#Diseases Dataset

In [75]:
```
#This is a machine learning project on different type of diseases.
#Are there any potential relationships with different epidemis issues.
#Does these diseases have anything in common with each other
#Are there any other factors that can cause or make these diseases worse
```

In [121…
```
!date
```

```
Sat 20 Jul 2024 02:16:40 PM EDT
```

In [46]:
```
from IPython import display
display.Image('/home/harlohutch77/python/python_master/rash.jpeg',height=400, width=800)
```

Out[46]:

In [90]:
```
# importing libraries
```

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
#This is a csv that is being made into a dataframe
df =pd.read_csv('/home/harlohutch77/python/python_master/d_training.csv')
df.head()
```

Out[2]:

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	…
0	1	1	1	0	0	0	0	0	0	0	…
1	0	1	1	0	0	0	0	0	0	0	…
2	1	0	1	0	0	0	0	0	0	0	…
3	1	1	0	0	0	0	0	0	0	0	…
4	1	1	1	0	0	0	0	0	0	0	…

5 rows × 134 columns

In [1]:
```
# other examples of removing columns
# Drop column 'B'
#df = df.drop('B', axis=1)

# Drop columns multiple column
#df = df.drop(['B', 'C'], axis=1)
```

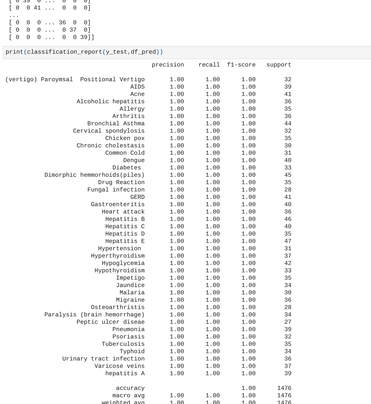
In [3]:
```
#del df['column_name'] deletes column from dataframe
del df['Unnamed: 133']
```

In [4]:
```
import seaborn as sns
import cufflinks as cf
```

In [5]:
```
from plotly import __version__
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot

print(__version__) # requires version >= 1.9.0

# For Notebooks - jupyter notebook
init_notebook_mode(connected=True)

cf.go_offline()
```

```
5.7.0
```

In [6]:
```
#This visualization show the prognosis as negative or postive; 0 = don't have and
#1 = does have
df[['prognosis','itching','skin_rash','nodal_skin_eruptions','blister','joint_pain']].iplot(kind='box')
```



Export to plot.ly »

Machine Learning

In [7]:
```
from sklearn.model_selection import train_test_split
```

In [8]:
```
X = df.drop('prognosis', axis=1)
y = df['prognosis']
```

In [9]:
```
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.30, random_state=42)
```

Random Forest Machine Learning

In [10]:
```
from sklearn.ensemble import RandomForestClassifier
df = RandomForestClassifier(n_estimators=100)
df.fit(X_train, y_train)
```

Out[10]:
```
RandomForestClassifier()
```

In [11]:
```
df_pred = df.predict(X_test)
```

In [12]:
```
from sklearn.metrics import classification_report, confusion_matrix
```

In [13]:
```
#Confusion Matrix
print(confusion_matrix(y_test,df_pred))
```

```
[[32  0  0 ...  0  0  0]
 [ 0 39  0 ...  0  0  0]
 [ 0  0 41 ...  0  0  0]
 ...
 [ 0  0  0 ... 36  0  0]
 [ 0  0  0 ...  0 37  0]
 [ 0  0  0 ...  0  0 39]]
```

In [14]:
```
print(classification_report(y_test,df_pred))
```

