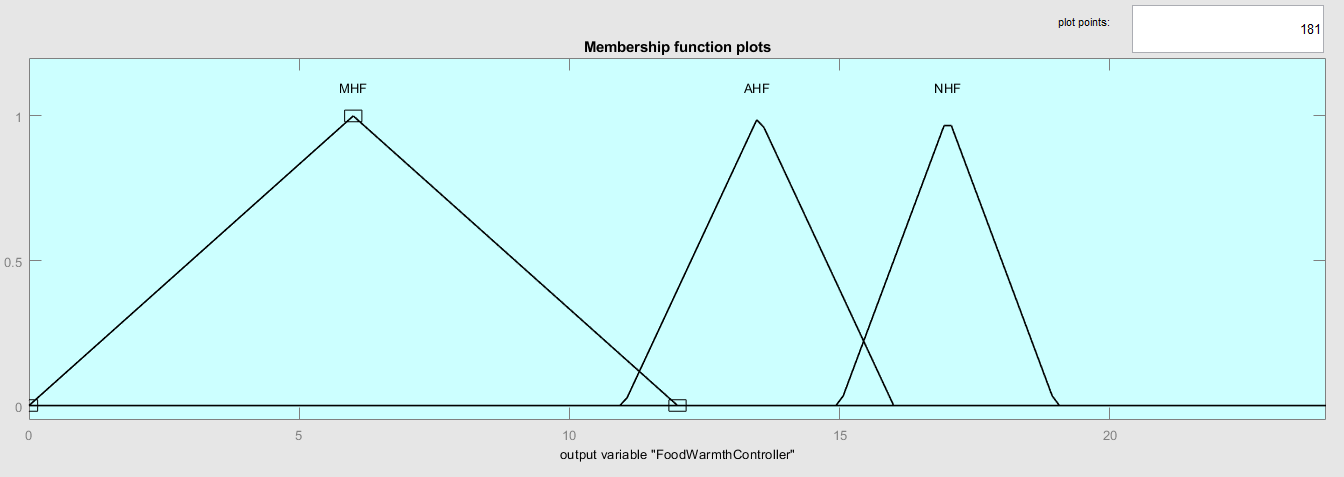
*Figure 20 GestureController Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| NA | 0 – 1 | Triangle |
| DA | 0.5 – 2.5 | Triangle |
| PA | 2.25 – 3.5 | Triangle |
| OA | 3 – 4 | Triangle |

*Table 17 GestureController Membership Function*

### FoodWarmthController

This controller can be used to periodically heat food in specific timing. The FoodWarmthController is divided into MHF (Morning Heat Food), AHF (Afternoon Heat Food), NHF (Night Heat Food).



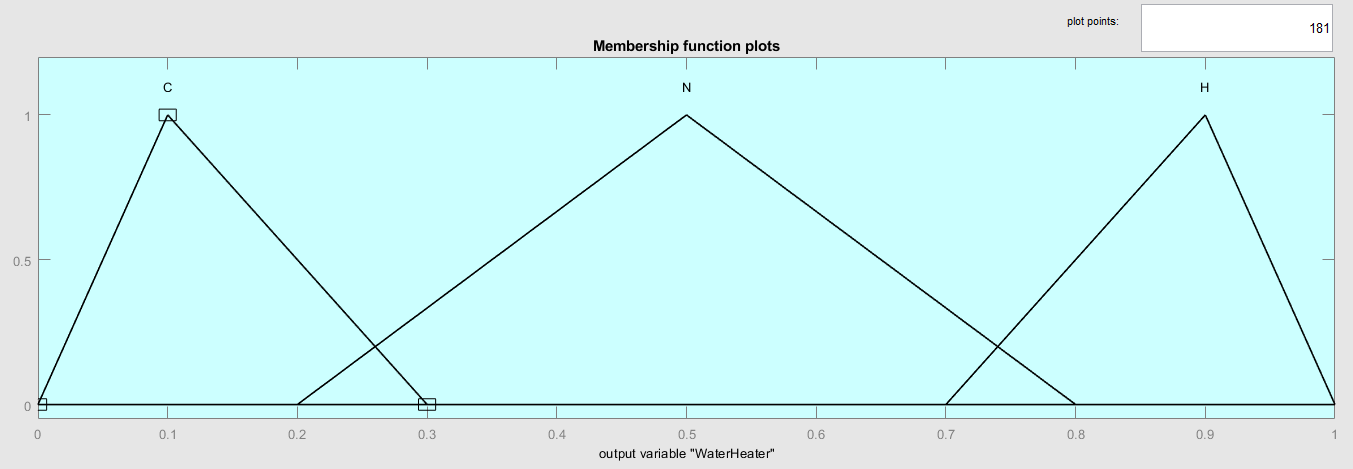
*Figure 21 FoodWarmthController Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range(Time in 24hr)** | **Type** |
| MHF | 0 – 12 | Triangle |
| AHF | 11 – 16 | Triangle |
| NHF | 15 – 19 | Triangle |

