```
In [ ]:
```

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

In []:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.style
from sklearn.model_selection import train_test_split,GridSearchCV
from sklearn.preprocessing import StandardScaler

import warnings
warnings.filterwarnings("ignore")
```

In []:

```
data = pd.read_csv('/content/drive/MyDrive/Training.csv')
```

In []:

```
data.head()
```

Out[83]:

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain :
0	1	1	1	0	0	0	0
1	0	1	1	0	0	0	0
2	1	0	1	0	0	0	0
3	1	1	0	0	0	0	0
4	1	1	1	0	0	0	0

5 rows × 134 columns

```
→
```

```
data=data.drop('Unnamed: 133',axis=1)
```

```
In [ ]:
```

```
data.head()
```

Out[85]:

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain :
0	1	1	1	0	0	0	0
1	0	1	1	0	0	0	0
2	1	0	1	0	0	0	0
3	1	1	0	0	0	0	0
4	1	1	1	0	0	0	0

5 rows × 133 columns

In []:

```
unique_disease=data['prognosis'].unique()
```

In []:

```
unique_disease
```

Out[87]:

```
In [ ]:
```

```
data.describe(include="all")
```

Out[88]:

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	
count	4920.000000	4920.000000	4920.000000	4920.000000	4920.000000	492
unique	NaN	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	NaN	
mean	0.137805	0.159756	0.021951	0.045122	0.021951	
std	0.344730	0.366417	0.146539	0.207593	0.146539	
min	0.000000	0.000000	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	0.000000	0.000000	
75%	0.000000	0.000000	0.000000	0.000000	0.000000	
max	1.000000	1.000000	1.000000	1.000000	1.000000	

11 rows × 133 columns

```
In []:

data.shape

Out[89]:
  (4920, 133)

In []:

data.isnull().sum().sum()

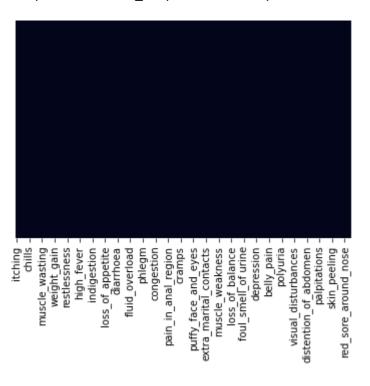
Out[90]:
```

0

```
sns.heatmap(data.isnull(),yticklabels=False,cbar=False)
```

Out[91]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f44494f03d0>



In []:

data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4920 entries, 0 to 4919

Columns: 133 entries, itching to prognosis

dtypes: int64(132), object(1)

memory usage: 5.0+ MB

data.corr()

Out[93]:

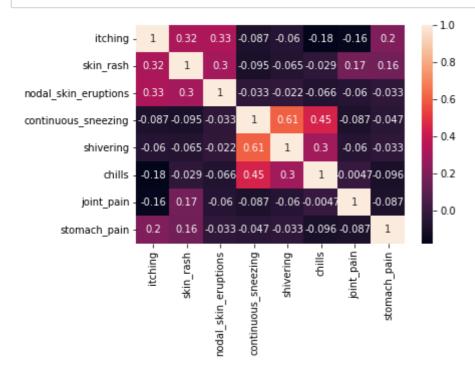
	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shiveri
itching	1.000000	0.318158	0.326439	-0.086906	-0.0598
skin_rash	0.318158	1.000000	0.298143	-0.094786	-0.0653
nodal_skin_eruptions	0.326439	0.298143	1.000000	-0.032566	-0.0224
continuous_sneezing	-0.086906	-0.094786	-0.032566	1.000000	0.6089
shivering	-0.059893	-0.065324	-0.022444	0.608981	1.0000
small_dents_in_nails	-0.061573	0.331087	-0.023073	-0.033480	-0.0230
inflammatory_nails	-0.061573	0.331087	-0.023073	-0.033480	-0.0230
blister	-0.061573	0.331087	-0.023073	-0.033480	-0.0230
red_sore_around_nose	-0.061573	0.331087	-0.023073	-0.033480	-0.0230
yellow_crust_ooze	-0.061573	0.331087	-0.023073	-0.033480	-0.0230

132 rows × 132 columns

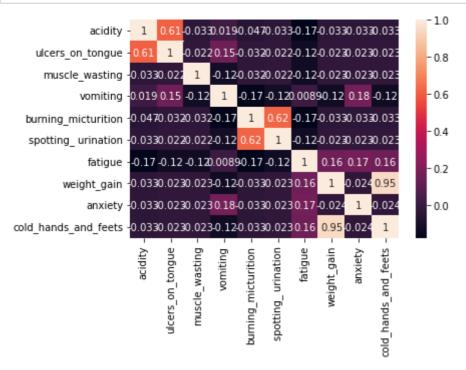
→

In []:

sns.heatmap(data.iloc[:, 0:8].corr(),annot=True)
plt.show()