```
                                   precision    recall  f1-score   support

(vertigo) Paroymsal  Positional Vertigo       1.00      1.00      1.00        32
                                    AIDS       1.00      1.00      1.00        39
                                    Acne       1.00      1.00      1.00        41
                  Alcoholic hepatitis          1.00      1.00      1.00        36
                                 Allergy       1.00      1.00      1.00        35
                               Arthritis       1.00      1.00      1.00        36
                  Bronchial Asthma             1.00      1.00      1.00        44
              Cervical spondylosis             1.00      1.00      1.00        32
                        Chicken pox            1.00      1.00      1.00        35
             Chronic cholestasis               1.00      1.00      1.00        30
                     Common Cold               1.00      1.00      1.00        31
                          Dengue               1.00      1.00      1.00        40
                        Diabetes               1.00      1.00      1.00        33
  Dimorphic hemmorhoids(piles)                 1.00      1.00      1.00        45
                    Drug Reaction              1.00      1.00      1.00        35
                Fungal infection               1.00      1.00      1.00        28
                            GERD               1.00      1.00      1.00        41
             Gastroenteritis                   1.00      1.00      1.00        40
                  Heart attack                 1.00      1.00      1.00        36
                  Hepatitis B                  1.00      1.00      1.00        46
                  Hepatitis C                  1.00      1.00      1.00        40
                  Hepatitis D                  1.00      1.00      1.00        35
                  Hepatitis E                  1.00      1.00      1.00        47
                 Hypertension                  1.00      1.00      1.00        31
             Hyperthyroidism                   1.00      1.00      1.00        37
                Hypoglycemia                   1.00      1.00      1.00        42
             Hypothyroidism                    1.00      1.00      1.00        33
                    Impetigo                   1.00      1.00      1.00        35
                    Jaundice                   1.00      1.00      1.00        34
                     Malaria                   1.00      1.00      1.00        30
                    Migraine                   1.00      1.00      1.00        36
          Osteoarthritis                       1.00      1.00      1.00        28
  Paralysis (brain hemorrhage)                 1.00      1.00      1.00        34
          Peptic ulcer disease                 1.00      1.00      1.00        27
                   Pneumonia                   1.00      1.00      1.00        39
                   Psoriasis                   1.00      1.00      1.00        32
                Tuberculosis                   1.00      1.00      1.00        35
                    Typhoid                    1.00      1.00      1.00        34
      Urinary tract infection                  1.00      1.00      1.00        36
             Varicose veins                    1.00      1.00      1.00        37
                 hepatitis A                   1.00      1.00      1.00        39

                    accuracy                                       1.00      1476
                   macro avg       1.00      1.00      1.00      1476
                weighted avg       1.00      1.00      1.00      1476
```

In [15]:
```
df =pd.read_csv('/home/harlohutch77/python/python_master/d_training.csv')
df.head()
```

Out[15]:

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue	…
0	1	1	1	0	0	0	0	0	0	0	…
1	0	1	1	0	0	0	0	0	0	0	…
2	1	0	1	0	0	0	0	0	0	0	…
3	1	1	0	0	0	0	0	0	0	0	…
4	1	1	1	0	0	0	0	0	0	0	…

5 rows × 134 columns

In [124…
```
#I am focusing on the epidermis issues to see if there is some kind of commonality
```

In [16]:
```
#the amount of patients with this issue
df['nodal_skin_eruptions'].sum()
```

Out[16]:
```
108
```

In [17]:
```
#this is how many patients that had this issue
df['small_dents_in_nails'].sum()
```

Out[17]:
```
114
```

In [18]:
```
#this is for the full count of patients accounted for
df['small_dents_in_nails'].count()
```

Out[18]:
```
4920
```

In [19]:
```
#this is the amount of patients that have a skin rash
df['skin_rash'].sum()
```

Out[19]:
```
786
```

In [20]:
```
#
df['skin_rash'].describe()
```

Out[20]:
```
count    4920.000000
mean        0.159756
std         0.366417
min         0.000000
25%         0.000000
50%         0.000000
75%         0.000000
max         1.000000
Name: skin_rash, dtype: float64
```

In [21]:
```
#description and explanation of prognosis
df['prognosis'].describe()
```

Out[21]:
```
count              4920
unique               41
top     Fungal infection
freq                120
Name: prognosis, dtype: object
```

In [22]:
```
#114 patients have Psoriasis and skin_rash
df[df['prognosis']=='Psoriasis']['skin_rash'].sum()
```

Out[22]:
```
114
```

In [23]:
```
df[df['prognosis']=='impetigo']['skin_rash'].sum()
```

Out[23]:
```
0
```

In [24]:
```
#this is to cross reference skin issues with fungal infection
df[df['prognosis']=='Fungal infection'][['skin_rash','itching','nodal_skin_eruptions']].sum()
```

Out[24]:
```
skin_rash             108
itching               108
nodal_skin_eruptions    108
dtype: int64
```

In [125…
```
# 4920 prognosis's in this dataset
# Is there a correlation between certain epidermis issues
# what other factors are at play
```

In [39]:
```
from IPython import display
display.Image('/home/harlohutch77/python/python_master/rash.jpeg',height=400, width=800)
```

Out[39]: