

EXPLORATION OF THE HIV DRUG RESISTANCE LANDSCAPE USING MACHINE LEARNING AND BIG DATA

Submitted in partial fulfillment of the requirements for the award of

Masters of Science degree in BIOINFORMATICS

By

BAVYAKEERTI. M (REG NO:40739002)



DEPARTMENT OF BIOINFORMATICS

SCHOOL OF BIO AND CHEMICAL ENGINEERING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL – 2022



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **BAVYAKEERTI. M** (Reg. No. 40739002) who carried out the project entitled "**EXPLORATION OF THE HIV DRUG RESISTANCE LANDSCAPE USING MACHINE LEARNING AND BIG DATA**" under my supervision from November 2021 to April 2022.

Internal Guide

Dr. DANIEL ALEX ANAND

Head of the Department

Dr. JEMMY CHRISTY

Submitted for Viva Voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I, BAVYAKEERTI (Reg No: 40739002), hereby declare that the Project Report entitled **“EXPLORATION OF THE HIV DRUG RESISTANCE LANDSCAPE USING MACHINE LEARNING AND BIG DATA”** done by me under the guidance of **Dr. DANIEL ALEX ANAND** is submitted in partial fulfilment of the requirements for the award of Master of science degree in Bioinformatics.

DATE:

SIGNATURE OF THE CANDIDATE

PLACE: CHENNAI

ACKNOWLEDGEMENT

I am pleased to acknowledge our sincere thanks to Board of management of **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. -----**, **Dean, School of Bioinformatics and Data Science** and **Dr. JEMMY CHRISTY MSC., MTECH., (PhD)**, Head of the Department and **Dr. DANIEL ALEX ANAND PhD., Dept. of Bioinformatics and Data Science** for providing me necessary support and details at the right time during the progressive reviews.

We would like to express our sincere and deep sense of gratitude to our Project Guide **Dr. DANIEL ALEX ANAND PhD., Dept of Bioinformatics** for his valuable guidance, suggestions and constant encouraged paved way for the successful completion of my Project work.

I wish to express my thanks to all the teaching and non-teaching staff members of the **Department of bioinformatics and Data science** who were helpful in many ways for the completion of project.

ABSTRACT

Human immunodeficiency virus (HIV) is one of the fastest developing pathogens known. The human immunodeficiency virus (HIV)/AIDS pandemic is one of the most important challenges facing humanity. More than 36 million people are currently living with HIV. The prediction of HIV resistance to a drug or a combination of drugs is an important issue for the development of new potent and safe antiretroviral drugs. The recommended treatment for HIV is a combination of three antiretroviral (ARV) drugs from two or more different drug groups and is known as highly active antiretroviral therapy (HAART). The Stanford HIV Drug Resistance Database (HIVDB) is an essential resource for public health officials monitoring ADR and TDR, for scientists developing new ARV drugs, and for HIV care providers managing patients with HIVDR. Developing a computational prediction of drug resistance phenotype will enable efficient and timely selection of the best treatment regimens. Machine learning methods are widely and successfully used in new drug discovery. Using machine learning approach that takes advantage of the hierarchical nature in the relationships among the sequence, structure, and function, an integrative analysis of these trajectories reveals key details of the resistance mechanism, including changes in the protein structure, hydrogen bonding, and protein-ligand contacts. Machine learning algorithms can learn from electronic registry data and help to predict HIV status with a fairly high level of accuracy. Machine learning algorithms are being implicated here to measure the mutation over genes of the affected person. Machine learning approach, like by using tool it's easier to interpret drug resistance dataset and predicting resistance scores of drugs and find out the best drug with less resistibility which performs better in the treatment of HIV. Mutation selection here plays a major role to categorize the inputs of the model and helps to calculate the drug resistance for each mutation in the model.

Keywords: HIV-AIDS, HIV drugs, Antiretroviral therapy, Drug resistance, Machine learning techniques.

INDEX

S.NO	TITLE	PAGE NO
1	CHAPTER 1- INTRODUCTION	
	1.1. OUTLINE OF THE PROJECT	
	1.1.1. HIV AIDS	
	1.1.2. HIV-DRUGS	
	1.1.3. DRUG RESISTANCE	
	1.1.4 HIV DRUG RESISTANCE PREDICTION	
	1.2. OUTSOURCE	
	1.2.1. TOOLS AND SOFTWARE USED TO PREDICT	
	1.2.2. HIV DRUG RESISTANCE STANFORD DATASET	
	1.2.3. MACHINE LEARNING APPROACH IN PREDICTION	
2	CHAPTER 2- LITERATURE SURVEY	
	2.1. REVIEW-LITERATURE WORK	
	2.1.1. WHAT IS DRUG RESISTANCE IN GENERAL?	
	2.1.2. HOW DOES A DRUG GET RESISTANT PERIODICALLY?	
	2.1.3. STUDY DRUG RESISTANCE DATABASE ONLINE	
	2.1.4. WHAT MACHINE LEARNING APPROACH IS USED TO PREDICT DRUG RESISTANCE?	
3	CHAPTER 3- AIM AND OBJECTIVES	

4	CHAPTER 4- METHODOLOGY
	4.1. HIV DRUG RESISTANCE DATABASES USED
	4.1.1. STANFORD HIV DRUG RESISTANCE DATABASE (HIVDB)
	4.1.2. HIV DRUG RESISTANCE DATABASE
	4.1.3. WHO HIV DRUG RESISTANCE DATABASE
	4.2. MACHINE LEARNING APPROACH
	4.2.1. FUZZY LOGIC TOOLBOX
	4.2.1.1. TRAINING DATA
	4.2.1.2. GENERATE OR LOAD FIS STRUCTURE
	4.2.1.3. TRAIN FIS
	4.2.1.4. SURFACE VIEWER
	4.2.1.5. RULE VIEWER
5	CHAPTER 5- RESULT AND DISCUSSION
6	CHAPTER 6- SUMMARY AND CONCLUSION
	6.1. SUMMARY
	6.2. CONCLUSION

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE
2.1	Online database as reference	
4.1	Protease dataset	
4.2	NRTI dataset	
4.3	NNRTI dataset	
4.4	MATLAB command window	
4.5	'anfisedit' command opens ANFIS Designer window	
4.6	Loading Data from file	
4.7	Data loaded in Neuro Fuzzy Designer	
4.8	View of FIS generated toolbox	
4.9	Trained FIS	
4.10	Surface viewer of NFV	
4.11	Rule viewer of NFV	
5.1	Excel sheet of Protease	
5.2	Final excel sheet of protease	
5.3	Rule viewer of Dataset NFV	
5.4	Input Data for NFV	
5.5	Plot points for NFV	
5.6	rule viewer of Dataset SQV	
5.7	Input Data for SQV	
5.8	Plot points for SQV	
5.9	Rule viewer of Dataset IDV	
5.10	Input Data for IDV	
5.11	Plot point for IDV	
5.12	Rule viewer of Dataset FPV	
5.13	Input data for FPV	

5.14	Plot points for FPV
5.15	Rule viewer of Dataset ATV
5.16	Input data for ATV
5.17	Plot points for ATV
5.18	Rule viewer of Dataset LPV
5.19	Input Data for LPV
5.20	Plot points for LPV
5.21	Rule viewer of Dataset TPV
5.22	Input Data for TPV
5.23	Plot points for TPV
5.24	Rule viewer of Dataset DRV
5.25	Input Data for DRV
5.26	Plot points for DRV
5.27	Surface plot of NFV
5.28	Surface plot of SQV
5.29	Surface plot of IDV
5.30	Surface plot of FPV
5.31	Surface plot of ATV
5.32	Surface plot of LPV
5.33	Surface plot of TPV
5.34	Surface plot of DRV
5.35	Excel Sheet of NRTI
5.36	Rule viewer of Dataset AZT
5.37	Input data for AZT
5.38	Plot points for AZT
5.39	Rule viewer of Dataset TDF
5.40	Input data for TDF

5.41	Plot points for TDF
5.42	Rule viewer of Dataset ABC
5.43	Input Data for ABC
5.44	Plot points for ABC
5.45	Rule viewer of Dataset 3TC
5.46	Input Data for 3TC
5.47	Plot points for 3TC
5.48	Surface plot of AZT
5.49	Surface plot of TDF
5.50	Surface plot of ABC
5.51	Surface plot of 3TC
5.52	Excel sheet of NNRTI
5.53	Rule viewer of Dataset NVP
5.54	Input data for NVP
5.55	Plot points for NVP
5.56	Rule viewer of Dataset EFV
5.57	Input Data for EFV
5.58	Plot Points for EFV
5.59	Rule viewer of Dataset ETR
5.60	Input data for ETR
5.61	Plot points for ETR
5.62	Rule viewer of dataset RPV
5.63	Input data for RPV
5.64	Plot points for RPV
5.65	Rule viewer of Dataset DOR
5.66	Input data for DOR
5.67	Plot points for DOR

5.68	Surface Plot of NVP
5.69	Surface plot of EFV
5.70	Surface plot of ETR
5.71	Surface plot of RPV
5.72	Surface plot of DOR

ABBREVIATION

HIV	- Human Immunodeficiency Virus
AIDS	- Acquired Immunodeficiency Syndrome
WHO	- World Health Organization
ART	- Anti Retroviral Therapy
NRTI	- Nucleoside Reverse Transcriptase Inhibitor
NNRTI	- Non-Nucleoside Reverse Transcriptase Inhibitor
INSTI	- Integrase Strand Transfer Inhibitor
PI	- Protease Inhibitor
SHDRDB	- Stanford HIV Drug Resistance Database
ADR	- Acquired Drug Resistance
TDR	- Transmitted Drug Resistance
PDR	- Pre-treatment Drug Resistance
FIS	- Fuzzy Inference System
ANFIS	- Adaptive-Network based Fuzzy Inference System
MF	- Membership Function
NFV	- Nelfinavir
SQV	- Saquinavir
IDV	- Indinavir
FPV	- Fosamprenavir
ATV	- Atazanavir
LPV	- Lopinavir
TPV	- Tipranavir
DRV	- Darunavir
APV	- Amprenavir
RTV	- Ritonavir
AZT	- Azidothymidine
TDF	- Tenofovir

ABC	- Abacavir
3TC	- Lamivudine
NVP	- Nevirapine
EFV	- Efavirenze
ETR	- Etravirine
RPV	- Rilpivirine
DOR	- Dolutegravir

CHAPTER – 1

INTRODUCTION

1.1. OUTLINE OF THE PROJECT

1.1.1. HIV AIDS:

HIV is a virus that damages the immune system. The immune system helps the body fight off infections. Untreated HIV infects and kills CD4 cells, which are a type of immune cell called T cells. Overtime, as HIV kills more CD4 cells, the body is more likely to get various types of infections and cancers.

Without treatment, a person with HIV is likely to develop a serious condition called AIDS. AIDS is a disease that can develop in people with HIV. It's the most advanced stage of HIV. However, treatment with antiretroviral drugs can prevent AIDS from developing.

1.1.2. HIV-DRUGS:

There's currently no known cure for HIV. However, medications can help people living with HIV manage the condition and lead healthy lives. These drugs work by stopping HIV from replicating.

This method used to treat and fight off the HIV infection is ARV (Anti Retro Viral therapy). There are six main types ('classes') of antiretroviral drugs. Each class of drug attacks HIV in a different way. Generally, drugs from two (or sometimes three) classes are combined to ensure a powerful attack on HIV.

Antiretroviral drugs are listed below with example:

a) Nucleoside/Nucleotide reverse transcriptase inhibitors (NRTIs):

Example: Abacavir, Lamivudin, Emtricitabine

b) Non- nucleoside reverse transcriptase inhibitors (NNRTIs):

Example: Doravirine, Etravirine, Nevirapine.

c) Protease inhibitor:

Example: Atazanavir.

d) *Entry inhibitors*: Entry inhibitor stops HIV from entering human cell, There are two types – CCR5 inhibitor, CD4 inhibitor.

Example: Maraviroc.

e) *Booster drugs*: Booster drugs are used to boost the effect of protease inhibitor. A small dose of booster drug effectively works well. It stays in the body for long time or at higher level. Without booster drug, the primary drugs are ineffective.

1.1.2.1. Protease Inhibitor:

Antiretroviral drugs are designed to treat HIV. Different drugs have different mechanisms of action. Protease inhibitors work by blocking the activity of HIV protease, which is an enzyme that HIV needs to multiply.

List of protease inhibitor drugs:

- Nelfinavir (NFV)
- Saquinavir (SQV)
- Indinavir (IDV)
- Fosamprenavir (FPV)
- Atazanavir (ATV)
- Lopinavir (LPV)
- Tipranavir (TPV)
- Darunavir (DRV)

1.1.2.2. Nucleoside reverse transcriptase inhibitors:

Nucleoside reverse transcriptase inhibitors (NRTIs) block reverse transcriptase (an HIV enzyme). HIV uses reverse transcriptase to convert its RNA into DNA (reverse transcription). Blocking reverse transcriptase and reverse transcription prevents HIV from replicating.

List of nucleoside reverse transcriptase inhibitors drugs:

- Zidovudine (AZT)

- **Tenofovir disoproxil fumarate (TDF)**
- **Abacavir sulfate (ABC)**
- **Lamivudine (3TC)**

1.1.2.3. Non-nucleoside reverse transcriptase inhibitor (NNRTIs):

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) are antiviral agents that bind non-competitively to HIV- 1's reverse transcriptase and prevents viral RNA conversion to DNA. They are used to treat human immunodeficiency virus (HIV infection) and acquired immune deficiency syndrome (AIDS).

List of non- nucleoside reverse transcriptase inhibitors drugs:

- **Efavirenz (EFV)**
- **Nevirapine (NVP)**
- **Etravirine (ETR)**
- **Doravirine (DOR)**
- **Rilpivirine (RPV)**

1.1.3. DRUG RESISTANCE:

Drug resistance is the reduction in effectiveness of a medication to cure a disease or condition.

- HIV DRUG RESISTANCE occurs when the virus starts to make changes (mutations) to its generic make-up (RNA) that are resistant to certain HIV drugs, or classes of HIV drugs.
- Once a person has HIV, the virus begins to multiply in the body. As HIV multiplies, it from sometimes changes form (mutates). Some HIV mutations that develop while a person is taking HIV medicines can lead to drug-resistant strains of HIV.
- Once drug resistance develops, HIV medicines that previously controlled the person's HIV are no longer effective. In other words, the HIV medicines can't prevent the drug resistant HIV from multiplying. Drug resistance can cause HIV treatment fail.

1.1.4. HIV DRUG RESISTANCE PREDICTION:

Which is briefing the process of prediction where drug resistance scores occurs of each drugs taken and predicted scores provides the concept of the better drug to be used against HIV infection. Resistance scores in this HIV drug resistance prediction process is in numerical format.

1.2. OUTSOURCE:

1.2.1. TOOLS AND SOFTWARE USED TO PREDICT:

The tool here used for HIV drug resistance prediction is ANFIS algorithm present in the MATLAB software, package of variety of tools. ANFIS is one of the good algorithms to analyze and predict.

1.2.2. HIV DRUG RESISTANCE- STANFORD DATASET:

PROTEASE INHIBITORS:

1453 patters of major drug resistance mutations from 163680 clinical isolates obtained from 155477 persons infected with HIV-1 virus are arranged in decreasing prevalence in the table that is updated in 2018. For each pattern, drug susceptibility data on isolates matching the specific mutation pattern is shown in that table. On the right side shown the fold resistance scores of drugs according to the mutation patterns.

NRTI:

1117 patterns of major drug resistance mutations from 163425 clinical isolates obtained from 154793 persons infected with HIV-1 virus are arranged in decreasing prevalence in the table that is updated in 2018. For each pattern drug susceptibility data on isolates matching the specific mutation pattern is shown on the left side. On the right side shown the fold resistance scores of drugs according to the mutation patterns.

NNRTI:

361 patterns of major drug resistance mutations from 152595 clinical isolates obtained from 145378 persons infected with HIV-1 virus arranged in decreasing prevalence in the table that is updated in 2018. For each pattern drug susceptibility data on isolates the specific matching the specific mutation pattern is shown on the left side. On the right side shown the fold resistance scores of drugs according to the mutation patterns.

1.2.3. MACHINE LEARNING APPROACH IN PREDICTION:

(Ian Witten, Eibe Frank et al, “Data Mining: Practical Machine Learning Tools and Techniques”, Jan 2005.)

- An adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS) is a kind of artificial neural network that is based on Takagi-Sugeno fuzzy inference system.
- The algorithm used, ANFIS help out to predict the drug susceptibility data's in the dataset.
- When using the ANFIS function, create or load the input data and pass it to training data input argument.
- When using Neuro-Fuzzy designer, in the load data section, select training, and then: To load data from a file, select file. To load data from the MATLAB workspace, select workspace.

CHAPTER – 2

LITERATURE SURVEY

2.1. REVIEW- LITERATURE WORK:

2.1.1. What is drug resistance in general?

When a bacteria, virus, or other microorganisms mutates (changes form) and becomes insensitive to (resistant to) a drug that was previously effective. When a drug is resistant cause of microorganism stops the activity of this so, the drug can't be effective towards the microorganisms (virus, bacteria etc.). Drug resistance can be a cause of HIV treatment failure.

2.1.2. How does a drug get resistant periodically?

- Drug resistance can happen when HIV changes, or mutates, in a person's body. Viruses do this to try and survive. When it mutates, it may become resistant to the medicines a person is taking. It is the reduction in effectiveness of a drug in curing a disease or condition.
- There are two types of virus particle in an infected person's body – wild type (non-mutated virus particle) and mutated virus particle. Mutation confers resistant to antiretroviral drugs so that the virus can replicate in the presence of ARV.
- Wild type of viruses is more fit to replicate than mutant virus.
- If the HIV viral load is increased under treatment, then HIV can replicate under selective condition and can develop resistance to treatment.
- In some cases, a person may contract a drug-resistant strain of HIV. This means the type of HIV they have does not respond to certain medications.
- Drug resistance is more likely in certain condition like treatment with less than three drugs, inappropriate selection of drugs, adding one drug to a failing regimen.

2.1.3. Study drug resistance database online:

Since the AIDS was pandemic first identified in the early 1980's HIV has killed more than 30 million individuals. HIV research ushered in a golden age of antiviral drug development.

National HIV research programs saved millions of lives, prevented millions of infections. Highly effective antiretroviral drugs (ARV) that inhibit HIV replication have been developed and made available at cost in many parts of the world. However, one of the main challenges in treating HIV is that it mutates at high rates and develops drug resistance. This poses an omnipresent threat of acquired HIV drug resistance (HIVDR) in persons receiving treatment. Both acquired (ADR) and transmitting drug resistance (TDR) increase the risk of treatment failure.

WHO is already developed five-year global action plan for 2017-2021 to support a coordinated international effort to prevent, monitor and respond to the emergence of HIV drug resistance, and to strengthen country efforts to achieve the global HIV targets. WHO recommends ART for all people with HIV as soon as possible after diagnosis without any restrictions of CD4 counts. It also recommends offer of pre-exposure prophylaxis to people at substantial risk of HIV infection as an additional prevention choice as part of comprehensive prevention. As a reference HIV Drug Resistance report 2019 made by scientist declared the concept of ADR (Acquired HIV drug resistance), TDR (Transmitted HIV drug resistance), PDR (Pre-treatment HIV drug resistance).

The Stanford HIV Drug Resistance Database (HIVDB) is an essential resource for public health officials monitoring different drug resistance, for scientists developing new ARV drugs, and for HIV care providers managing patients with HIVDR. The presence of HIVDR prior to the start of an ARV-treatment regimen is a good predictor of the success or failure of that regimen. HIVDB contains a free publicly available genotypic resistance interpretation system that interpretation system has been integrated into the workflows of the World Health Organization's HIV drug resistance network, the NIH-funded AIDS Clinical Trials Group, and several hundred research and more worldwide.

SHDRDB collaborated with researchers from around the world to track global trends in drug resistance and to characterize the genetic mechanisms of HIVDR to the ARVs for which more data are needed including those belonging to the nucleoside RT inhibitor (NRTI), nonnucleoside RT inhibitor (NNRTI), protease inhibitor (PI).

Protease Inhibitor:

1453 patterns of major drug resistance mutations from 163680 clinical isolates obtained from 155477 persons infected with HIV-1 Group M virus are arranged in decreasing prevalence in the table. Sequences containing a mixture at one of the major drug resistance positions were excluded from the table. For each pattern, drug susceptibility data on isolates matching the specific mutation pattern is shown on the left side. The analysis was originally published in [Rhee et al. 2004](#). The data was described in [Rhee et al. 2010](#).

The number of sequences with each pattern is indicated, as well as the median fold resistance, as assayed by the Virologic Pheno-Sense assay, to each of several individual members of the selected drug class.

NRTI:

1117 patterns of major drug resistance mutations from 163425 clinical isolates obtained from 154793 persons infected with HIV-1 virus are arranged in decreasing prevalence in the table. Sequences containing a mixture at one of the major drug resistance positions were excluded from the table. For each pattern drug susceptibility data on isolates matching the specific mutation pattern is shown on the left side. The analysis was originally published in [Rhee et al. 2004](#). The data was described in [Melikian et al. 2012](#).

NNRTI:

361 patterns of major drug resistance mutations from 152595 clinical isolates obtained from 145378 persons infected with HIV-1 virus arranged in decreasing prevalence in the table. Sequences containing a mixture at one of the major drug resistance positions were excluded from the table. For each pattern drug susceptibility data on isolates the specific matching the specific mutation pattern is shown on the left side. The analysis was originally published in [Rhee et al. 2004](#).

The databases used as reference is given below in figure – 2.1.

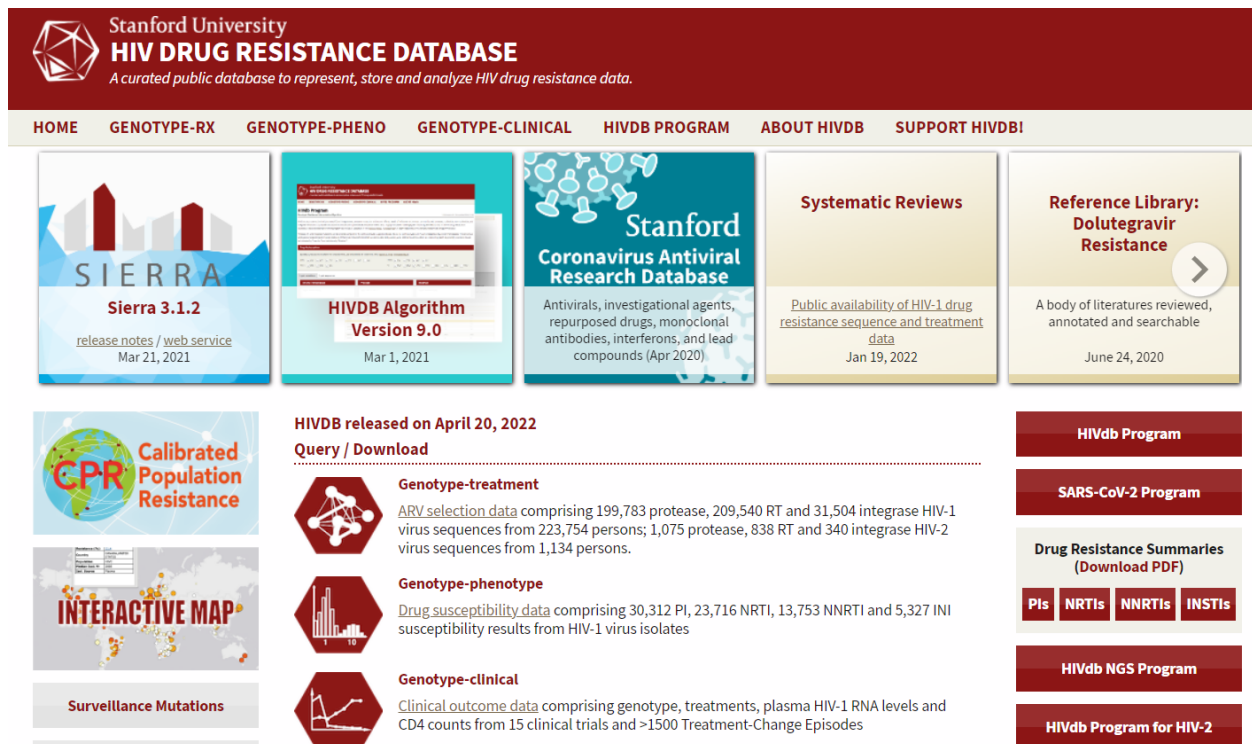


Figure – 2.1: Online Database as reference

2.1.4. What machine learning approach use to predict drug resistance:

Several antiretroviral drugs have been approved for use in HIV infected people. Despite efforts made by the scientific community, an effective drug that kills the virus has not been developed yet. A lot of computational algorithms have been used for finding mutations associated with drug resistance as well as for prediction of HIV resistance.

Antiretroviral drugs are a very effective therapy against HIV infection. However, the high mutation rate of HIV permits the emergence of variants that can be resistant to the drug treatment.

Predicting drug resistance to previously unobserved variants is therefore very important for an optimum medical treatment. In this paper, we propose the use of ANFIS designer (MATLAB tool) functions to predict drug resistance from virus sequence data.

These tool functions are very simple to implement and are able to take into account HIV data particularities, such as membership function, rules and to predict the different drug resistance scores of each mutation pattern, as it is known that not all positions contribute equally to the resistance.

To train a fuzzy system using neuro-adaptive methods, you must collect input/output training data using experiments or simulations of the system you want to model. In general, ANFIS training works well if the training data is fully representative of the features of the data that the trained FIS is intended to model.

To specify your training data, you can create an array in the MATLAB workspace. Each row contains a data point, with the final column containing the output value and the remaining columns containing input values. You can then pass this data to the training Data input argument of the ANFIS function or load it into the Neuro- Fuzzy Designer app.

Load the data from file. Each line of the file contains a data point with values separated by white space. The final value on each line is the output, and the remaining values are the inputs.

When using the ANFIS function, create or load the input data and pass it to the training Data input argument. When using Neuro-Fuzzy Designer, in the Load data section, select Training, and then to load data from a file, select file. To load data from the MATLAB workspace, select workspace.

CHAPTER – 3

AIM AND OBJECTIVES

The primary aim for the study is to determine and predict the resistance in drugs towards HIV-AIDS.