```
sns.heatmap(data.iloc[:, 8:18].corr(),annot=True)
plt.show()
```



In []:

```
X = data.drop(['prognosis'], axis=1)
y = data['prognosis'].values
```

In []:

```
 X\_train, \ X\_test, \ y\_train, \ y\_test = train\_test\_split(X, \ y, \ test\_size=0.30 \ , \ random\_state=1)
```

In []:

```
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

```
(3444, 132)
(3444,)
(1476, 132)
(1476,)
```

KNN

```
In [ ]:
```

```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.model_selection import RandomizedSearchCV
from sklearn.metrics import roc_auc_score,classification_report
#,roc_curve,classification_report,confusion_matrix,plot_confusion_matrix
```

```
neigh = KNeighborsClassifier()
params ={'n_neighbors':[3,5,7,9,11,13,15,17,19,21,23]}
folds = 3

clf = RandomizedSearchCV(neigh, params, cv=folds, scoring='accuracy', n_iter=10, random_stasearch = clf.fit(X_train,y_train)
best_k=search.best_params_['n_neighbors']
print(best_k)
```

13

In []:

```
neigh = KNeighborsClassifier(n_neighbors=best_k)
neigh.fit(X_train, y_train)
```

Out[101]:

KNeighborsClassifier(n_neighbors=13)

In []:

```
neigh.score(X_train,y_train)
```

Out[102]:

1.0

In []:

```
neigh.score(X_test,y_test)
```

Out[103]:

1.0

In []:

```
ytrain_predict_probknn=neigh.predict_proba(X_train)
ytest_predict_probknn=neigh.predict_proba(X_test)
```

```
ytrain_predict_knn=neigh.predict(X_train)
ytest_predict_knn=neigh.predict(X_test)
```

```
train_auc = roc_auc_score(y_train, ytrain_predict_probknn,multi_class='ovr')
print("Train AUC:",train_auc)
```

Train AUC: 1.0

In []:

```
test_auc = roc_auc_score(y_test, ytest_predict_probknn,multi_class='ovr')
print("Test AUC:",test_auc)
```

Test AUC: 1.0

In []:

print(classification_report(y_test, ytest_predict_knn))

port		precision	recall	f1-score	sup
	Paroymsal Positional Vertigo	1.00	1.00	1.00	
43	AIDS	1.00	1.00	1.00	
31	Acne	1.00	1.00	1.00	
39	Alcoholic hepatitis	1.00	1.00	1.00	
29	Allergy	1.00	1.00	1.00	
40	Arthritis	1.00	1.00	1.00	
35	Bronchial Asthma	1.00	1.00	1.00	
33	Cervical spondylosis	1.00	1.00	1.00	
34	Chicken pox	1.00	1.00	1.00	
32	Chronic cholestasis	1.00	1.00	1.00	
31	Common Cold	1.00	1.00	1.00	
35	Dengue	1.00	1.00	1.00	
41	Diabetes	1.00	1.00	1.00	
36	Dimorphic hemmorhoids(piles)	1.00	1.00	1.00	
41	Drug Reaction	1.00	1.00	1.00	
28	Fungal infection	1.00	1.00	1.00	
33	GERD	1.00	1.00	1.00	
34	Gastroenteritis	1.00	1.00	1.00	
38	Heart attack	1.00	1.00	1.00	
44	Hepatitis B	1.00	1.00	1.00	
35	Hepatitis C	1.00	1.00	1.00	
45	Hepatitis D	1.00	1.00	1.00	
40	Hepatitis E	1.00	1.00	1.00	
34	Hypertension	1.00	1.00	1.00	
37	Hyperthyroidism	1.00	1.00	1.00	
37	Hypoglycemia	1.00	1.00	1.00	
37	Hypothyroidism	1.00	1.00	1.00	
35					

