1) Overview:

Number of records: 150

Number of features: 11

First 5 records:

	Name	Age	Gender	Weight	Symptom_1	Symptom_2	Symptom_3	Duration	\
0	Surya	20	Male	60.3	Fever	Nausea	Sneezing	3 days	
1	Pooja	64	Female	56.8	Back pain	Chest pain	Headache	5 days	
2	Vikram	49	Male	75.0	Sore throat	Fatigue	Rash	2 days	
3	Anjali	63	Female	76.9	Rash	Sore throat	Sneezing	4 days	
4	Neha	48	Female	72.6	Chest pain	Vomiting	Headache	3 days	

	Body_Temperature	Heart_Rate	Disease_Prediction
0	100.3°F	107 bpm	Flu
1	103.2°F	83 bpm	Dengue
2	102.7°F	77 bpm	COVID-19
3	101.0°F	108 bpm	Flu
4	100.4°F	94 bpm	Cold

Basic Statistics:

	Age	Weight	Duration	Body_Temperature	Heart_Rate
count	150.000000	150.000000	150.000000	150.000000	150.00000
mean	42.186667	69.085333	4.146667	100.508667	91.14000
std	14.991893	12.502029	1.987838	1.839096	11.35813
min	18.000000	45.800000	1.000000	97.500000	70.00000
25%	28.000000	59.175000	2.000000	98.825000	82.25000
50%	43.000000	70.950000	4.000000	100.600000	92.00000
75%	54.000000	79.300000	6.000000	102.100000	99.00000

max 69.000000 89.800000 7.000000 103.400000 110.00000

Disease Distribution:

Disease_Prediction

Malaria 21

Flu 17

Cold 17

Dengue 16

Viral Infection 16

COVID-19 15

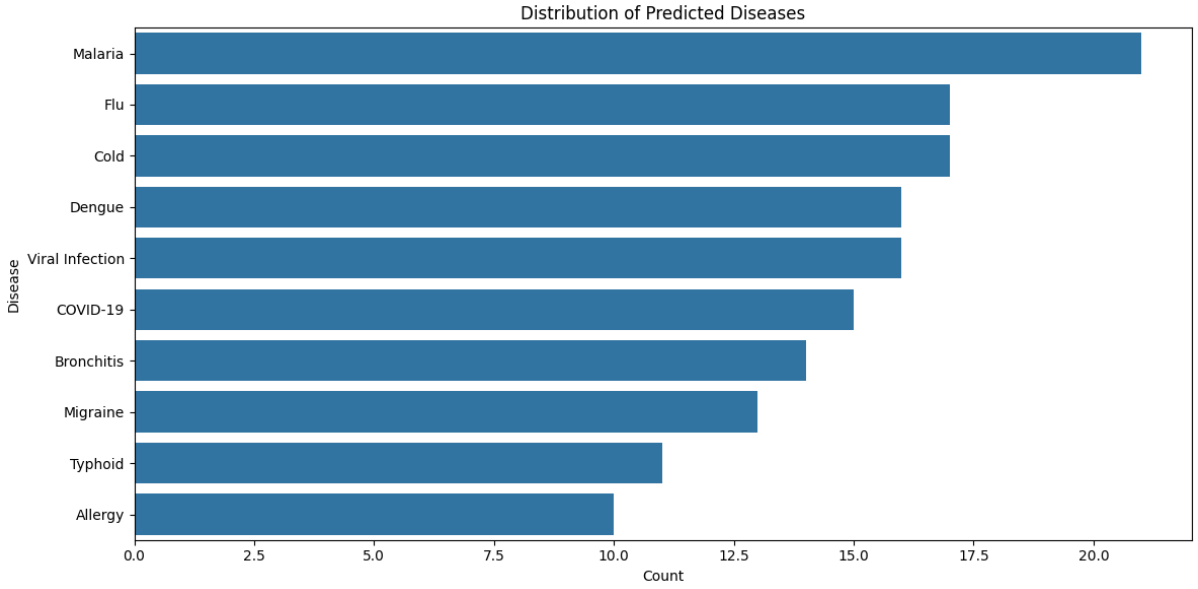
Bronchitis 14

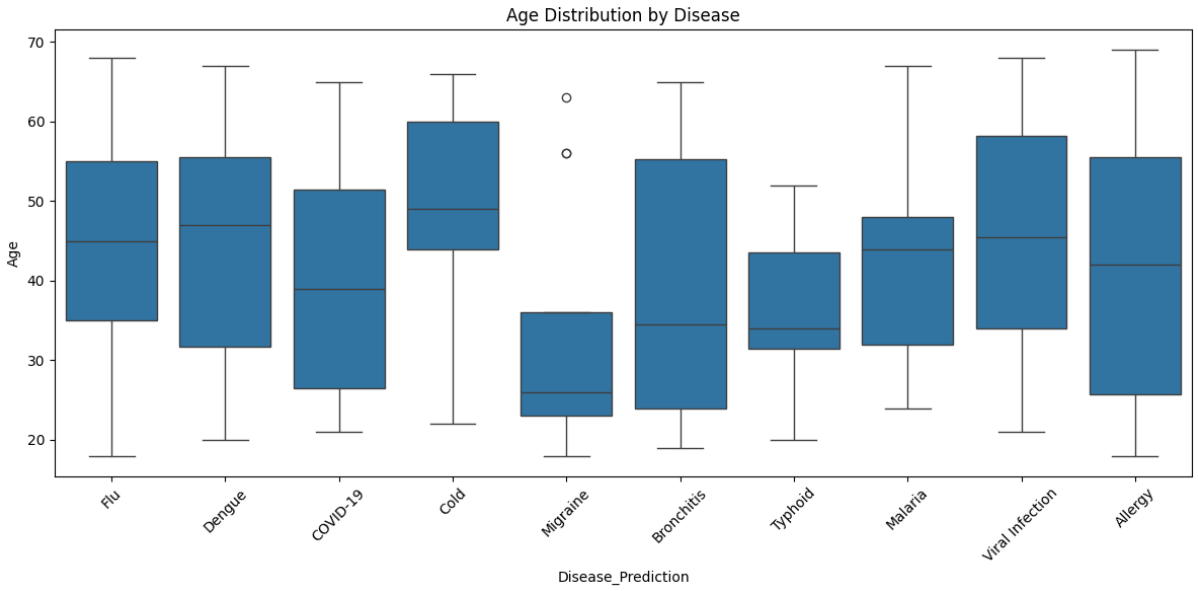
Migraine 13

Typhoid 11

Allergy 10

Name: count, dtype: int64





Most Common Symptoms:

Vomiting 38

Headache 37

Sneezing 35

Nausea 35

Chest pain 35

Body Pain 35

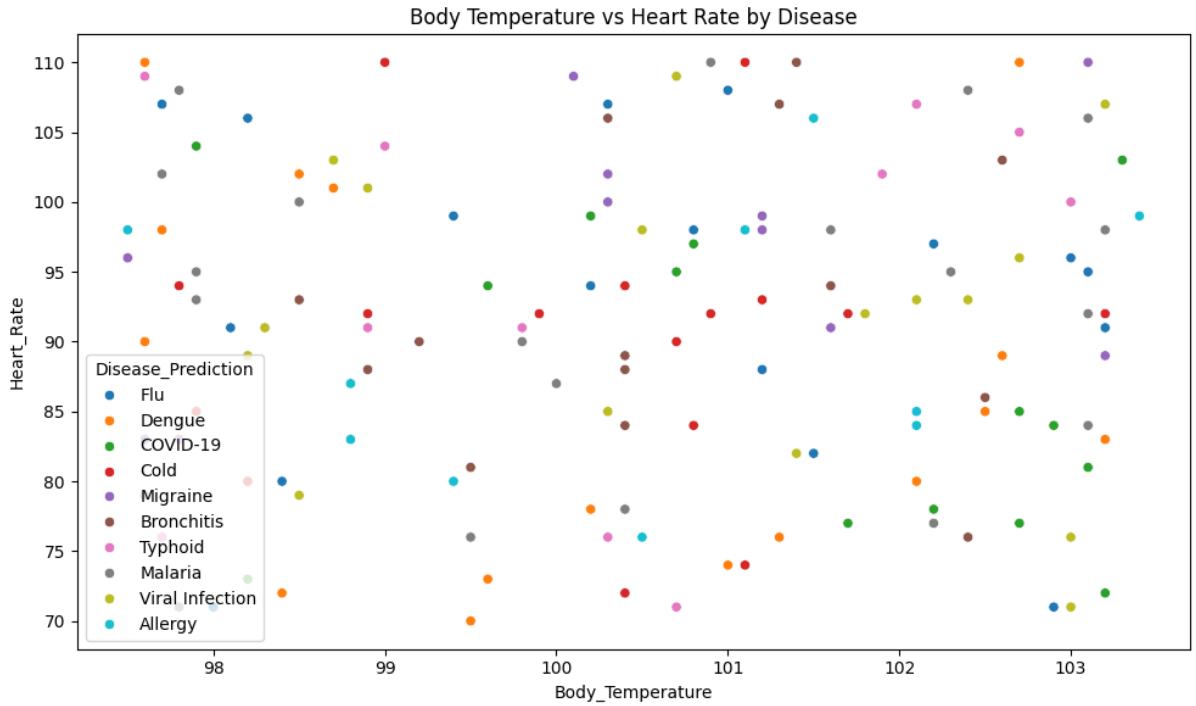
Back pain 34

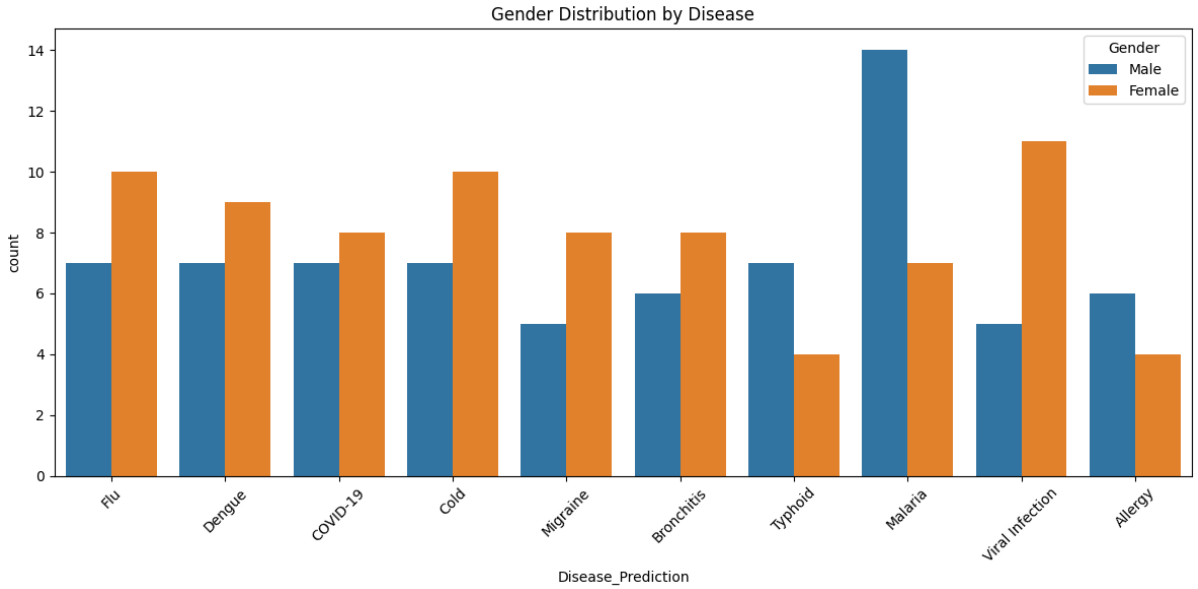
Rash 34

Fever 31

Runny nose 31

Name: count, dtype: int64





Prediction for symptoms ['Fever', 'Headache']:

Disease_Prediction

Cold 9

Dengue 8

Malaria 8

Viral Infection 7

Flu 6

Bronchitis 6

COVID-19 5

Typhoid 5

Migraine 4

Allergy 4

Name: count, dtype: int64

2. Missing values before handling:

Name 0

Age 0

Gender 0

Weight 0

Symptom_1 0

Symptom_2 0

Symptom_3 0

Duration 0

Body_Temperature 0

Heart_Rate 0

Disease_Prediction 0

dtype: int64

Missing values after handling:

Name 0

Age 0

Gender 0

Weight 0

Symptom_1 0

Symptom_2 0

Symptom_3 0

Duration 0

Body_Temperature 0

Heart_Rate 0

Disease_Prediction 0

dtype: int64

Preprocessing Summary:

Original data shape: (150, 11)

Training data shape: (105, 9)

Test data shape: (45, 9)

First 5 rows of processed training data:

	0	1	2	3	4	5	6	7	8
0	0.074313	-0.832796	0.220916	1.001820	-0.180014	-0.025825	-0.498559	-1.663040	0.389378
1	0.746973	1.200775	1.124457	-1.592579	1.538299	-0.272339	0.955572	-0.068093	-1.224490
2	0.074313	-0.832796	0.340855	-1.592579	-0.670960	-1.258396	-0.983270	1.471856	-0.058919
3	-0.127485	1.200775	0.676684	0.294256	-0.670960	-0.765368	0.470862	-1.168056	-1.045172
4	-1.607338	-0.832796	-1.154384	1.473529	-0.180014	-0.765368	-0.983270	-1.663040	0.568696

3. === Dataset Overview ===

Number of records: 150

Number of features: 11

First 5 records:

```
   Name  Age  Gender  Weight   Symptom_1   Symptom_2 Symptom_3 Duration  \
0  Surya   20    Male    60.3      Fever      Nausea  Sneezing   3 days
1   Pooja   64  Female    56.8   Back pain   Chest pain  Headache   5 days
2 Vikram   49    Male    75.0  Sore throat     Fatigue     Rash   2 days
3  Anjali   63  Female    76.9       Rash  Sore throat  Sneezing   4 days
4   Neha   48  Female    72.6   Chest pain    Vomiting  Headache   3 days

  Body_Temperature Heart_Rate Disease_Prediction
0        100.3°F    107 bpm            Flu
1        103.2°F     83 bpm         Dengue
2        102.7°F     77 bpm       COVID-19
3        101.0°F    108 bpm            Flu
4        100.4°F     94 bpm           Cold
```

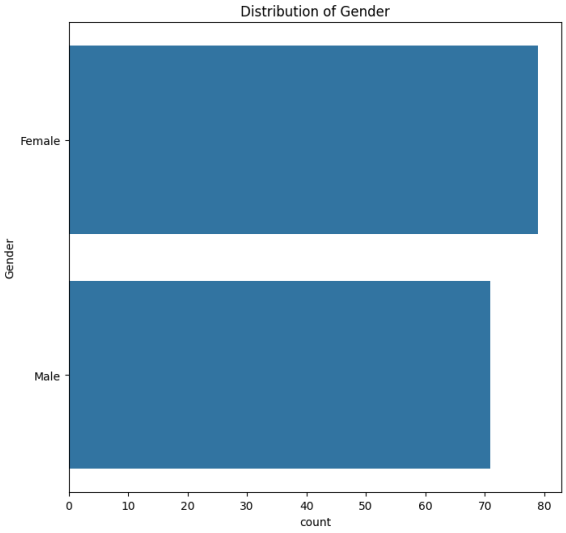
Dataset information:

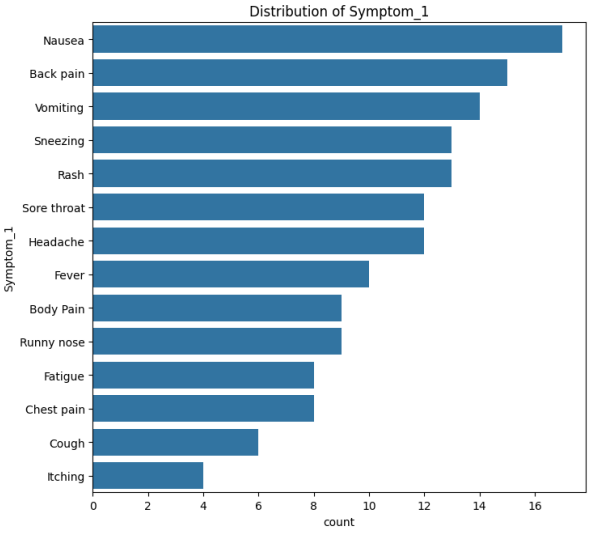
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Name            150 non-null    object
 1   Age             150 non-null    int64
 2   Gender          150 non-null    object
 3   Weight          150 non-null    float64
 4   Symptom_1       150 non-null    object
```

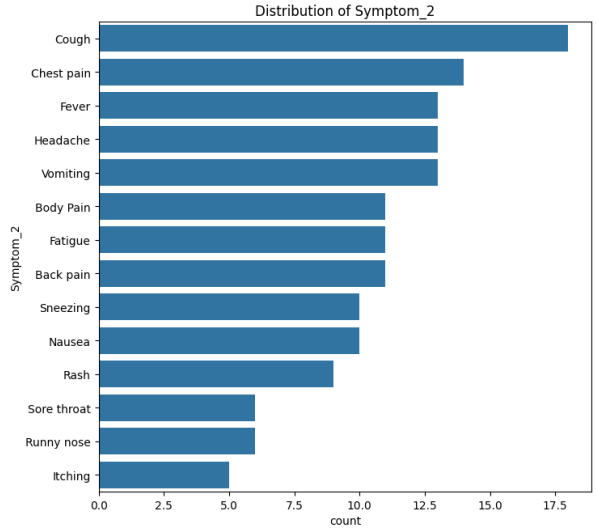
```
 5  Symptom_2         150 non-null    object
 6  Symptom_3         150 non-null    object
 7  Duration          150 non-null    object
 8  Body_Temperature  150 non-null    object
 9  Heart_Rate        150 non-null    object
 10 Disease_Prediction 150 non-null   object
dtypes: float64(1), int64(1), object(9)
memory usage: 13.0+ KB
None

=== Missing Values ===
Name               0
Age                0
Gender             0
Weight             0
Symptom_1          0
Symptom_2          0
Symptom_3          0
Duration           0
Body_Temperature   0
Heart_Rate         0
Disease_Prediction 0
dtype: int64