*Table 18 FoodWarmthController Membership Function*

### WaterHeater

Water heater controller can be used to control the level of temperature in water. This can be handy as different residents require different level of water temperature. WaterHeater is divided into C (Cold), Normal (Normal) and H (Heat).



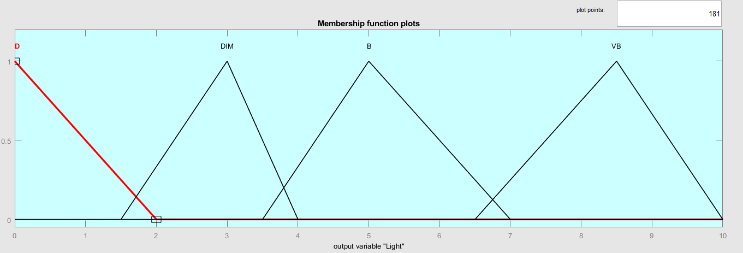
*Figure 22 WaterHeater Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| C | 0 – 0.3 | Triangle |
| N | 0.2 – 0.8 | Triangle |
| H | 0.7 – 1 | Triangle |

*Table 19 WaterHeater Membership Function*

### Light

A modern method of illuminating the home is with smart lights. LDR and PIR sensor input are used to adjust the smart light. Light is divided into D (Dark), DIM, B (Bright) and VB (very bright).



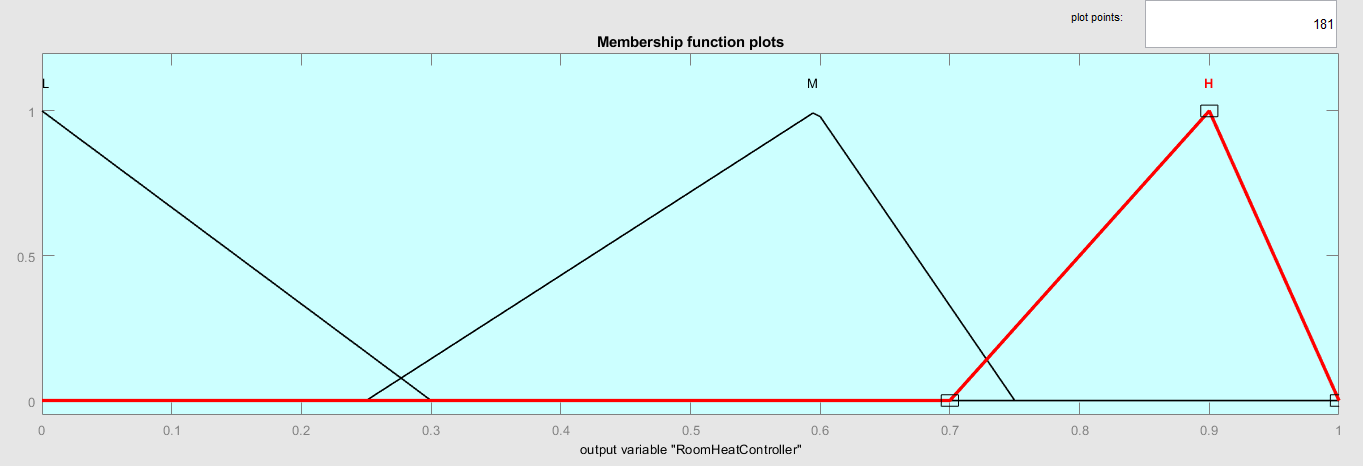
*Table 23 Light Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| D | 0 - 2 | Triangle |
| DIM | 1.5 - 4 | Triangle |
| B | 3.5 - 7 | Triangle |
| VB | 6.5 - 10 | Triangle |

*Table 20 Light Membership Function*

### RoomHeatController

This controller can be used to control the level of heat in individual’s room. RoomHeatController is divided into L (Low), M (Medium) and H (Heat)



*Figure 24 RoomHeatController Membership Function*

|  |  |  |
| --- | --- | --- |
| **Membership Function** | **Range** | **Type** |
| L | 0 – 0.3 | Triangle |
| M | 0.25 – 0.75 | Triangle |
| VB | 0.7 - 1 | Triangle |

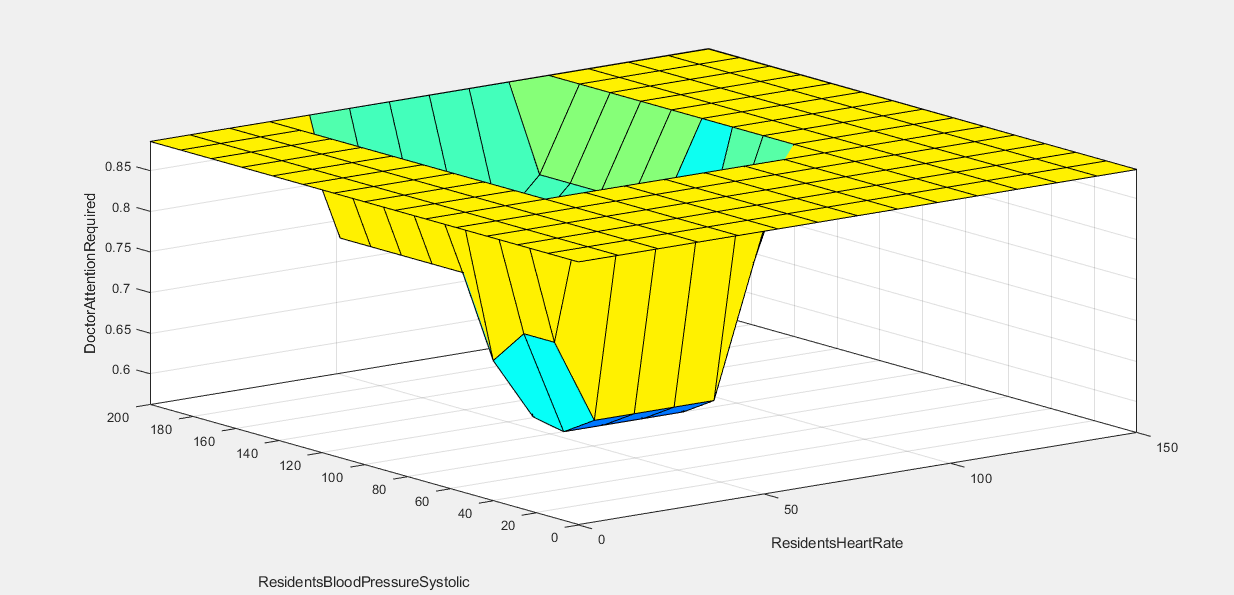
*Table 21 RoomHeatController Membership Function*