1/30/21, 8:28 PM		Disease_Prediction	n_Final - Jupyter	Notebook
29	Impetigo	1.00	1.00	1.00
	Jaundice	1.00	1.00	1.00
39	Malaria	1.00	1.00	1.00
31	Migraine	1.00	1.00	1.00
38	Osteoarthristis	1.00	1.00	1.00
41	Paralysis (brain hemorrhage)	1.00	1.00	1.00
42	, , ,			
37	Peptic ulcer diseae	1.00	1.00	1.00
34	Pneumonia	1.00	1.00	1.00
32	Psoriasis	1.00	1.00	1.00
33	Tuberculosis	1.00	1.00	1.00
36	Typhoid	1.00	1.00	1.00
	Urinary tract infection	1.00	1.00	1.00
30	Varicose veins	1.00	1.00	1.00
36	hepatitis A	1.00	1.00	1.00
41				
1476	accuracy			1.00
	macro avg	1.00	1.00	1.00
1476	weighted avg	1.00	1.00	1.00
1476				

Decision Tree

In []:

from sklearn.tree import DecisionTreeClassifier

In []:

```
dtree = DecisionTreeClassifier()
dtree.fit(X_train, y_train)
```

Out[110]:

DecisionTreeClassifier()

In []:

ytrain_predict_probdt=dtree.predict_proba(X_train)
ytest_predict_probdt=dtree.predict_proba(X_test)

```
In [ ]:
```

```
train_auc = roc_auc_score(y_train, ytrain_predict_probdt,multi_class='ovr')
print("Train AUC:",train_auc)
```

Train AUC: 1.0

In []:

```
test_auc = roc_auc_score(y_test, ytest_predict_probdt,multi_class='ovr')
print("Test AUC:",test_auc)
```

Test AUC: 1.0

In []:

```
ytrain_predict_dt=dtree.predict(X_train)
ytest_predict_dt=dtree.predict(X_test)
```

In []:

```
dtree.score(X_train,y_train)
dtree.score(X_test,y_test)
```

Out[115]:

In []:

	<pre>print(classification_report(y_test,</pre>	vtest predict dt))
ı	print(crassification_report(y_test)	y cest_predict_dt//

port		precision	recall	f1-score	sup
	Paroymsal Positional Vertigo	1.00	1.00	1.00	
43	AIDS	1.00	1.00	1.00	
31	Acne	1.00	1.00	1.00	
39	Alcoholic hepatitis	1.00	1.00	1.00	
29	Allergy	1.00	1.00	1.00	
40	Arthritis	1.00	1.00	1.00	
35	Bronchial Asthma	1.00	1.00	1.00	
33	Cervical spondylosis	1.00	1.00	1.00	
34	Chicken pox	1.00	1.00	1.00	
32	Chronic cholestasis	1.00	1.00	1.00	
31	Common Cold	1.00	1.00	1.00	
35	Dengue	1.00	1.00	1.00	
41	_				
36	Diabetes	1.00	1.00	1.00	
41	Dimorphic hemmorhoids(piles)	1.00	1.00	1.00	
28	Drug Reaction	1.00	1.00	1.00	
33	Fungal infection	1.00	1.00	1.00	
34	GERD	1.00	1.00	1.00	
38	Gastroenteritis	1.00	1.00	1.00	
44	Heart attack	1.00	1.00	1.00	
35	Hepatitis B	1.00	1.00	1.00	
45	Hepatitis C	1.00	1.00	1.00	
40	Hepatitis D	1.00	1.00	1.00	
34	Hepatitis E	1.00	1.00	1.00	
	Hypertension	1.00	1.00	1.00	
37	Hyperthyroidism	1.00	1.00	1.00	
37	Hypoglycemia	1.00	1.00	1.00	
37	Hypothyroidism	1.00	1.00	1.00	
35	shoots (Doots of Disease Boots that on Final issue)				

1/30/21, 8:28 PM		Disease_Prediction	ı_Final - Jupyter N	lotebook
29	Impetigo	1.00	1.00	1.00
39	Jaundice	1.00	1.00	1.00
	Malaria	1.00	1.00	1.00
31	Migraine	1.00	1.00	1.00
38	Osteoarthristis	1.00	1.00	1.00
41	Paralysis (brain hemorrhage)	1.00	1.00	1.00
42	Peptic ulcer diseae	1.00	1.00	1.00
37	Pneumonia	1.00	1.00	1.00
34	Psoriasis	1.00	1.00	1.00
32	Tuberculosis	1.00	1.00	1.00
33	Typhoid	1.00	1.00	1.00
36	Urinary tract infection	1.00	1.00	1.00
30	Varicose veins	1.00	1.00	1.00
36	hepatitis A	1.00	1.00	1.00
41	nepatitis A	1.00	1.00	1.00
4.7-6	accuracy			1.00
1476	macro avg	1.00	1.00	1.00
1476	weighted avg	1.00	1.00	1.00
1476				

RandomForest

In []:

from sklearn.ensemble import RandomForestClassifier

In []:

```
rf = RandomForestClassifier()
rf.fit(X_train, y_train)
```

Out[118]:

RandomForestClassifier()

```
ytrain_predict_probrf=rf.predict_proba(X_train)
ytest_predict_probrf=rf.predict_proba(X_test)
```

```
In [ ]:
```

```
train_auc = roc_auc_score(y_train, ytrain_predict_probrf,multi_class='ovr')
print("Train AUC:",train_auc)
```

Train AUC: 1.0

In []:

```
test_auc = roc_auc_score(y_test, ytest_predict_probrf,multi_class='ovr')
print("Test AUC:",test_auc)
```

Test AUC: 1.0

In []:

```
ytrain_predict_rf=rf.predict(X_train)
ytest_predict_rf=rf.predict(X_test)
```

In []:

```
rf.score(X_train,y_train)
rf.score(X_test,y_test)
```

Out[123]:

In []:

print(classification_report(y_test, ytest_predict_rf))

upport		precision	recall	f1-score	S	
	Paroymsal Positional Vertigo	1.00	1.00	1.00		
43	AIDS	1.00	1.00	1.00		
31	Acne	1.00	1.00	1.00		
39	Alcoholic hepatitis	1.00	1.00	1.00		
29	Allergy	1.00	1.00	1.00		
40	Arthritis	1.00	1.00	1.00		
35	Bronchial Asthma	1.00	1.00	1.00		
33	Cervical spondylosis	1.00	1.00	1.00		
34	Chicken pox	1.00	1.00	1.00		
32	Chronic cholestasis	1.00	1.00	1.00		
31	Common Cold	1.00	1.00	1.00		
35	Dengue	1.00	1.00	1.00		
41	Diabetes	1.00	1.00	1.00		
36	Dimorphic hemmorhoids(piles)	1.00	1.00	1.00		
41	Drug Reaction	1.00	1.00	1.00		
28	Fungal infection	1.00	1.00	1.00		
33	GERD	1.00	1.00	1.00		
34	Gastroenteritis	1.00	1.00	1.00		
38	Heart attack	1.00	1.00	1.00		
44	Hepatitis B	1.00	1.00	1.00		
35	Hepatitis C	1.00	1.00	1.00		
45	Hepatitis D	1.00	1.00	1.00		
40	Hepatitis E	1.00	1.00	1.00		
34	Hypertension	1.00	1.00	1.00		
37	Hyperthyroidism	1.00	1.00	1.00		
37	Hypoglycemia	1.00	1.00	1.00		
37	Hypothyroidism	1.00	1.00	1.00		
35	ypochyroddism	1.00	1.00	1.00		

1/30/21, 8:28 PM		Disease_Prediction_Fina	al - Jupyter I	Notebook	
29	Impetigo	1.00	1.00	1.00	
	Jaundice	1.00	1.00	1.00	
39	Malaria	1.00	1.00	1.00	
31	Migraine	1.00	1.00	1.00	
38	Osteoarthristis	1.00	1.00	1.00	
41	Paralysis (brain hemorrhage)	1.00	1.00	1.00	
42	Peptic ulcer diseae	1.00	1.00	1.00	
37	Pneumonia	1.00	1.00	1.00	
34	Psoriasis	1.00	1.00	1.00	
32					
33	Tuberculosis	1.00	1.00	1.00	
36	Typhoid	1.00	1.00	1.00	
30	Urinary tract infection	1.00	1.00	1.00	
36	Varicose veins	1.00	1.00	1.00	
41	hepatitis A	1.00	1.00	1.00	
	accuracy			1.00	
1476	-	1.00	1.00	1.00	
1476	macro avg				
1476	weighted avg	1.00	1.00	1.00	
					-

GradientBoosting

In []:

from sklearn.ensemble import GradientBoostingClassifier

In []:

```
gradientclass = GradientBoostingClassifier(n_estimators=100, learning_rate=1.0, max_depth=1,
```

```
ytrain_predict_probgb=gradientclass.predict_proba(X_train)
ytest_predict_probgb=gradientclass.predict_proba(X_test)
```

```
train_auc = roc_auc_score(y_train, ytrain_predict_probgb,multi_class='ovr')
print("Train AUC:",train_auc)
```

Train AUC: 0.48896891338601567

In []:

```
test_auc = roc_auc_score(y_test, ytest_predict_probgb,multi_class='ovr')
print("Test AUC:",test_auc)
```

Test AUC: 0.48801191826233714

In []:

```
ytrain_predict_gradientclass=gradientclass.predict(X_train)
ytest_predict_gradientclass=gradientclass.predict(X_test)
```

In []:

```
gradientclass.score(X_train,y_train)
gradientclass.score(X_test,y_test)
```

Out[131]:

In []:

print(classification_report(y_test, ytest_predict_gradientclass))

port		precision	recall	f1-score	sup
	Paroymsal Positional Vertigo	0.00	0.00	0.00	
43	AIDS	0.00	0.00	0.00	
31	Acne	0.00	0.00	0.00	
39	Alcoholic hepatitis	0.00	0.00	0.00	
29	Allergy	0.00	0.00	0.00	
40	Arthritis	0.00	0.00	0.00	
35	Bronchial Asthma	0.00	0.00	0.00	
33	Cervical spondylosis	0.00	0.00	0.00	
34	Chicken pox	0.00	0.00	0.00	
32	Chronic cholestasis	0.00	0.00	0.00	
31	Common Cold	0.00	0.00	0.00	
35	Dengue	0.00	0.00	0.00	
41	Diabetes	0.00	0.00	0.00	
36	Dimorphic hemmorhoids(piles)	0.00	0.00	0.00	
41	Drug Reaction	0.00	0.04	0.00	
28	Fungal infection	0.00	0.00	0.00	
33	GERD	0.00	0.00	0.00	
34	Gastroenteritis	0.00	0.00	0.00	
38	Heart attack	0.00	0.00	0.00	
44	Hepatitis B	0.00	0.00	0.00	
35	Hepatitis C	0.00	0.00	0.00	
45	Hepatitis D	0.00	0.00	0.00	
40	Hepatitis E	0.00	0.00	0.00	
34	Hypertension	0.00	0.00	0.00	
37	Hyperthyroidism	0.00	0.00	0.00	
37	Hypoglycemia	0.00	0.00	0.00	
37	Hypothyroidism	0.00	0.00	0.00	
35					

1/30/21, 8:28 PM	Disease_Prediction_Final - Jupyter Notebo			
29	Impetigo	0.00	0.00	0.00
	Jaundice	0.00	0.00	0.00
39	Malaria	0.00	0.00	0.00
31	Migraine	0.00	0.00	0.00
38	Osteoarthristis	0.00	0.00	0.00
41				
42	Paralysis (brain hemorrhage)	0.00	0.00	0.00
37	Peptic ulcer diseae	0.00	0.00	0.00
34	Pneumonia	0.00	0.00	0.00
	Psoriasis	0.00	0.00	0.00
32	Tuberculosis	0.00	0.00	0.00
33	Typhoid	0.00	0.00	0.00
36	Urinary tract infection	0.00	0.00	0.00
30	•			
36	Varicose veins	0.00	0.00	0.00
41	hepatitis A	0.00	0.00	0.00
	accuracy			0.00
1476	accuracy	0.00	0.05	
1476	macro avg	0.00	0.00	0.00
1476	weighted avg	0.00	0.00	0.00

LogisticRegression

In []:

```
from sklearn.linear_model import LogisticRegression
```

In []:

```
model = LogisticRegression()
model = model.fit(X_train, y_train)
```

```
ytrain_predict_problr=model.predict_proba(X_train)
ytest_predict_problr=model.predict_proba(X_test)
```

```
In [ ]:
```

```
train_auc = roc_auc_score(y_train, ytrain_predict_problr,multi_class='ovr')
print("Train AUC:",train_auc)
```

Train AUC: 1.0

In []:

```
test_auc = roc_auc_score(y_test, ytest_predict_problr,multi_class='ovr')
print("Test AUC:",test_auc)
```

Test AUC: 1.0

In []:

```
ytrain_predict_model=model.predict(X_train)
ytest_predict_model=model.predict(X_test)
```

In []:

```
model.score(X_train,y_train)
model.score(X_test,y_test)
```

Out[139]:

In []:

print(classification_report(y_test, ytest_predict_model))

port		precision	recall	f1-score	sup
(vertigo) 43	Paroymsal Positional Vertigo	1.00	1.00	1.00	
31	AIDS	1.00	1.00	1.00	
39	Acne	1.00	1.00	1.00	
29	Alcoholic hepatitis	1.00	1.00	1.00	
40	Allergy	1.00	1.00	1.00	
35	Arthritis	1.00	1.00	1.00	
	Bronchial Asthma	1.00	1.00	1.00	
33	Cervical spondylosis	1.00	1.00	1.00	
34	Chicken pox	1.00	1.00	1.00	
32	Chronic cholestasis	1.00	1.00	1.00	
31	Common Cold	1.00	1.00	1.00	
35	Dengue	1.00	1.00	1.00	
41	Diabetes	1.00	1.00	1.00	
36	Dimorphic hemmorhoids(piles)	1.00	1.00	1.00	
41	Drug Reaction	1.00	1.00	1.00	
28	Fungal infection	1.00	1.00	1.00	
33	GERD	1.00	1.00	1.00	
34	Gastroenteritis	1.00	1.00	1.00	
38	Heart attack	1.00	1.00	1.00	
44	Hepatitis B	1.00	1.00	1.00	
35	Hepatitis C	1.00	1.00	1.00	
45	Hepatitis D	1.00	1.00	1.00	
40	Hepatitis E	1.00	1.00	1.00	
34	Hypertension	1.00	1.00	1.00	
37	Hyperthyroidism	1.00	1.00	1.00	
37	Hypoglycemia	1.00	1.00	1.00	
37	Hypothyroidism	1.00	1.00	1.00	
35	Trypochyr otatsii	1.00	1.00	1.00	

1/30/21, 8:28 PM	Disease_Prediction_Final - Jupyter Notebook				
20	Impetigo	1.00	1.00	1.00	
29	Jaundice	1.00	1.00	1.00	
39	Malaria	1.00	1.00	1.00	
31					
38	Migraine	1.00	1.00	1.00	
41	Osteoarthristis	1.00	1.00	1.00	
	Paralysis (brain hemorrhage)	1.00	1.00	1.00	
42	Peptic ulcer diseae	1.00	1.00	1.00	
37	Pneumonia	1.00	1.00	1.00	
34					
32	Psoriasis	1.00	1.00	1.00	
33	Tuberculosis	1.00	1.00	1.00	
	Typhoid	1.00	1.00	1.00	
36	Urinary tract infection	1.00	1.00	1.00	
30	Varicose veins	1.00	1.00	1.00	
36					
41	hepatitis A	1.00	1.00	1.00	
	accuracy			1.00	
1476	•	1 00	1 00		
1476	macro avg	1.00	1.00	1.00	
1476	weighted avg	1.00	1.00	1.00	
±-+/ U					