=== Descriptive Statistics ===
        Name        Age  Gender     Weight Symptom_1 Symptom_2  Symptom_3  \
count    150  150.000000    150  150.000000      150      150        150
unique    20         NaN      2         NaN       14       14         14
```

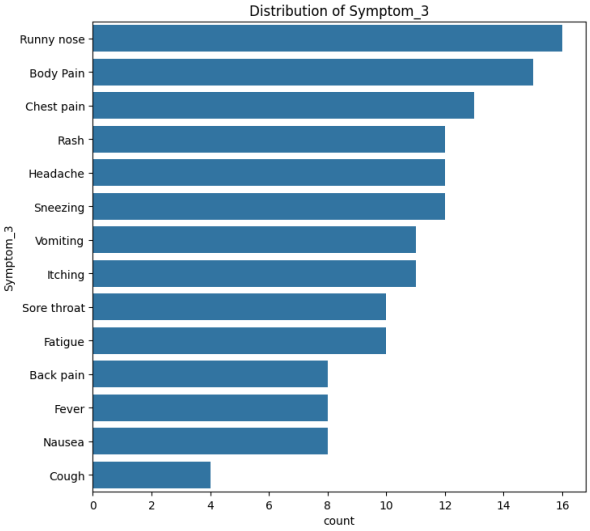
top	Arjun	NaN	Female	NaN	Nausea	Cough	Runny nose
freq	13	NaN	79	NaN	17	18	16
mean	NaN	42.186667	NaN	69.085333	NaN	NaN	NaN
std	NaN	14.991893	NaN	12.502029	NaN	NaN	NaN
min	NaN	18.000000	NaN	45.800000	NaN	NaN	NaN
25%	NaN	28.000000	NaN	59.175000	NaN	NaN	NaN
50%	NaN	43.000000	NaN	70.950000	NaN	NaN	NaN
75%	NaN	54.000000	NaN	79.300000	NaN	NaN	NaN
max	NaN	69.000000	NaN	89.800000	NaN	NaN	NaN

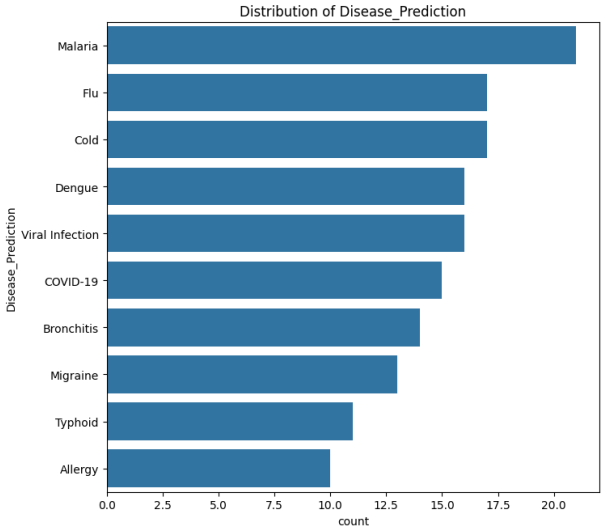
	Duration	Body_Temperature	Heart_Rate	Disease_Prediction
count	150.000000	150.000000	150.00000	150
unique	NaN	NaN	NaN	10
top	NaN	NaN	NaN	Malaria
freq	NaN	NaN	NaN	21
mean	4.146667	100.508667	91.14000	NaN
std	1.987838	1.839096	11.35813	NaN
min	1.000000	97.500000	70.00000	NaN
25%	2.000000	98.825000	82.25000	NaN
50%	4.000000	100.600000	92.00000	NaN
75%	6.000000	102.100000	99.00000	NaN
max	7.000000	103.400000	110.00000	NaN

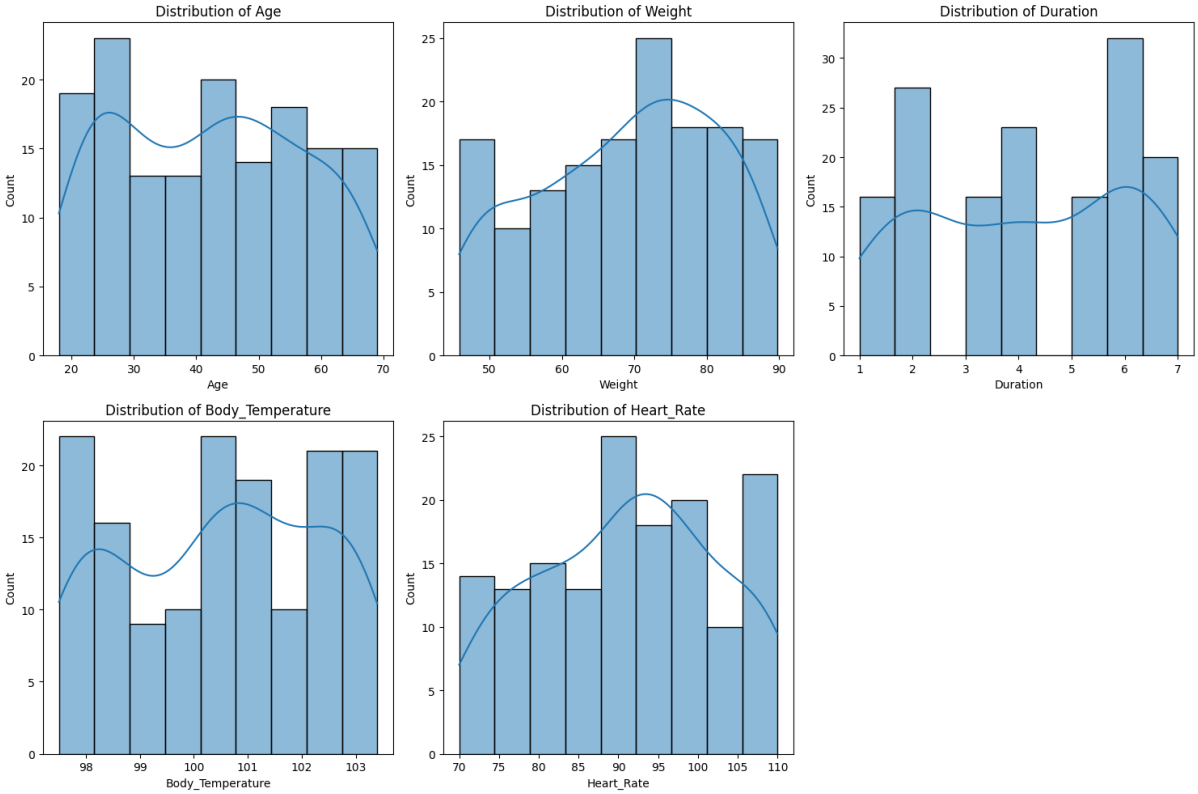

Distribution of Gender
Gender
Female
Male
0
10
20
30
40
50
60
70
80
count

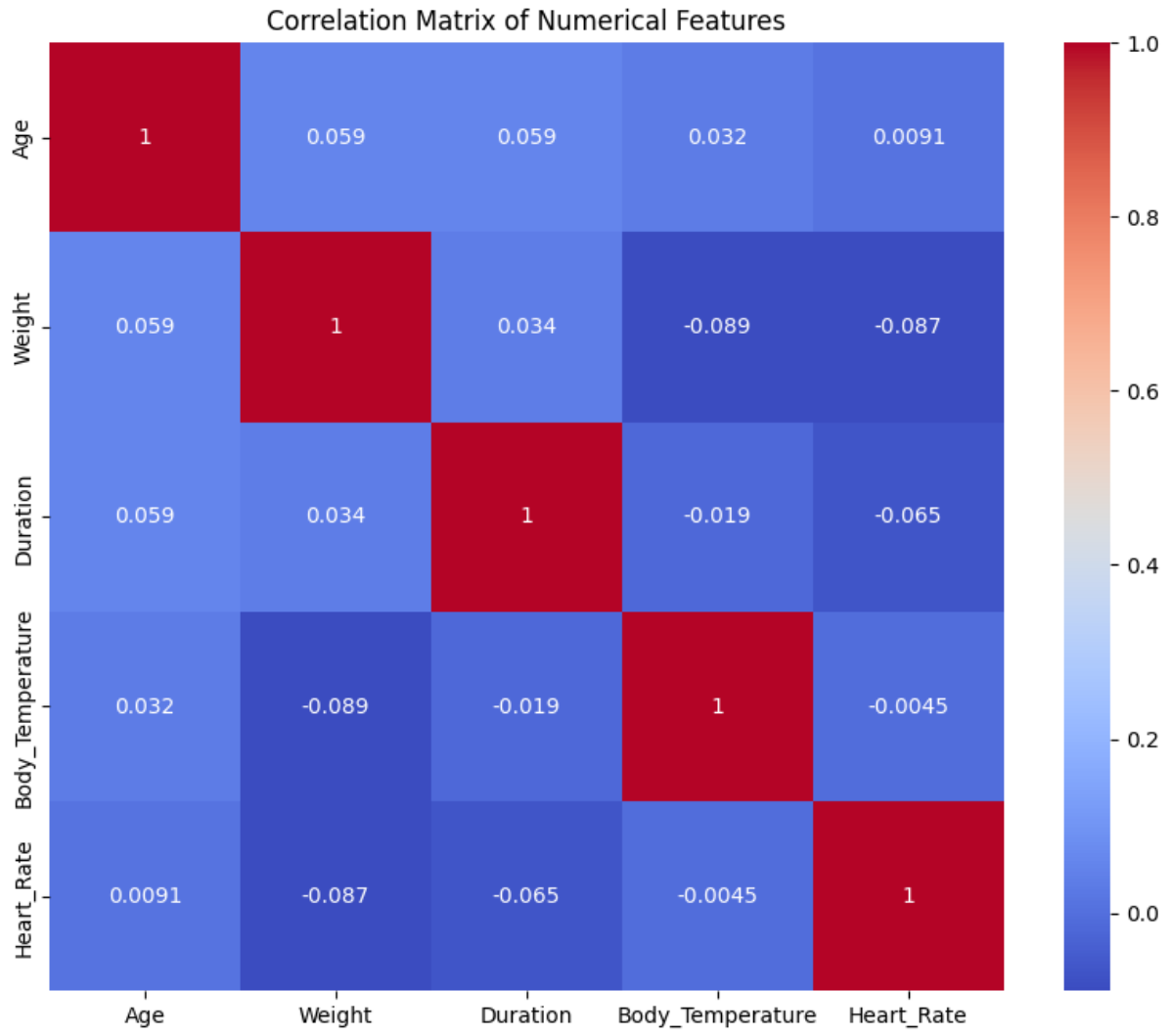

Distribution of Symptom_1
Symptom_1
Nausea
Back pain
Vomiting
Sneezing
Rash
Sore throat
Headache
Fever
Body Pain
Runny nose
Fatigue
Chest pain
Cough
Itching
0
2
4
6
8
10
12
14
16
count

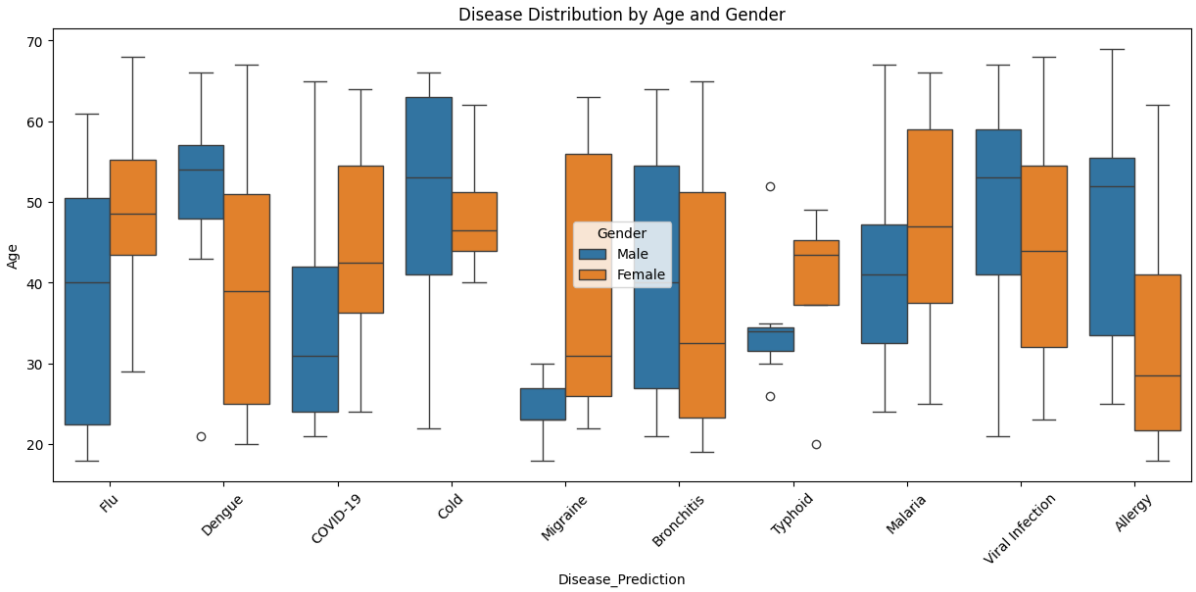

Distribution of Symptom_2
Symptom_2
Cough
Chest pain
Fever
Headache
Vomiting
Body Pain
Fatigue
Back pain
Sneezing
Nausea
Rash
Sore throat
Runny nose
Itching
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
count

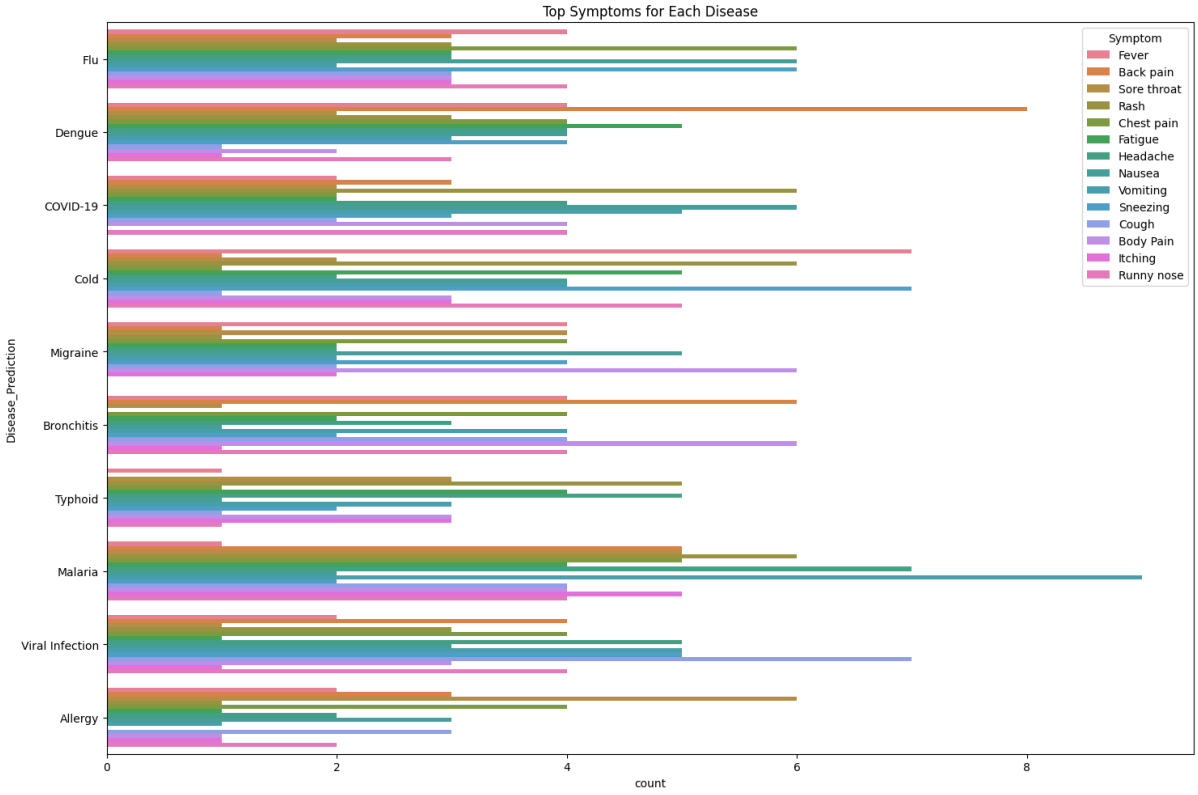

Distribution of Symptom_3
Symptom_3
Runny nose
Body Pain
Chest pain
Rash
Headache
Sneezing
Vomiting
Itching
Sore throat
Fatigue
Back pain
Fever
Nausea
Cough
0
2
4
6
8
10
12
14
16
count

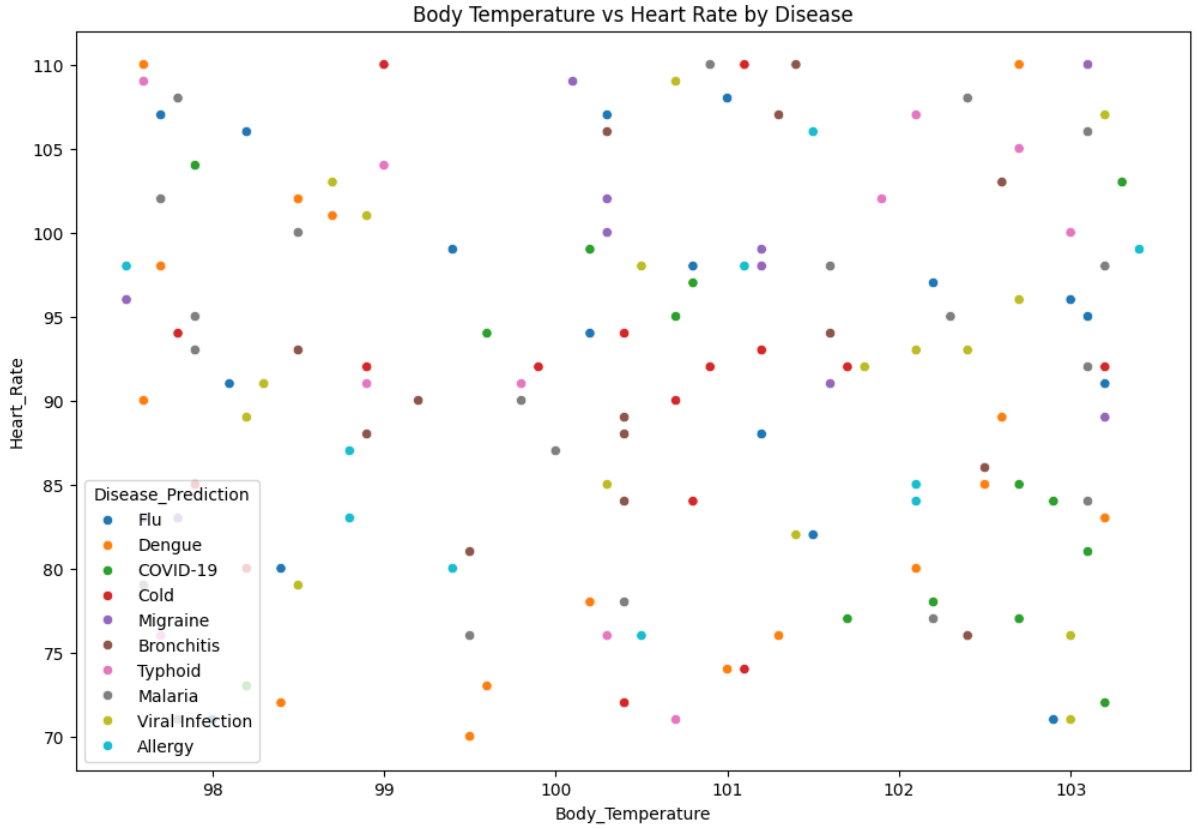

Distribution of Disease_Prediction
Disease_Prediction
Malaria
Flu
Cold
Dengue
Viral Infection
COVID-19
Bronchitis
Migraine
Typhoid
Allergy
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
count

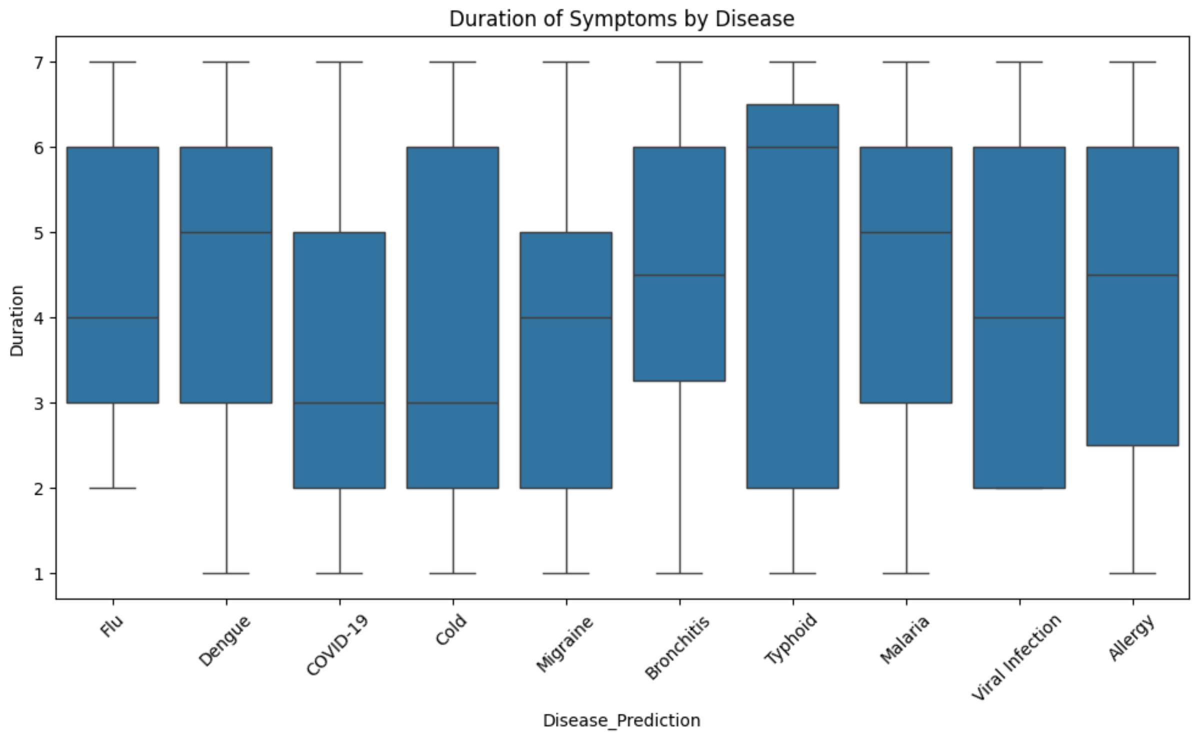

Distribution of Age
Count
Age
Distribution of Weight
Count
Weight
Distribution of Duration
Count
Duration
Distribution of Body_Temperature
Count
Body_Temperature
Distribution of Heart_Rate
Count
Heart_Rate


Correlation Matrix of Numerical Features
Age
Weight
Duration
Body_Temperature
Heart_Rate
1
0.059
0.059
0.032
0.0091
0.059
1
0.034
-0.089
-0.087
0.059
0.034
1
-0.019
-0.065
0.032
-0.089
-0.019
1
-0.0045
0.0091
-0.087
-0.065
-0.0045
1
1.0
0.8
0.6
0.4
0.2
0.0


Disease Distribution by Age and Gender
Age
Disease_Prediction
Gender
Male
Female
Flu
Dengue
COVID-19
Cold
Migraine
Bronchitis
Typhoid
Malaria
Viral Infection
Allergy


Top Symptoms for Each Disease
Symptom
Fever
Back pain
Sore throat
Rash
Chest pain
Fatigue
Headache
Nausea
Vomiting
Sneezing
Cough
Body Pain
Itching
Runny nose
Disease_Prediction
Flu
Dengue
COVID-19
Cold
Migraine
Bronchitis
Typhoid
Malaria
Viral Infection
Allergy
0
2
4
6
8
count


Body Temperature vs Heart Rate by Disease
Heart_Rate
110
105
100
95
90
85
80
75
70
Disease_Prediction
Flu
Dengue
COVID-19
Cold
Migraine
Bronchitis
Typhoid
Malaria
Viral Infection
Allergy
98
99
100
101
102
103
Body_Temperature



/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

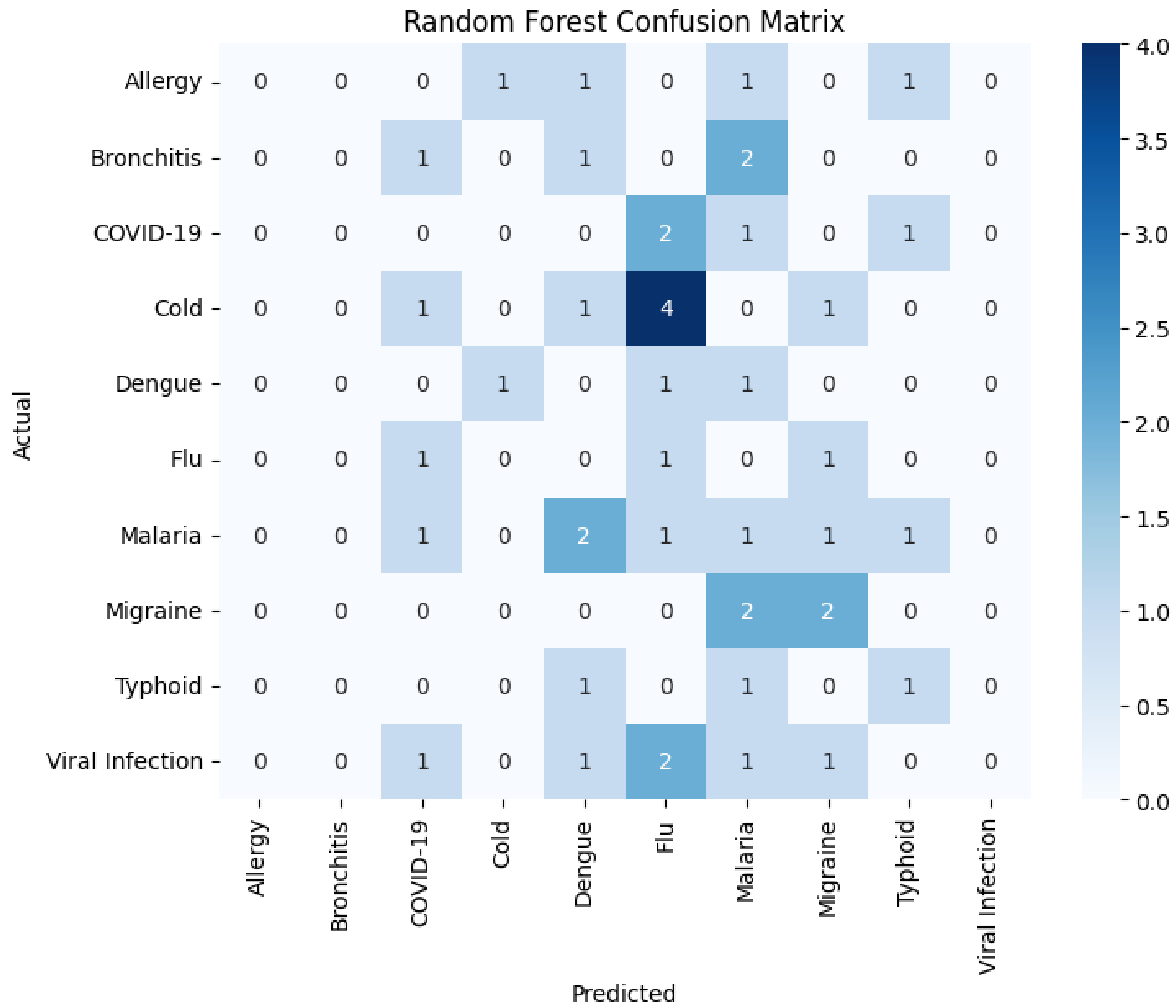
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

Random Forest Classification Report:

precision recall f1-score support

Allergy	0.00	0.00	0.00	4
Bronchitis	0.00	0.00	0.00	4
COVID-19	0.00	0.00	0.00	4
Cold	0.00	0.00	0.00	7
Dengue	0.00	0.00	0.00	3
Flu	0.09	0.33	0.14	3
Malaria	0.10	0.14	0.12	7
Migraine	0.33	0.50	0.40	4
Typhoid	0.25	0.33	0.29	3
Viral Infection	0.00	0.00	0.00	6
accuracy			0.11	45
macro avg	0.08	0.13	0.09	45
weighted avg	0.07	0.11	0.08	45



/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

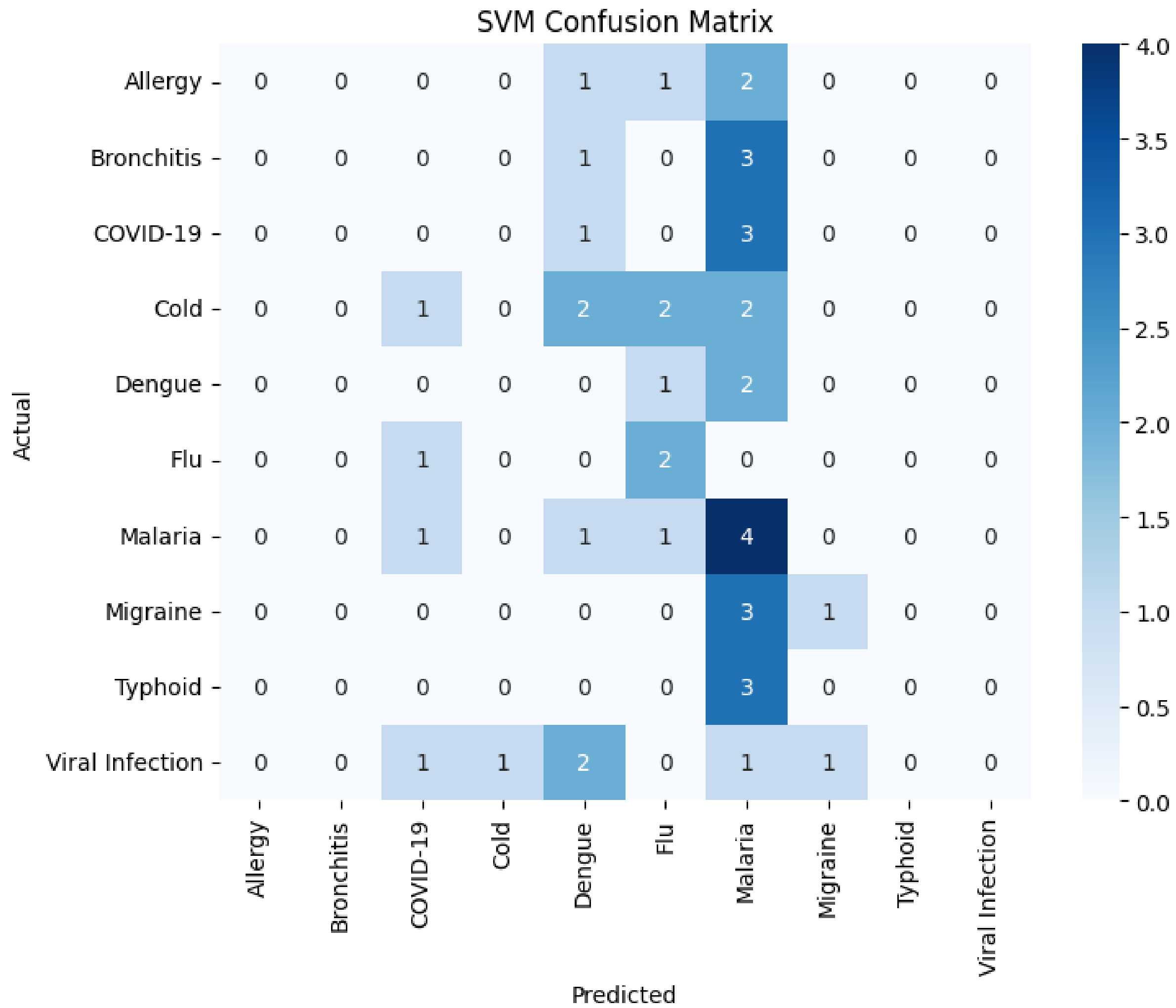
_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

SVM Classification Report:

	precision	recall	f1-score	support
Allergy	0.00	0.00	0.00	4
Bronchitis	0.00	0.00	0.00	4
COVID-19	0.00	0.00	0.00	4
Cold	0.00	0.00	0.00	7
Dengue	0.00	0.00	0.00	3
Flu	0.29	0.67	0.40	3
Malaria	0.17	0.57	0.27	7
Migraine	0.50	0.25	0.33	4
Typhoid	0.00	0.00	0.00	3
Viral Infection	0.00	0.00	0.00	6
accuracy			0.16	45
macro avg	0.10	0.15	0.10	45
weighted avg	0.09	0.16	0.10	45



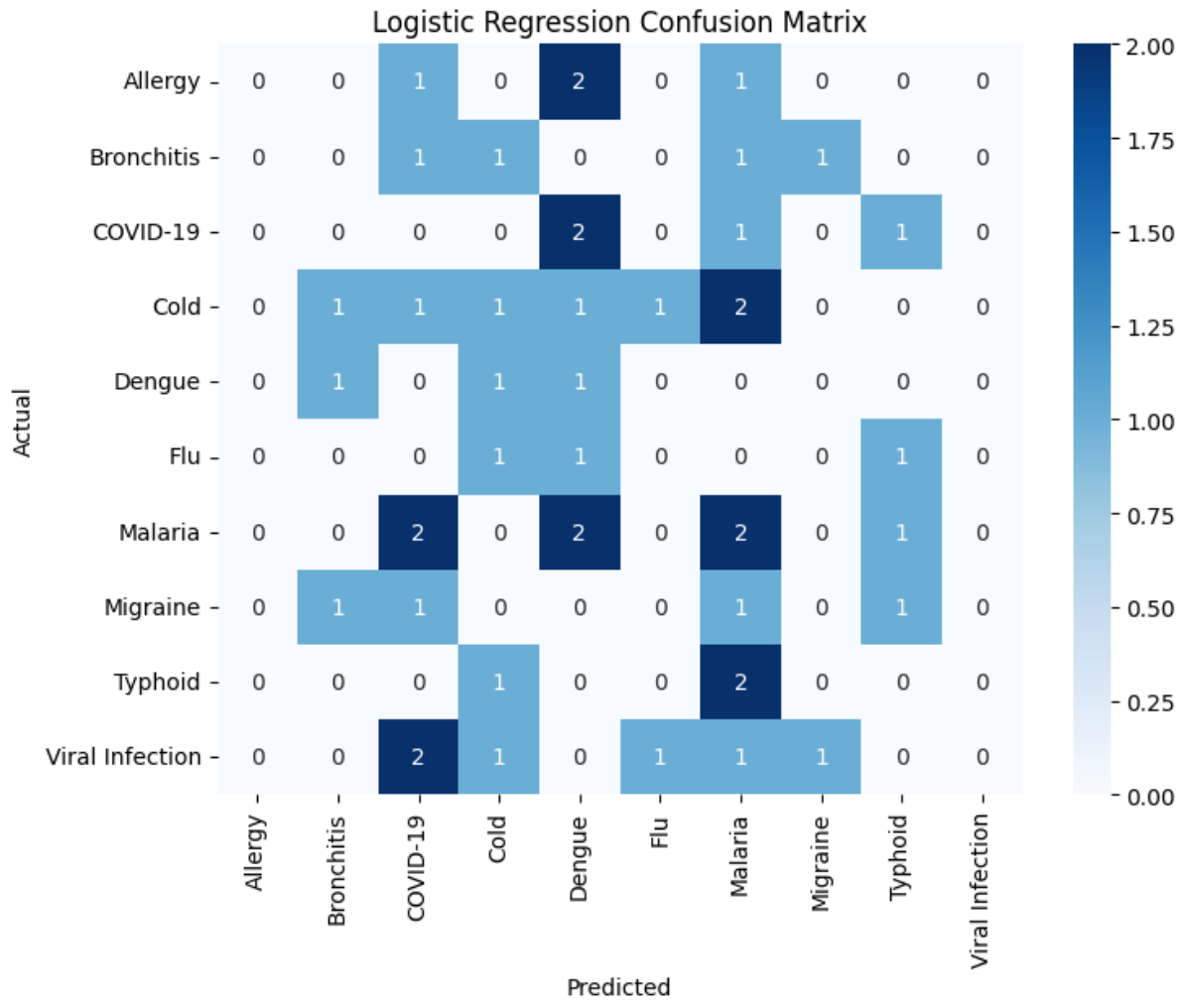

```
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

  _warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

Logistic Regression Classification Report:

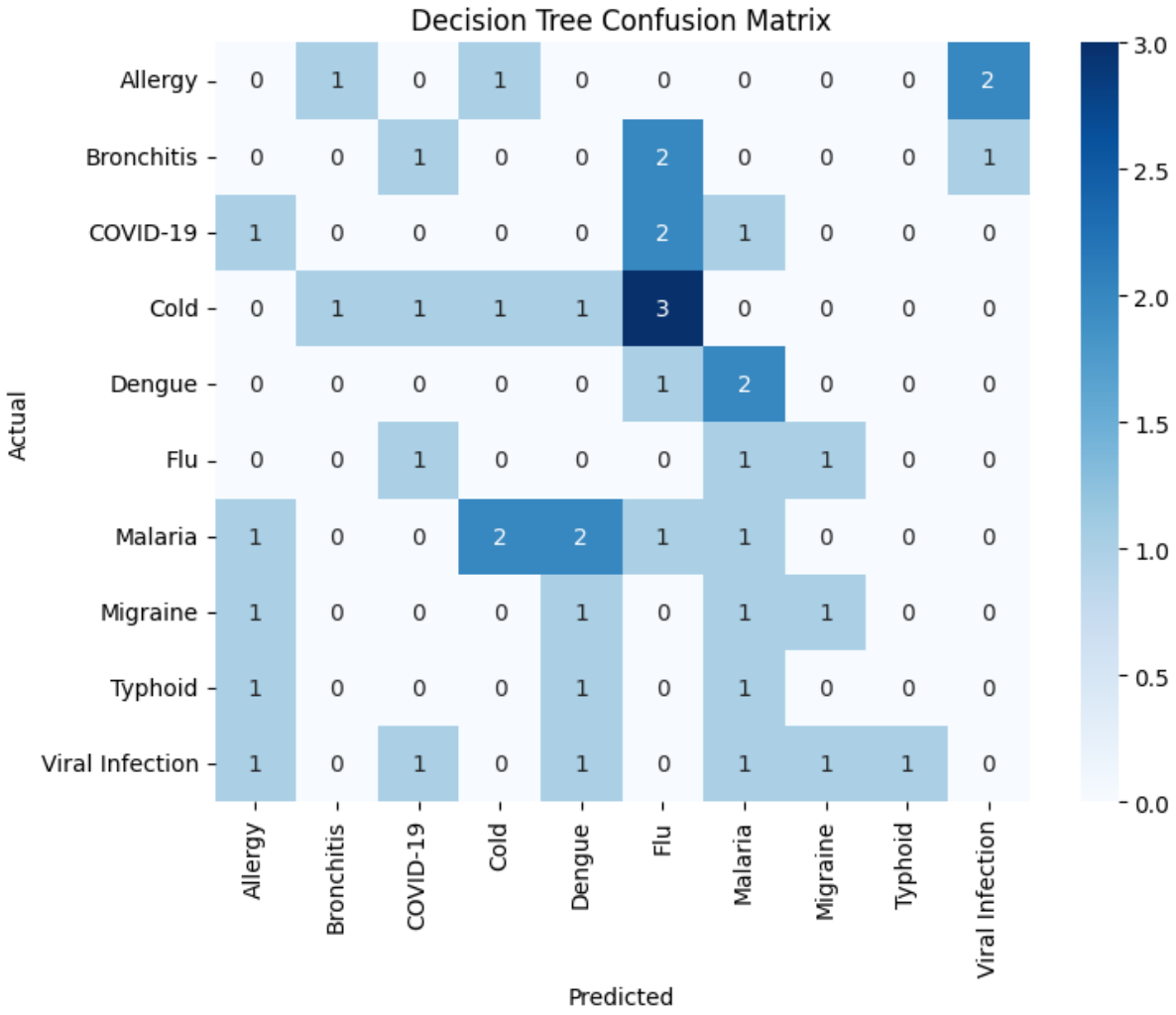
	precision	recall	f1-score	support
Allergy	0.00	0.00	0.00	4
Bronchitis	0.00	0.00	0.00	4
COVID-19	0.00	0.00	0.00	4
Cold	0.17	0.14	0.15	7
Dengue	0.11	0.33	0.17	3
Flu	0.00	0.00	0.00	3
Malaria	0.18	0.29	0.22	7
Migraine	0.00	0.00	0.00	4
Typhoid	0.00	0.00	0.00	3
Viral Infection	0.00	0.00	0.00	6
accuracy			0.09	45
macro avg	0.05	0.08	0.05	45
weighted avg	0.06	0.09	0.07	45



Decision Tree Classification Report:

	precision	recall	f1-score	support
Allergy	0.00	0.00	0.00	4
Bronchitis	0.00	0.00	0.00	4
COVID-19	0.00	0.00	0.00	4
Cold	0.25	0.14	0.18	7
Dengue	0.00	0.00	0.00	3
Flu	0.00	0.00	0.00	3
Malaria	0.12	0.14	0.13	7
Migraine	0.33	0.25	0.29	4

Typhoid	0.00	0.00	0.00	3
Viral Infection	0.00	0.00	0.00	6
accuracy			0.07	45
macro avg	0.07	0.05	0.06	45
weighted avg	0.09	0.07	0.07	45



/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

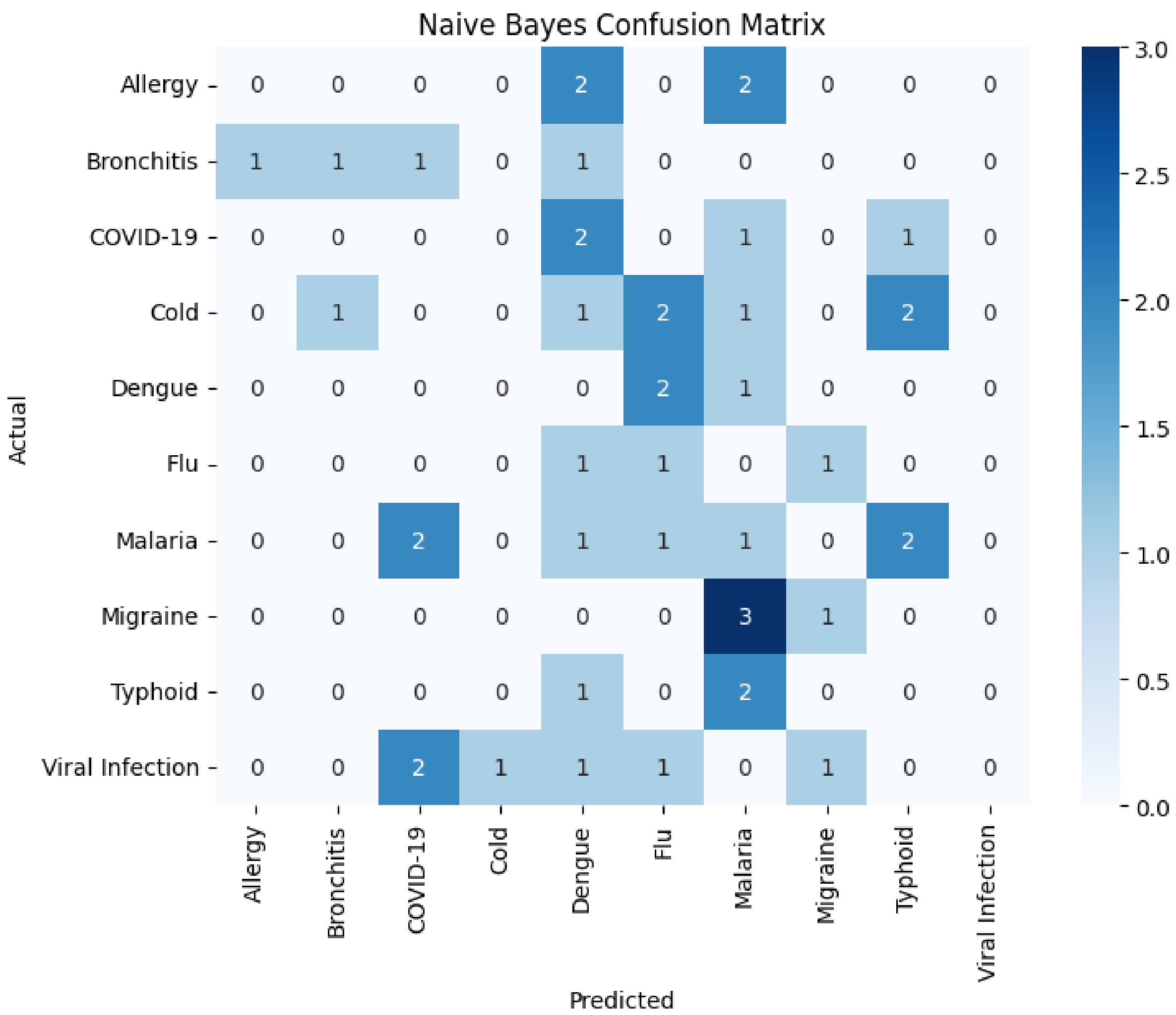
_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

Naive Bayes Classification Report:

	precision	recall	f1-score	support
Allergy	0.00	0.00	0.00	4
Bronchitis	0.50	0.25	0.33	4
COVID-19	0.00	0.00	0.00	4
Cold	0.00	0.00	0.00	7
Dengue	0.00	0.00	0.00	3
Flu	0.14	0.33	0.20	3
Malaria	0.09	0.14	0.11	7
Migraine	0.33	0.25	0.29	4
Typhoid	0.00	0.00	0.00	3
Viral Infection	0.00	0.00	0.00	6
accuracy			0.09	45
macro avg	0.11	0.10	0.09	45
weighted avg	0.10	0.09	0.09	45



/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

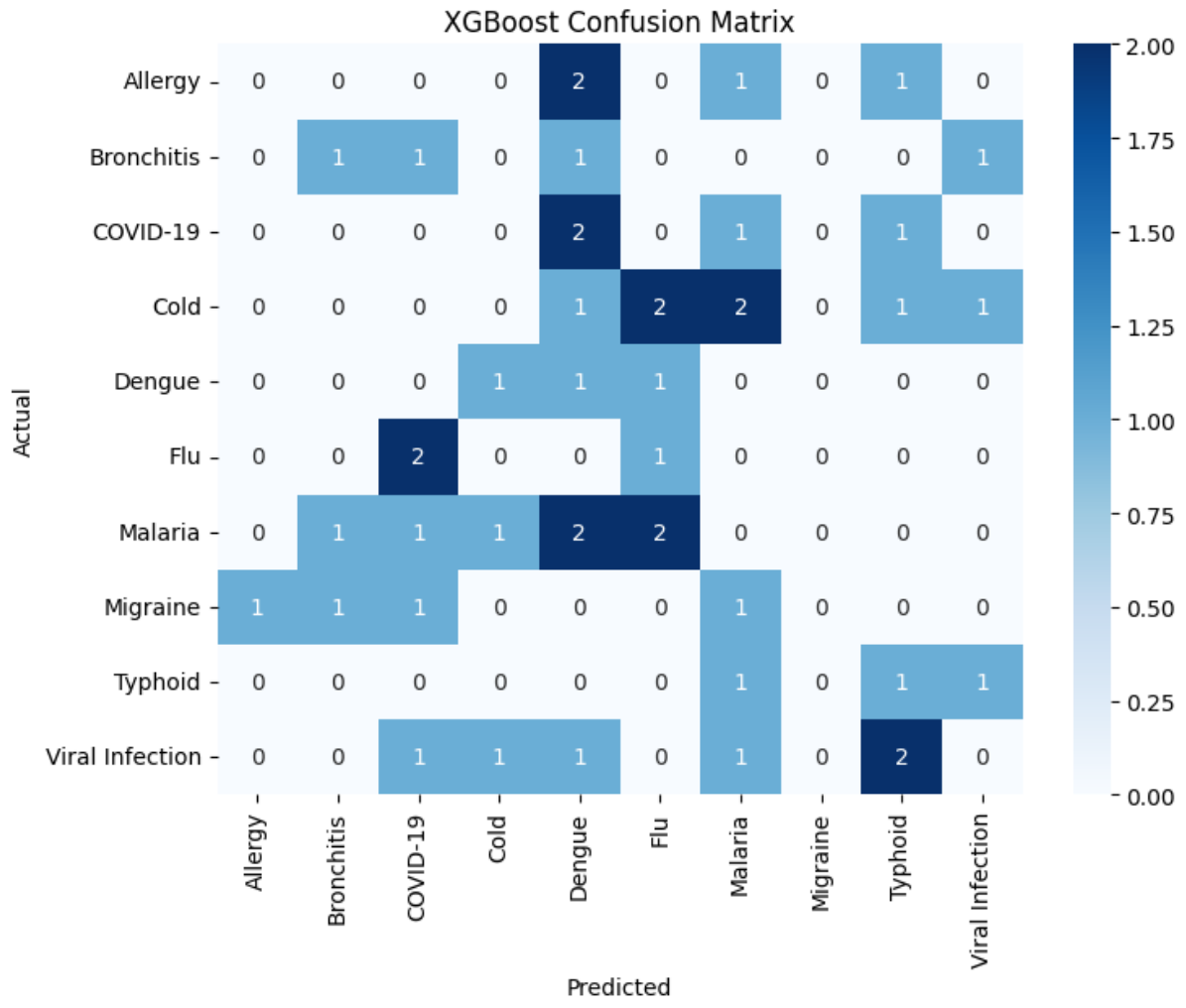
_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.

_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))

XGBoost Classification Report:

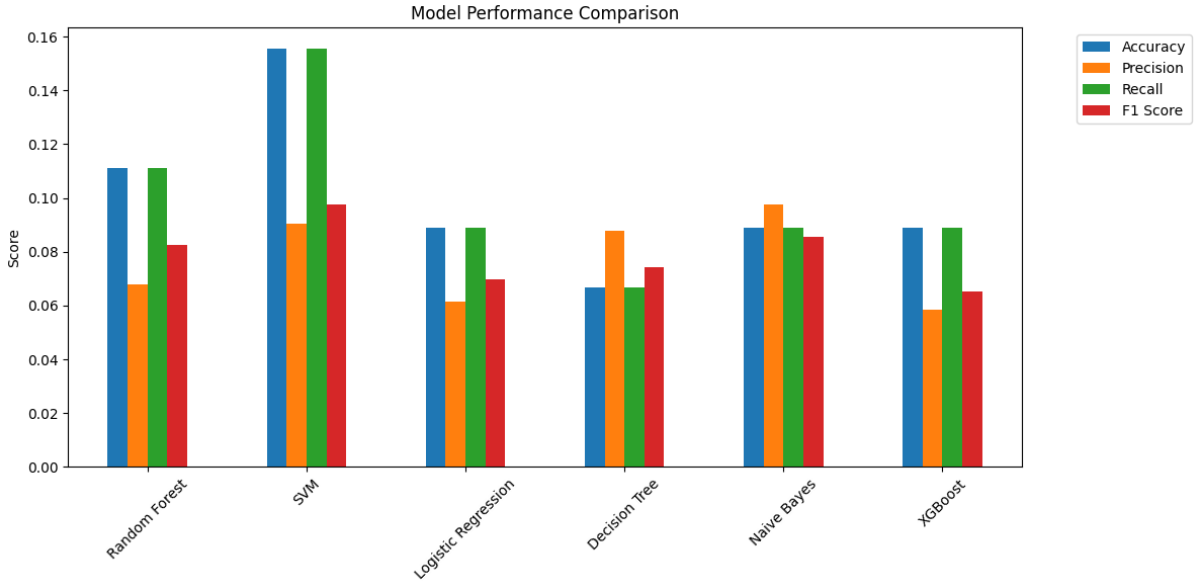
	precision	recall	f1-score	support
Allergy	0.00	0.00	0.00	4
Bronchitis	0.33	0.25	0.29	4
COVID-19	0.00	0.00	0.00	4
Cold	0.00	0.00	0.00	7
Dengue	0.10	0.33	0.15	3
Flu	0.17	0.33	0.22	3
Malaria	0.00	0.00	0.00	7
Migraine	0.00	0.00	0.00	4
Typhoid	0.17	0.33	0.22	3
Viral Infection	0.00	0.00	0.00	6
accuracy			0.09	45
macro avg	0.08	0.12	0.09	45
weighted avg	0.06	0.09	0.07	45

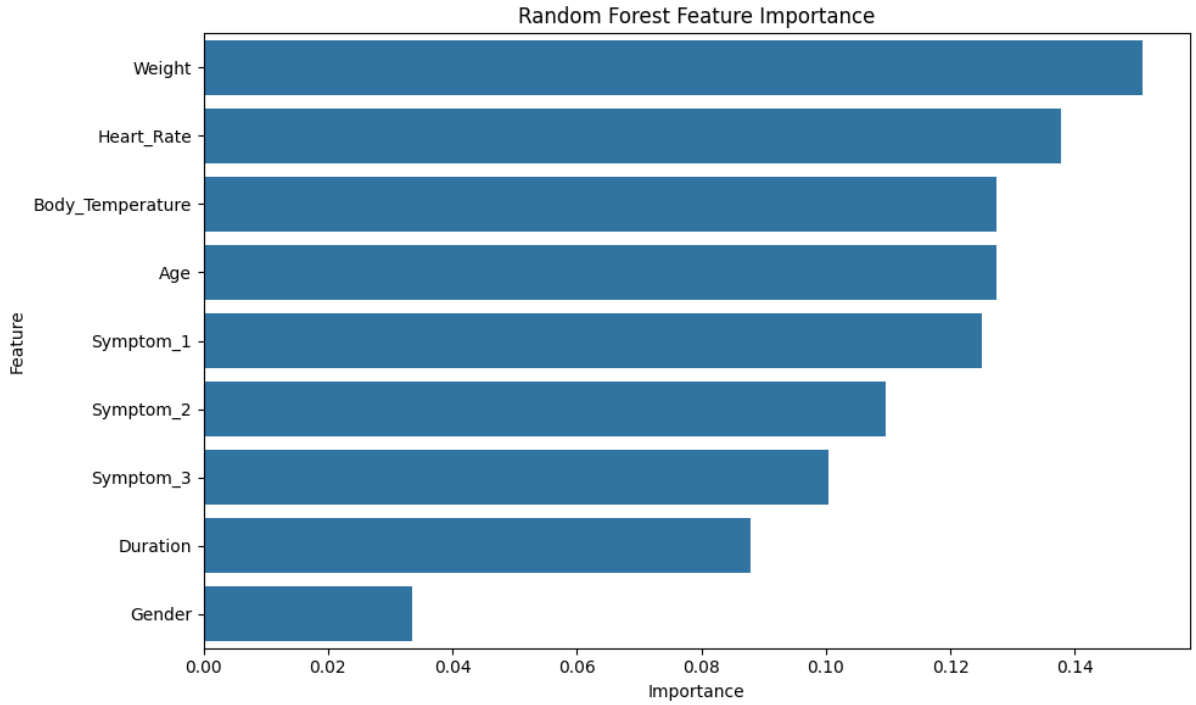


Model Performance Comparison:

	Accuracy	Precision	Recall	F1 Score
Random Forest	0.111111	0.067912	0.111111	0.082428
SVM	0.155556	0.090545	0.155556	0.097778
Logistic Regression	0.088889	0.061616	0.088889	0.069611
Decision Tree	0.066667	0.087963	0.066667	0.074420
Naive Bayes	0.088889	0.097739	0.088889	0.085644
XGBoost	0.088889	0.058519	0.088889	0.065283

<Figure size 1200x600 with 0 Axes>


Model Performance Comparison
Accuracy
Precision
Recall
F1 Score
Score
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Random Forest
SVM
Logistic Regression
Decision Tree
Naive Bayes
XGBoost


Random Forest Feature Importance
Feature
Weight
Heart_Rate
Body_Temperature
Age
Symptom_1
Symptom_2
Symptom_3
Duration
Gender
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Importance