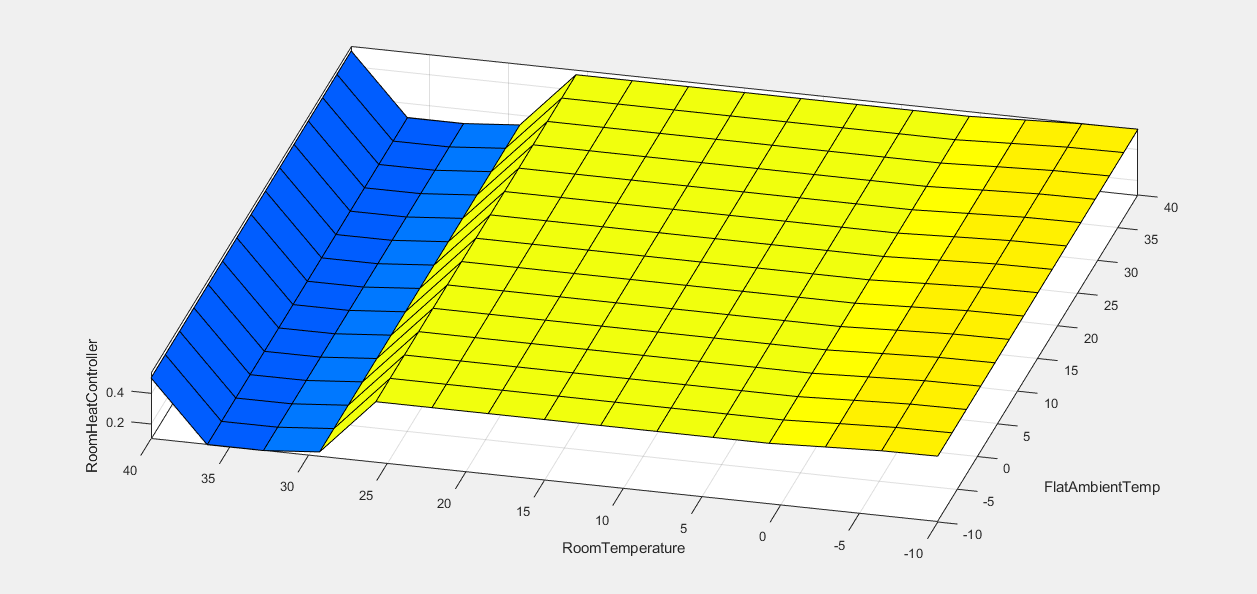
## Justification For the FLC Designed

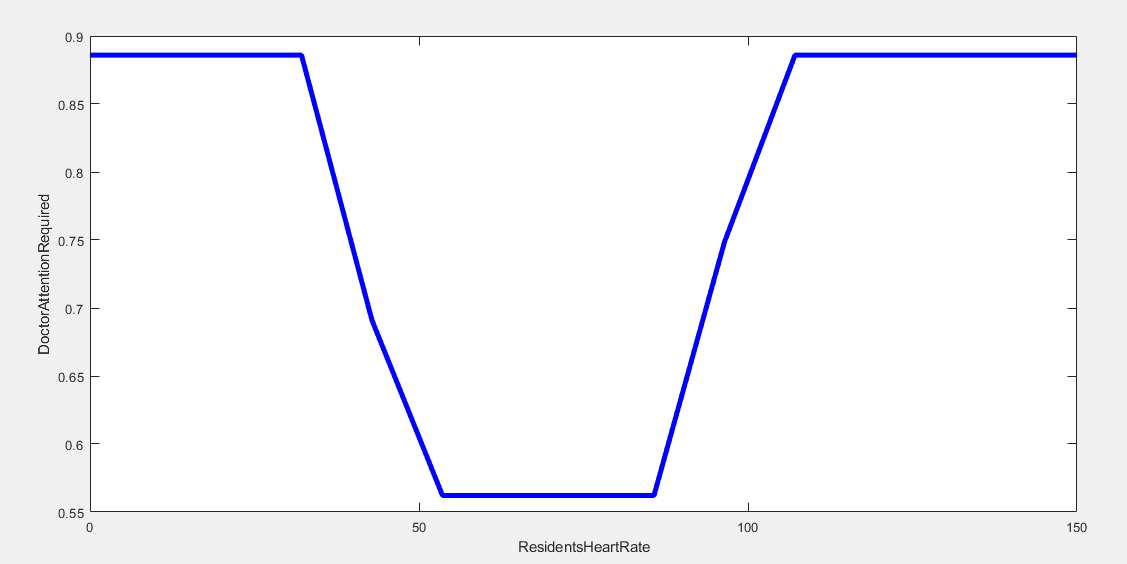
The inputs which we take in are the residents Heart Rate, Blood Pressure, Temperature, Volume Intensity etc. We have also have created membership function which we suit the variable. For example the Residents Heart Rate is ranged between 0-150 with 3 classes and its output functions are also assigned accordingly that is for the heart rate variable directly contribute to the output variable on whether we should call the doctor. Similarly all the input and output variables are designed.

## Results

The input versus output surface is shown below.



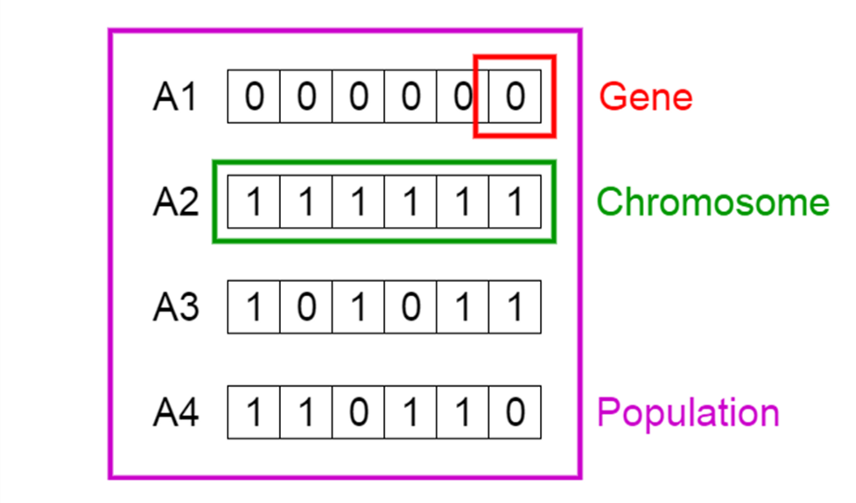




# OPTIMIZATION USING GENETIC ALGORITHM

We have used the help of Genetic Algorithm to optimize the Fuzzy design. A genetic algorithm is a search heuristic inspired by Charles Darwin's theory of natural evolution.

This algorithm reflects the process of natural selection, in which the fittest individuals are chosen for reproduction in order to produce offspring of the next generation. The steps involved with genetic algorithms is Initial Population, Fitness Function, Selection, Cross-over and Mutation.



*Figure 25 Genetic Algorithm (Source: towardsdatascience)*

To initiate Genetic Algorithm, first we need change the genes into a binary format. So these will be represented by an eight bit number (00000000). The genes presents a small unit of a chromosome and a collection of such a chromosome can be called a population. So we need to convert the membership function into binary value. For example we take the ResidentsHeartRate as input its three membership functions LR, OR, HR, as this are of the type triangle we will have 3 parameters and 4 if it was a trapezoids. These parameters represents the genes and our heart rate input has 3+3+3 = 9 chromosomes with it. So if we consider the whole input membership function we have 156 chromosomes. For example the heart rate input can be expressed as shown below.

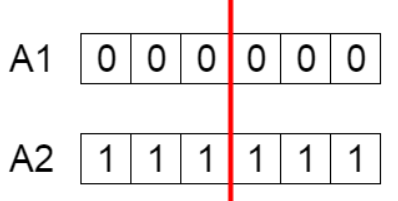
|  |  |  |
| --- | --- | --- |
| **ResidentsHeartRate** | | |
| LR | | |
| *0* | *30* | *60* |
| 0000 | 11110 | 111100 |
| OR | | |
| *40* | *60* | *100* |
| 101000 | 111100 | 1100100 |
| HR | | |
| *90* | *120* | *150* |
| 1011010 | 1111000 | 10010110 |

*Table 21 Binary Form of input Membership Function for ResidentsHeartRate*

Similarly we convert all the membership function into binary. The length of the chromosome is decided by the length of genes. So here the genes are of 8-bit. We also convert the output membership to binary.

Then we use the help of Fitness function to see who among the individual is best for the purpose of reproduction so we can have great offspring’s. The fitness function will provide a score for all the individuals on the basis of selection process. Some of the selection process examples are Rank selection, Roulette Wheel Selection, Stochastic Universal Sampling (SUS), Tournament Selection etc.

The next step is called crossover an important phase in genetic algorithms where we select a cross over point for each pair of parents which are to be mated. There are different types of crossover such as are One Point Crossover, Multi Point Crossover, Uniform Crossover, Whole Arithmetic Recombination, Davis’ Order Crossover.



*Figure 26 Crossover point*

The offspring’s are developed by exchanging the genes in individuals which we have selected. The new offspring is after crossover is shown in figure below.



*Figure 27 New Offspring*

The following step is called mutation, and it involves flipping the bits in the string. Mutations are the source of evolution.

They are a source of variation which increases an organism's resilience to its surroundings and hinders deterioration in environmental stress.

# References

E.H.Mamdani, & S.Assilian. (1975). An experiment in linguistic synthesis with a fuzzy logic controller. 1-13.

Galliakis, M., Skourlas, C., Galiotou, E., & Voyiatiz, I. (2018). A low-cost smart home for the assistance of elderly persons and patients. *ACM*, 93-98.

Leski, J., Michal Jezewski , & Robert Czabanski. (2017). Introduction to Fuzzy Systems. In J. Leski, M. J., & R. C., *Theory and Applications of Ordered Fuzzy Numbers* (pp. 23-43). Springer.

MathWorks. (n.d.). *mathworks.* Retrieved from mathworks: https://in.mathworks.com/help/fuzzy/types-of-fuzzy-inference-systems.html

tutoialspoint. (n.d.). *tutoialspoint*. Retrieved from tutoialspoint: https://www.tutorialspoint.com/fuzzy\_logic/fuzzy\_logic\_inference\_system.htm#

Verma, N. (2021, March 01). *timesofinida.* Retrieved from https://timesofindia.indiatimes.com/readersblog/technical-mitra/how-to-technology-changes-human-life-30071/

Wilson, C. (2014). Smart homes and their users: A systematic analysis and key challenges. *Springer*, 463-476.

**APPENDIX**

Github Link : https://github.com/AleenaAlby/Modelling-and-Optimisation-Under-Uncertainty