Disease Checker Bot – Using Evolutionary Algorithms

Term Project of Genetic Algorithms

Parv Maheshwari

19MA20033

Table of Contents

Abstract

In this project I have tried to develop a Disease checker bot which is currently being trained on data extracted from SymCat. The bot is being trained through a neural net which is a combination of traditional neural nets and some important steps of evolutionary algorithms like – Selection, Crossover and Mutation. I have presented a detailed approach of all the solutions which I tried and their results have been thoroughly compared.

Problem Statement

The problem statement consisted of developing a primary healthcare disease checker bot that will try to predict a possible disease based on symptoms shown by the patient. This also includes generating a customizable dataset which can be changed according to the need of the healthcare professionals. The motivation behind this would be that this can be further extended to an Indian demographic specific bot which would be among the first ones in the country.

Data Set Extraction and Generation

In table 1, all the datasets along with their advantages and disadvantages have been listed.

S. No.	Data Set Name	Advantages	Disadvantages
1	Disease Symptom – Kaggle	1. 24 common diseases with possible combinations of symptoms from among a set of 40 symptoms. 2. Severity associated with symptoms.	1. Very less number of diseases and symptoms included. 2. No description provided for the diseases and symptoms.
2	Health Analytics – Kaggle	1. Has state & district wise distribution data of common health diseases. 2. Demographic distribution present for the same.	1. No symptoms described for the diseases. 2. Description of the diseases very generic.
3	SymCat	1. Has 801 diseases with correlated 474 symptoms. 2. Probability for each symptom in a disease given along with vice-versa. 3. Demographic data (age group, gender, race) available for the diseases & symptoms.	1. No formatted data present. All data has to be scraped from the website. 2. Getting the data ready for a model will also have to be done. 3. Diseases too extensive.

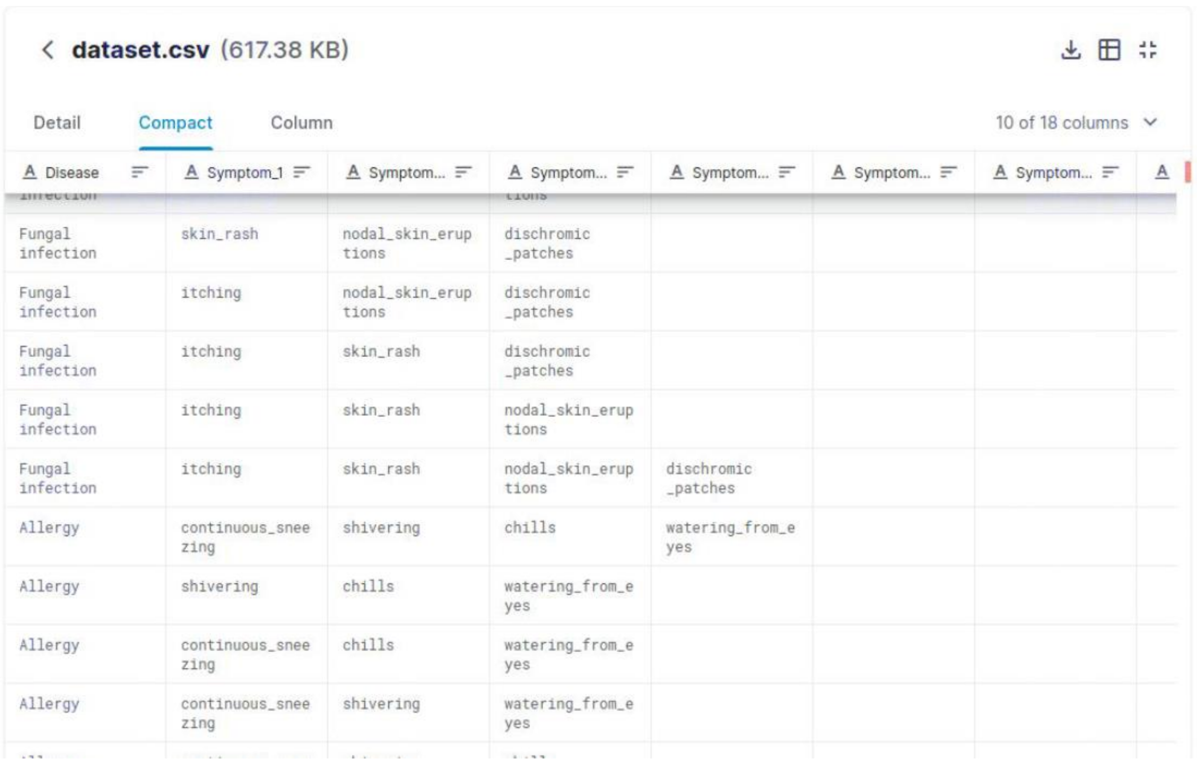
Table 1 – Various datasets explored during the project along with their advantages and disadvantages.

Seeing the need of a much more extensive database than the simple – "Disease Symptom – by Kaggle", further datasets were explored and finally Symcat was chosen as one of the most extensive sources of information which can be converted to a training dataset with feasible labor. It originally has 801 diseases with correlated 474 symptoms along with probability for each symptom in a disease given along with vice-versa along with Demographic data (age group, gender, race) for the diseases & symptoms.

This is very useful when tuning the disease bot checker according to a particular location and expected demographic of the users is already known. Due to time constraints, this tuning based on demographic had not been done.

One of the major disadvantages of Symcat as already mentioned in Table 1 and also visible from Fig. 3 and Fig. 4 are that no formatted data is present. Therefore, to solve this the following methodology was used –

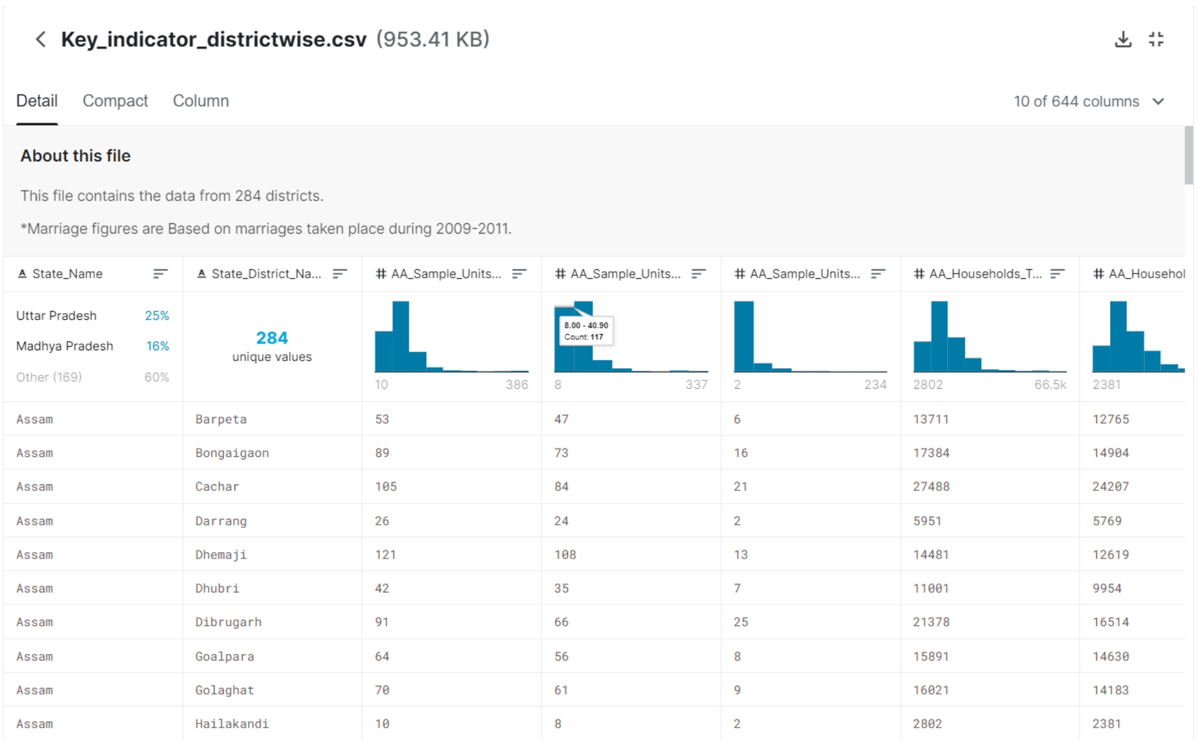
- Data was scraped using a python script from the Symcat website and converted to json file. (Fig. 5)
- Diseases that come under primary healthcare, were shortlisted.
- Data was converted from the JSON file to the standard traning.csv (Fig. 6) in the following steps
    - First all possible symptoms of the shortlisted diseases were extracted along with their probabilities. Here the probabilities represent the probability of showing a symptom give you have a particular disease. (Code 1)
    - Now this extracted data was used to generate a larger data set using data augmentation in which probabilities of the symptoms were used and then a dataset was generated which consisted of symptoms as inputs and the expected disease. (Code 2)


dataset.csv (617.38 KB)