Listed are the prominent objectives of the study:

- To study all the HIV-AIDS drug resistance database online.
- To study HIV-AIDS drug resistance dataset – international as well as national.
- To create a worksheet based on the HIV drug resistance database.
- To work-on predicting drug resistance score using machine learning technique.
- To work-on ANFIS algorithm to simulate the given dataset.
- To analyze the statistics of HIV drug resistance dataset.
- To co-construct the rule editor and interpret the data accordingly in rule viewer.
- To interpret the three- dimensional graphical representation in surface viewer.

CHAPTER – 4

METHODOLOGY

4.1. HIV DRUG RESISTANCE DATABASE USED

4.1.1. STANFORD HIV DRUG RESISTANCE DATABASE (HIVDB):

It is an essential resource for public health officials monitoring ADR and TDR, for scientists developing new ARV drugs, and for HIV care providers managing patients with HIVDR. Researchers studying HIVDR also use HIVDB to place their findings in the context of previous studies and to perform meta-analysis that require data from large numbers of studies, making it possible to discover new knowledge that cannot be discovered from a single study. Figure – 4.1, 4.2, and 4.3 shows the source of the dataset.

4.1.2. HIV DRUG RESISTANCE DATABASE:

PROTEASE INHIBITORS:



Stanford University HIV DRUG RESISTANCE DATABASE

A curated public database to represent, store and analyze HIV drug resistance data.

[HOME](#) [GENOTYPE-RX](#) [GENOTYPE-PHENO](#) [GENOTYPE-CLINICAL](#) [HIVDB PROGRAM](#) [ABOUT HIVDB](#) [SUPPORT HIVDB!](#)

PI Mutation Pattern & Susceptibility

1453 patterns of major drug resistance mutations (defined [here](#)) from 163680 clinical isolates obtained from 155477 persons infected with HIV-1 Group M virus are arranged in decreasing prevalence in the table (updated on 2018-9-28). Sequences containing a mixture at one of the major drug resistance positions were excluded from the table. For each pattern, drug susceptibility data on isolates matching the specific mutation pattern is shown on the left side. The analysis was originally published in [Rhee et al. 2004](#). The data was described in [Rhee et al. 2010](#).

The number of sequences with each pattern is indicated, as well as the median fold resistance, as assayed by the ViroLogic PhenoSense assay, to each of several individual members of the selected drug class. Subscripts indicate the number of sequences assayed for resistance to the designated drug. Interquartile ranges are available by clicking [here](#). Data on the 30 most common mutation patterns is also available in a [PDF format](#).

Fold Resistance		
NFV	3	6
SQV	3	15
IDV	3	15
FPV	4	11
ATV	3	15
LPV	9	55
TPV	2	8
DRV	10	90

Mutation Patterns	Number of Sequences	NFV fold _n	SQV fold _n	IDV fold _n	FPV fold _n	ATV fold _n	LPV fold _n	TPV fold _n	DRV fold _n
None	141417	1.1 ₇₅₁	0.7 ₇₃₀	0.8 ₇₄₃	0.7 ₇₀₈	0.9 ₄₆₄	0.7 ₅₉₂	0.9 ₃₈₃	0.7 ₃₆₅
90M	2399	7.3 ₅₀	3.2 ₅₁	3.1 ₅₁	1.3 ₄₇	3.1 ₃₃	1.6 ₄₂	1.2 ₂₄	0.9 ₁₉
30N	1704	40 ₄₆	1.3 ₄₄	1.3 ₄₅	0.8 ₃₉	2.5 ₃₀	1.2 ₇	1.2 ₁₂	1 ₈
46I,90M	743	22 ₁₆	4.2 ₃₆	9.1 ₃₄	3.2 ₂₉	4.9 ₁₇	2.7 ₂₃	0.9 ₁₃	0.8 ₆
46I	651	11 ₅	0.9 ₅	4.4 ₅	2.1 ₄	2.2 ₃	2.7 ₃	1.1 ₂	0.8 ₂
54V,82A,90M	642	57 ₁₆	27 ₂₆	24 ₂₆	3.6 ₂₄	36 ₇	39 ₂₄	2.4 ₄	1.6 ₁
54V,82A	601	9.4 ₂₁	1.4 ₂₁	8 ₁₉	1.5 ₁₉	2.3 ₁₉	11 ₁₇	1.6 ₆	0.8 ₃
46L	558	8.2 ₁₂	1.5 ₂	2.6 ₂	1.1 ₁	2.4 ₁	1.6 ₂		
46L,54V,82A	503	39 ₂₀	6.2 ₂₀	19 ₁₉	4.6 ₁₇	31 ₅	37 ₁₇	1.1 ₁₀	1.7 ₄
46I,54V,82A	502	16 ₁₄	1.6 ₂₄	15 ₂₂	2.5 ₂₀	7.5 ₅	27 ₁₈	1.6 ₅	0.8 ₃
46I,84V,90M	452	48 ₁₃	43 ₃₁	28 ₃₂	7.7 ₃₂	22 ₁₈	17 ₂₇	2 ₁₁	4 ₄

Figure – 4.1: Protease dataset

NRTI:



Stanford University HIV DRUG RESISTANCE DATABASE

A curated public database to represent, store and analyze HIV drug resistance data.

[HOME](#) [GENOTYPE-RX](#) [GENOTYPE-PHENO](#) [GENOTYPE-CLINICAL](#) [HIVDB PROGRAM](#) [ABOUT HIVDB](#) [SUPPORT HIVDB!](#)

NRTI Mutation Pattern & Susceptibility

1117 patterns of major drug resistance mutations (defined [here](#)) from 163425 clinical isolates obtained from 154793 persons infected with HIV-1 Group M virus are arranged in decreasing prevalence in the table (updated on 2018-9-28). Sequences containing a mixture at one of the major drug resistance positions were excluded from the table. For each pattern, drug susceptibility data on isolates matching the specific mutation pattern is shown on the left side. The analysis was originally published in [Rhee et al. 2004](#). The data was described in [Melikyan et al. 2012](#).

The number of sequences with each pattern is indicated, as well as the median fold resistance, as assayed by the ViroLogic PhenoSense assay, to each of several individual members of the selected drug class. Subscripts indicate the number of sequences assayed for resistance to the designated drug. Interquartile ranges are available by clicking [here](#). Data on the 30 most common mutation patterns is also available in a [PDF format](#).

Fold Resistance		
AZT	3	10
D4T	1.5	2
TDF	1.5	4
ABC	3	6
DDI	1.5	2
3TC	3	20

Mutation Patterns	Number of Sequences	AZT fold _n	TDF fold _n	ABC fold _n	3TC fold _n
None	114551	0.8 ₄₁₉	0.8 ₃₄₆	0.9 ₄₁₃	1 ₇₀₉
184V	12156	0.5 ₁₂₄	0.5 ₆₃	3.1 ₁₂₅	200 ₁₁₅
67N,70R,184V	1955	3.7 ₃₂	1.2 ₂₆	4.5 ₃₁	200 ₅₉
41L,184V,210W,215Y	1399	18 ₅₁	1.6 ₃₈	6.5 ₄₈	200 ₆₉
41L,184V,215Y	1381	6 ₄₁	1.1 ₂₄	5.1 ₄₁	200 ₅₅
41L,67N,184V,210W,215Y	1279	30 ₅₃	1.6 ₄₁	6.5 ₄₆	200 ₇₂
65R,184V	1258	0.4 ₁₆	1.2 ₁₆	8.4 ₁₆	200 ₂₇
70R,184V	1109	0.8 ₁₄	0.7 ₇	3.4 ₁₅	200 ₂₁
41L	940	0.9 ₉	1 ₇	1 ₈	1.5 ₁₄
41L,67N,210W,215Y	793	40 ₃₅	4.7 ₂₅	5.3 ₂₉	6.2 ₄₇
65R	763	0.5 ₁₉	1.8 ₁₆	2.4 ₁₉	8.7 ₂₈

Figure – 4.2: NRTI dataset

NNRTI:

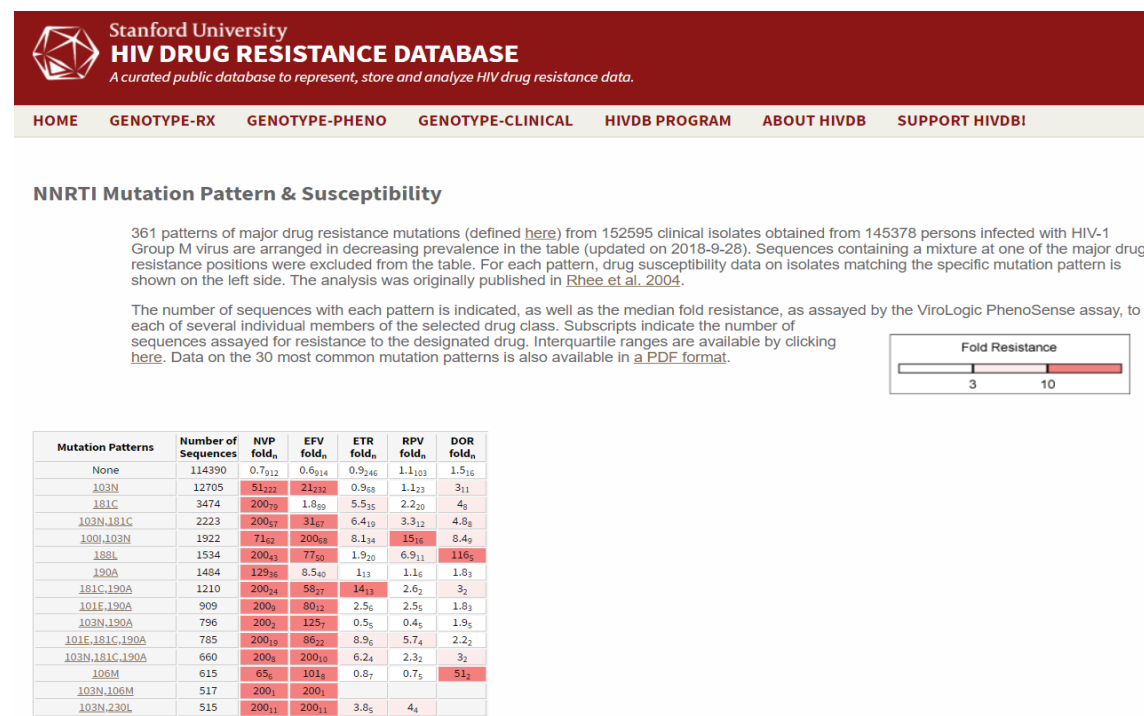


Figure – 4.3: NNRTI dataset

This collects, annotate and analyze genotypic drug resistance data from previously untreated (ARV- naive) and previously treated (ARV- experienced) persons and systematically aggregate data from many published studies to analyze temporal global trends in the drug – resistance mutations responsible for ADR and TDR, thus creating a resource for clinical and molecular epidemiologists. Our work is made challenging by the extensive genetic sequence heterogeneity of HIV, including the many different HIV subtypes and recombinant forms. Nonetheless, by collecting well-characterized HIV sequences from more than 150,000 individuals, we have been able to distinguish naturally occurring HIV mutations from those that are selected by ARV therapy and contribute to HIVDR.

4.1.3. WHO HIV Drug Resistance database:

This collaborate with researchers from around the world to track global trends in ADR and TDR and to characterize the genetic mechanisms of HIVDR to the ARVs for which more data are needed including those belonging to the nucleoside RT inhibitor (NRTI),

non-nucleoside RT inhibitor (NNRTI), protease inhibitor (PI) class of drugs. **(WHO. Antiretroviral treatment working group treatment white paper. 2010.)**

4.2. MACHINE LEARNING APPROACH:

With the rise in big data, machine learning has become a key technique for solving problems.

Machine learning is a data analytics technique that teaches computers to do what comes naturally to humans and animals: learn from experience. Machine learning algorithms use computational methods to “learn” information directly from data without relying on a predetermined equation as a model. The algorithms adaptively improve their performance as the number of samples available for learning increases.

4.2.1. Fuzzy Logic Toolbox:

Fuzzy Logic Toolbox software provides a command-line function (anfis) and an interactive app (Neuro-Fuzzy Designer) for training an adaptive neuro-fuzzy inference system (ANFIS).

The architecture and learning procedure underlying ANFIS (adaptive-network – based fuzzy inference system) is presented, which is a fuzzy inference system implemented in the framework of adaptive networks. By using a hybrid learning procedure, the proposed ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input-output data pairs.

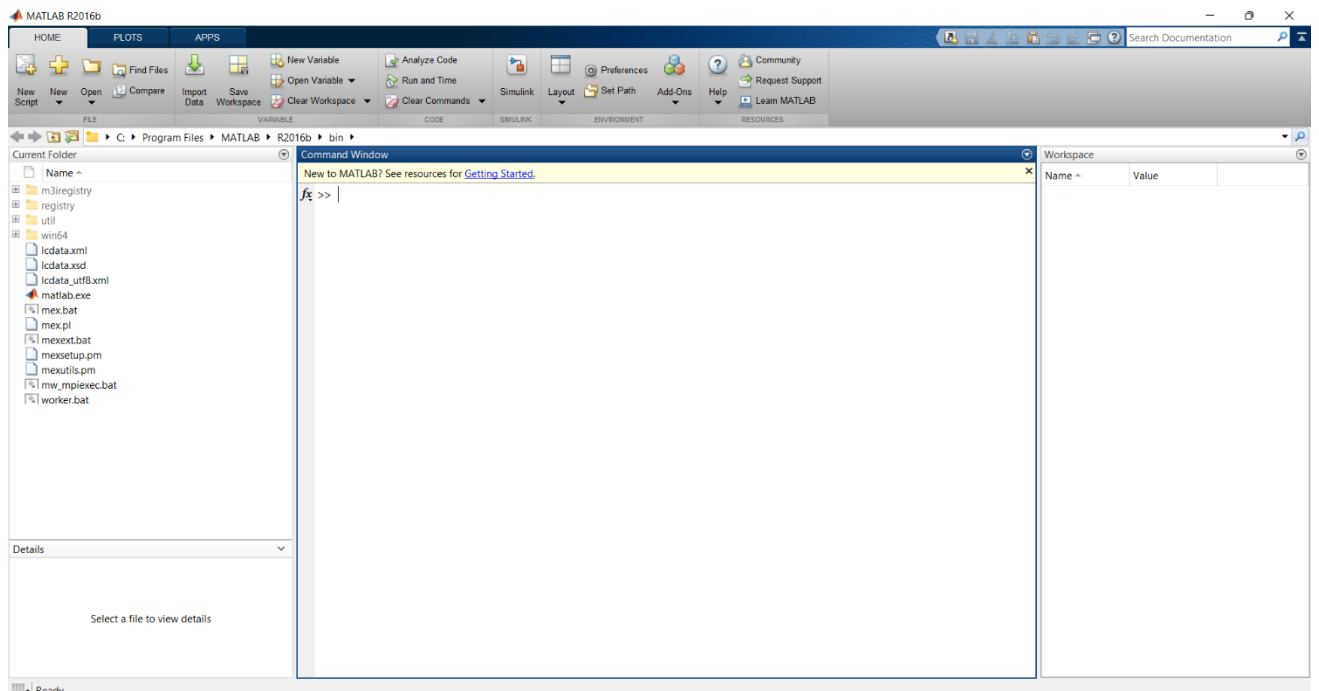


Figure - 4.4: MATLAB Command Window

The toolbox make the model complex system behaviours using simple logic rules, and then implement these rules in a ANFIS system interference.

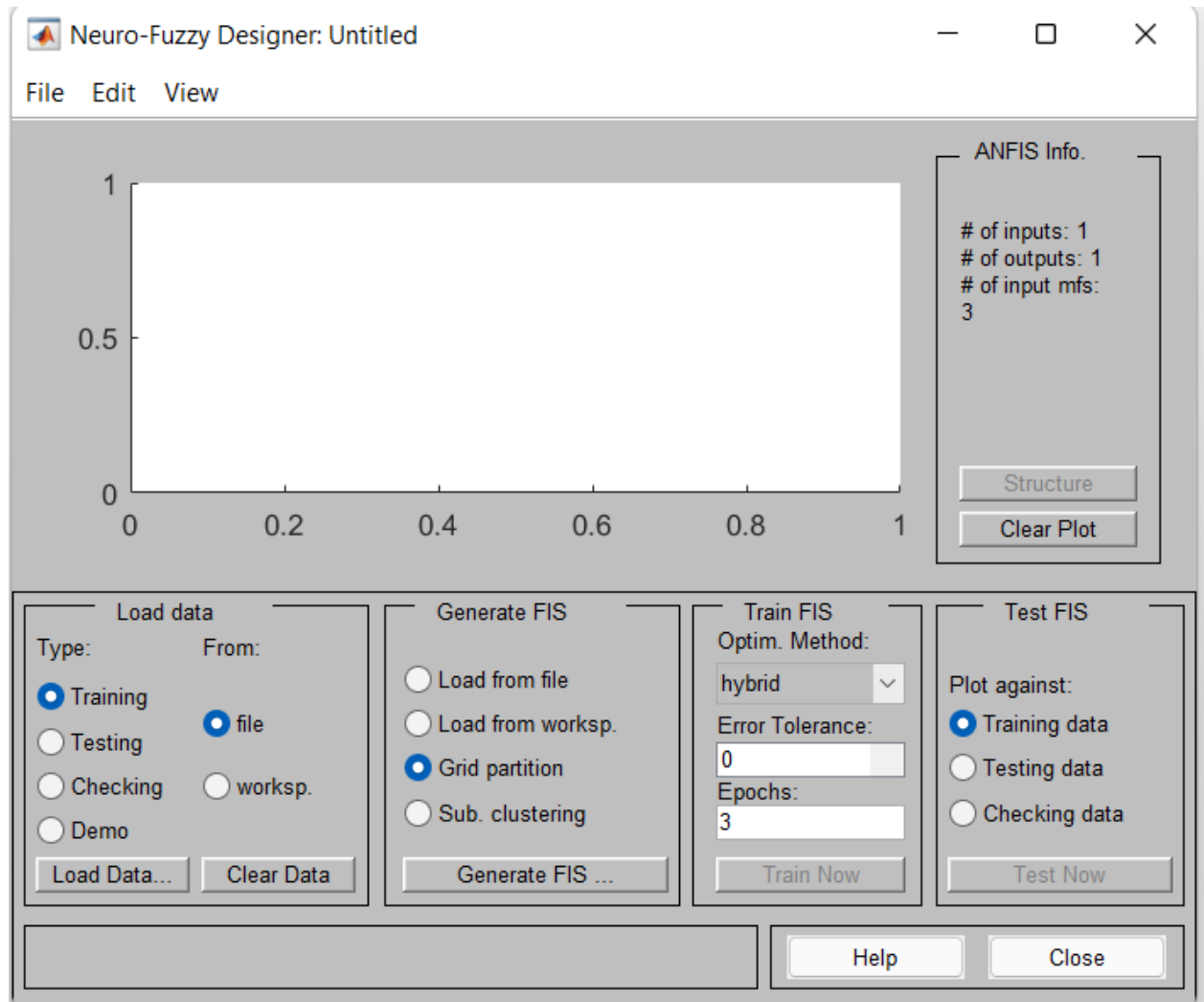


Figure – 4.5: ‘anfisedit’ command opens ANFIS Designer window

4.2.1.1. TRAINING DATA:

To train a fuzzy system using neuro-adaptive methods, need to collect input/output training data using experiments or simulations of the system we want to model. In general, ANFIS training works well if the training data is fully representative of the features of the data that the trained FIS is intended to model.

- Load the data from a file. Each line of the file contains a data point with

values separated by white space. The final value on each line is the output, and the remaining values are the inputs.

When using the anfis function, create or load the input data and pass it to the training data input argument. When using Neuro-Fuzzy Designer, in the load data section, select Training, and then load data from file.

	A	B	C	D	E	F	G	H	I	J	K	L	M
	ProteaseAnfisNFV1												
1	0	0	0	0	0	0	0	0	0	0	0	1	1.1751
2	2	0	0	0	0	0	0	0	0	0	0	0	7.3500
3	0	0	3	0	0	0	0	0	0	0	0	1	4046
4	0	0	3	0	0	0	0	0	0	0	0	0	2236
5	0	0	0	0	0	0	4	0	5	0	0	1	115
6	0	0	0	0	0	0	4	0	5	0	0	0	5726
7	0	0	6	0	0	0	0	0	0	0	0	0	9.4210
8	0	0	6	0	0	0	4	0	5	0	0	0	8.2200
9	0	0	3	0	0	0	4	0	5	0	0	0	3920
10	0	0	3	0	0	0	4	0	0	4	0	1	1624
11	0	0	6	0	0	0	4	0	5	0	0	1	4833
12	0	0	0	0	0	0	0	0	0	4	0	1	4017
13	0	0	3	0	0	0	4	0	0	4	0	1	1924
14	0	0	3	0	0	0	4	0	5	0	0	1	8621
15	0	0	0	0	0	0	0	0	5	0	0	0	4922
16	0	0	0	0	0	0	0	0	0	0	7	0	2.1800
17	0	0	3	0	0	0	4	4	5	5	0	0	7.7150
18	0	0	0	0	0	0	4	0	0	4	0	1	118
19	0	0	0	0	0	0	4	0	0	0	0	1	1446
20	2	0	0	0	0	0	0	0	0	0	0	1	343
21	0	0	0	0	0	6	0	0	0	0	0	0	7415
22	0	0	0	0	0	0	0	0	0	4	0	0	0.3150
23	0	0	3	0	0	0	0	0	0	0	7	0	2.5100
24	0	0	6	0	0	0	0	0	5	0	0	0	188
25	0	0	0	0	0	0	4	0	0	0	0	0	6.1300
26	2	0	3	0	0	0	0	0	0	0	0	0	182
27	0	0	0	0	0	0	0	0	5	0	0	1	255
28	0	0	3	0	0	0	0	0	0	4	0	0	6.1200

Figure – 4.6: Loading data from file

- In the load data section, select Training and workspace. Click load data. In the load from workspace dialog box, enter the variable name.

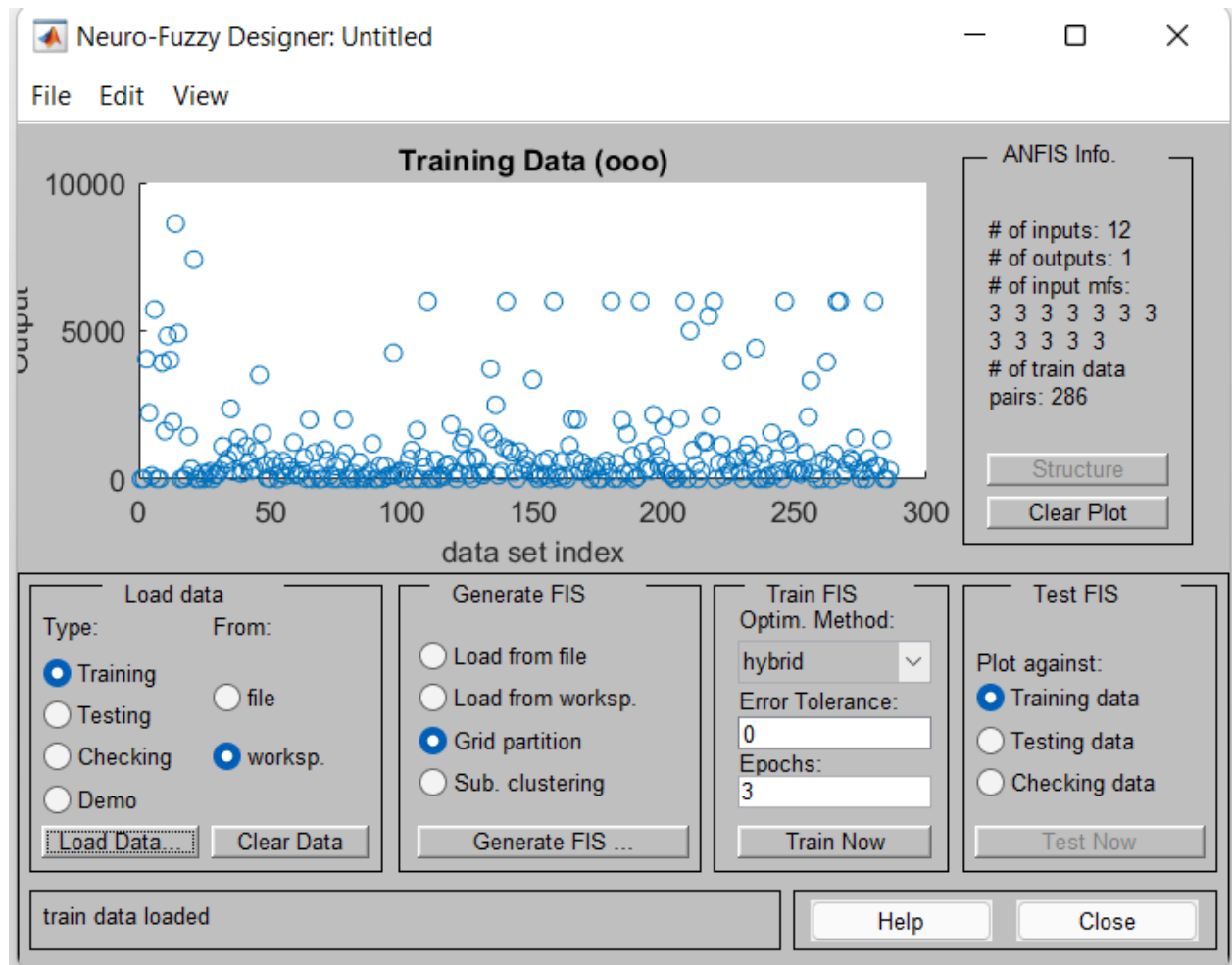


Figure – 4.7: Data loaded in Neuro-Fuzzy Designer

4.2.1.2. GENERATE OR LOAD FIS STRUCTURE:

Before starting the FIS training, you must specify an initial FIS model structure.

To specify the model structure, perform one of the following tasks:

- Load a previously saved FIS object from a file or the MATLAB workspace.
- Generate the initial FIS model using Sub-clustering.

Click Generate FIS. In the functions dialog box do not change the options given.

Now, Click OK and we can view the FIS is generated.

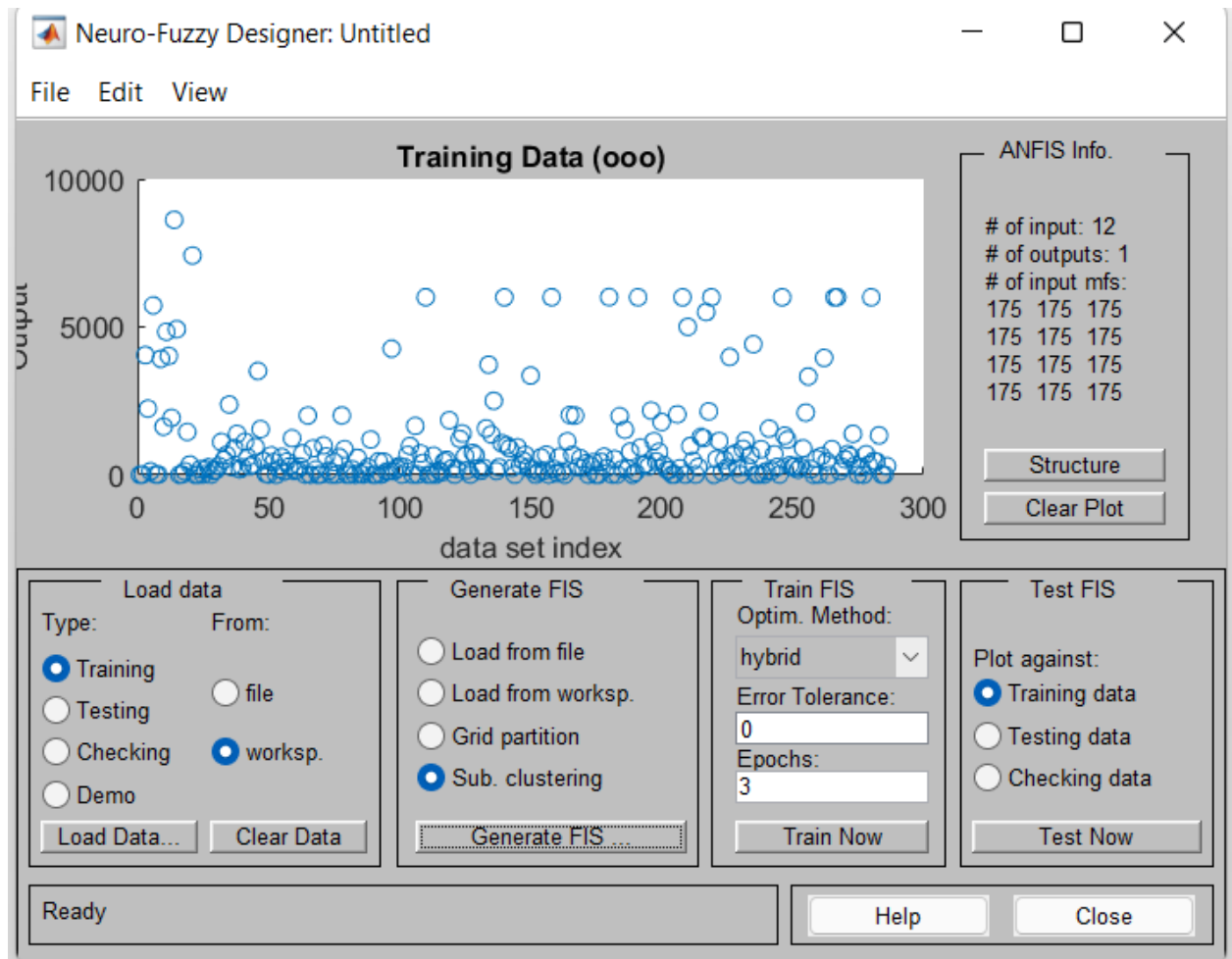


Figure- 4.8: View of FIS generated toolbox

4.2.1.3. TRAIN FIS:

After loading the training data and generating the initial FIS structure, you can train your FIS. To do so, in Neuro- Fuzzy Designer, in the train FIS section, specify the following parameters.

- Optim. Methods – Optimization method. For example, select the hybrid mode, which uses a combination of back propagation and least – squares regression to tune the FIS parameters.
- Epochs – number of training epochs. For this example, 200 to 300 epochs.
- Error tolerance – Error tolerance stopping condition. For this example,

specify a value of 0, which indicates the training will stop when the number of training epochs is reached.

- To train the FIS, click Train NOW.

This app trains the FIS and plots the training error (as stars) and checking error for each training epoch. The checking error decreases up to a certain point in the training, and then it decreases. This increase occurs at the point where the training starts overfitting the training data. The app selects the FIS associated with his overfitting point as the trained ANFIS model.

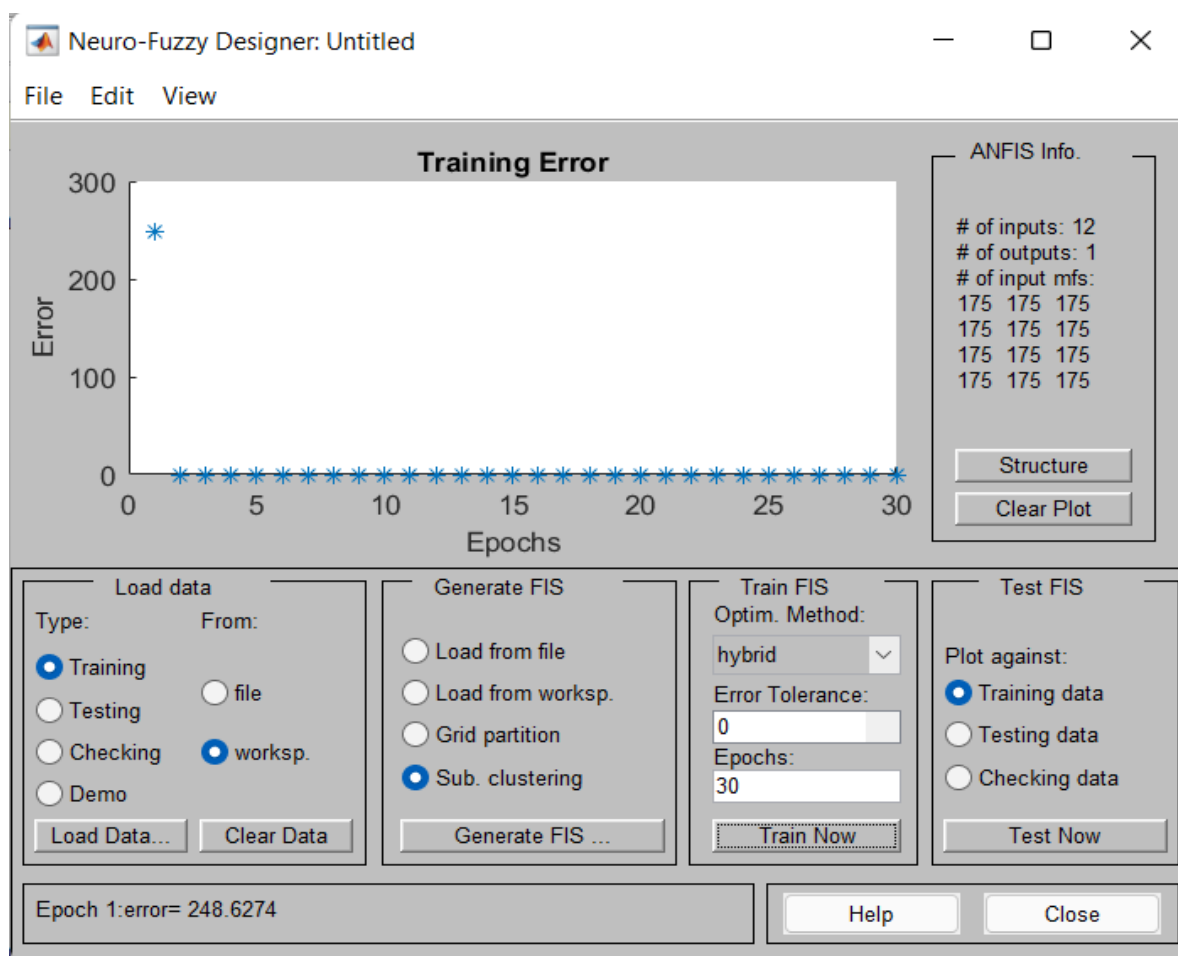


Figure – 4.9: TRAINED FIS

4.2.1.4. SURFACE VIEWER:

Surface viewer is used to view the output surface for your fuzzy system. To view the output surface, you must specify the input and output variables of your FIS, their corresponding membership functions, and the fuzzy rules for your

system.

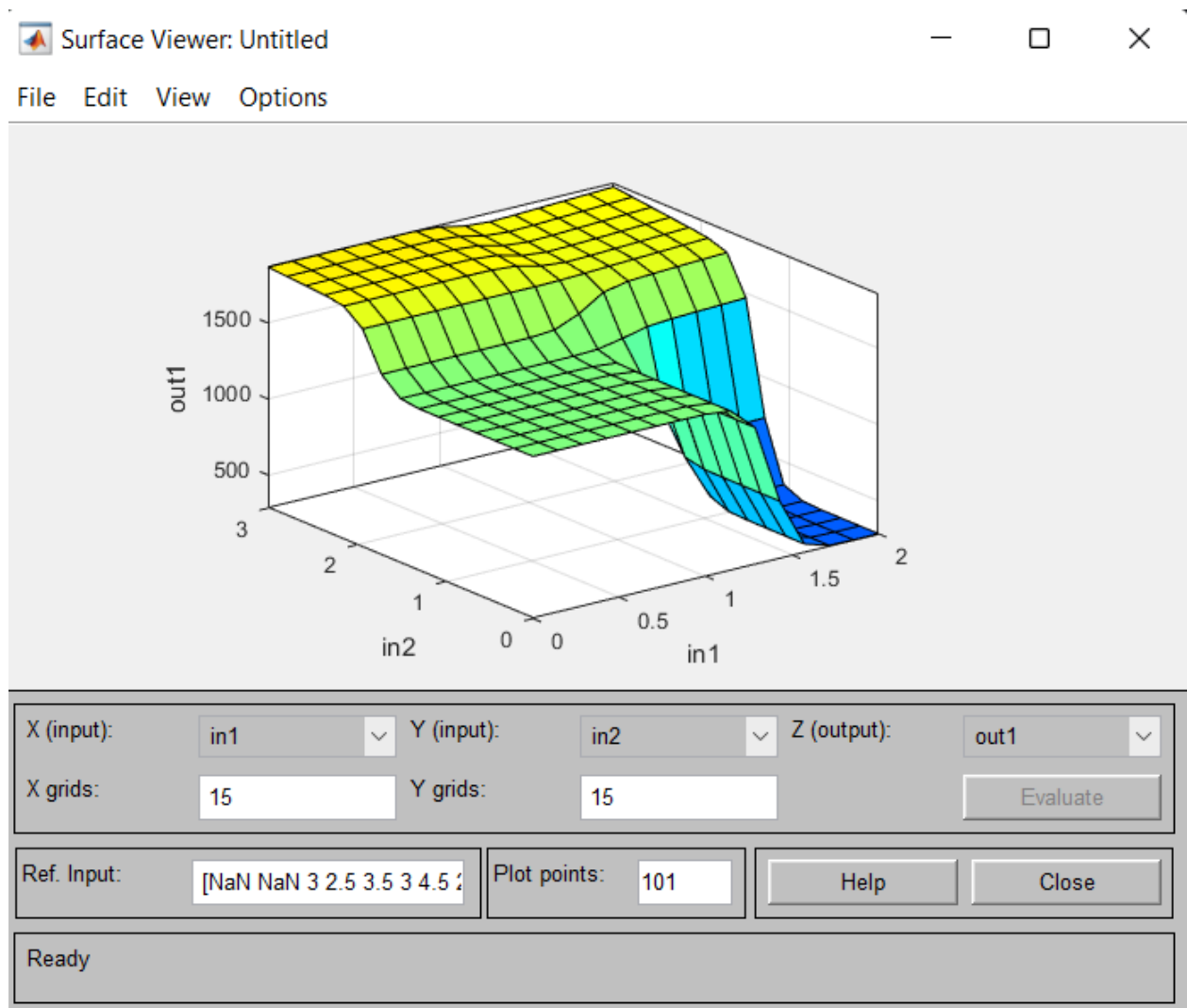


FIGURE – 4.10: SURFACE VIEWER OF NFV

4.2.1.5. RULE VIEWER:

Rule viewer is used to inference process for your fuzzy system. You can adjust the input values and view the corresponding output of each fuzzy rule, the aggregated output fuzzy set, and the defuzzified output value. To view the inference process, you must specify the input and output variables of your FIS, their corresponding membership functions, and the fuzzy rules for your system.

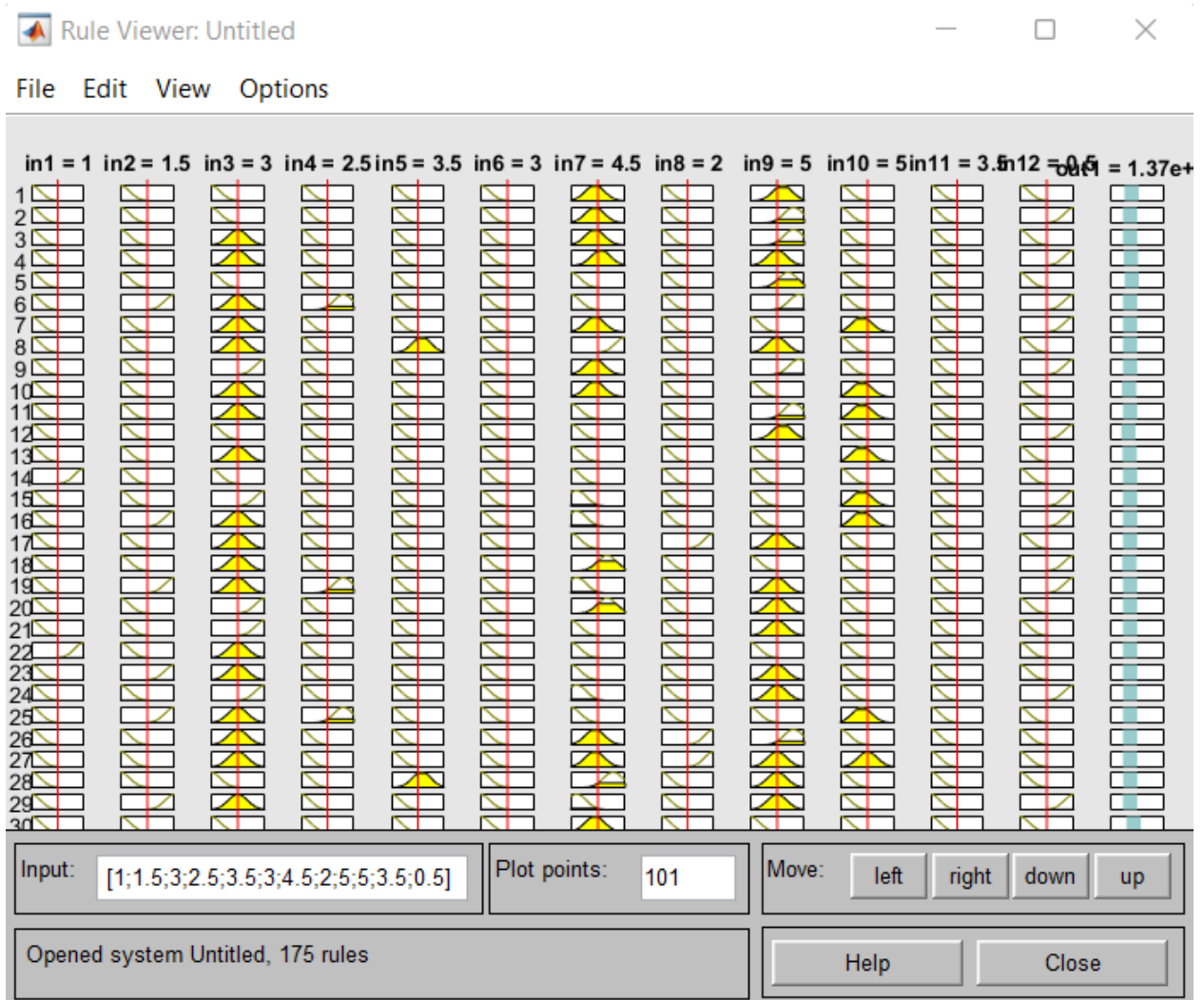


FIGURE – 4.11: RULE VIEWER OF NFV

CHAPTER – 5

RESULT AND DISCUSSION

5.1. EXCEL SHEET OF PROTEASE TABLE:

From the HIV drug resistance database, we have made excel sheet for protease table:

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	30	32	46	47	48	50	54	76	82	84	88	90		NFV foldn	SQV foldn	IDV foldn	FPV foldn	ATV foldn	LPV foldn	TPV foldn	DRV foldn
2	0	0	0	0	0	0	0	0	0	0	0	1		1.1751	0.773	0.8743	0.7708	0.9464	0.7592	0.9383	0.7365
3	2	0	0	0	0	0	0	0	0	0	0	0		7.35	3.251	3.151	1.347	3.133	1.642	1.224	0.919
4	0	0	3	0	0	0	0	0	0	0	0	0	1 M = 1	4046	1.344	1.345	0.839	2.53	1.27	1.212	18
5	0	0	3	0	0	0	0	0	0	0	0	0	0 N = 2	2236	4.236	9.134	3.229	4.917	2.723	0.913	0.86
6	0	0	0	0	0	0	4	0	5	0	0	0	1 I = 3	115	0.95	4.45	2.14	2.23	2.73	1.12	0.82
7	0	0	0	0	0	0	4	0	5	0	0	0	0 V = 4	5726	2726	2426	3.624	367	3924	2.44	1.61
8	0	0	6	0	0	0	0	0	0	0	0	0	0 A = 5	9.421	1.421	820	1.519	2.31	1117	1.66	0.83
9	0	0	6	0	0	0	4	0	5	0	0	0	0 L = 6	8.22	1.52	2.62	1.12	2.41	1.62		
10	0	0	3	0	0	0	4	0	5	0	0	0	0 S = 7	3920	6.22	1919	4.617	319	3717	1.11	1.74
11	0	0	3	0	0	0	4	0	0	4	0	0	1 F = 8	1624	1.624	1522	2.52	7.58	2718	1.65	0.83
12	0	0	6	0	0	0	4	0	5	0	0	0	1 T = 9	4833	4331	2832	7.732	2219	1727	211	44
13	0	0	0	0	0	0	0	0	0	4	0	0	1 C = 10	4017	1117	2216	8.217	177	5015	4.44	4.13
14	0	0	3	0	0	0	4	0	0	4	0	1		1924	2524	1125	4.524	1512	6.423	29	2.16
15	0	0	3	0	0	0	4	0	5	0	0	1		8621	16118	5321	2021	17112	9620	4.712	119
16	0	0	0	0	0	0	0	0	5	0	0	0		4922	2222	2922	6.422	2416	5822	2.614	1.88
17	0	0	0	0	0	0	0	0	0	0	7	0		2.18	0.88	2.18	18	1.97	2.28	0.43	
18	0	0	3	0	0	0	4	4	5	5	0	0		7.715	1.215	2.515	0.115	7.24	1.43	0.61	0.31
19	0	0	0	0	0	0	4	0	0	4	0	1		118	1.68	373	6.58	1.48	828	0.51	
20	0	0	0	0	0	0	4	0	0	0	0	1		1446	7736	506	136	1552	433	172	132
21	2	0	0	0	0	0	0	0	0	0	0	1		343	603	53	1.33	721	282	231	2.11
22	0	0	0	0	0	0	6	0	0	0	0	0		7415	615	2.815	1.315	4.85	1.85	0.51	0.41
23	0	0	0	0	0	0	0	0	0	4	0	0		0.315	0.38	0.315	0.315	619	0.35		
24	0	0	3	0	0	0	0	0	0	0	7	0		2.51	2.81	1.99	3.99	2.34	2.69	3.23	2.62
25	0	0	6	0	0	0	0	0	5	0	0	0		188	1.38	5.88	0.27	6.11	0.82		
26	0	0	0	0	0	0	4	0	0	0	0	0		6.13	1.33	7.43	5.83	2.52	73		
27	2	0	3	0	0	0	0	0	0	0	0	0		182	8.12	6.22	32	8.62	5.42	2.71	0.61
28	0	0	0	0	0	0	0	0	5	0	0	1		255	0.65	1.25	0.85	1.53	0.84	0.42	0.51
29	0	0	3	0	0	0	0	0	0	4	0	0		6.12	1.62	7.12	1.62	5.51	2.72		
30	0	0	6	0	0	0	0	0	0	0	0	1		131	111	5.61	6.71	141	141	111	
31	0	0	0	0	4	0	4	0	5	0	0	0		157	3.47	7.57	1.77	5.43	2.85	0.62	0.72
32	0	0	3	0	0	0	4	0	8	0	0	1		706	1526	736	2.56	776	336	13	0.82

Figure 5.1: Excel sheet of protease table with 8 drugs

After this, we made a final excel sheet based on the grouping of amino acids to perform the machine learning analysis.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	0	0	0	0	0	0	0	0	0	0	0	0	1.1751
2	2	0	0	0	0	0	0	0	0	0	0	0	7.35
3	0	0	3	0	0	0	0	0	0	0	0	1	4046
4	0	0	3	0	0	0	0	0	0	0	0	0	2236
5	0	0	0	0	0	0	4	0	5	0	0	1	115
6	0	0	0	0	0	0	4	0	5	0	0	0	5726
7	0	0	6	0	0	0	0	0	0	0	0	0	9.421
8	0	0	6	0	0	0	4	0	5	0	0	0	8.22
9	0	0	3	0	0	0	4	0	5	0	0	0	3920
10	0	0	3	0	0	0	4	0	0	4	0	1	1624
11	0	0	6	0	0	0	4	0	5	0	0	1	4833
12	0	0	0	0	0	0	0	0	0	4	0	1	4017
13	0	0	3	0	0	0	4	0	0	4	0	1	1924
14	0	0	3	0	0	0	4	0	5	0	0	1	8621
15	0	0	0	0	0	0	0	0	5	0	0	0	4922
16	0	0	0	0	0	0	0	0	0	0	7	0	2.18
17	0	0	3	0	0	0	4	4	5	5	0	0	7.715
18	0	0	0	0	0	0	4	0	0	4	0	1	118
19	0	0	0	0	0	0	4	0	0	0	0	1	1446
20	2	0	0	0	0	0	0	0	0	0	0	1	343
21	0	0	0	0	0	6	0	0	0	0	0	0	7415
22	0	0	0	0	0	0	0	0	0	4	0	0	0.315
23	0	0	3	0	0	0	0	0	0	0	7	0	2.51
24	0	0	6	0	0	0	0	0	5	0	0	0	188
25	0	0	0	0	0	0	4	0	0	0	0	0	6.13
26	2	0	3	0	0	0	0	0	0	0	0	0	182
27	0	0	0	0	0	0	0	0	5	0	0	1	255
28	0	0	3	0	0	0	0	0	0	4	0	0	6.12
29	0	0	6	0	0	0	0	0	0	0	0	1	131
30	0	0	0	0	4	0	4	0	5	0	0	0	157
31	0	0	3	0	0	0	4	0	8	0	0	1	296
32	0	0	0	0	0	0	6	0	0	4	0	1	1108

Figure - 5.2: Final excel sheet for analysis:

5.1.1. RULE VIEWER OF STANFORD HIV DRUG RESISTANCE DATASET (PROTEASE):

As datasheet of 1453 patterns of major drug resistance mutations is created in rule editor. So, rule viewer interpreted the median fold drug resistance score for mutation positions as well as drugs.

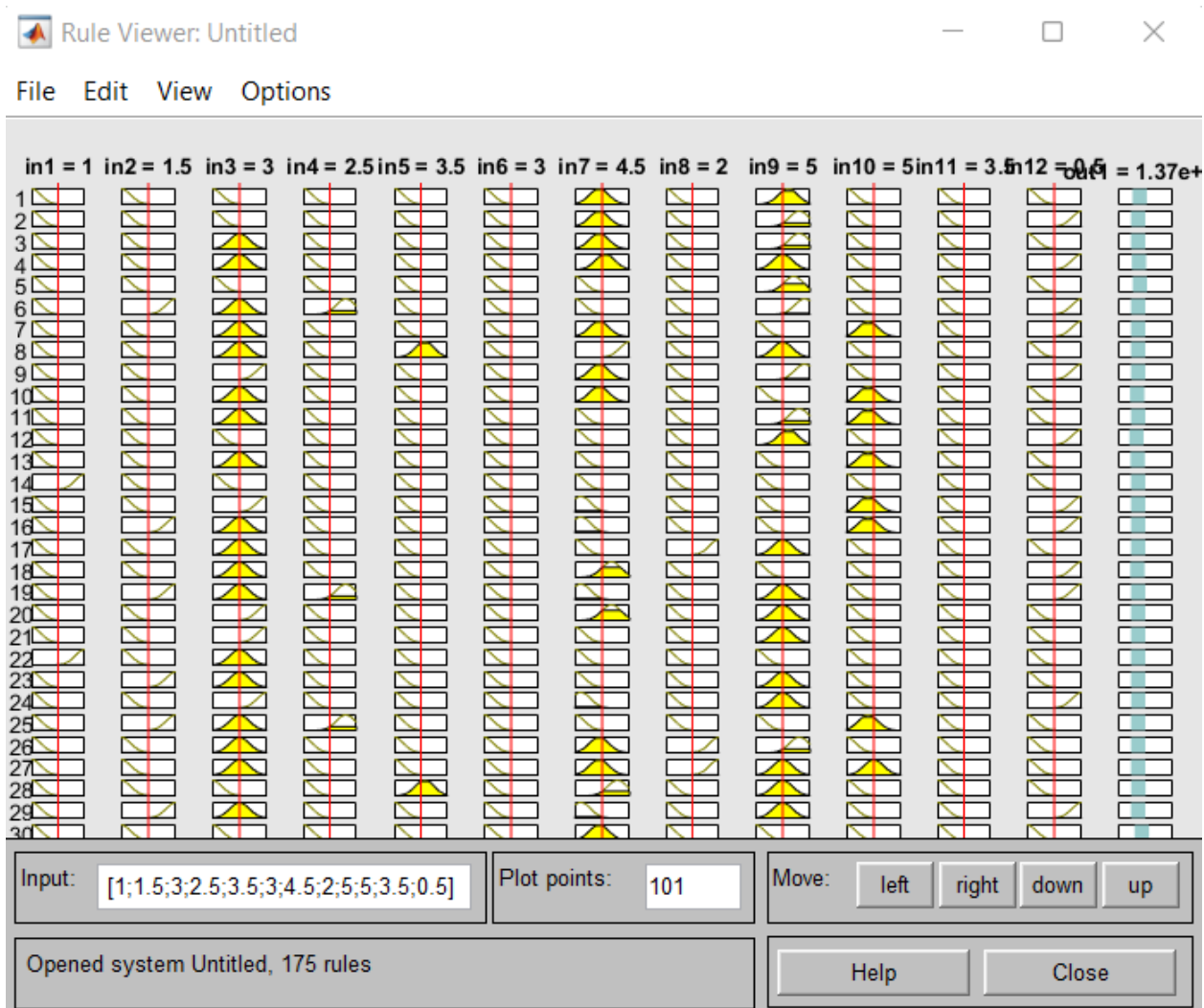


Figure – 5.3: Rule viewer of Dataset NFV

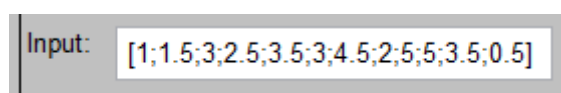


Figure – 5.4: Input Data's for NFV

In figure – 5.3 rule viewer represented the median fold drug resistance score and in figure – 5.4 where the input values are given for mutation position [30- 1, 32- 1.5, 46- 3, 47- 2.5, 48- 3.5, 50- 3, 54- 4.5, 76- 2, 82- 5, 84- 5, 88- 3.5, 90- 0.5]. Plot points of input values is in figure – 5.5.

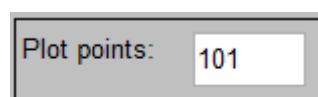


Figure – 5.5: Plot points for NFV

We input one set of values in figure 5.4. But what if we change input values or replace with another values, so in figure 5.6 input values are replaced by another values and drug's median fold resistance scores are interpreted accordingly.



Figure – 5.6: Rule viewer of dataset SQV

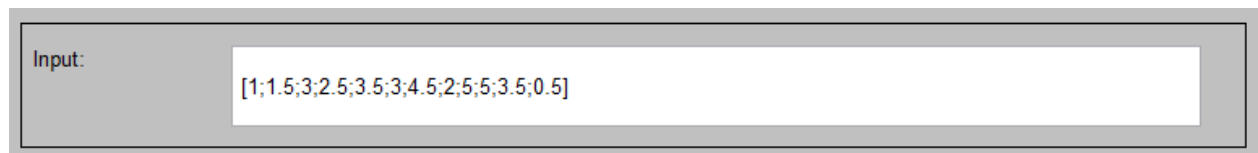


Figure – 5.7: Input values for SQV

In figure – 5.6 rule viewer represented the median fold drug resistance score and in figure - 5.7 where input values are given for mutation position. Plot points selected for this set of values.

Plot points:

101

Figure – 5.8: Plot points for SQV

We do the same for the entire protease sheet and for each drug.

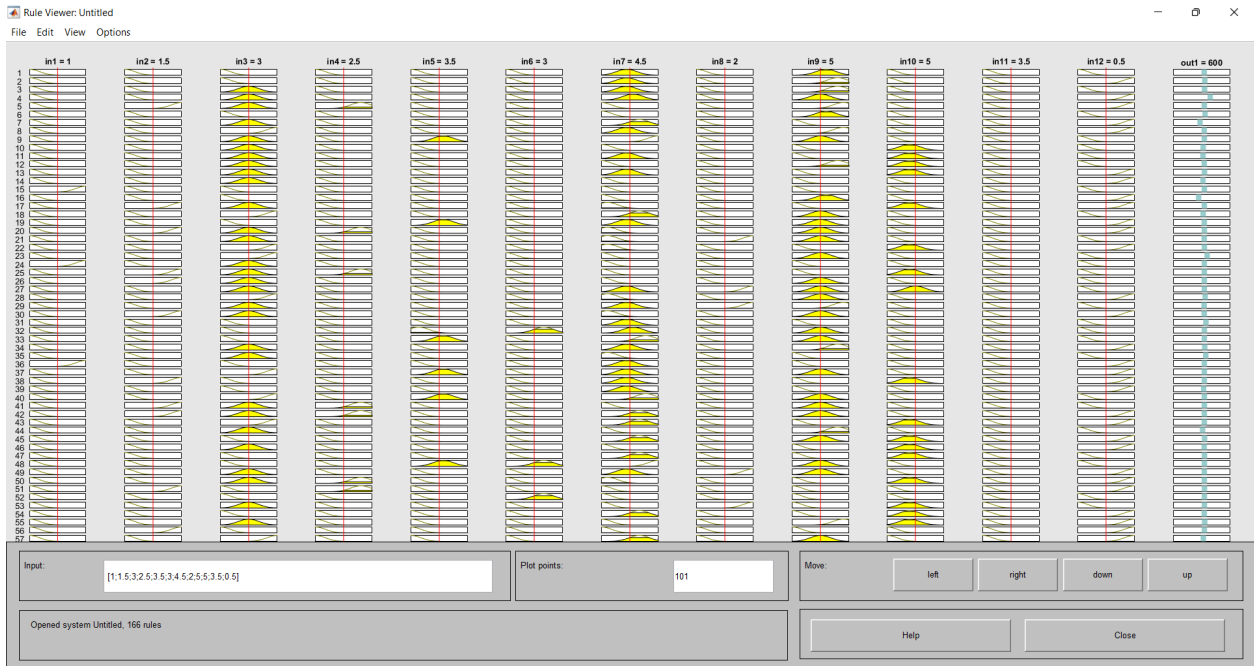


Figure – 5.9: Rule viewer of dataset IDV

Input:

[1;1.5;3;2.5;3.5;3.4;5;2.5;5.3;5;0.5]

Figure – 5.10: Input values for IDV

In figure – 5.9 rule viewer represented the median fold drug resistance score and in figure - 5.10 where input values are given for mutation position. Plot points selected for this set of values.

Plot points:

101

Figure – 5.11: Plot points for IDV



Figure – 5.12: Rule viewer of dataset FPV

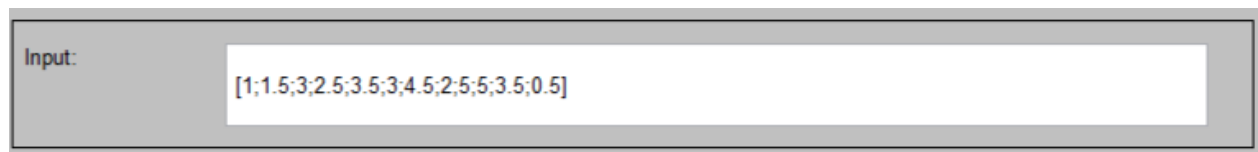


Figure – 5.13: Input values for FPV

In figure – 5.12 rule viewer represented the median fold drug resistance score and in figure - 5.13 where input values are given for mutation position. Plot points selected for this set of values.



Figure – 5.14: Plot points for FPV

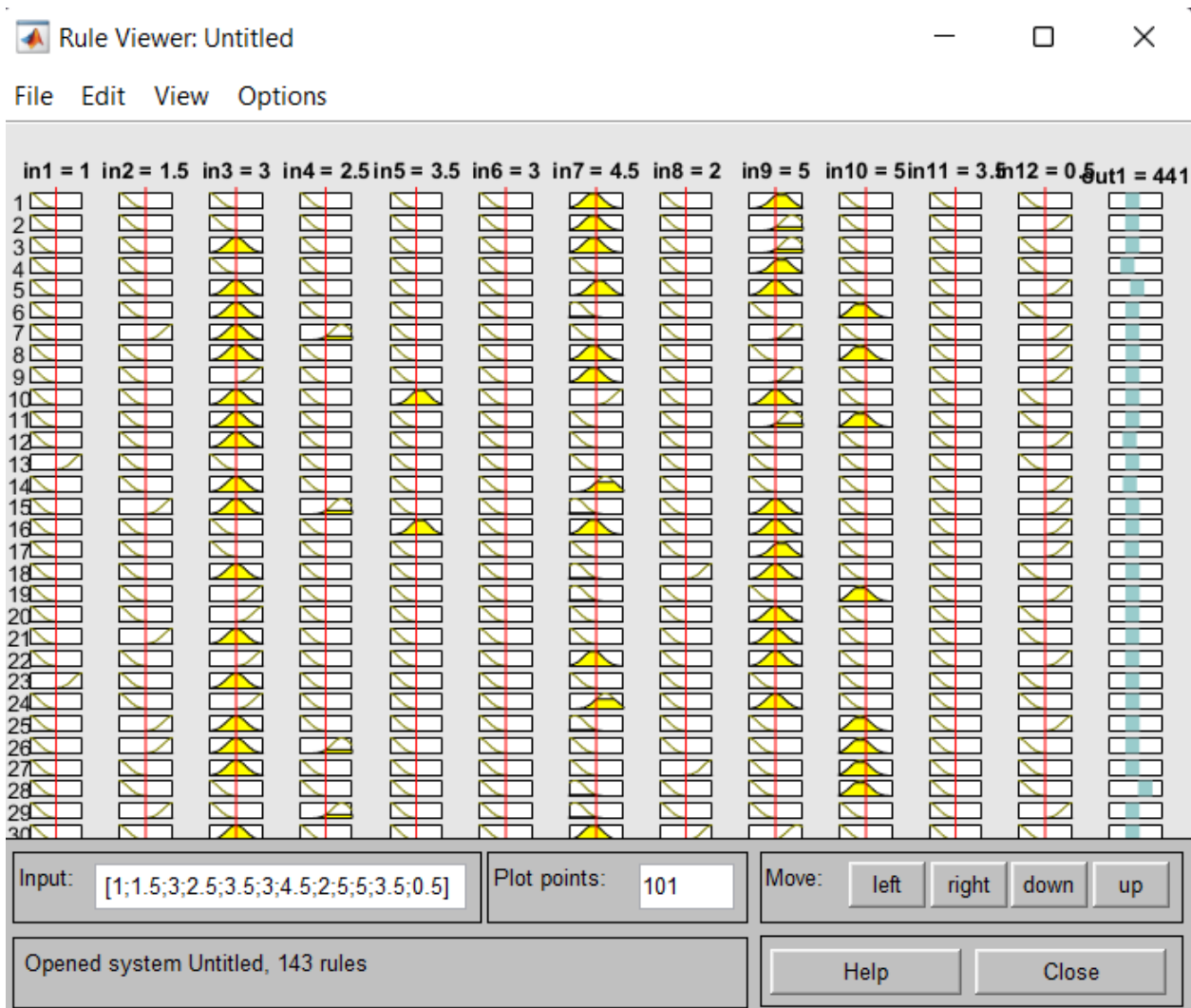


Figure – 5.15: Rule viewer of dataset ATV

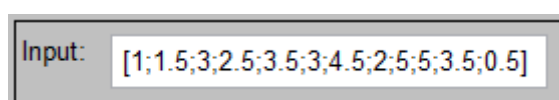


Figure – 5.16: Input values for ATV

In figure – 5.15 rule viewer represented the median fold drug resistance score and in figure - 5.16 where input values are given for mutation position. Plot points selected for this set of values.

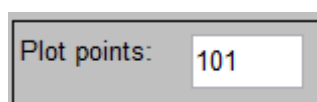


Figure – 5.17: Plot points for ATV

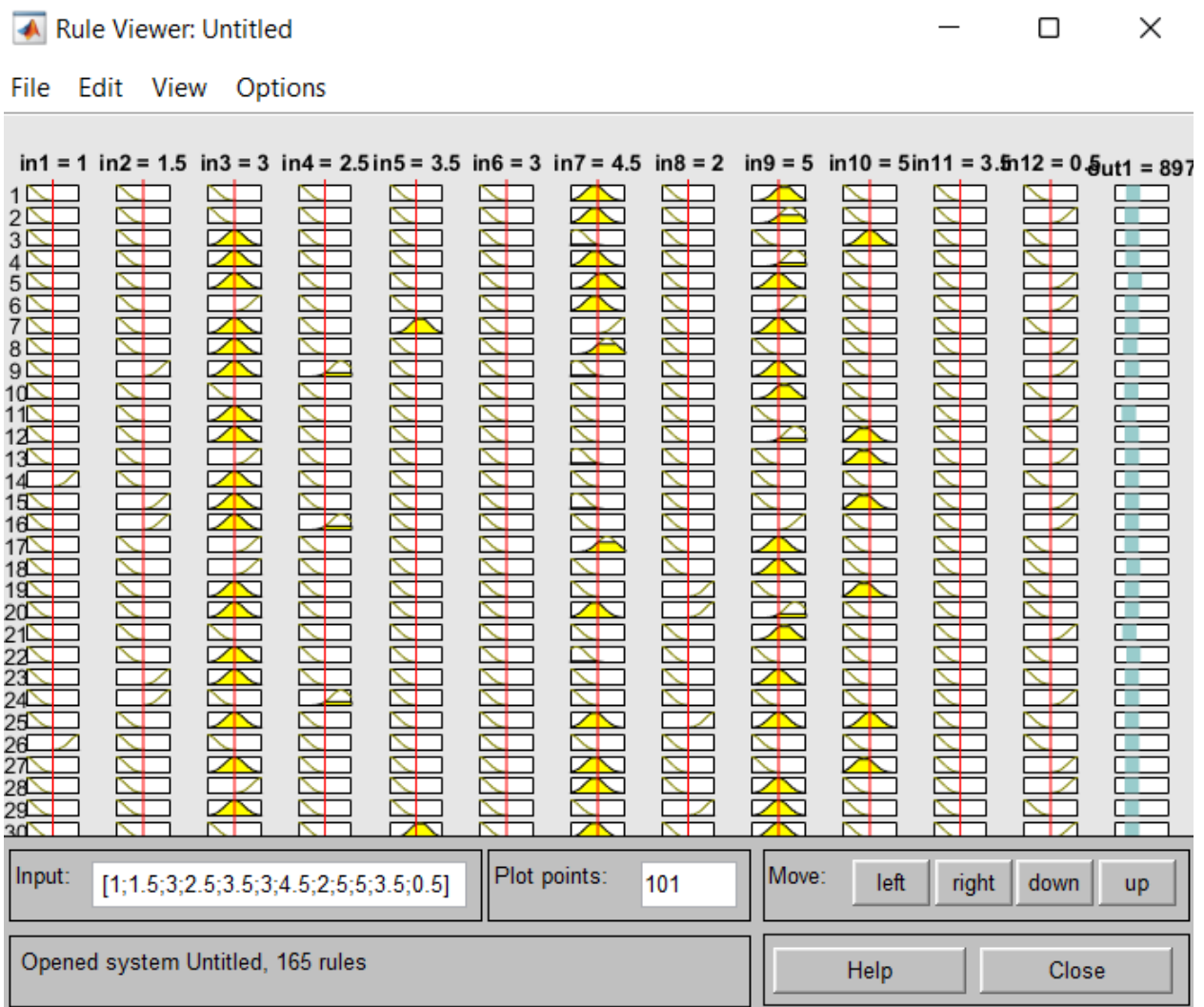


Figure – 5.18: Rule viewer of dataset LPV

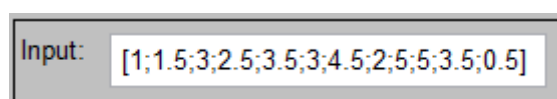


Figure – 5.19: Input values for LPV

In figure – 5.18 rule viewer represented the median fold drug resistance score and in figure - 5.19 where input values are given for mutation position. Plot points selected for this set of values.

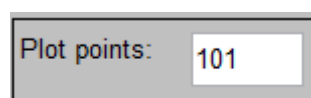


Figure – 5.20: Plot points for LPV

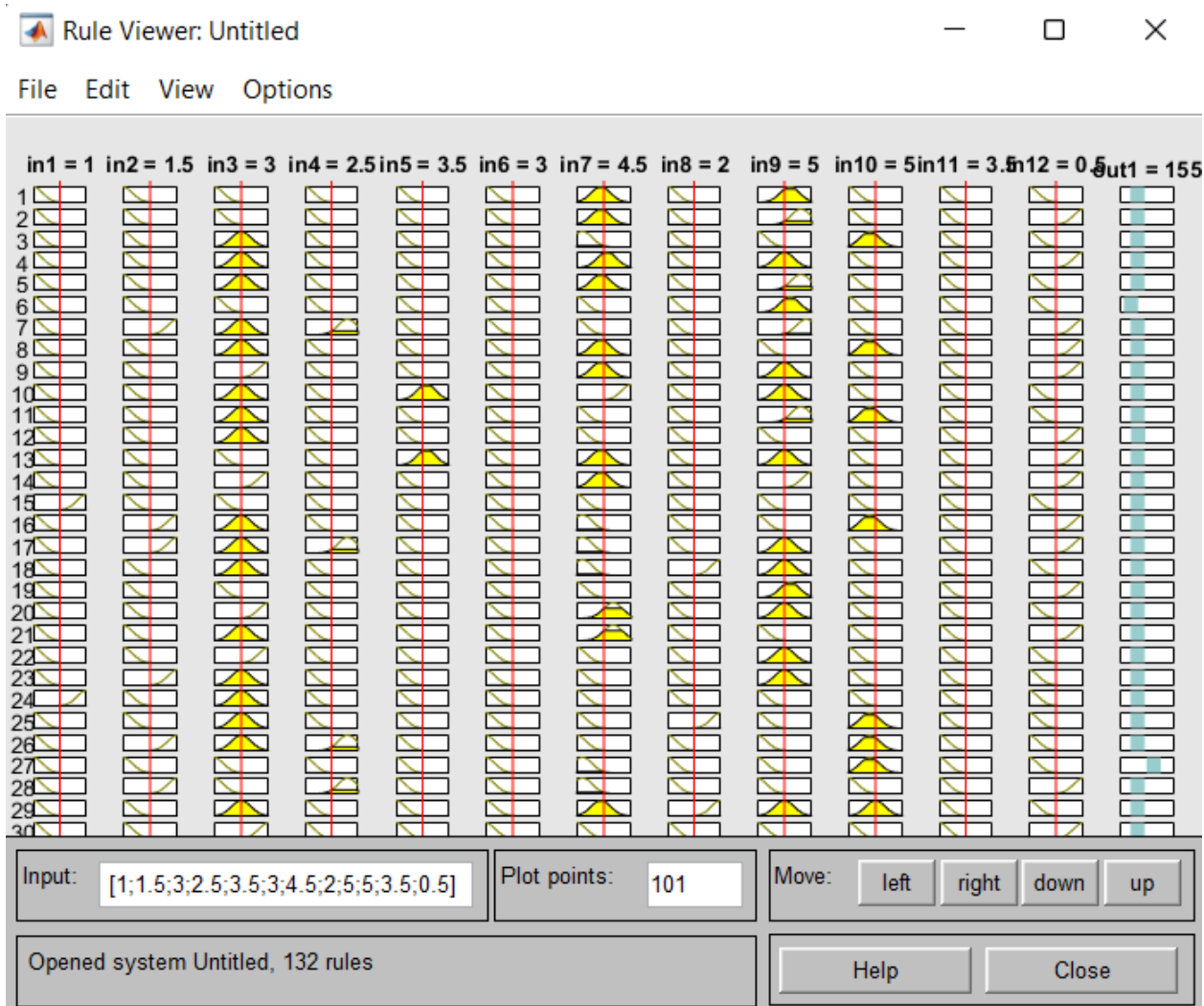


Figure – 5.21: Rule viewer of dataset TPV

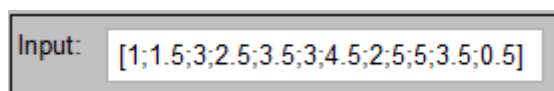


Figure – 5.22: Input values for TPV

In figure – 5.18 rule viewer represented the median fold drug resistance score and in figure - 5.19 where input values are given for mutation position. Plot points selected for this set of values.

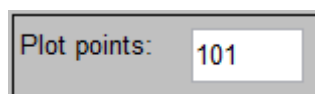


Figure – 5.23: Plot points for TPV

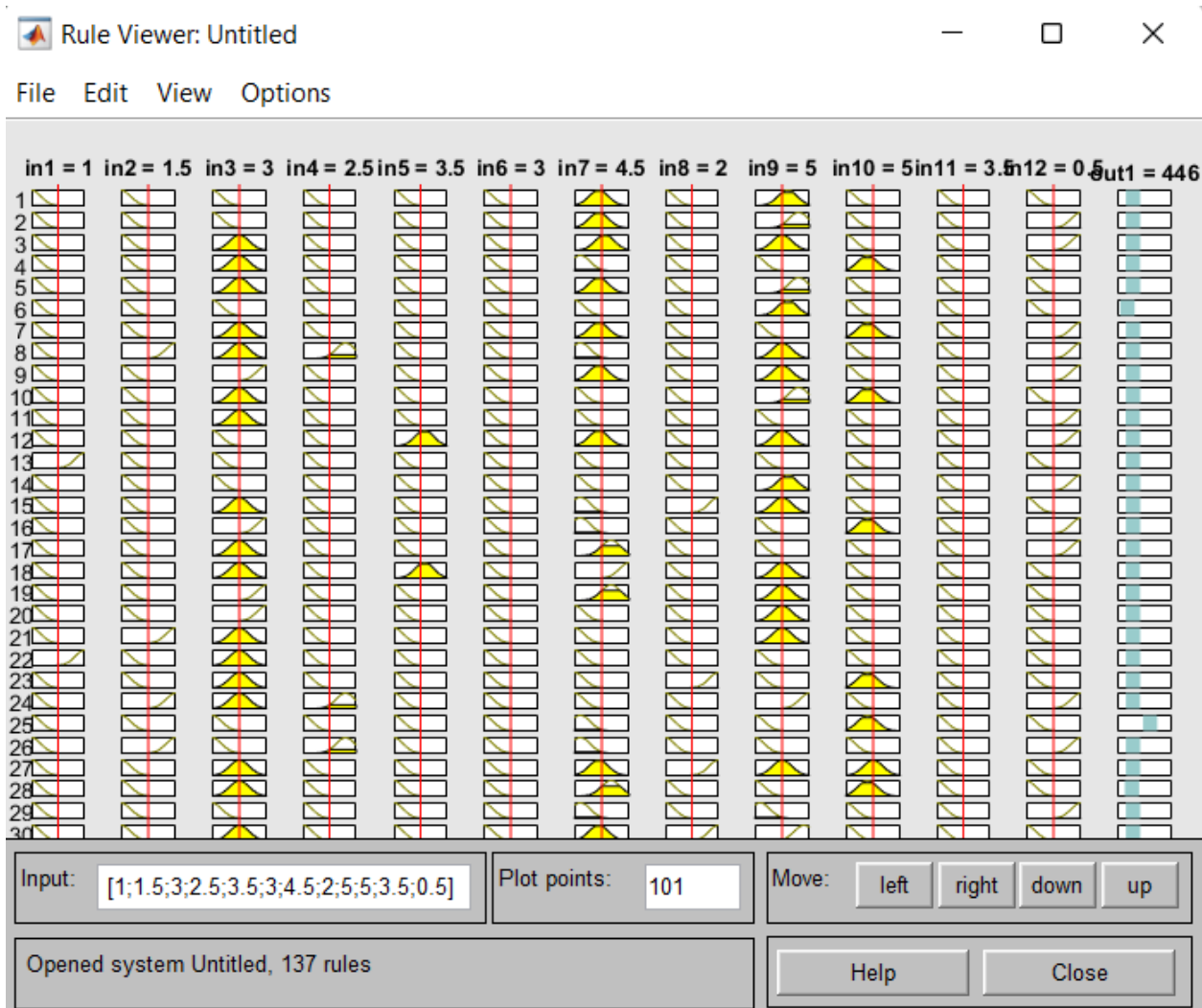


Figure – 5.24: Rule viewer of dataset of DRV

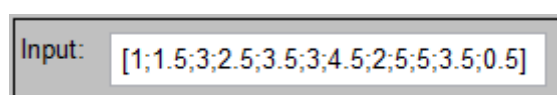


Figure – 5.25: Input values for DRV

In figure – 5.18 rule viewer represented the median fold drug resistance score and in figure - 5.19 where input values are given for mutation position. Plot points selected for this set of values.

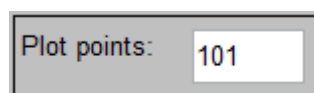


Figure – 5.26: Plot points for DRV

5.1.2. SURFACE VIEW OF STANFORD HIV DRUG RESISTANCE DATASET (PROTEASE):

Surface view tool provides graphical representation as in three – dimensional structure provided median fold drug resistance score and mutation positions. The figures () given below represents surface structure. According to the scores of x, y, z axis, peaks are performed, and peaks represent higher median fold drug resistance.

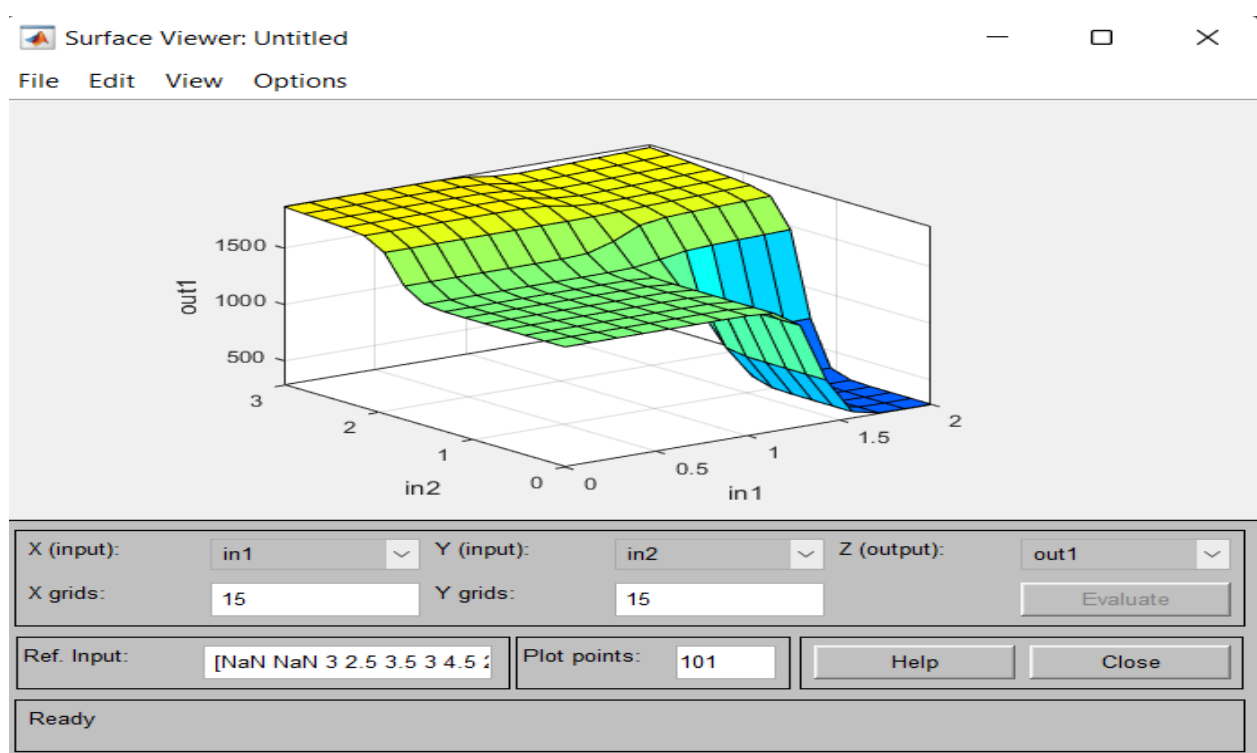


Figure – 5.27: Surface plot of NFV

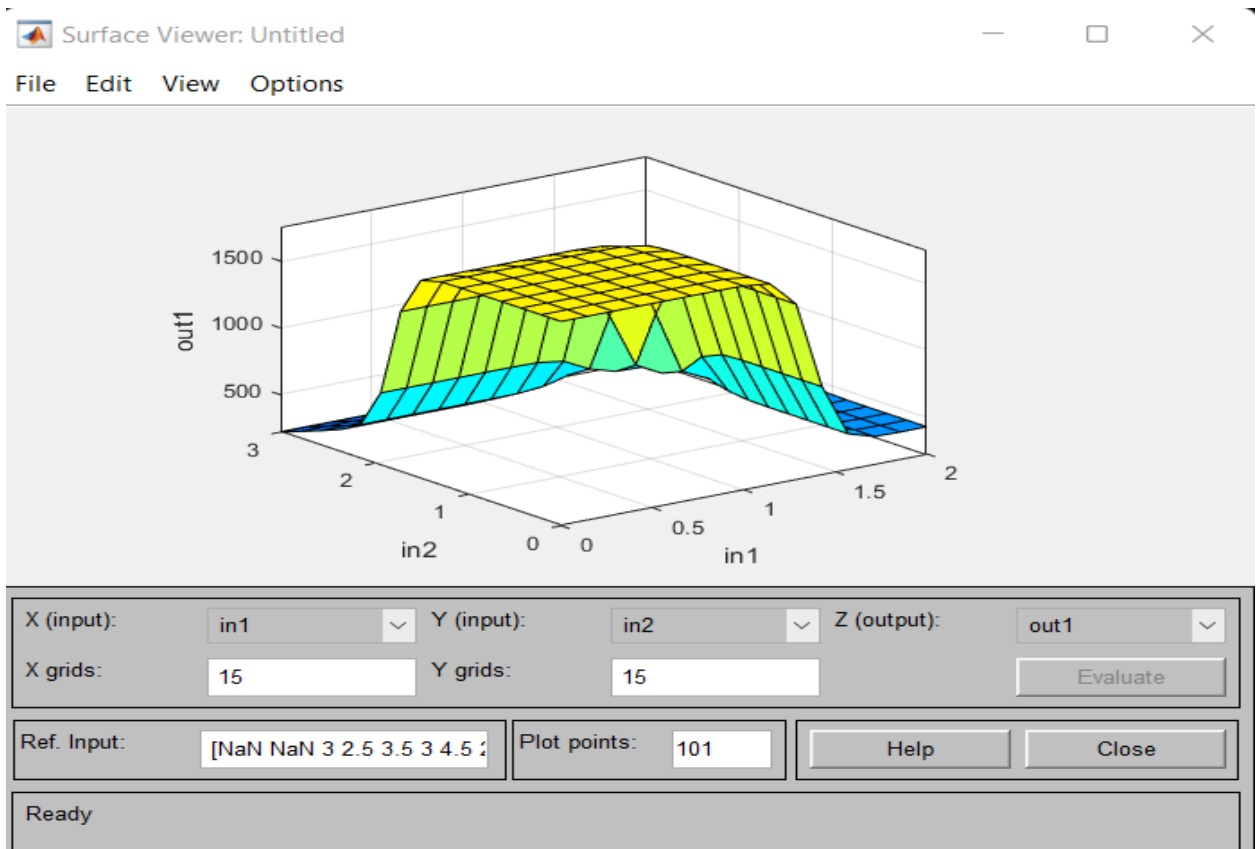


Figure – 5.28: Surface plot of SQV

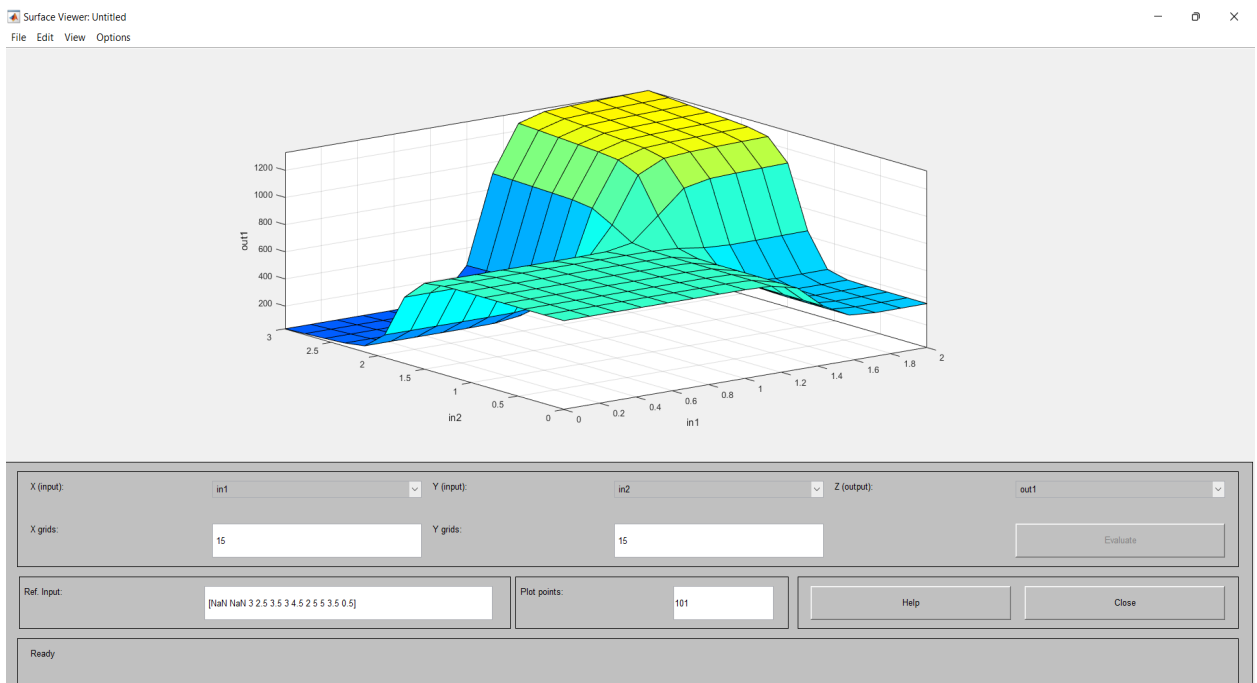


Figure – 5.29: Surface plot of IDV

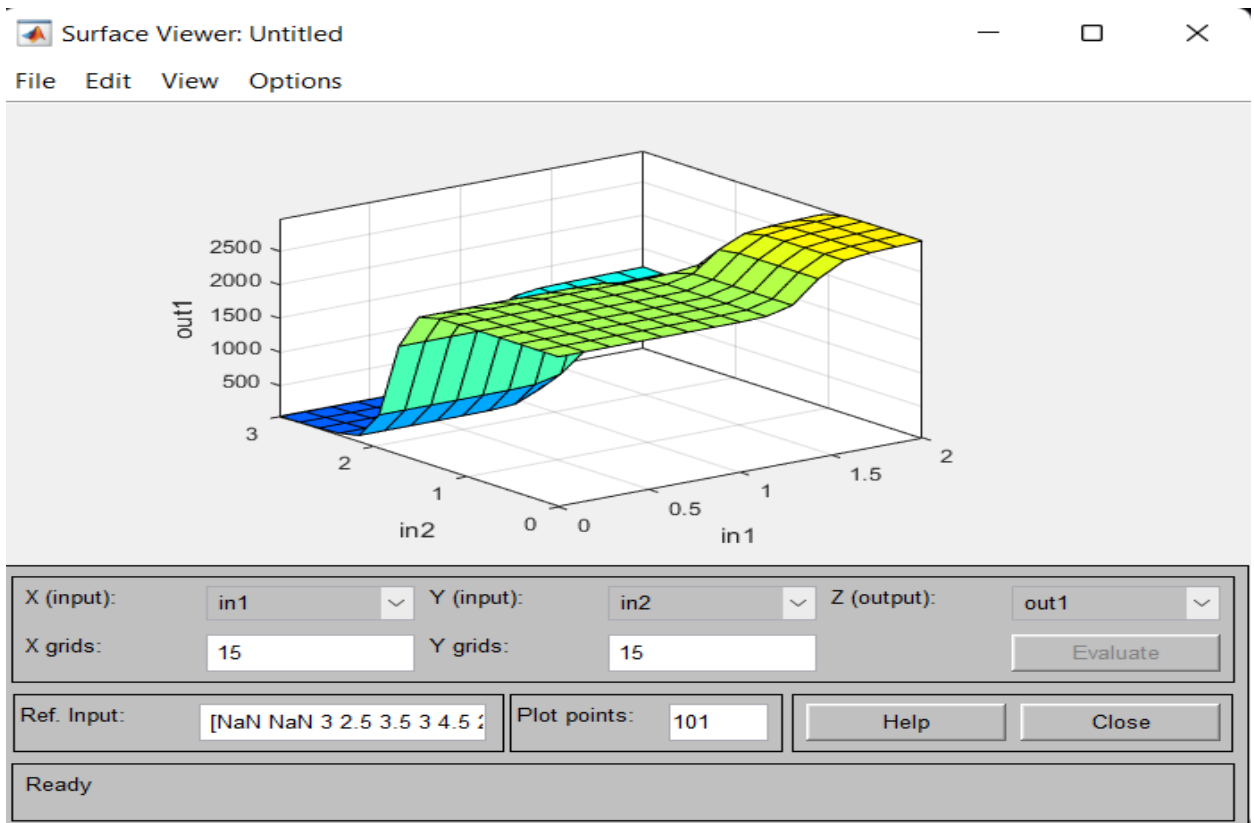


Figure – 5.30: Surface plot of FPV

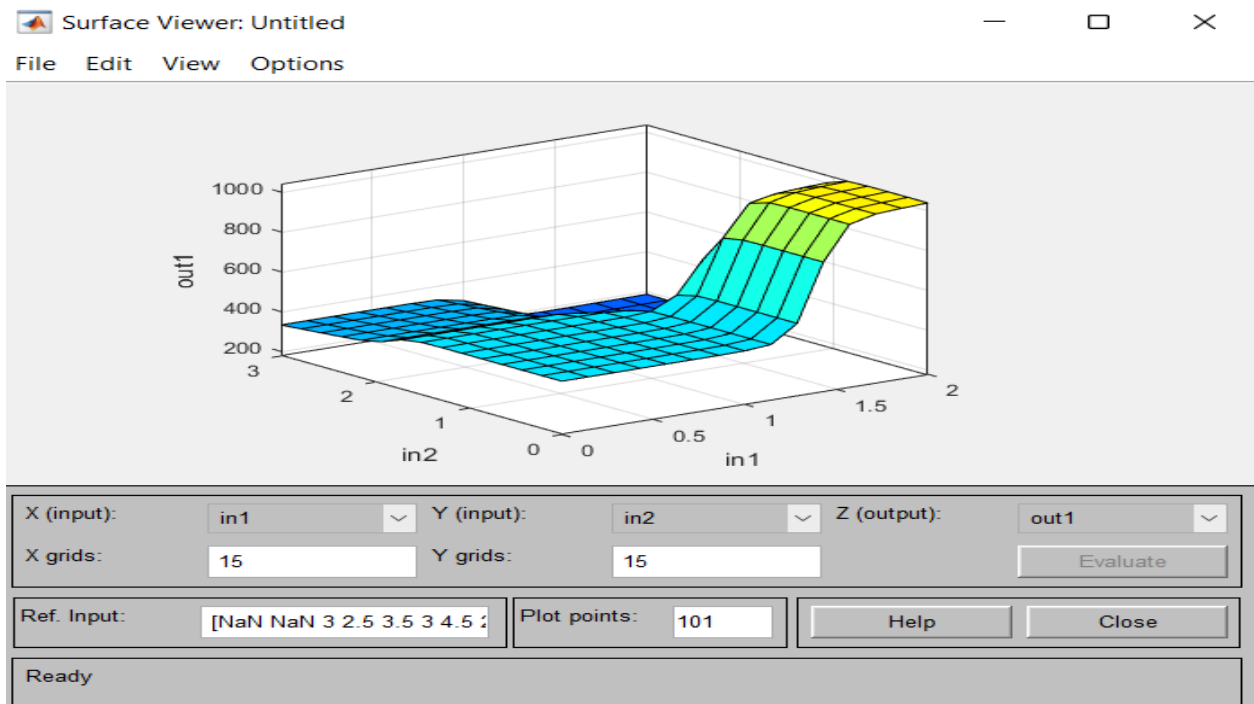


Figure – 5.31: Surface plot of ATV

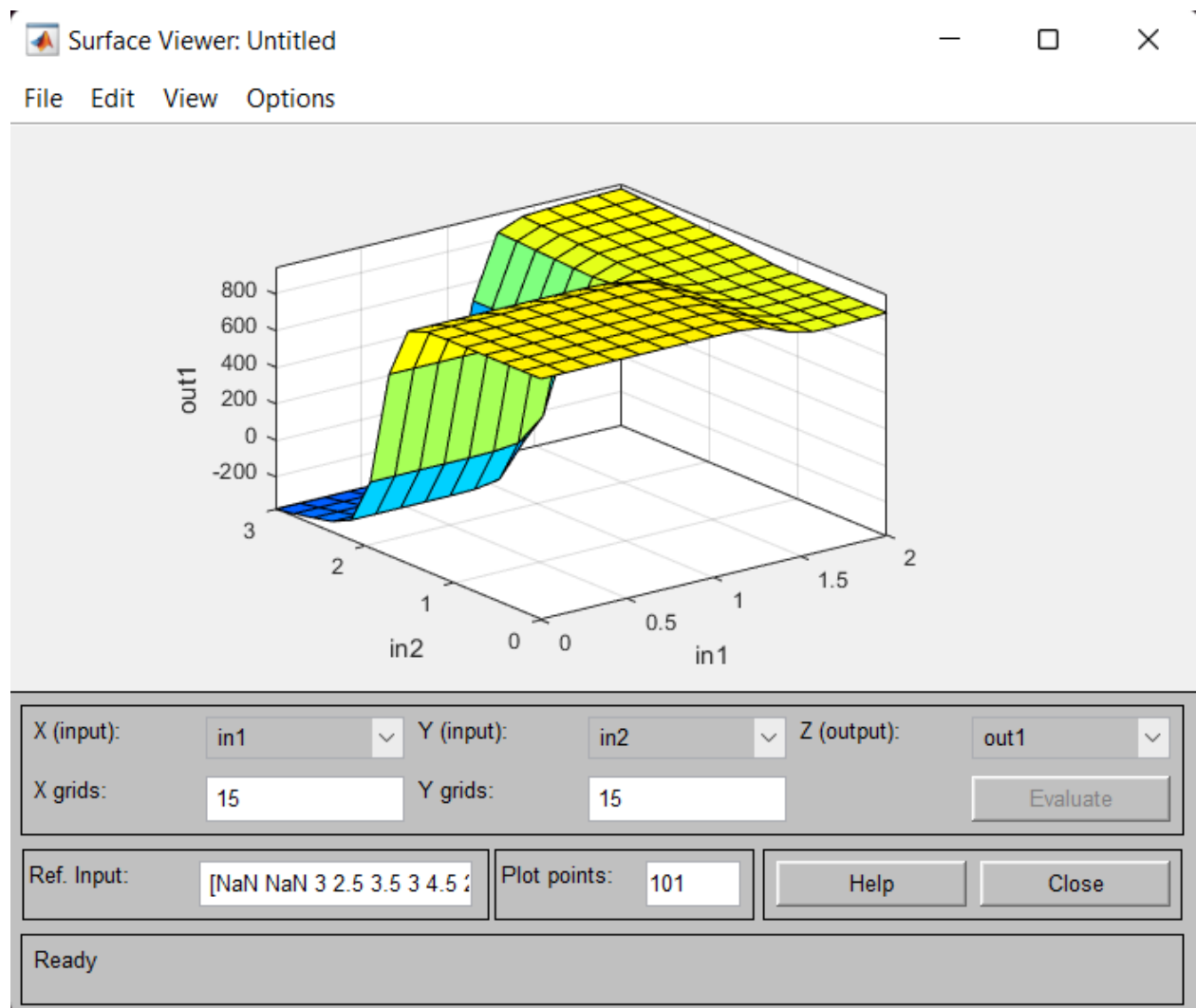


Figure – 5.32: Surface plot of LPV

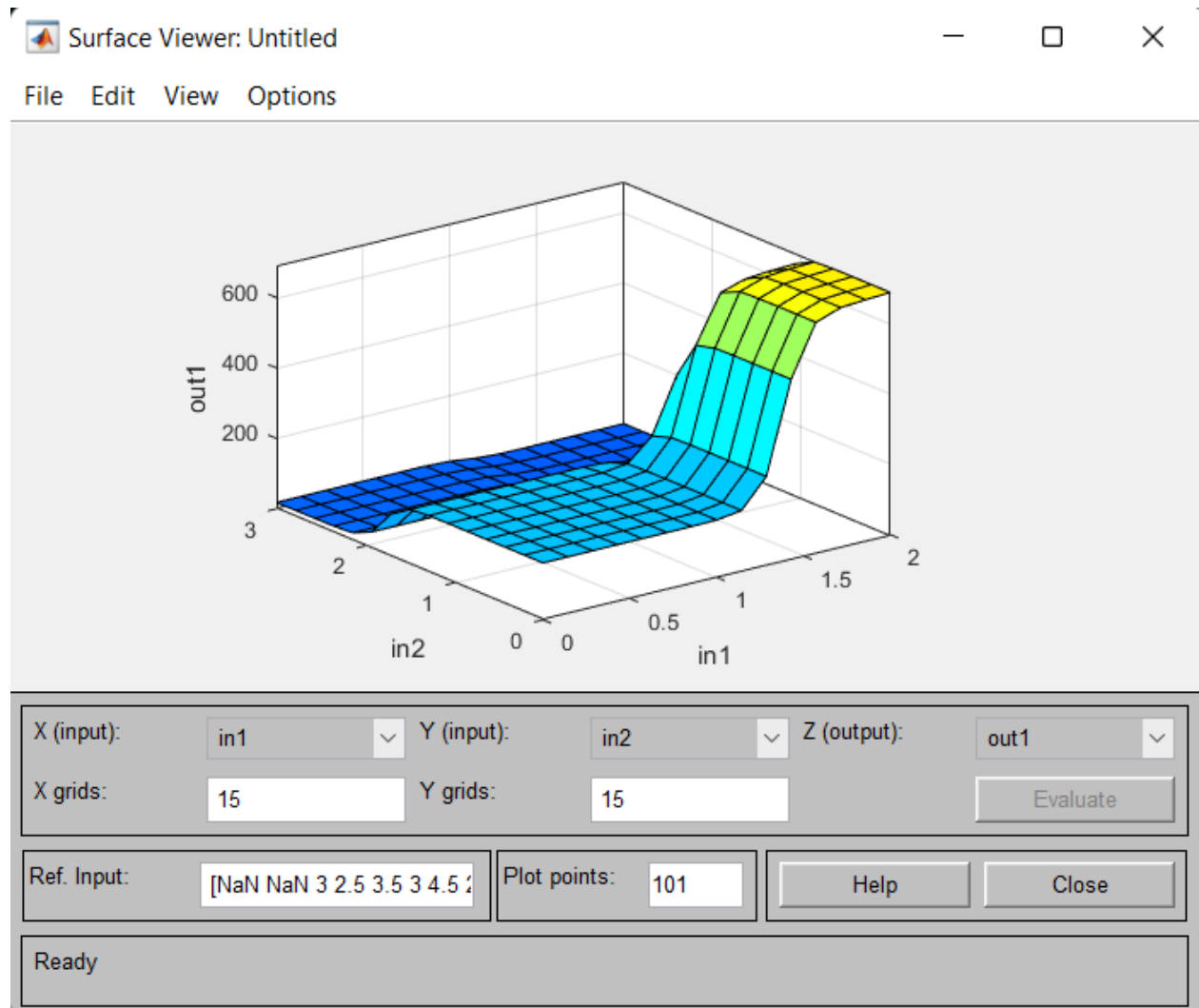


Figure – 5.33: Surface plot of TPV

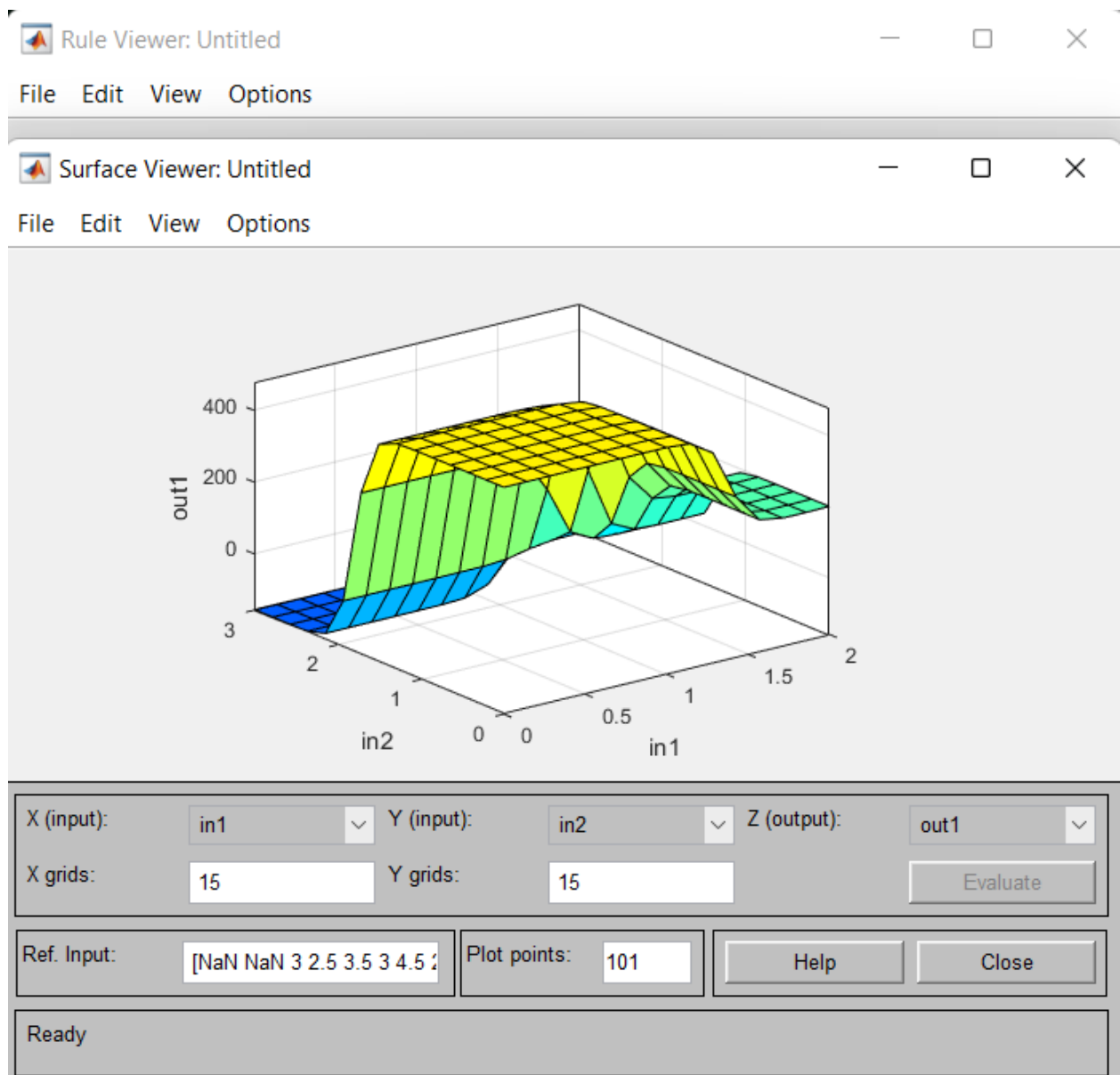


Figure – 5.34: Surface plot of DRV

5.2 EXCEL SHEET OF NRTI:

	A	B	C	D	E	F	G	H	I	J	K	L
1	0	0	0	0	0	0	0	0	1	0	0	0.8419
2	0	0	2	0	3	0	0	0	1	0	0	0.5124
3	4	0	0	0	0	0	0	0	1	5	6	3.732
4	4	0	0	0	0	0	0	0	1	0	6	1851
5	4	0	2	0	0	0	0	0	1	5	6	641
6	0	3	0	0	0	0	0	0	1	0	0	3053
7	0	0	0	0	3	0	0	0	1	0	0	0.416
8	4	0	0	0	0	0	0	0	0	0	0	0.814
9	4	0	2	0	0	0	0	0	0	5	6	0.99
10	0	3	0	0	0	0	0	0	0	0	0	40135
11	4	0	0	0	0	0	0	0	0	5	6	0.519
12	0	0	0	0	0	0	0	0	1	0	6	16421
13	0	0	2	0	3	0	0	0	1	0	7	1.316
14	0	0	2	0	3	0	0	0	0	0	0	7.77
15	0	0	0	0	0	1	0	0	1	0	0	2823
16	4	0	0	0	0	0	0	0	0	0	6	0.39
17	0	0	0	0	3	0	0	0	0	0	0	1212
18	4	0	2	8	0	0	0	0	1	5	6	5.315
19	4	0	0	0	0	0	0	0	1	0	0	4328
20	0	0	0	0	0	0	0	0	1	0	7	0.57
21	4	0	0	0	0	0	0	0	1	0	7	2.33
22	0	0	0	0	0	0	0	0	9	0	0	3.57
23	0	0	2	0	3	0	0	0	0	0	7	0.317
24	0	0	2	0	0	0	0	0	1	0	0	10215
25	4	0	2	0	3	0	0	0	1	0	7	0.84
26	0	0	2	0	0	0	0	0	0	0	0	5919
27	0	0	0	0	0	0	0	0	0	0	6	1.27
28	4	0	2	0	0	1	0	0	1	5	6	7.415

Figure – 5.35: Final excel sheet for analysis

5.1.3. RULE VIEWER OF STANFORD HIV DRUG RESISTANCE DATABASE (NRTI):

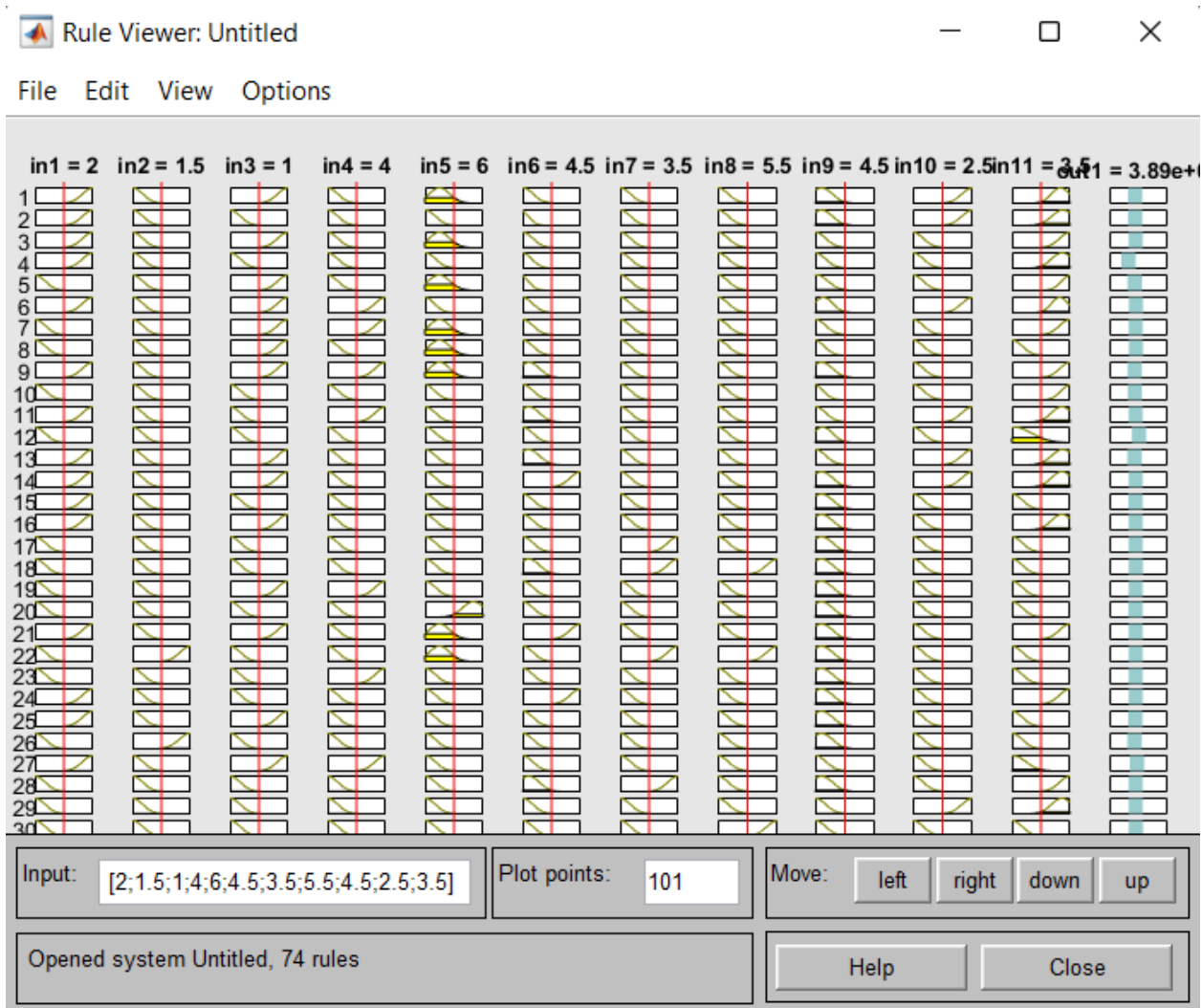


Figure – 5.36: Rule viewer of Dataset AZT

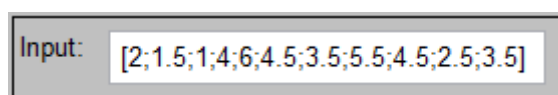


Figure – 5.37: Input Values

In figure – 5.36 rule viewer represented the median fold drug resistance score and in figure – 5.37 where input values are given for mutation position. Plot points selected for this set of values.

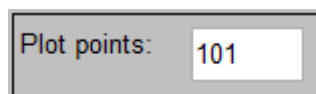


Figure – 5.38 Plot point

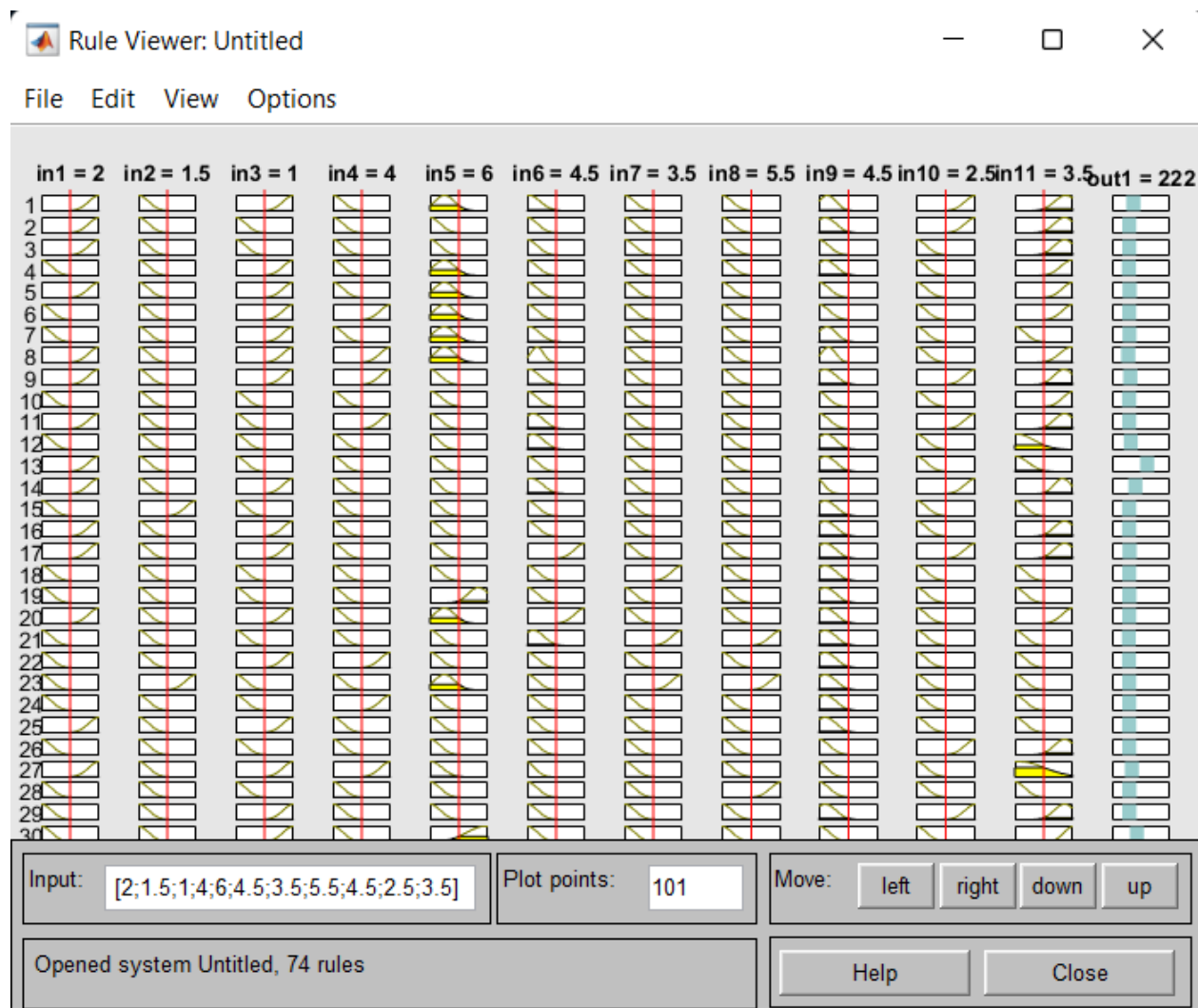


Figure – 5.39: Rule viewer of Dataset TDF

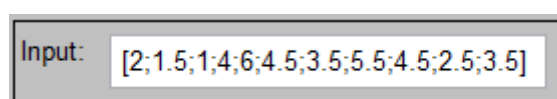


Figure – 5.40: Input values

In figure – 5.39 rule viewer represented the median fold drug resistance score and in figure – 5.40 where input values are given for mutation position. Plot points selected for this set of values.

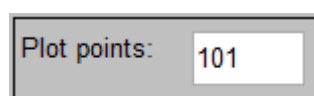


Figure – 5.41 Plot points

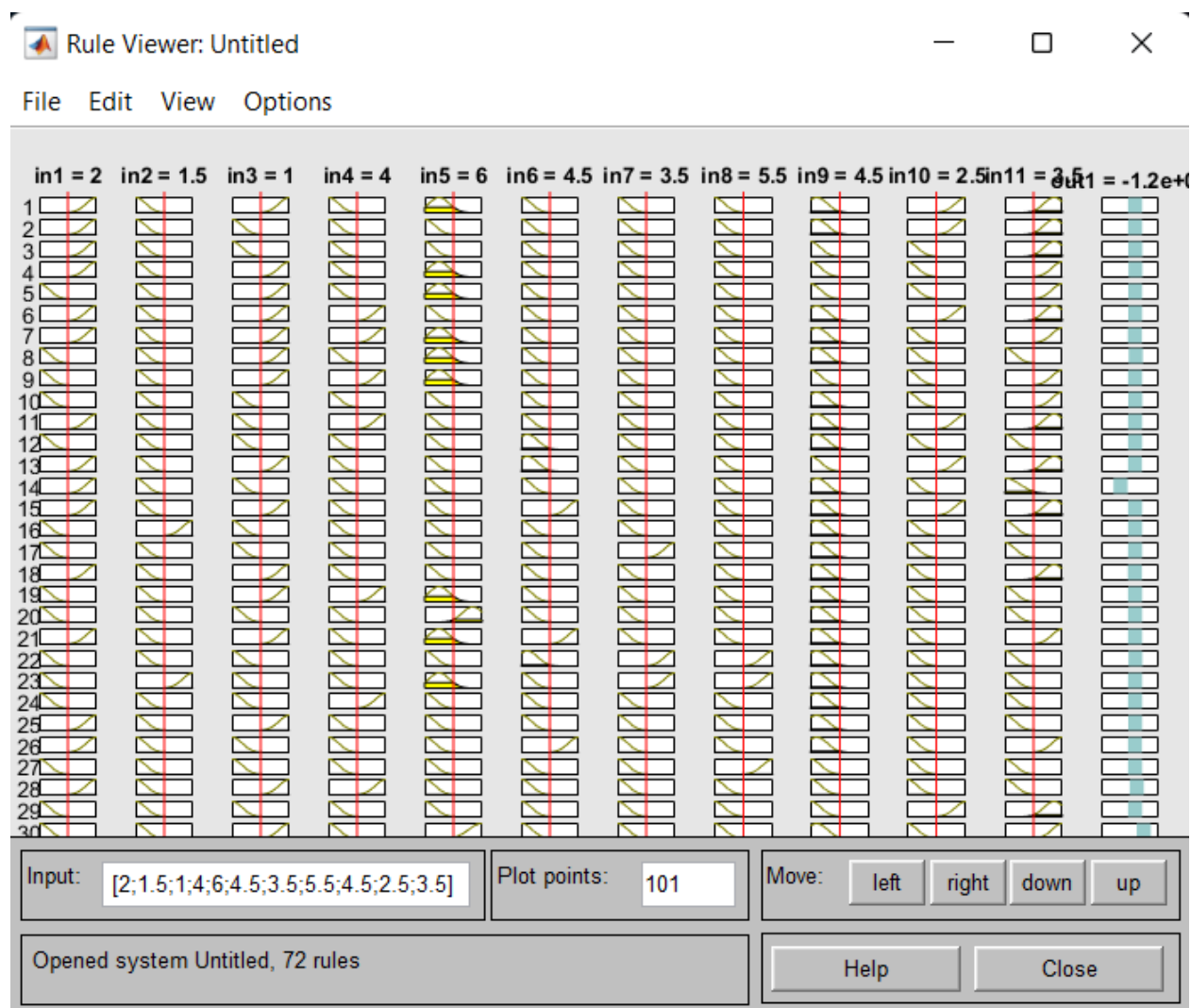


Figure – 5.42 Rule viewer of Dataset ABC

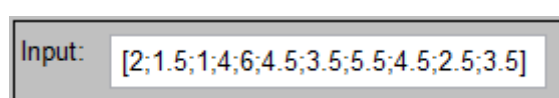


Figure – 5.43 Input values

In figure – 5.42 rule viewer represented the median fold drug resistance score and in figure – 5.43 where input values are given for mutation position. Plot points selected for this set of values.

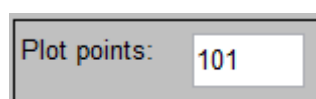


Figure – 5.44: Plot points

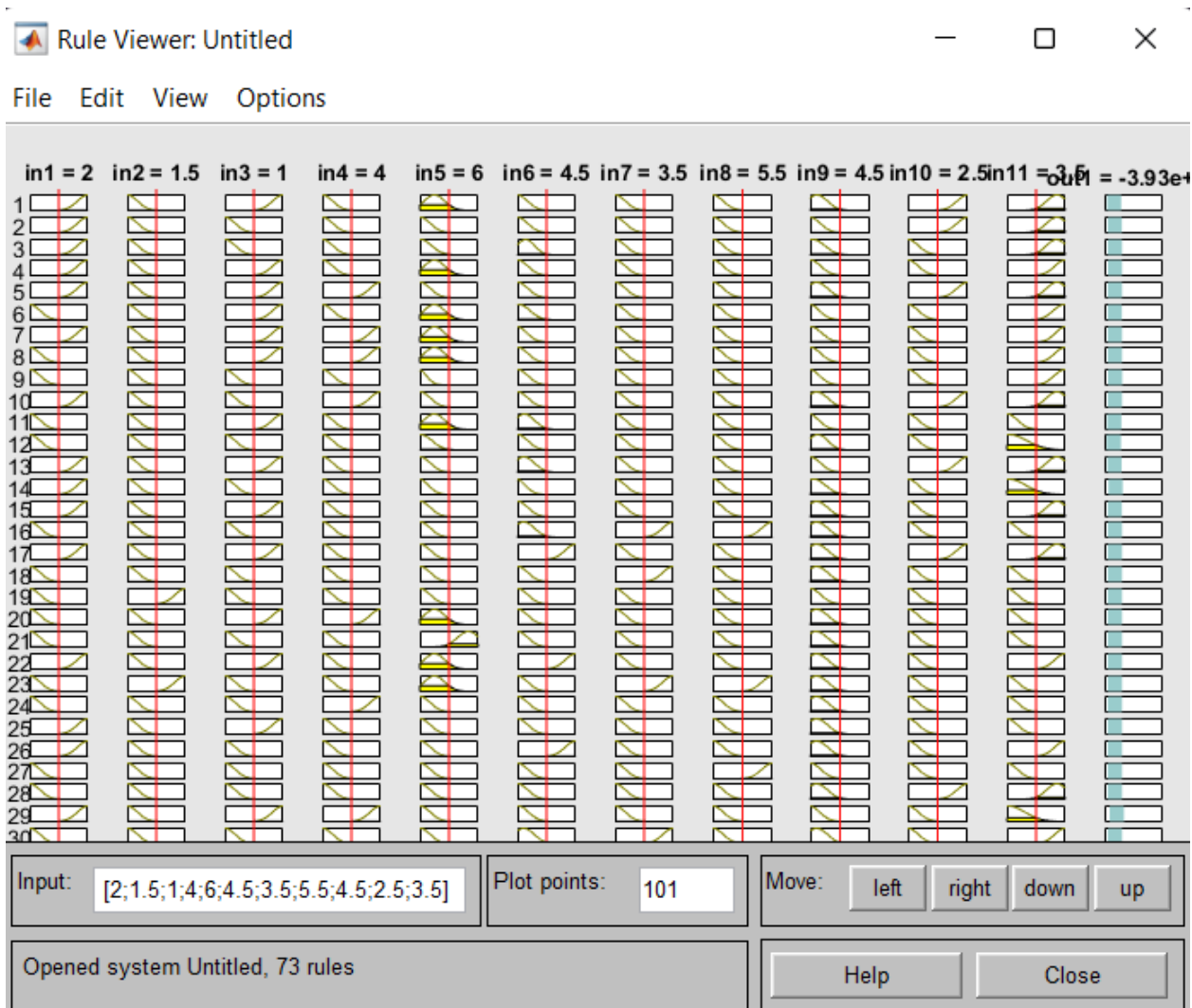


Figure – 5.45: Rule viewer of Dataset 3TC

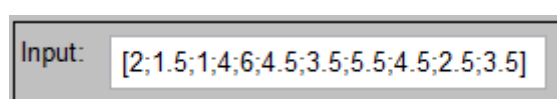


Figure – 5.46: Input values

In figure – 5.45 rule viewer represented the median fold drug resistance score and in figure – 5.46 where input values are given for mutation position. Plot points selected for this set of values.

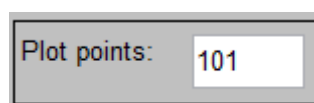


Figure – 5.47: Plot points

5.1.4 SURFACE VIEWER OF STANFORD HIV DRUG RESISTANCE DATASET (NRTI):

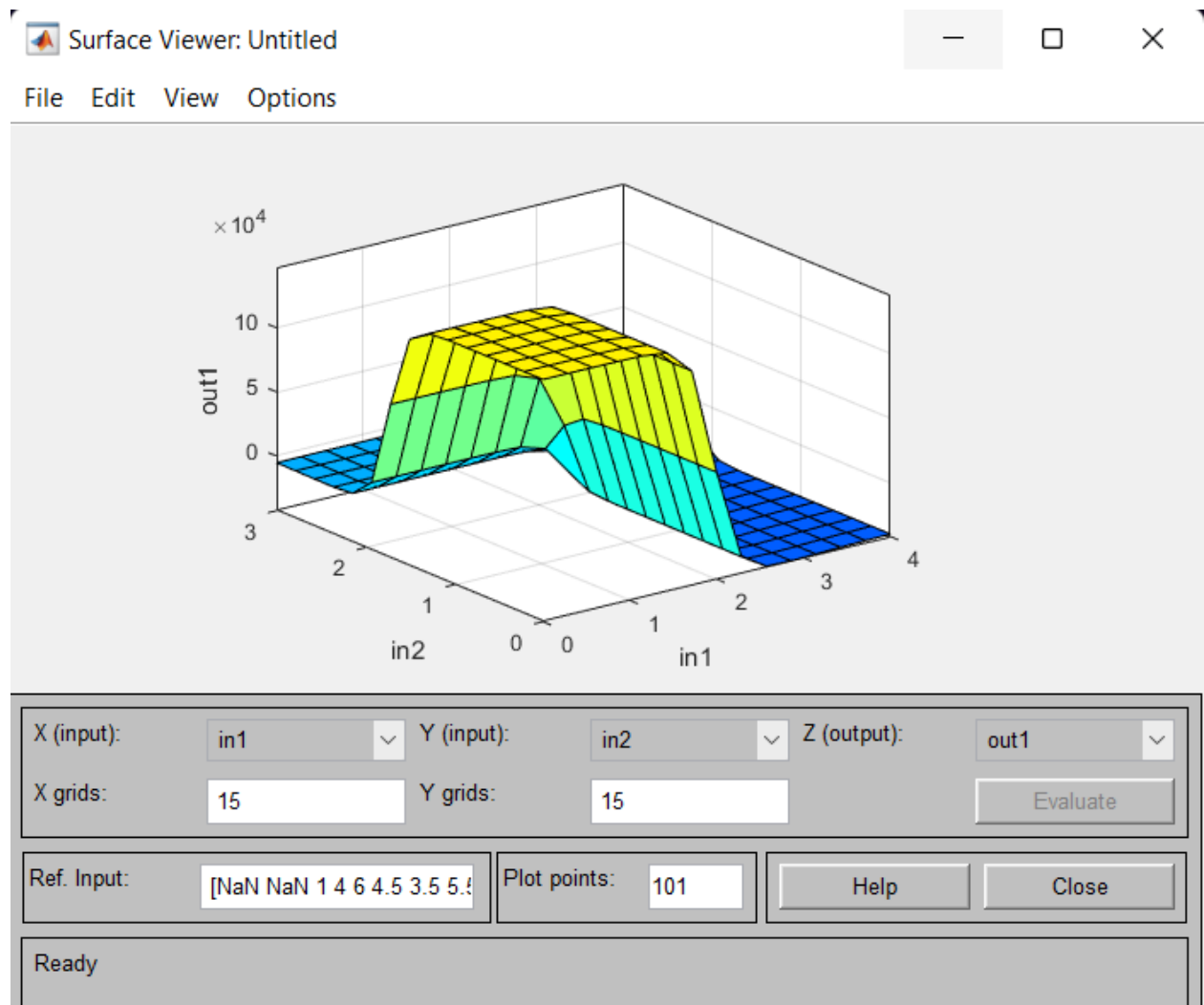


Figure – 5.48: Surface viewer of Dataset AZT

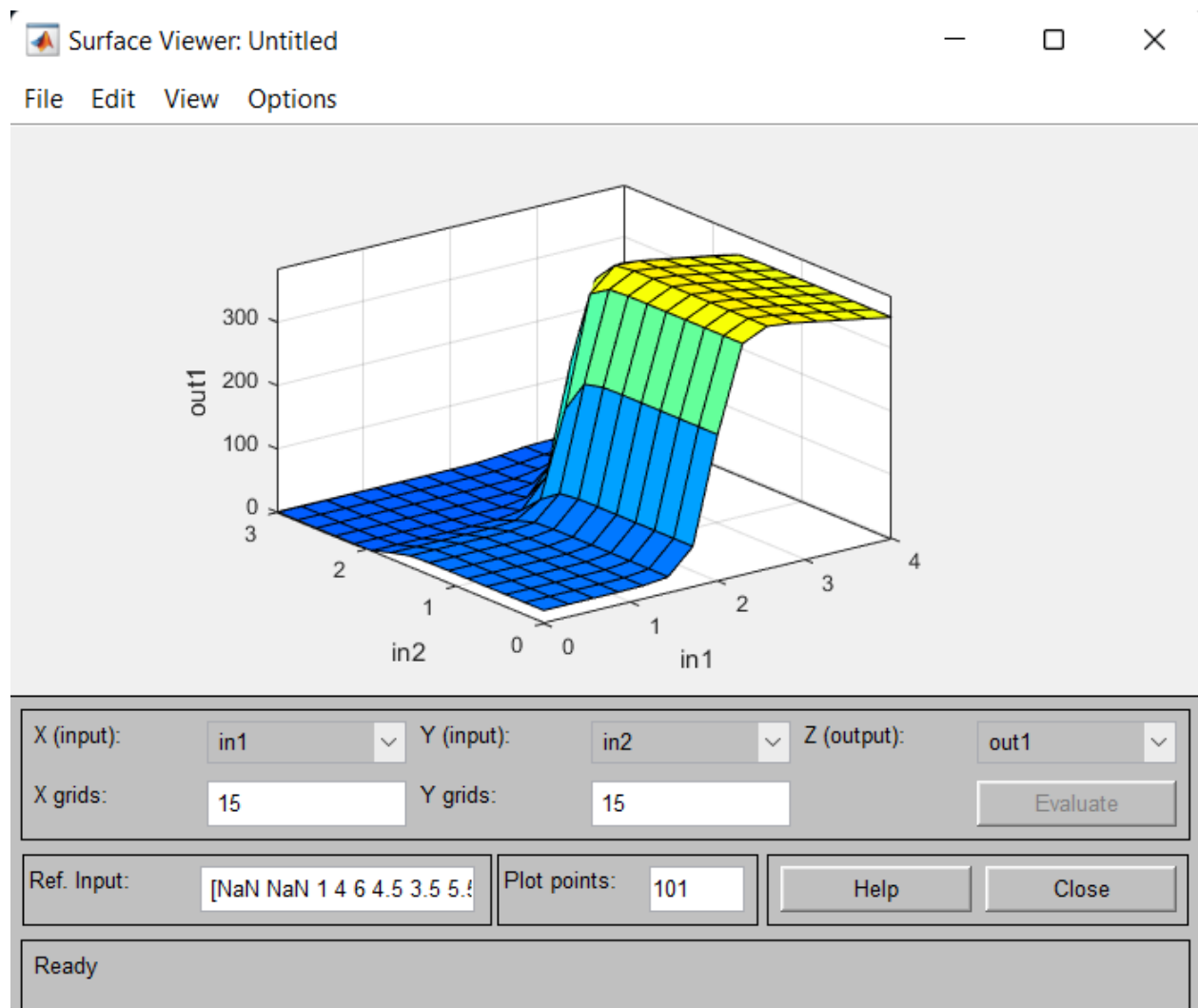


Figure – 5.49: Surface viewer of Dataset TDF

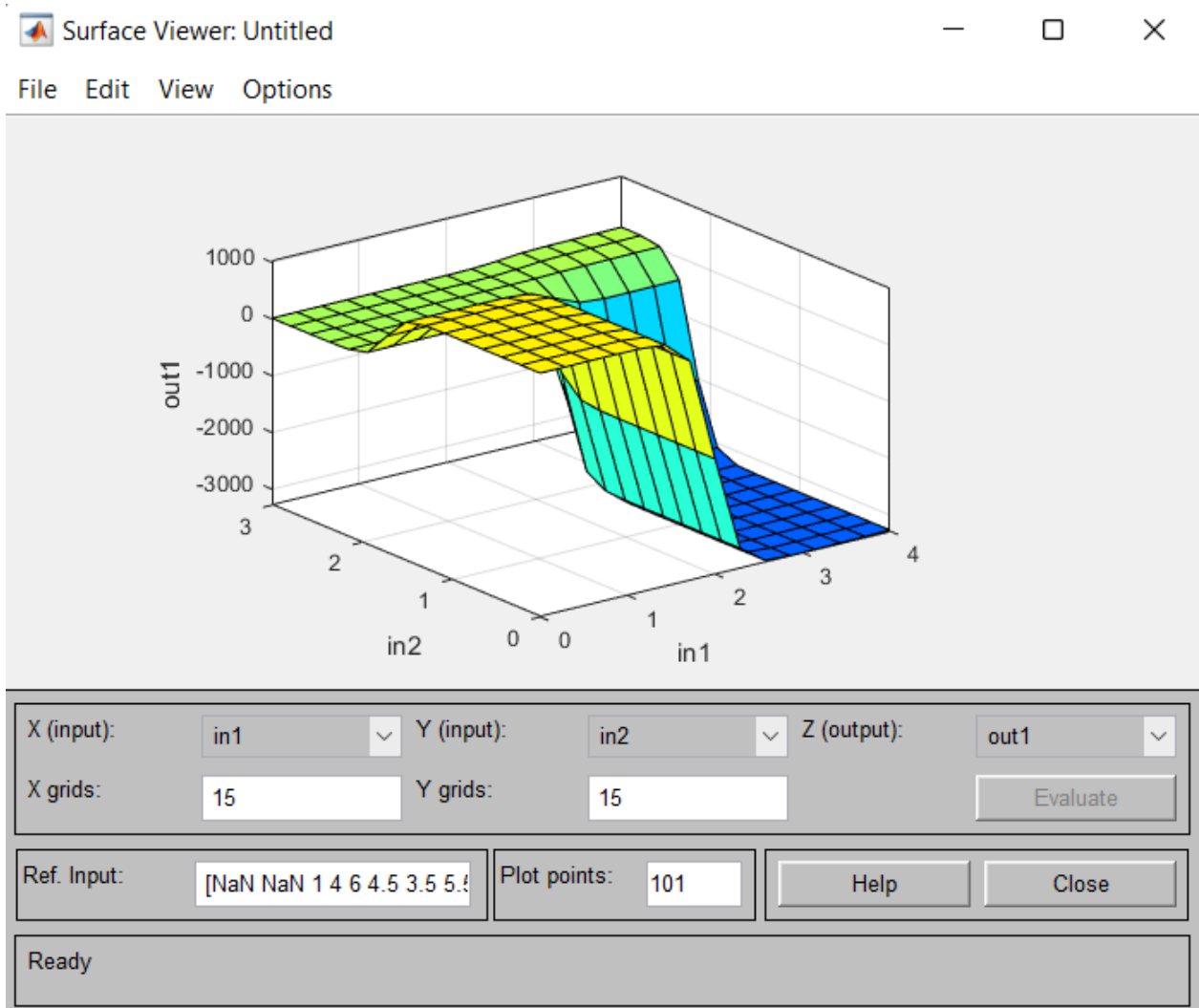


Figure – 5.50: Surface viewer of Dataset ABC

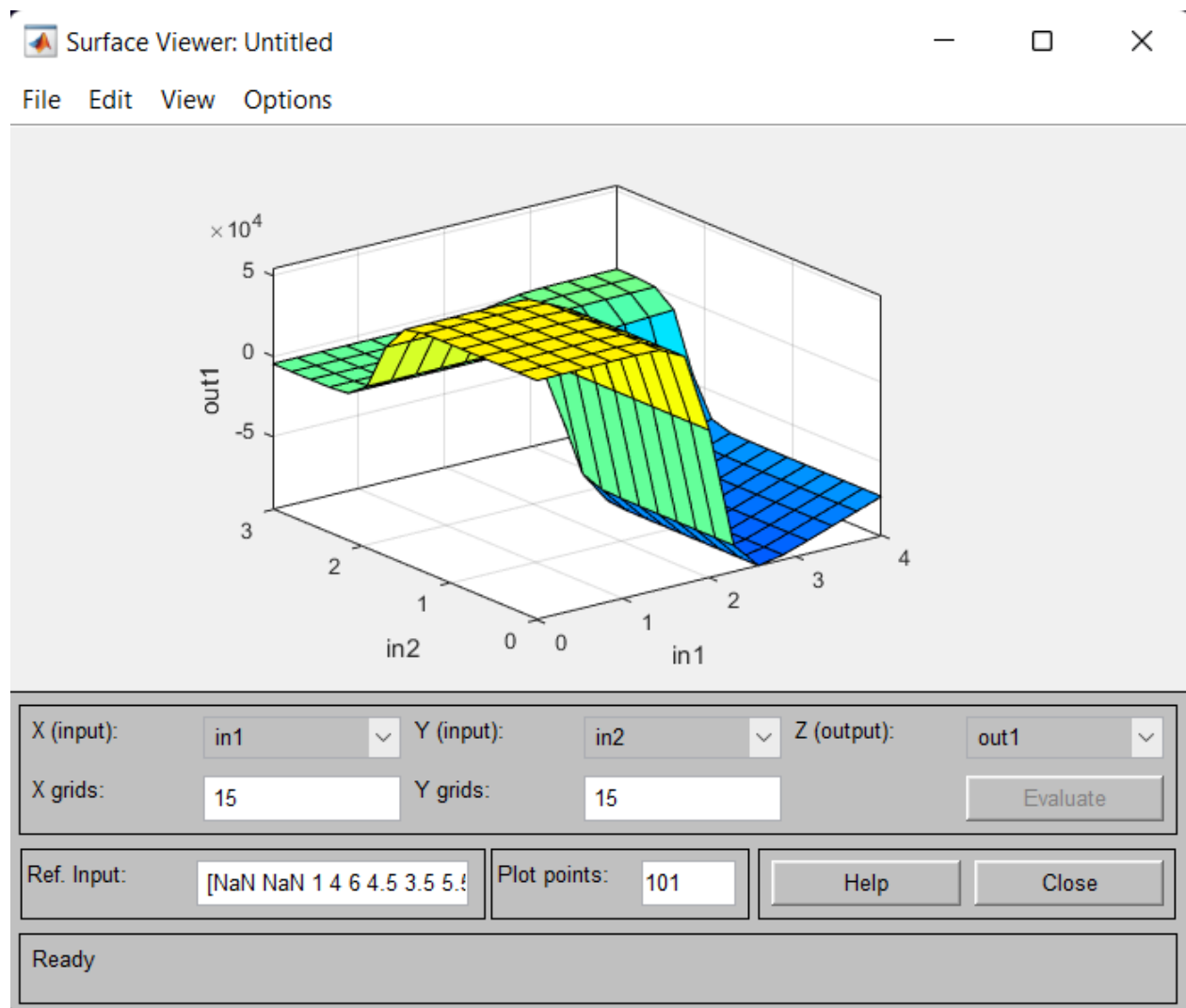


Figure – 5.51: Surface viewer of Dataset 3TC

5.3 EXCEL SHEET OF NNRTI:

	A	B	C	D	E	F	G	H	I	J
1	0	0	1	0	0	0	0	0	0	0.7912
2	0	0	0	0	2	0	0	0	0	51222
3	0	0	1	0	2	0	0	0	0	20079
4	3	0	1	0	0	0	0	0	0	20057
5	0	0	0	0	0	4	0	0	0	7162
6	0	0	0	0	0	0	5	0	0	20043
7	0	0	0	0	2	0	5	0	0	12936
8	0	6	0	0	0	0	5	0	0	20024
9	0	0	1	0	0	0	5	0	0	2009
10	0	6	0	0	2	0	5	0	0	2002
11	0	0	1	0	2	0	5	0	0	20019
12	0	0	0	7	0	0	0	0	0	2008
13	0	0	1	7	0	0	0	0	0	656
14	0	0	1	0	0	0	0	4	0	2001
15	0	0	0	0	0	0	8	0	0	20011
16	0	6	0	0	0	0	0	0	0	20015
17	0	6	0	0	0	0	8	0	0	5.212
18	0	0	8	0	0	0	5	0	0	2008
19	0	0	0	5	0	0	0	0	0	2007
20	0	0	1	0	0	4	0	0	0	20014
21	0	9	1	0	0	0	0	0	0	2009
22	0	6	0	0	2	0	8	0	0	20015
23	0	6	0	0	2	0	0	0	0	2002
24	0	0	0	0	0	0	6	0	0	2005
25	0	0	0	0	2	0	8	0	0	2008
26	0	0	0	0	3	0	0	0	0	2007
27	0	0	0	0	10	0	0	0	0	20013
28	0	6	0	7	0	0	5	0	0	2005
29	0	0	0	0	0	0	11	0	0	2001
30	0	6	1	0	0	0	0	0	0	2006
31	0	0	1	5	0	0	0	0	0	2001
32	0	0	0	5	0	0	5	0	0	2006

Figure – 5.52: Final excel sheet for analysis

5.1.4. RULE VIEWER OF STANFORD HIV DRUG RESISTANCE DATABASE (NNRTI):

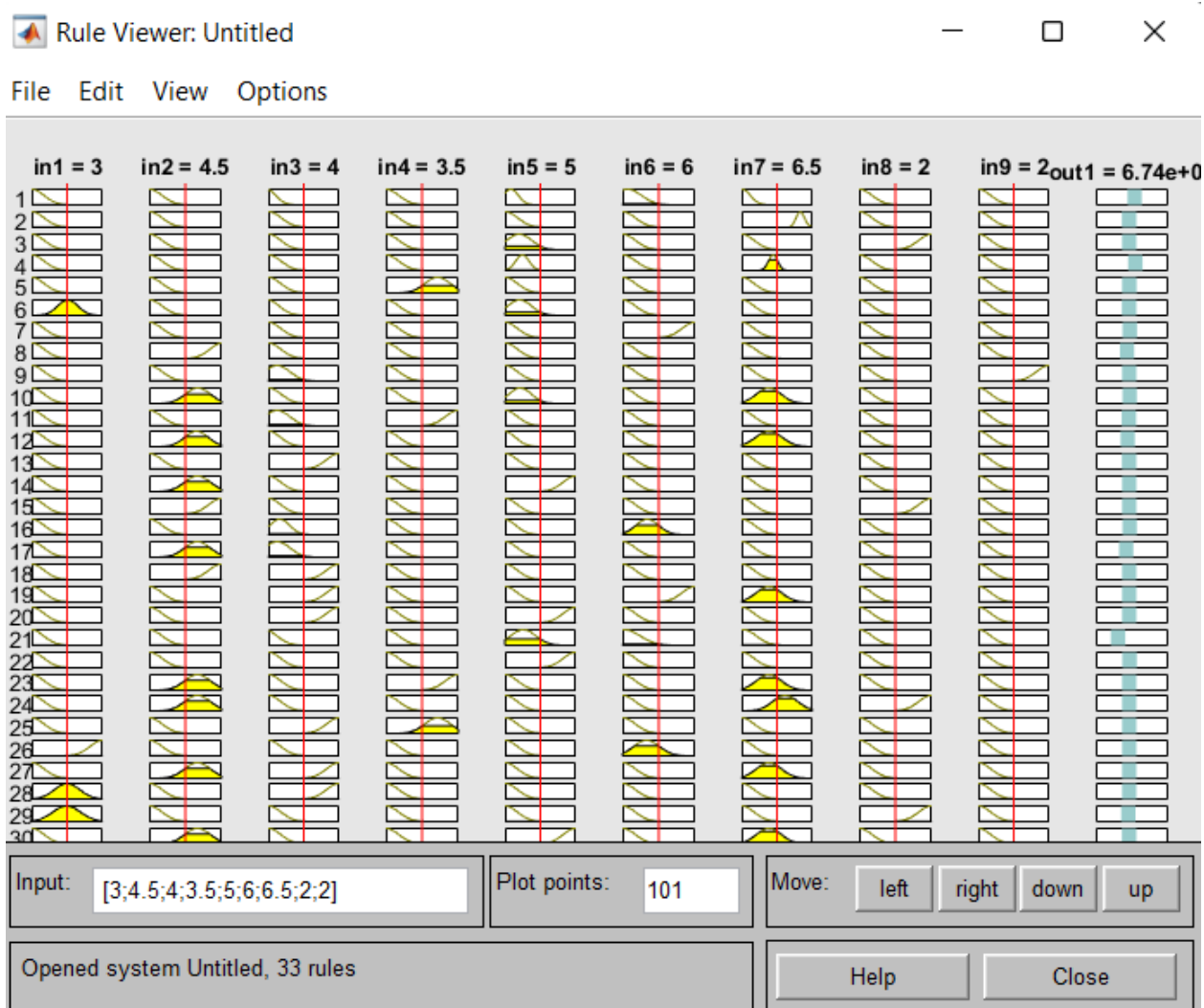


Figure – 5.53: Rule viewer of Dataset NVP

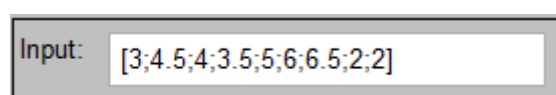


Figure – 5.54: Input Values

In figure – 5.53 rule viewer represented the median fold drug resistance score and in figure – 5.54 where input values are given for mutation position. Plot points selected for this set of values.

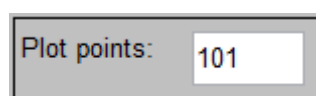


Figure – 5.55 Plot point

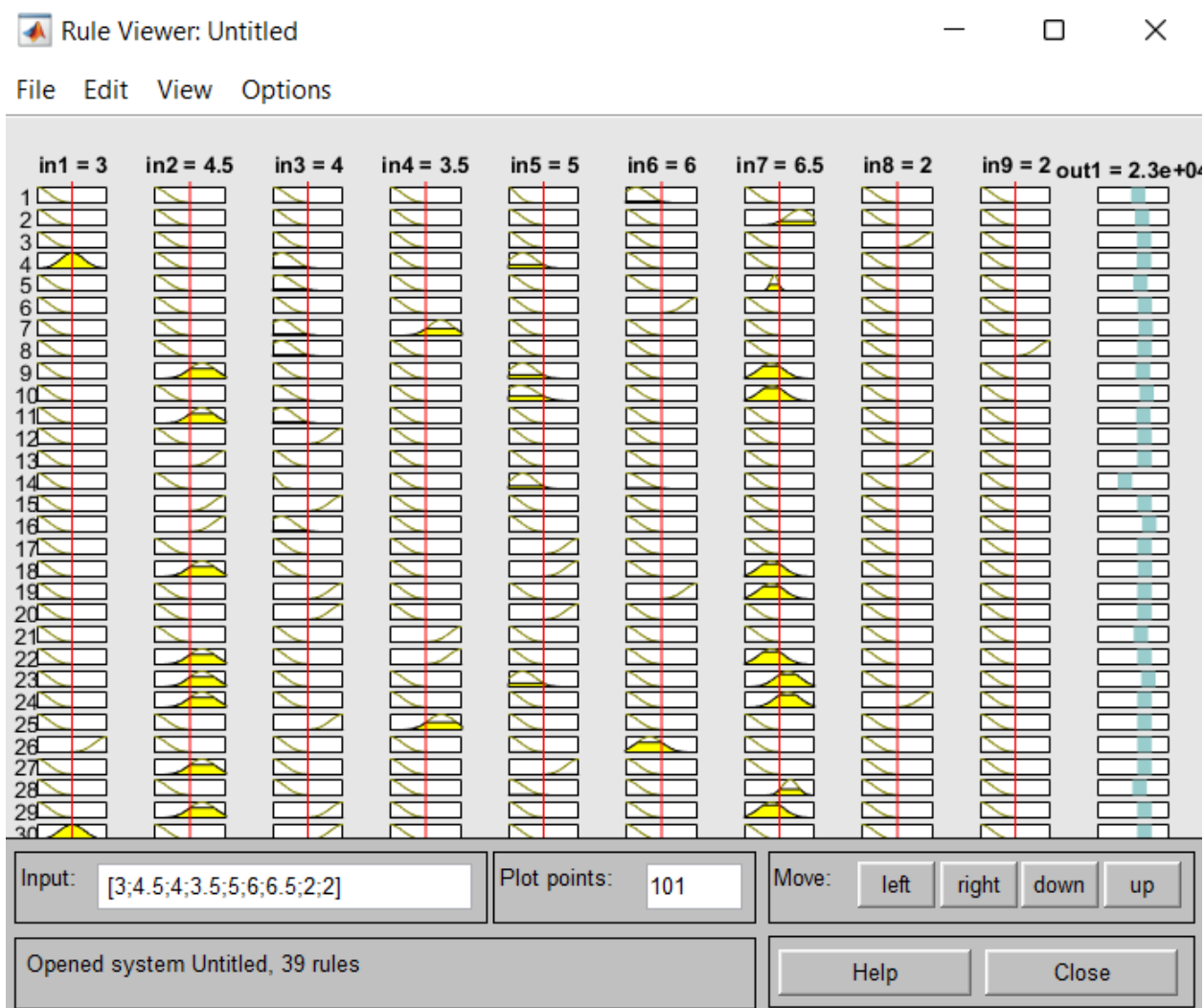


Figure – 5.56: Rule viewer of Dataset EFV

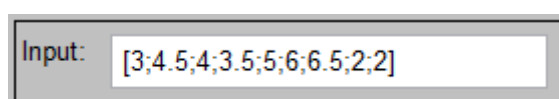


Figure – 5.57: Input values

In figure – 5.56 rule viewer represented the median fold drug resistance score and in figure – 5.57 where input values are given for mutation position. Plot points selected for this set of values.

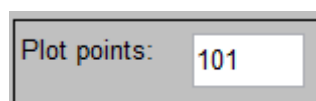


Figure – 5.58 Plot points

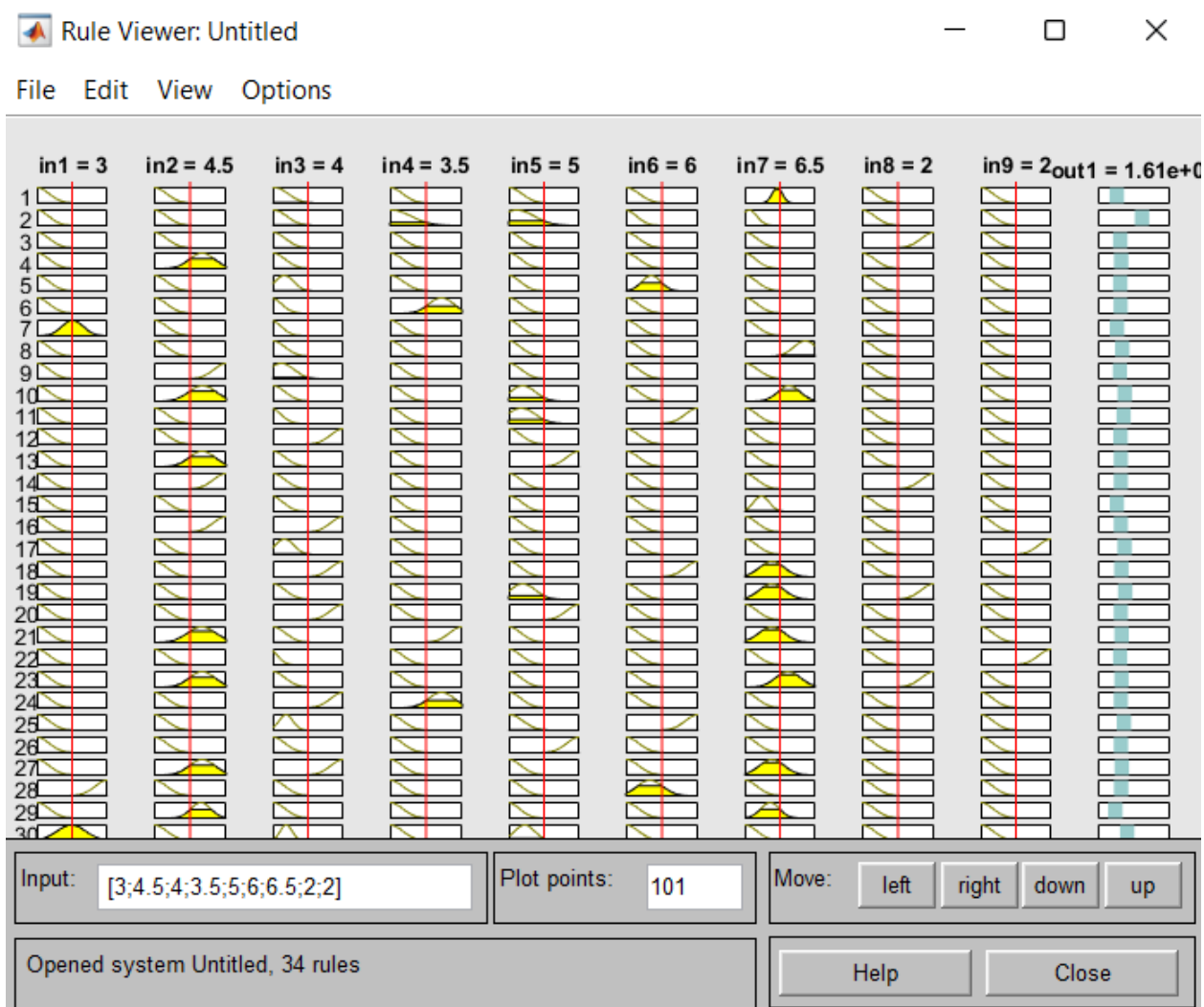


Figure – 5.59 Rule viewer of Dataset ETR

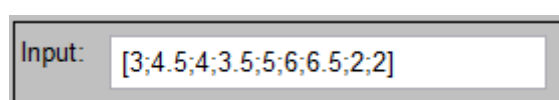


Figure – 5.60 Input values

In figure – 5.59 rule viewer represented the median fold drug resistance score and in figure – 5.60 where input values are given for mutation position. Plot points selected for this set of values.

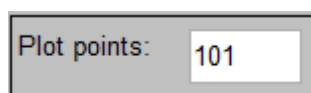


Figure – 5.61: Plot points

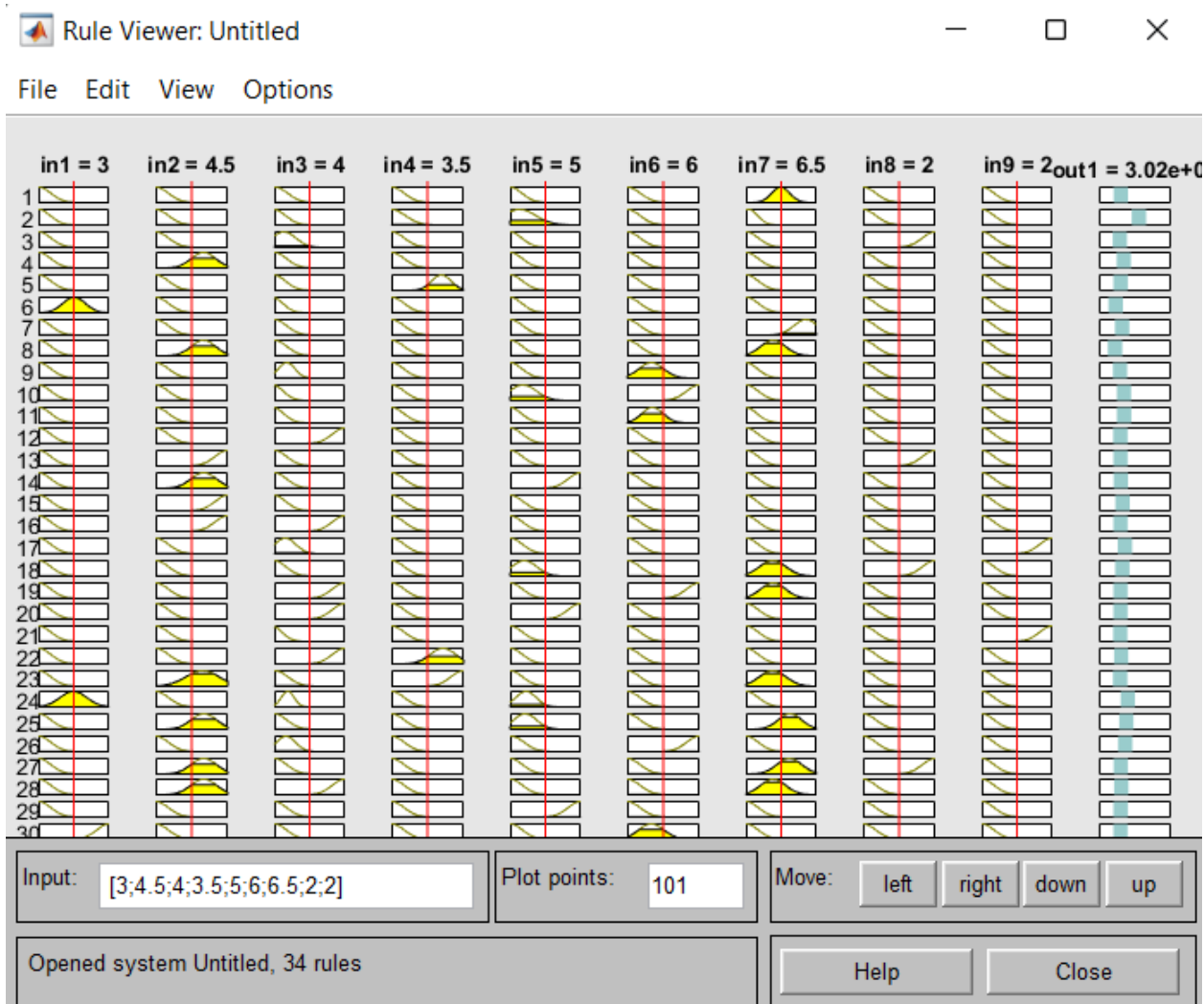


Figure – 5.62: Rule viewer of Dataset RPV

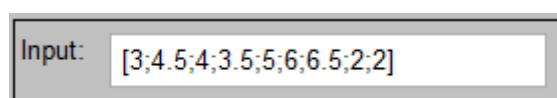


Figure – 5.63: Input values

In figure – 5.62 rule viewer represented the median fold drug resistance score and in figure – 5.63 where input values are given for mutation position. Plot points selected for this set of values.

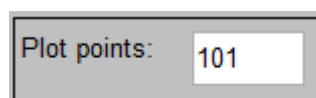


Figure – 5.64: Plot points

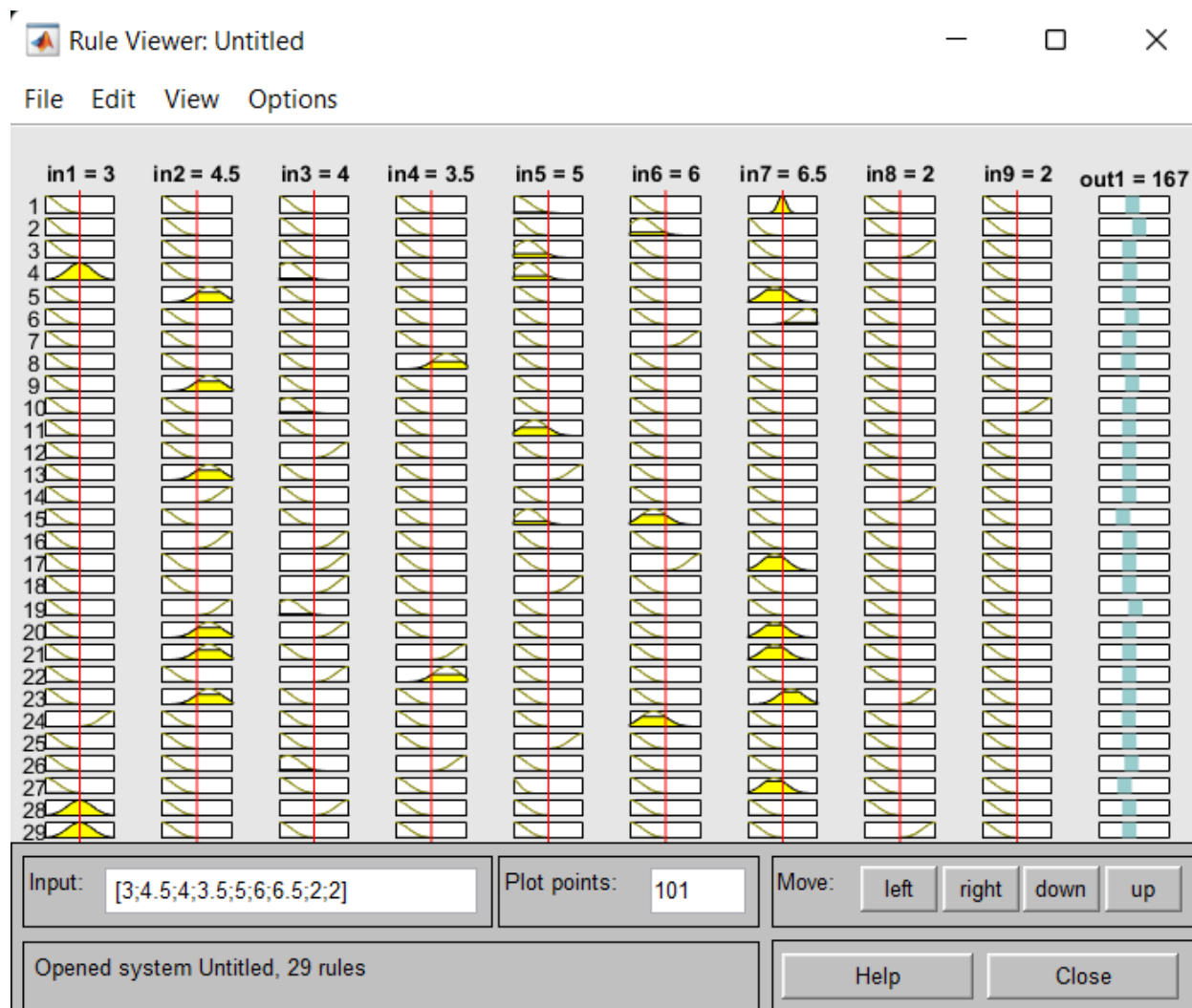


Figure – 5.65: Rule viewer of Dataset DOR

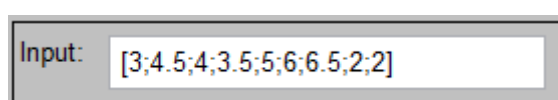


Figure – 5.66: Input values

In figure – 5.65 rule viewer represented the median fold drug resistance score and in figure – 5.66 where input values are given for mutation position. Plot points selected for this set of values

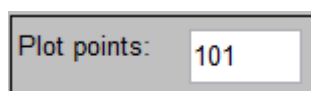


Figure – 5.67: Plot points

5.1.4 SURFACE VIEWER OF STANFORD HIV DRUG RESISTANCE

DATASET (NRTI):

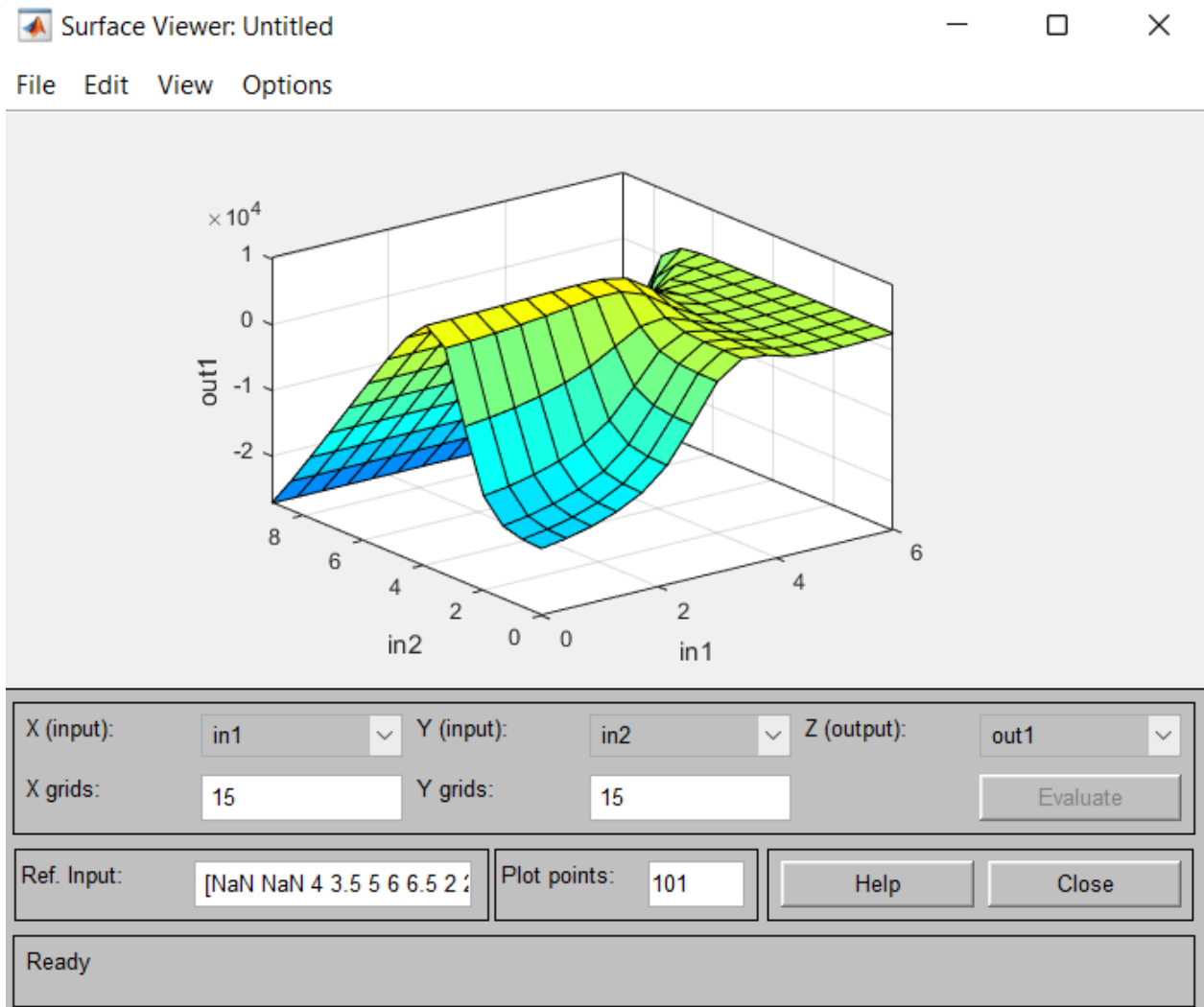


Figure – 5.68: Surface viewer of Dataset NVP

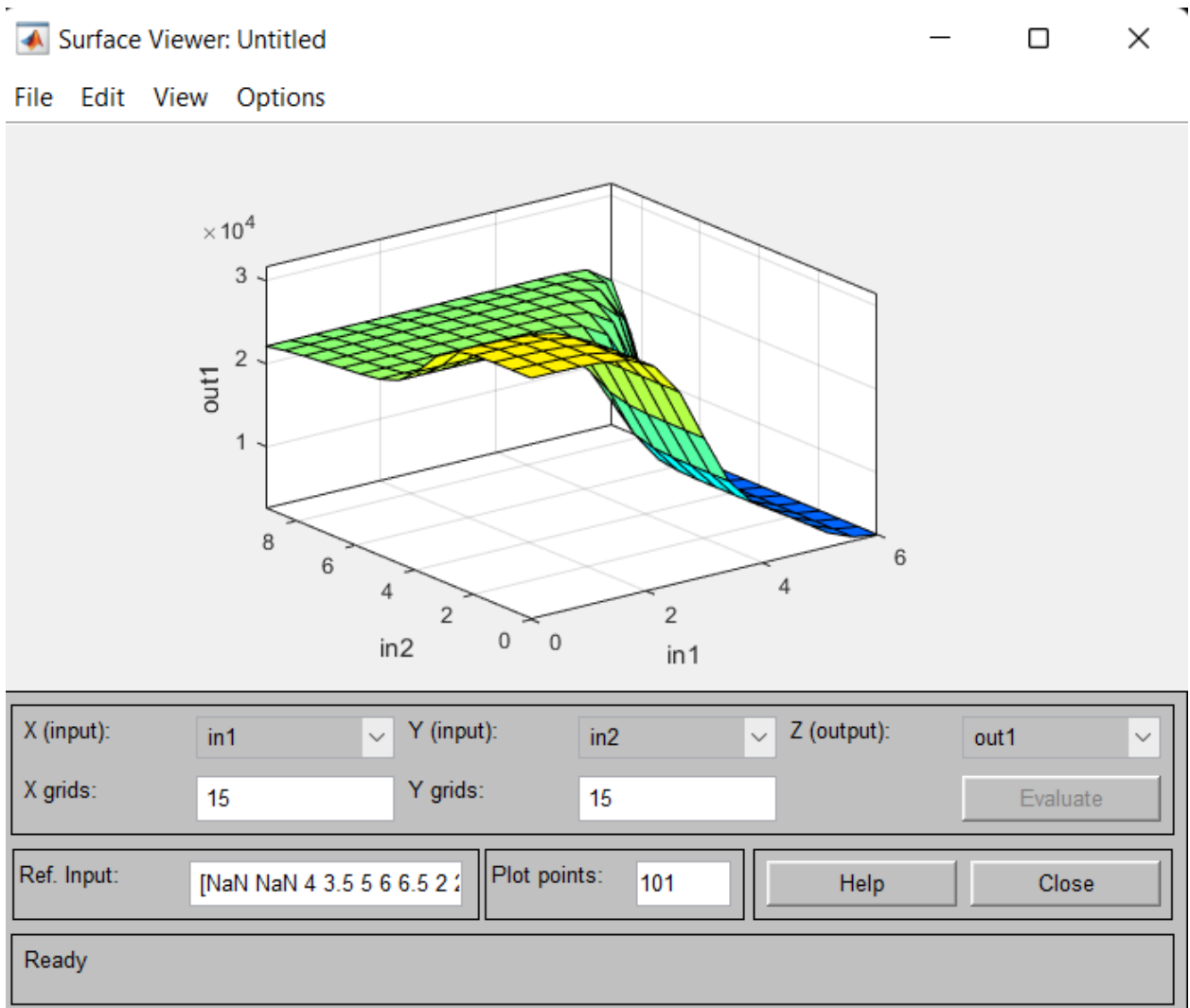


Figure – 5.69: Surface viewer of Dataset EFV

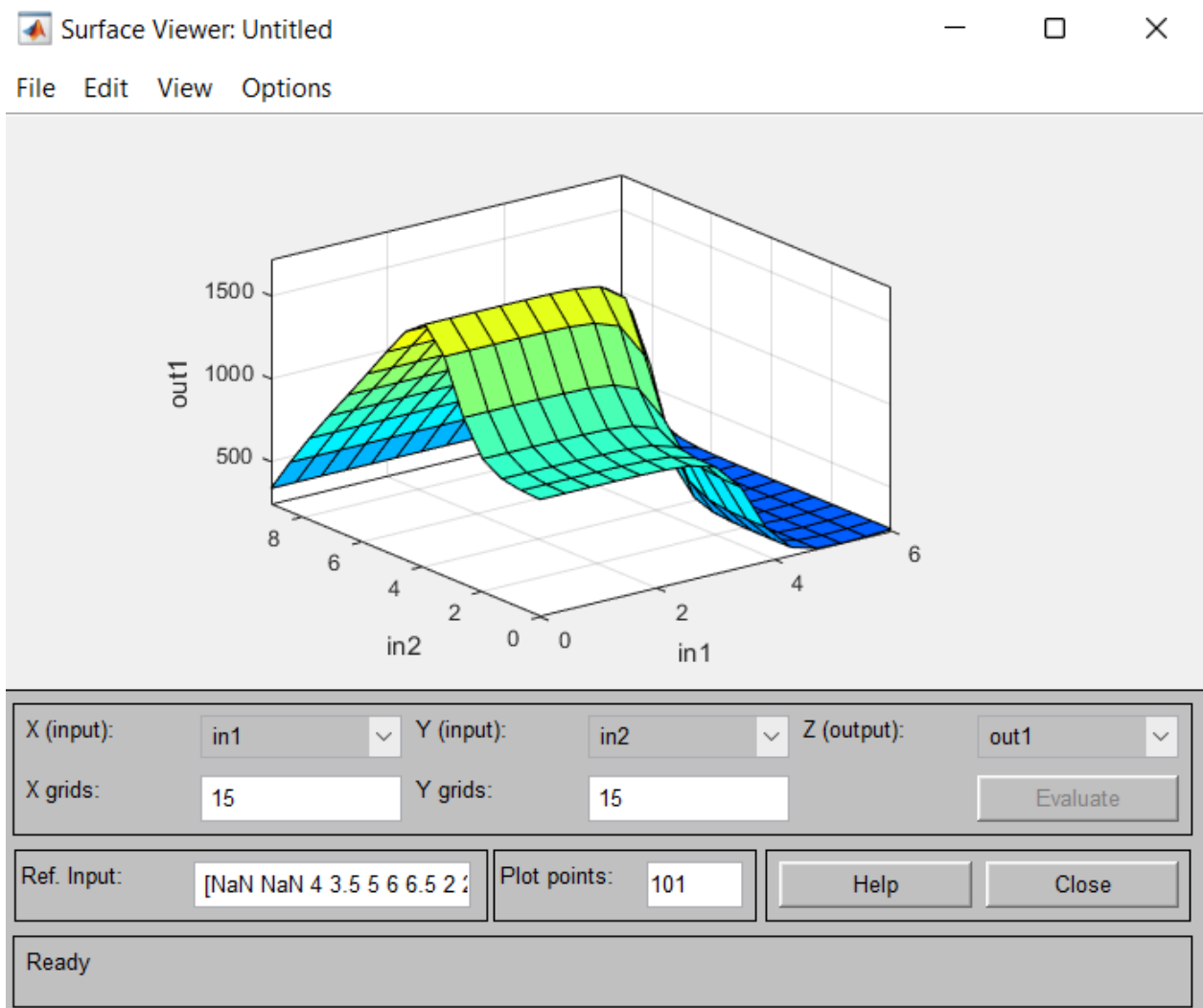


Figure – 5.70: Surface viewer of Dataset ETR

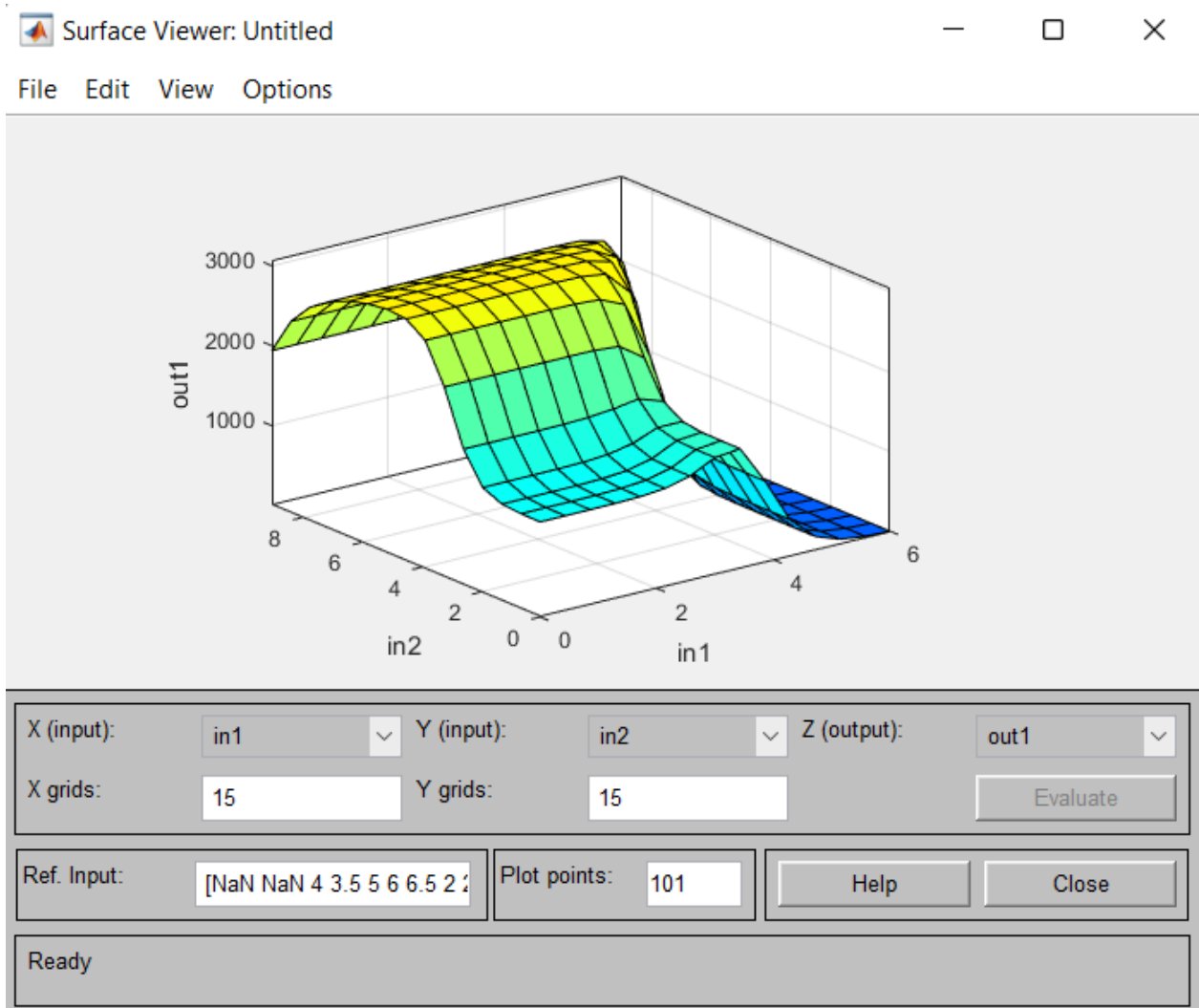


Figure – 5.71: Surface viewer of Dataset RPV

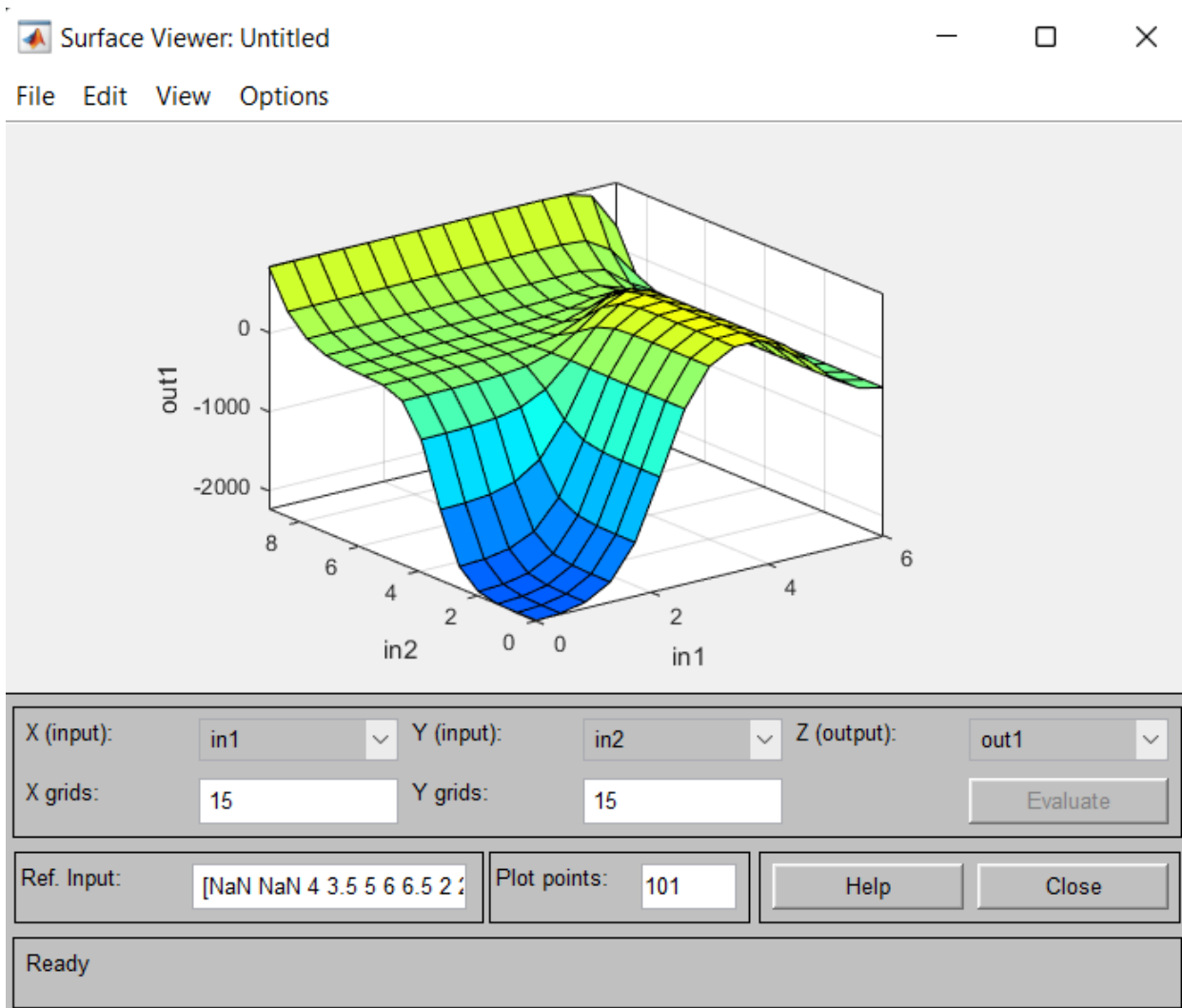


Figure – 5.72: Surface viewer of Dataset DOR

CHAPTER - 6

SUMMARY AND CONCLUSION

6.1. SUMMARY:

The summary for this was twelve dataset of mutation position and eight dataset of median fold drug resistance score (Protease Dataset).

The summary for this was eleven dataset of mutation position and eight four dataset of median fold drug resistance score (NRTI).

The summary for this was nine dataset of mutation position and five dataset of median fold drug resistance score (NNRTI).

The Raw data which we grouped, filtered and analyzed for our project has been taken from Stanford HIV drug resistance database.

We select Protease, NRTI and NNRTI as our main source for the raw data from the database

After that we convert the data of protease, NRTI and NNRTI table into a meaningful mutational data in the Microsoft excel sheet, then we have checked the drug resistance scores from the table.

Then we filtered each raw data

Then we remove all the outliers in all the sheet (Removing outliers is nothing but removing the blank spaces of drugs).

Now, In MATLAB we need to import the final table and select as numeric matrix and import it in the workspace and then in command we need to type "anfisedit".

Later one new interface will flash on the screen and there we must select the workspace and load the data.

After loading the data, we have to click on Generate FIS button after selecting Sub – clustering.

In Epochs, we gave 1000 and train the data for every drugs.

After completing the training wave click on view and we have to check the Rule and Surface viewer for the given drugs which give us the

proper explanation about the given drugs.

Mostly we need to focus on our input and output functional data to get more accurate and good drugs prediction results. We must observe the rule viewer and surface viewer data for each and every drug which is also provided drug resistance scores and different mutational patterns. The drugs which have the lesser drug resistance score is good to be used.

6.2 CONCLUSION:

There were no significance outliers in the model, and we could observe interactions between various specific mutations in the same sequence responding to the drug resistance.

The model does not follow linear and non-linear regression model and here we performed the modelling of the data of Protease inhibitor, NRTI and NNRTI drug resistance and its susceptibility.

Our dataset is trained with Sub - Clustering algorithm including all the mutations. The dataset was set to train for 1000 epochs.

For each and every drug resistance prediction the maximal number of mutation pattern was analyzed, given variation in mutation position with mutual accuracy of drug resistance prediction. For a better and accurate prediction of various drug resistance, there is provided an increase in membership function for the input and output data.

The median fold drug resistance score with each source data was fully analyzed; thoroughly observed rule view and surface view, provided mutation pattern with different drug resistance. A variation in drug resistance fold is observed, that is interpreted and based on the

resultants, the resistance of HIV drug is predicted and called out for lesser median fold drug resistance score. The drug with lesser resistance score is more to be used, in general.