Detail | Compact | Column — 10 of 18 columns

Disease	Symptom_1	Symptom...	Symptom...	Symptom...	Symptom...	Symptom...
Fungal infection	skin_rash	nodal_skin_eruptions	dischromic_patches			
Fungal infection	itching	nodal_skin_eruptions	dischromic_patches			
Fungal infection	itching	skin_rash	dischromic_patches			
Fungal infection	itching	skin_rash	nodal_skin_eruptions			
Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic_patches		
Allergy	continuous_sneezing	shivering	chills	watering_from_eyes		
Allergy	shivering	chills	watering_from_eyes			
Allergy	continuous_sneezing	chills	watering_from_eyes			
Allergy	continuous_sneezing	shivering	watering_from_eyes			

Fig 1 – Sample of data available in the Disease Symptom dataset available on Kaggle.


Key_indicator_districtwise.csv (953.41 KB)

Detail | Compact | Column — 10 of 644 columns

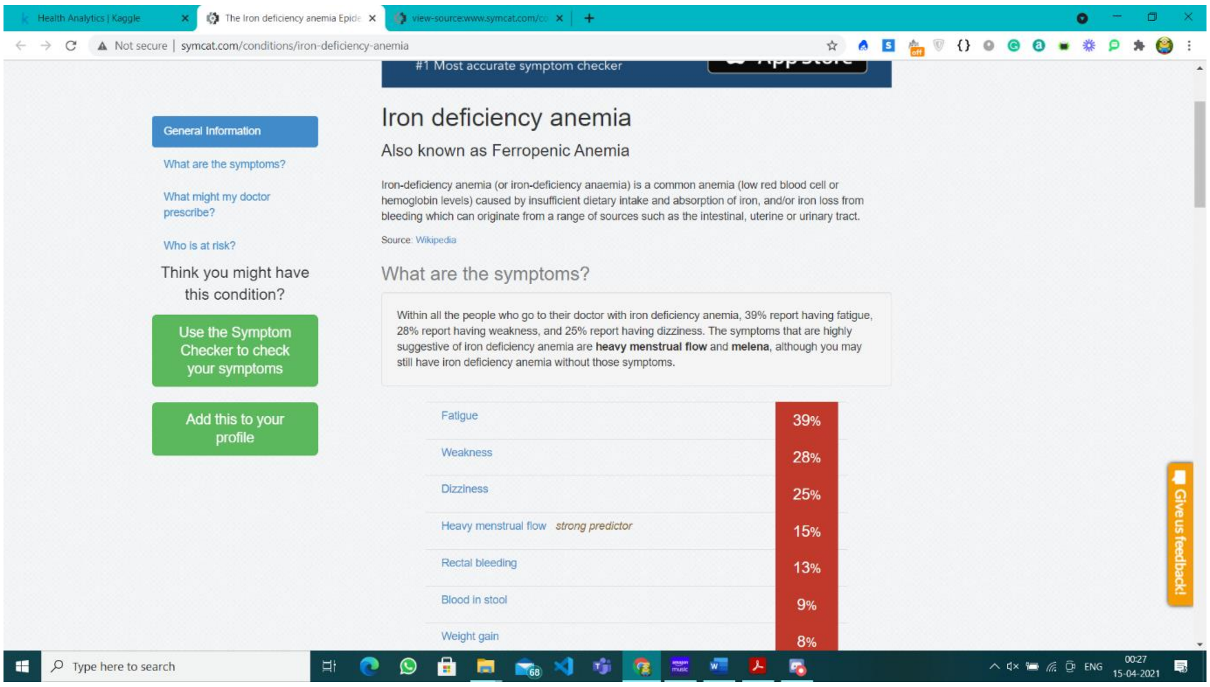
About this file

This file contains the data from 284 districts.

*Marriage figures are Based on marriages taken place during 2009-2011.

State_Name	State_District_Na...	AA_Sample_Units...	AA_Sample_Units...	AA_Sample_Units...	AA_Households_T...	AA_Househol
Uttar Pradesh 25%; Madhya Pradesh 16%; Other (169) 60%	284 unique values	10 – 386	8 – 337	2 – 234	2802 – 66.5k	2381
Assam	Barpeta	53	47	6	13711	12765
Assam	Bongaigaon	89	73	16	17384	14904
Assam	Cachar	105	84	21	27488	24207
Assam	Darrang	26	24	2	5951	5769
Assam	Dhemaji	121	108	13	14481	12619
Assam	Dhubri	42	35	7	11001	9954
Assam	Dibrugarh	91	66	25	21378	16514
Assam	Goalpara	64	56	8	15891	14630
Assam	Golaghat	70	61	9	16021	14183
Assam	Hailakandi	10	8	2	2802	2381

Fig 2 – Sample of data available in the Health Analytics dataset available on Kaggle.


General Information
What are the symptoms?
What might my doctor prescribe?
Who is at risk?
Think you might have this condition?
Use the Symptom Checker to check your symptoms
Add this to your profile
Give us feedback!

Iron deficiency anemia

Also known as Ferropenic Anemia

Iron-deficiency anemia (or iron-deficiency anaemia) is a common anemia (low red blood cell or hemoglobin levels) caused by insufficient dietary intake and absorption of iron, and/or iron loss from bleeding which can originate from a range of sources such as the intestinal, uterine or urinary tract.

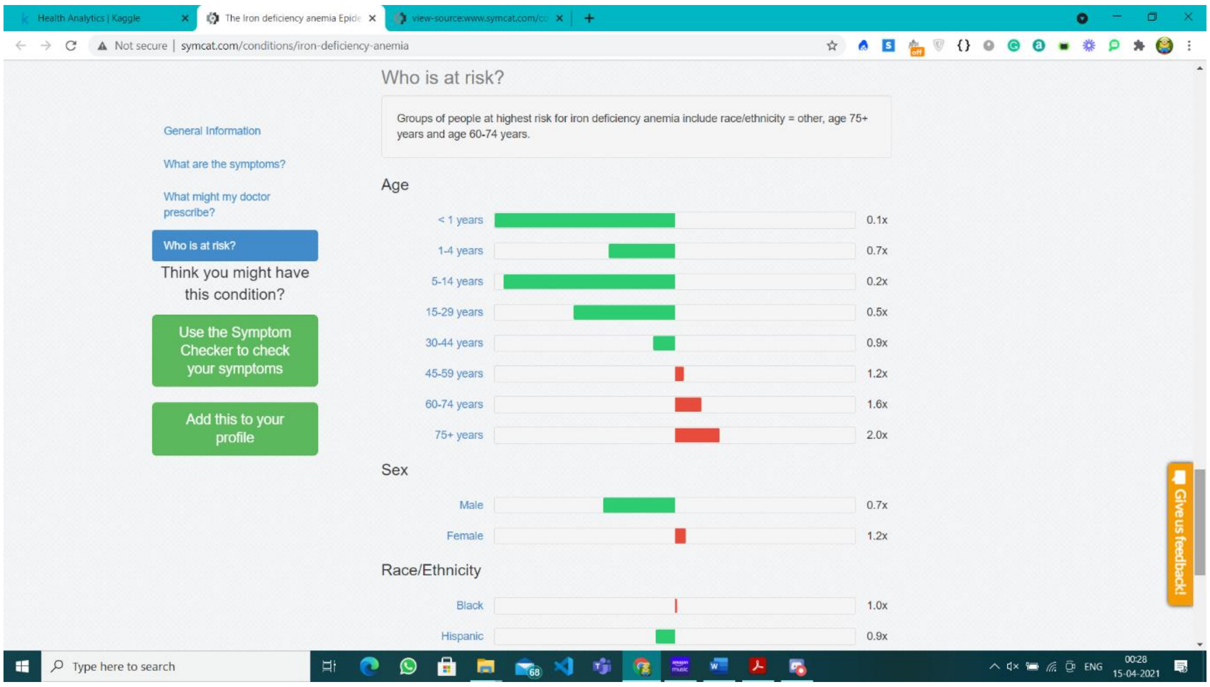
Source: Wikipedia

What are the symptoms?

Within all the people who go to their doctor with iron deficiency anemia, 39% report having fatigue, 28% report having weakness, and 25% report having dizziness. The symptoms that are highly suggestive of iron deficiency anemia are **heavy menstrual flow** and **melena**, although you may still have iron deficiency anemia without those symptoms.

Fatigue	39%
Weakness	28%
Dizziness	25%
Heavy menstrual flow *strong predictor*	15%
Rectal bleeding	13%
Blood in stool	9%
Weight gain	8%

Fig 3 – Sample of data available on the Symcat website. Here we can see a disease (For example - Iron deficiency anemia in the above figure) and its possible symptoms along with their probabilities.


General Information
What are the symptoms?
What might my doctor prescribe?
Who is at risk?
Think you might have this condition?
Use the Symptom Checker to check your symptoms
Add this to your profile
Give us feedback!

Who is at risk?

Groups of people at highest risk for iron deficiency anemia include race/ethnicity = other, age 75+ years and age 60-74 years.

Age

< 1 years	0.1x
1-4 years	0.7x
5-14 years	0.2x
15-29 years	0.5x
30-44 years	0.9x
45-59 years	1.2x
60-74 years	1.6x
75+ years	2.0x

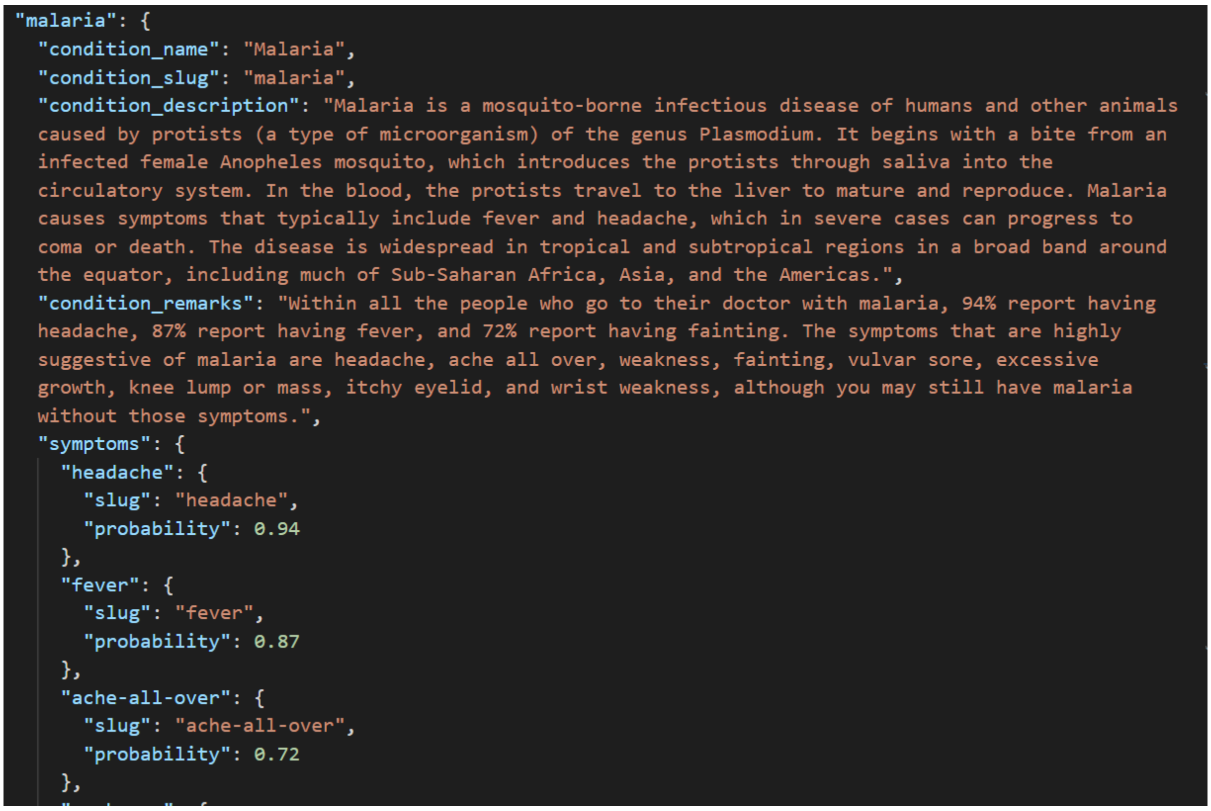
Sex

Male	0.7x
Female	1.2x

Race/Ethnicity

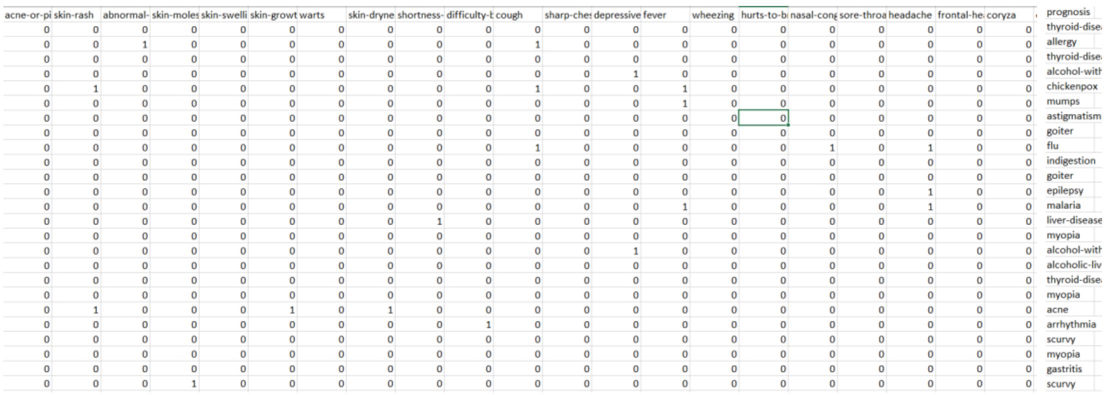
Black	1.0x
Hispanic	0.9x

Fig 4 – Sample of data available on the Symcat website. Here we can see a disease (For example - Iron deficiency anemia in the above figure) and its expected demographic distribution.



```json
"malaria": {
  "condition_name": "Malaria",
  "condition_slug": "malaria",
  "condition_description": "Malaria is a mosquito-borne infectious disease of humans and other animals
  caused by protists (a type of microorganism) of the genus Plasmodium. It begins with a bite from an
  infected female Anopheles mosquito, which introduces the protists through saliva into the
  circulatory system. In the blood, the protists travel to the liver to mature and reproduce. Malaria
  causes symptoms that typically include fever and headache, which in severe cases can progress to
  coma or death. The disease is widespread in tropical and subtropical regions in a broad band around
  the equator, including much of Sub-Saharan Africa, Asia, and the Americas.",
  "condition_remarks": "Within all the people who go to their doctor with malaria, 94% report having
  headache, 87% report having fever, and 72% report having fainting. The symptoms that are highly
  suggestive of malaria are headache, ache all over, weakness, fainting, vulvar sore, excessive
  growth, knee lump or mass, itchy eyelid, and wrist weakness, although you may still have malaria
  without those symptoms.",
  "symptoms": {
    "headache": {
      "slug": "headache",
      "probability": 0.94
    },
    "fever": {
      "slug": "fever",
      "probability": 0.87
    },
    "ache-all-over": {
      "slug": "ache-all-over",
      "probability": 0.72
    },
```

Fig 5 – Data scraped from the Symcat website in json format. Here we can see a disease (For example - malaria in the above figure), its description and probable symptoms along with their probabilities.



acne-or-pi	skin-rash	abnormal-	skin-moles	skin-swelli	skin-growt	warts	skin-dryne	shortness-	difficulty-t	cough	sharp-ches	depressive	fever	wheezing	hurts-to-b	nasal-cong	sore-throa	headache	frontal-he	coryza	prognosis
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	thyroid-dise
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	allergy
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	thyroid-dise
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	alcohol-with
0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	chickenpox
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	mumps
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	astigmatism
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	goiter
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	flu
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	indigestion
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	goiter
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	epilepsy
0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	malaria
0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	liver-disease
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	myopia
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	alcohol-with
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	alcoholic-liv
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	thyroid-dise
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	myopia
0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	acne
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	arrhythmia
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	scurvy
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	myopia
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	gastritis
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	scurvy

Fig 6 – Sample of Training.csv which has been generated from the data augmentation step after cleaning of the dataset from the data scraped from the Symcat website.

Approach

Several approaches to train a artificial neural net using evolutionary algorithms –

- Using PyGAD - which is an open-source Python 3 library for building the genetic algorithm and optimizing machine learning algorithms.
- Using existing frameworks to build a model along with handwritten selection, crossover and mutation function.Following frameworks were explored –
    - Keras
    - Pytorch

PyGad

PyGAD is an open-source Python 3 library which is used for building the genetic algorithm and optimizing machine learning algorithms. PyGAD supports different types of crossover, mutation, and parent selection. PyGAD allows different types of problems to be optimized using the genetic algorithm by customizing the fitness function.

The pygad.gann module was used which builds and trains neural networks (for either classification or regression) using the genetic algorithm.

The steps to use PyGad for building and training a neural network using the genetic algorithm are as follows:

- Prepare the training data. (Code 3)
- Create an instance of the pygad.gann.GANN class. (Code 4)
    - this is done with the __init__() function.
    - The architecture of the created network has the following layers:
        - An input layer with 144 neurons (i.e. inputs (symptoms))
        - A single hidden layer with 200 neurons.
        - An output layer with 48 neurons (i.e. classes (diseases)).
- Fetch the population weights as vectors. (Code 5)

    - For the genetic algorithm, the parameters (i.e. genes) of each solution are represented as a single vector.

- Prepare the fitness function. (Code 6)
    - The fitness function for training a neural network uses the pygad.nn.predict() function to predict the class labels based on the current solution’s weights.
    - Based on such predictions, the classification accuracy is calculated. This accuracy is used as the fitness value of the solution.

- Prepare the generation callback function. (Code 7)
    - After each generation of the genetic algorithm, the fitness function will be called to calculate the fitness value of each solution.
    - This callback function can be used to update the trained_weights attribute of layers of each network in the population.
    - Code 7 shows the implementation for a function that updates the trained_weights attribute of the layers of the population networks. It works by converting the current population from the vector form to the matric form using the pygad.gann.population_as_matrices() function. It accepts the population as vectors and returns it as matrices.

      The population matrices are then passed to the update_population_trained_weights() method in the pygad.gann module to update the trained_weights attribute of all layers for all solutions within the population.

- Create an instance of the pygad.GA class. (Code 8)
- Run the created instance of the pygad.GA class.

This completes our training step of the popualtion =. Now for the results we -

- Plot the Fitness Values
- Retrieve Information about the best solution.
- Make predictions on test set using the trained weights.

- Calculate some statistics.

```python
df = pd.read_csv(file_dir + 'Training.csv')


# Random in place shuffling of the original input data before splitting it into train and test set

df = df.sample(frac=1).reset_index(drop=True)

rows = len(df.index)

test_ratio =0.1
train_ratio = 1-test_ratio
df_train = df.iloc[:int(train_ratio*rows),:]
df_test = df.iloc[int(train_ratio*rows+1):,:]

from sklearn import preprocessing
le = preprocessing.LabelEncoder()
arn = pd.concat([df_train['prognosis'], df_test['prognosis']])
le.fit(arn)

x_train = df_train[df_train.columns.difference(['prognosis'])]
y_train = le.fit_transform(df_train['prognosis'])

x_test = df_test[df_test.columns.difference(['prognosis'])]
y_test = le.fit_transform(df_test['prognosis'])
```

Code 3 – Prepare the training data by using pandas and preprocessing module from sklearn.

```python
# Reading the input data.
data_inputs = x_train.to_numpy()

# Reading the output data.
data_outputs = y_train

# The length of the input vector for each sample (i.e. number of neurons in the input layer).
num_inputs = data_inputs.shape[1]
# The number of neurons in the output layer (i.e. number of classes).
num_classes = 48

# Creating an initial population of neural networks. The return of the initial_population()
function holds references to the networks, not their weights. Using such references, the weights
of all networks can be fetched.
num_solutions = 20  # A solution or a network can be used interchangeably.
GANN_instance = pygad.gann.GANN(num_solutions=num_solutions,
                                num_neurons_input=num_inputs,
                                num_neurons_hidden_layers=[200],
                                num_neurons_output=num_classes,
                                hidden_activations=["relu"],
                                output_activation="softmax")
```

Code 4 – Create an instance of the pygad.gann.GANN class with number of neurons in model architecture as [144 , 200, 48].

```python
population_vectors =
pygad.gann.population_as_vectors(population_networks=GANN_instance.po
pulation_networks)
```

Code 5 - Fetch the population weights as vectors.

```
def fitness_func(solution, sol_idx):
    global GANN_instance, data_inputs, data_outputs

    predictions = 
pygad.nn.predict(last_layer=GANN_instance.population_networks[sol
_idx],
                                   data_inputs=data_inputs)
    correct_predictions = numpy.where(predictions == 
data_outputs)[0].size
    solution_fitness = 
(correct_predictions/data_outputs.size)*100

    return solution_fitness
```
Code 6 - Fitness function.

```
def callback_generation(ga_instance):
    global GANN_instance

    population_matrices = 
pygad.gann.population_as_matrices(population_networks=GANN_instan
ce.population_networks, 
population_vectors=ga_instance.population)

GANN_instance.update_population_trained_weights(population_traine
d_weights=population_matrices)

    print("Generation = 
{generation}".format(generation=ga_instance.generations_completed
))
    print("Fitness    = 
{fitness}".format(fitness=ga_instance.best_solution()[1]))
```
Code 7 – Callback function which is called at the end of every generation. Major component of this function includes the weight update step which is similar to back propagation in simple artificial neural nets.

```
initial_population = population_vectors.copy()

num_parents_mating = 4

num_generations = 500

mutation_percent_genes = 5

parent_selection_type = "sss"

crossover_type = "single_point"

mutation_type = "random"

keep_parents = 1


ga_instance = pygad.GA(num_generations=num_generations,
                       num_parents_mating=num_parents_mating,
                       initial_population=initial_population,
                       fitness_func=fitness_func,
mutation_percent_genes = mutation_percent_genes,
parent_selection_type = parent_selection_type,
                        crossover_type=crossover_type,
                        mutation_type=mutation_type,
                        keep_parents=keep_parents,
                        on_generation=callback_generation)
```

Code 8 – Initializing a random initial population and then initializing an instance of pygad.ga class. This instance acts as a bundle for all the population along with various fitness and callback functions.

```
ga_instance.run()
```

Code 9 – starts to train the population

```
ga_instance.plot_result()
```

Code 10 – plots the best_fitness_in_genration vs number of generation graph

```
solution, solution_fitness, solution_idx =
ga_instance.best_solution()
print("Parameters of the best solution :
{solution}".format(solution=solution))
print("Fitness value of the best solution =
{solution_fitness}".format(solution_fitness=solution_fitness))
print("Index of the best solution :
{solution_idx}".format(solution_idx=solution_idx))
```

Code 11 – Information about the best solution is retrieved from the ga_instance.

```
predictions =
pygad.nn.predict(last_layer=GANN_instance.population_networks[sol
ution_idx], data_inputs=data_inputs)
print("Predictions of the trained network :
{predictions}".format(predictions=predictions))
```

Code 12 – Prediction step on the training or test data according to the need of the user.

Simple ANN model + Steps of Evolutianory Algorithms written from Scratch

On seeing the low performance of PyGad, I decied to manually build a model and then evolve it using both the standard mathematical gradient sdescent along with evolutionary algorithms supporting convergence of the population. The following steps were applied in sequential order –

- Step 1 – Preparing the data. This invlolved converting numpy arrays to tensors for faster processing on GPU
- Step 2 - An artifical neural net (ANN) was build. In case of PyTorch , a separate function for traininf of the ANN was also required.
- Step 3 - Hyperparameters were initialised
- Step 4 - Selection, crossover and mutation functions were defined

- Step 5 - Random initial population was initialised
- Step 6 - For each generation, the following steps are being performed
    - Step 6.1 - Each member of the population undergoes usual neural network training. But the difference from normal neural net training is that for a single generation, the members are only trained for a very small number of epochs (this is controlled by the hyperparameter - epochs_per_nn)
    - Step 6.2 - Fitness is calculated on the basis of classification accuracy
    - Step 6.3 - Some parents form previous generation our added to the pool to maintain sustainance of parents.
    - Step 6.4 - Population is sorted in decreasing order of their fitness values and then the population is resized to the original population size.
    - Step 6.5 - The best member among the population at the current genration is considered and commpared with the so far best member which has been produced.
    - Step 6.6 - Selection function is called which returns a list of parents among the current populaiton which will undergo crossover to produce children.
    - Step 6.7 - Crossover is applied on parents
    - Step 6.8 - The resultant children of crossover undergo mutation.
    - Step 6.9 - After this the children networks are added to the population and then next generation starts.
- Step 7 - After all the genrations are ovevr, we get the best model form the variable - `best_model.` This is used as the final model in predictions and calculating the accuracies.

The followinf types of selection, crossover and mutations were used

- Selection –

    - Fittest_selection – members in decreasing order of their fitness are selected as parents, i.e. fitter the member more is its possibilty to be a parent.

- Crossover –

    - Uniform Crossover – In this crossover two parents result in a single child. Here each parent has an equal probability (p_cross = 0.5) to contribute to the weights of each layer of the resulting child.

- Mutation –

    - Bitwise Mutation – Each layer individually can go under mutation depending on p_mut.

Now each of these steps were applied for two different neural network frameworks – Keras and PyTorch. While Keras was preferreed for its ease of access of layer weights which are the main subject for crossover and mutation, along with a more readble code, PyTorch helped in using GPU in a much more efficient way by transfereing all the data stored in tensors along with the whole popluatin of models being transfred to and trained on the GPU devidce side using CUDA. This resulted in an exponential decrease in amout of training time. Hence this allowed me to do perform much more experimenttionwith the hyperparametters, and types of selection , crossover and mutation functions.

Now the two implementations are provided –

Keras with Steps of Evolutionary Algorithms

Step 1 –

```
symp_train_data = tf.convert_to_tensor(x_train.values,
dtype=tf.int64)
train_outputs =  tf.convert_to_tensor(y_train)

symp_test_data =  tf.convert_to_tensor(x_test.values,
dtype=tf.int64)
test_outputs =  tf.convert_to_tensor(y_test)
```

Step 2 –

```
class ANN(Sequential):

    def __init__(self, child_weights=None):
        super().__init__()
        self.fitness = 0
        if child_weights is None:
            layer1 = Dense(200, input_shape=(144,),
activation='relu')
            layer2 = Dense(48, activation=None)
            self.add(layer1)
            self.add(layer2)

    def forward_propagation(self, train_feature, train_label):
        predict_label = self.predict(train_feature)
```

Step 3 –

```
p_cross = 0.5
p_mut = 0.3
num_parents = 10
population = 30
epochs = 50
epochs_per_nn = 1
```

Step 4 - selection not defined

```
def selection (pool):
    global num_parents

    #coosing te fittest num_parent members of te population
    parents = copy.deepcopy(pool[:num_parents])
    return parents

def crossover(nn1, nn2):
    global p_cross , symp_train_data  , train_outputs
    nn1_weights = []
    nn2_weights = []
    child_weights = []

    for layer in nn1.layers:
        nn1_weights.append(layer.get_weights()[0])
    # print(len(nn1_weights[1]))
    for layer in nn2.layers:
        nn2_weights.append(layer.get_weights()[0])

    for i in range(len(nn1_weights)):
        for j in range(len(np.shape(nn1_weights[i]))):
            cross = random.random()          #uniform crossover
            if cross<p_cross:
                nn1_weights[i][j] = nn2_weights[i][j];

        child_weights.append(nn1_weights[i])

    mutation(child_weights)
    final_child_weights=nn1.get_weights()

    final_child_weights[0] = child_weights[0]
    final_child_weights[2] = child_weights[1]

    child = ANN()
    child.set_weights(final_child_weights)
    predict_label =  child.predict(symp_train_data)
    preds_classes = np.argmax(predict_label, axis=-1)
    child.fitness = accuracy_score(train_outputs, preds_classes)
    print("Fitness of the child is ",str(child.fitness))
    return child

def mutation(child_weights):
    global p_mut
    for i in range(len(child_weights)):
    # selection = random.randint(0, len(child_weights)-1)
        mut = random.random()
        if mut <= p_mut:
            child_weights[i] *= random.uniform(0.8,1.2)
        else:
            pass
```

Step 5 -

```

# store all active ANNs
networks = []
pool = []
# Generation counter
generation = 0

# Initial Population
for i in range(population):
    pool.append(ANN())
# Track Max Fitness
max_fitness = 0
next_gen_confirmed = []
fitness_per_generation = []


best_model = ANN()
```

Step 6 -

```
# Evolution Loop
for i in range(epochs):
    generation += 1
    logging.debug("Generation: " + str(generation) + "\r\n")
    networks = []
    for ann in pool:
        # Propagate to calculate fitness score
         # Step 6.1
        ann.compile(optimizer=optimizers.Adam(),
loss=keras.losses.SparseCategoricalCrossentropy(from_logits=True),
metrics=[keras.metrics.SparseCategoricalAccuracy()])
        ann.fit(x=symp_train_data, y = train_outputs , epochs =
epochs_per_nn, verbose=0)
       # Step 6.2
        predict_label =  ann.predict(symp_train_data,)
        preds_classes = np.argmax(predict_label, axis=-1)
        ann.fitness = accuracy_score(train_outputs,
preds_classes)

        # Add to pool after calculating fitness
        networks.append(ann)
    pool = networks
    # Sort the population by fitness
    # Step 6.3
    pool+= next_gen_confirmed

    # Step 6.4
    pool = sorted(pool, key=lambda x: x.fitness)
    pool.reverse()
    pool = pool[:population]
    next_gen_confirmed = pool[:3]
    fitness_per_generation.append(pool[0].fitness)
    logging.debug("Max Fitness of generation : " +
str(pool[0].fitness) + "\r\n")

    # Step 6.5
    if pool[0].fitness > max_fitness:
        max_fitness = pool[0].fitness
        logging.debug("Max Fitness: " + str(max_fitness) +
"\r\n")
        best_model = pool[0]

    # Step 6.6
    parents = selection(pool)

    # Step 6.7 and step 6.8 (Mutation function is called inside the crossover function)
    for i in range(len(parents)):
        target = (i+1) % num_parents
        child = crossover(parents[i], parents[target])

        # Step 6.9
        pool.append(child)
```

PyTorch with Steps of Evolutionary Algorithms

Step 1 –

```
device = "cuda" if torch.cuda.is_available() else "cpu"

symp_train_data = torch.tensor(x_train.values,
dtype=torch.int64).to(device)
train_outputs = torch.tensor(y_train).to(device)

symp_test_data = torch.tensor(x_test.values,
dtype=torch.int64).to(device)
test_outputs = torch.tensor(y_test).to(device)

categorical_column_sizes = [2 for column in
df.columns.difference(['prognosis'])]
categorical_embedding_sizes = [(2, min(50, (2+1)//2)) for column in
df.columns.difference(['prognosis'])]
```

Step 2 –

```
class Model(nn.Module):

    def _init_(self, embedding_size, output_size = 48, layers =
[200], p=0.1):
        self.fitness = 0
        super().__init__()
        self.all_embeddings = nn.ModuleList([nn.Embedding(ni, nf)
for ni, nf in embedding_size])
        self.embedding_dropout = nn.Dropout(p)

        all_layers = []
        num_categorical_cols = sum((nf for ni, nf in
embedding_size))
        input_size = num_categorical_cols

        for i in layers:
            all_layers.append(nn.Linear(input_size, i))
            all_layers.append(nn.ReLU(inplace=True))
            all_layers.append(nn.BatchNorm1d(i))
            all_layers.append(nn.Dropout(p))
            input_size = i

        all_layers.append(nn.Linear(layers[-1], output_size))

        self.layers = nn.Sequential(*all_layers)

    def forward(self, x_categorical):
        embeddings = []
        for i,e in enumerate(self.all_embeddings):
            embeddings.append(e(x_categorical[:,i]))
        x = torch.cat(embeddings, 1)

        x = self.embedding_dropout(x)
        x = self.layers(x)

        return x

def train(model,epochs):
    optimizer = torch.optim.Adam(model.parameters(), lr=0.001)
    loss_function = nn.CrossEntropyLoss()

    for i in range(epochs):
        y_pred = model(symp_train_data)

        single_loss = loss_function(y_pred,train_outputs)

        optimizer.zero_grad()
        single_loss.backward()
        optimizer.step()
```

Step 3 –

```
p_cross = 0.5
p_mut = 0.3
```

```
num_parents = 10
population = 30
epochs = 50
epochs_per_nn = 1
```

Step 4 –

```
def selection (pool):
    global num_parents

    #coosing te fittest num_parent members of te population
    parents = copy.deepcopy(pool[:num_parents])
    return parents

def crossover(nn1, nn2):
    global p_cross
    child = Model(categorical_embedding_sizes).to(device)

    for i in range(len(nn1.all_embeddings)):
        if hasattr(nn1.all_embeddings[i] , 'weight'):
            for j in range(len(nn1.all_embeddings[i].weight)):
                cross = random.random()          #uniform crossover
                if cross<p_cross:
                    child.all_embeddings[i].weight.data[j] =
nn2.all_embeddings[i].weight.data[j];
                else:
                    child.all_embeddings[i].weight.data[j] =
nn1.all_embeddings[i].weight.data[j];

    for i in range(len(nn1.layers)):
        if hasattr(nn1.layers[i] , 'weight'):
            for j in range(len(nn1.layers[i].weight)):
                cross = random.random()          #uniform crossover
                if cross<p_cross:
                    child.layers[i].weight.data[j] =
nn2.layers[i].weight.data[j];
                else:
                    child.layers[i].weight.data[j] =
nn1.layers[i].weight.data[j];

    child = mutation(child)

    return child

def mutation(child):
    global p_mut
    for i in range(len(child.all_embeddings)):
        if hasattr(child.all_embeddings[i] , 'weight'):
            for j in range(len(child.all_embeddings[i].weight)):
                mut = random.random()
                if mut <= p_mut:
                    child.all_embeddings[i].weight.data[j] *=
random.uniform(0.8,1.2)
                else:
                    pass
    for i in range(len(child.layers)):
        if hasattr(child.layers[i] , 'weight'):
            for j in range(len(child.layers[i].weight)):
                mut = random.random()
                if mut <= p_mut:
                    child.layers[i].weight.data[j] *=
random.uniform(0.7,1.3)
                else:
                    pass

    return child
```

Step 5 –

```
# store all active ANNs
networks = []
pool = []
# Generation counter
generation = 0

# Initial Population
for i in range(population):
    pool.append(Model(categorical_embedding_sizes).to(device))
# Track Max Fitness
max_fitness = 0
next_gen_confirmed = []
fitness_per_generation = []

best_model = Model(categorical_embedding_sizes).to(device)
```

Step 6 –

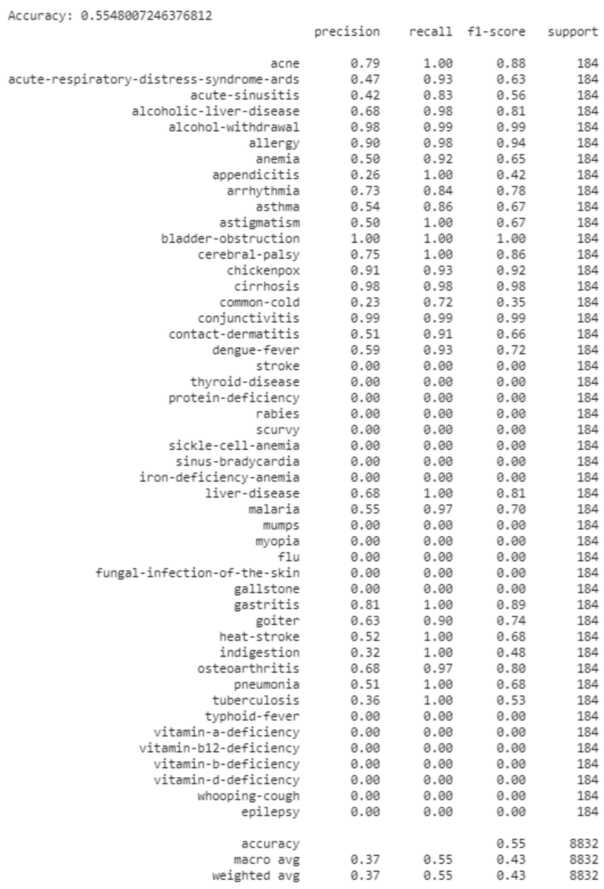
```
for i in range(epochs):
    generation += 1
    logging.debug("Generation: " + str(generation) + "\r\n")
    networks = []
    for ann in pool:
        # Propagate to calculate fitness score
        #Step 6.1
        train(ann,epochs_per_nn

        #Step 6.2
        predict_label =
ann(symp_train_data).detach().cpu().numpy()
        predict_label = np.argmax(predict_label, axis=-1)
        ann.fitness = accuracy_score(train_outputs.cpu(),
predict_label)
        # Add to pool after calculating fitness
        networks.append(ann)

    pool = networks

    # Step 6.3
    pool+= next_gen_confirmed
    # Step 6.4
    pool = sorted(pool, key=lambda x: x.fitness)
    pool.reverse()
    pool = pool[:population]
    next_gen_confirmed = copy.deepcopy(pool[:3])
    fitness_per_generation.append(pool[0].fitness)
    logging.debug("Max Fitness of generation : " +
str(pool[0].fitness) + "\r\n")

    # Step 6.5
    if pool[0].fitness > max_fitness:
        max_fitness = pool[0].fitness
        logging.debug("Max Fitness: " + str(max_fitness) +
"\r\n")
        best_model = pool[0]
    # Step 6.6
    parents = selection(pool)

    # Step 6.7 and step 6.8 (Mutation function is called inside the crossover function)
    for i in range(len(parents)):
        target = (i+1) % num_parents
        child = crossover(parents[i], parents[target])

        # Step 6.9
        pool.append(child)
```

Observations

CLASSIFIER ACCURACY

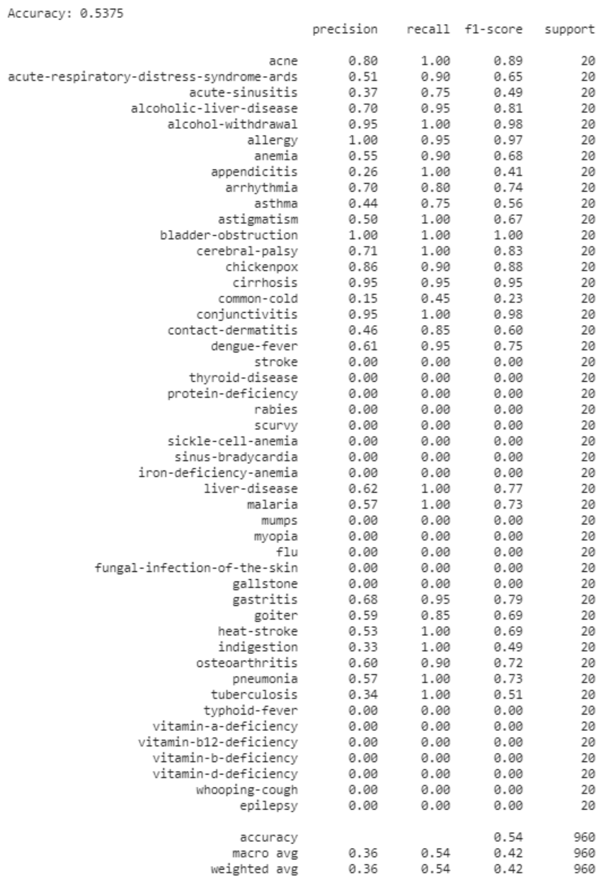
- Naïve bayes
  - Train Accuracy – 55.48%
  - Test Accuracy – 53.75%

On Training Data



```
Accuracy: 0.5548007246376812
                                          precision    recall  f1-score   support

                                    acne       0.79      1.00      0.88       184
acute-respiratory-distress-syndrome-ards       0.47      0.93      0.63       184
                         acute-sinusitis       0.42      0.83      0.56       184
                 alcoholic-liver-disease       0.68      0.98      0.81       184
                      alcohol-withdrawal       0.98      0.99      0.99       184
                                 allergy       0.90      0.98      0.94       184
                                  anemia       0.50      0.92      0.65       184
                            appendicitis       0.26      1.00      0.42       184
                              arrhythmia       0.73      0.84      0.78       184
                                  asthma       0.54      0.86      0.67       184
                             astigmatism       0.50      1.00      0.67       184
                     bladder-obstruction       1.00      1.00      1.00       184
                          cerebral-palsy       0.75      1.00      0.86       184
                              chickenpox       0.91      0.93      0.92       184
                               cirrhosis       0.98      0.98      0.98       184
                             common-cold       0.23      0.72      0.35       184
                          conjunctivitis       0.99      0.99      0.99       184
                      contact-dermatitis       0.51      0.91      0.66       184
                            dengue-fever       0.59      0.93      0.72       184
                                  stroke       0.00      0.00      0.00       184
                         thyroid-disease       0.00      0.00      0.00       184
                      protein-deficiency       0.00      0.00      0.00       184
                                  rabies       0.00      0.00      0.00       184
                                  scurvy       0.00      0.00      0.00       184
                      sickle-cell-anemia       0.00      0.00      0.00       184
                       sinus-bradycardia       0.00      0.00      0.00       184
                  iron-deficiency-anemia       0.00      0.00      0.00       184
                           liver-disease       0.68      1.00      0.81       184
                                 malaria       0.55      0.97      0.70       184
                                   mumps       0.00      0.00      0.00       184
                                  myopia       0.00      0.00      0.00       184
                                     flu       0.00      0.00      0.00       184
            fungal-infection-of-the-skin       0.00      0.00      0.00       184
                               gallstone       0.00      0.00      0.00       184
                               gastritis       0.81      1.00      0.89       184
                                  goiter       0.63      0.90      0.74       184
                             heat-stroke       0.52      1.00      0.68       184
                             indigestion       0.32      1.00      0.48       184
                          osteoarthritis       0.68      0.97      0.80       184
                               pneumonia       0.51      1.00      0.68       184
                            tuberculosis       0.36      1.00      0.53       184
                           typhoid-fever       0.00      0.00      0.00       184
                    vitamin-a-deficiency       0.00      0.00      0.00       184
                  vitamin-b12-deficiency       0.00      0.00      0.00       184
                    vitamin-b-deficiency       0.00      0.00      0.00       184
                    vitamin-d-deficiency       0.00      0.00      0.00       184
                          whooping-cough       0.00      0.00      0.00       184
                                epilepsy       0.00      0.00      0.00       184

                                accuracy                           0.55      8832
                               macro avg       0.37      0.55      0.43      8832
                            weighted avg       0.37      0.55      0.43      8832
```

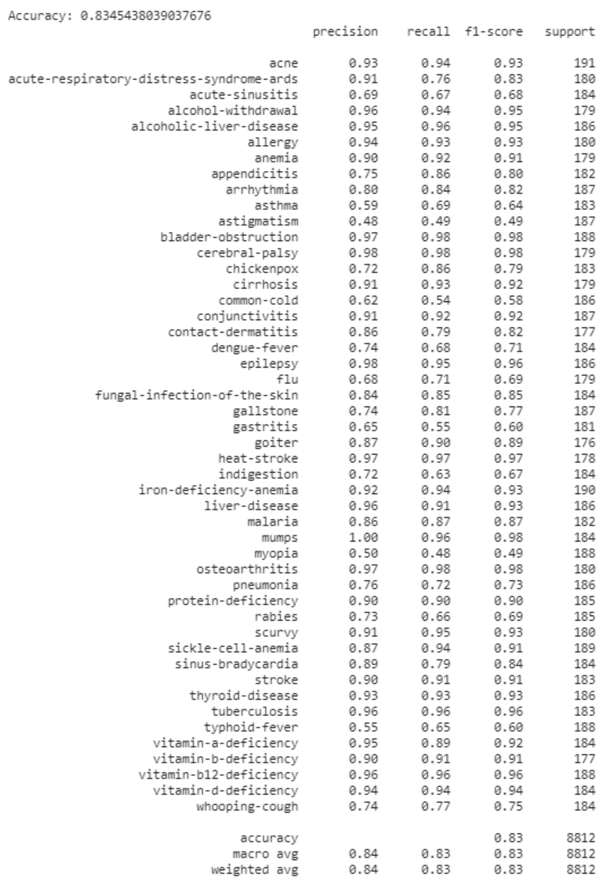
On Test Data



```
Accuracy: 0.5375
                                          precision    recall  f1-score   support

                                    acne       0.80      1.00      0.89        20
acute-respiratory-distress-syndrome-ards       0.51      0.90      0.65        20
                         acute-sinusitis       0.37      0.75      0.49        20
                 alcoholic-liver-disease       0.70      0.95      0.81        20
                      alcohol-withdrawal       0.95      1.00      0.98        20
                                 allergy       1.00      0.95      0.97        20
                                  anemia       0.55      0.90      0.68        20
                            appendicitis       0.26      1.00      0.41        20
                              arrhythmia       0.70      0.80      0.74        20
                                  asthma       0.44      0.75      0.56        20
                             astigmatism       0.50      1.00      0.67        20
                     bladder-obstruction       1.00      1.00      1.00        20
                          cerebral-palsy       0.71      1.00      0.83        20
                              chickenpox       0.86      0.90      0.88        20
                               cirrhosis       0.95      0.95      0.95        20
                             common-cold       0.15      0.45      0.23        20
                          conjunctivitis       0.95      1.00      0.98        20
                      contact-dermatitis       0.46      0.85      0.60        20
                            dengue-fever       0.61      0.95      0.75        20
                                  stroke       0.00      0.00      0.00        20
                         thyroid-disease       0.00      0.00      0.00        20
                      protein-deficiency       0.00      0.00      0.00        20
                                  rabies       0.00      0.00      0.00        20
                                  scurvy       0.00      0.00      0.00        20
                      sickle-cell-anemia       0.00      0.00      0.00        20
                       sinus-bradycardia       0.00      0.00      0.00        20
                  iron-deficiency-anemia       0.00      0.00      0.00        20
                           liver-disease       0.62      1.00      0.77        20
                                 malaria       0.57      1.00      0.73        20
                                   mumps       0.00      0.00      0.00        20
                                  myopia       0.00      0.00      0.00        20
                                     flu       0.00      0.00      0.00        20
            fungal-infection-of-the-skin       0.00      0.00      0.00        20
                               gallstone       0.00      0.00      0.00        20
                               gastritis       0.68      0.95      0.79        20
                                  goiter       0.59      0.85      0.69        20
                             heat-stroke       0.53      1.00      0.69        20
                             indigestion       0.33      1.00      0.49        20
                          osteoarthritis       0.60      0.90      0.72        20
                               pneumonia       0.57      1.00      0.73        20
                            tuberculosis       0.34      1.00      0.51        20
                           typhoid-fever       0.00      0.00      0.00        20
                    vitamin-a-deficiency       0.00      0.00      0.00        20
                  vitamin-b12-deficiency       0.00      0.00      0.00        20
                    vitamin-b-deficiency       0.00      0.00      0.00        20
                    vitamin-d-deficiency       0.00      0.00      0.00        20
                          whooping-cough       0.00      0.00      0.00        20
                                epilepsy       0.00      0.00      0.00        20

                                accuracy                           0.54       960
                               macro avg       0.36      0.54      0.42       960
                            weighted avg       0.36      0.54      0.42       960
```

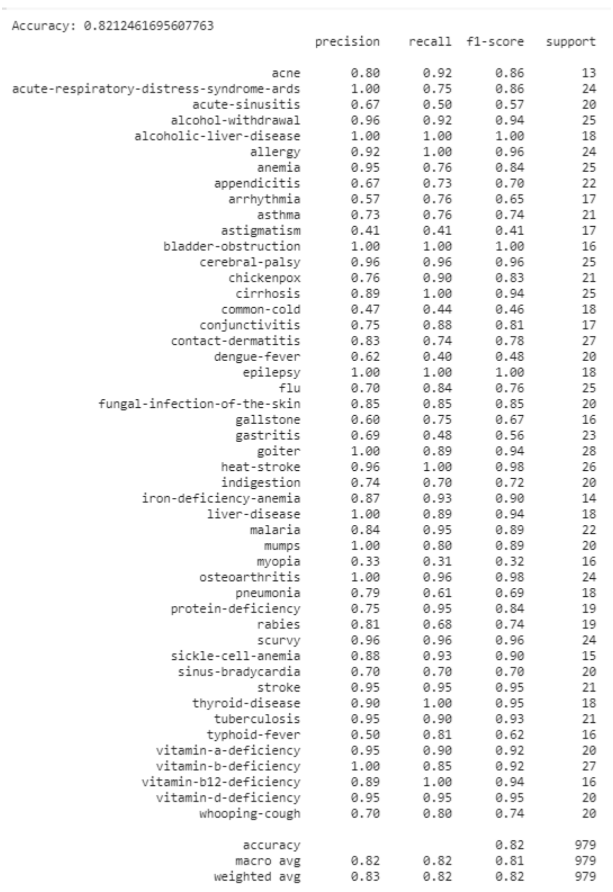
- Normal NN
  - Train Accuracy – 83.45%
  - Test Accuracy – 82.12%

On Training Data



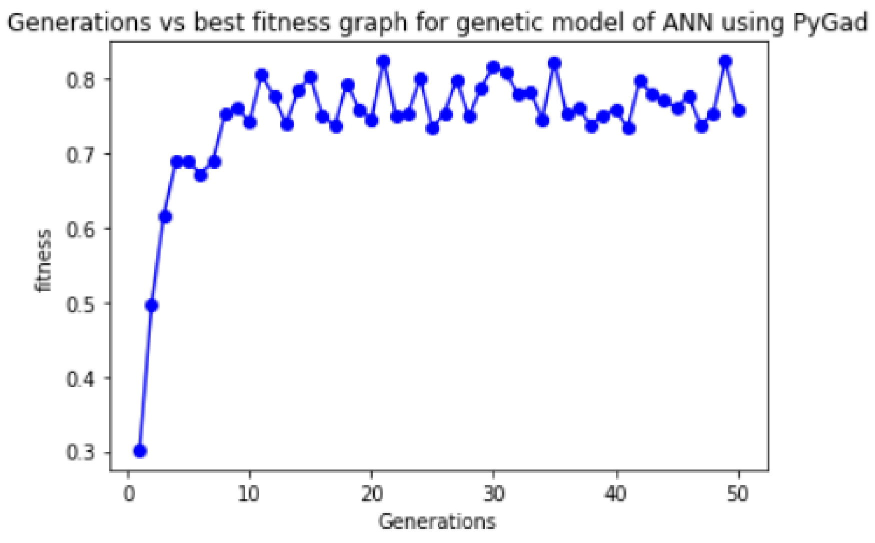
```
Accuracy: 0.8345438039037676
                                          precision    recall  f1-score   support

                                    acne       0.93      0.94      0.93       191
acute-respiratory-distress-syndrome-ards       0.91      0.76      0.83       180
                         acute-sinusitis       0.69      0.67      0.68       184
                      alcohol-withdrawal       0.96      0.94      0.95       179
                 alcoholic-liver-disease       0.95      0.96      0.95       186
                                 allergy       0.94      0.93      0.93       180
                                  anemia       0.90      0.92      0.91       179
                            appendicitis       0.75      0.86      0.80       182
                              arrhythmia       0.80      0.84      0.82       187
                                  asthma       0.59      0.69      0.64       183
                             astigmatism       0.48      0.49      0.49       187
                     bladder-obstruction       0.97      0.98      0.98       188
                          cerebral-palsy       0.98      0.98      0.98       179
                              chickenpox       0.72      0.86      0.79       183
                               cirrhosis       0.91      0.93      0.92       179
                             common-cold       0.62      0.54      0.58       186
                          conjunctivitis       0.91      0.92      0.92       187
                      contact-dermatitis       0.86      0.79      0.82       177
                            dengue-fever       0.74      0.68      0.71       184
                                epilepsy       0.98      0.95      0.96       186
                                     flu       0.68      0.71      0.69       179
            fungal-infection-of-the-skin       0.84      0.85      0.85       184
                               gallstone       0.74      0.81      0.77       187
                               gastritis       0.65      0.55      0.60       181
                                  goiter       0.87      0.90      0.89       176
                             heat-stroke       0.97      0.97      0.97       178
                             indigestion       0.72      0.63      0.67       184
                  iron-deficiency-anemia       0.92      0.94      0.93       190
                           liver-disease       0.96      0.91      0.93       186
                                 malaria       0.86      0.87      0.87       182
                                   mumps       1.00      0.96      0.98       184
                                  myopia       0.50      0.48      0.49       188
                          osteoarthritis       0.97      0.98      0.98       180
                               pneumonia       0.76      0.72      0.73       186
                      protein-deficiency       0.90      0.90      0.90       185
                                  rabies       0.73      0.66      0.69       185
                                  scurvy       0.91      0.95      0.93       180
                      sickle-cell-anemia       0.87      0.94      0.91       189
                       sinus-bradycardia       0.89      0.79      0.84       184
                                  stroke       0.90      0.91      0.91       183
                         thyroid-disease       0.93      0.93      0.93       186
                            tuberculosis       0.96      0.96      0.96       183
                           typhoid-fever       0.55      0.65      0.60       188
                    vitamin-a-deficiency       0.95      0.89      0.92       184
                    vitamin-b-deficiency       0.90      0.91      0.91       177
                  vitamin-b12-deficiency       0.96      0.96      0.96       188
                    vitamin-d-deficiency       0.94      0.94      0.94       184
                          whooping-cough       0.74      0.77      0.75       184

                                accuracy                           0.83      8812
                               macro avg       0.84      0.83      0.83      8812
                            weighted avg       0.84      0.83      0.83      8812
```

On Test Data

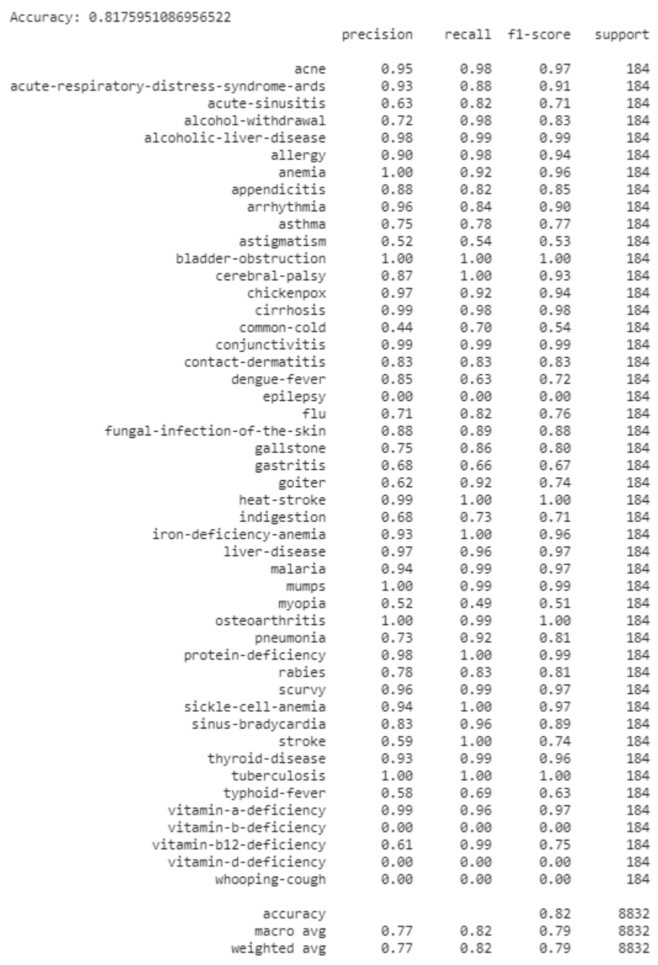


```
Accuracy: 0.8212461695607763
                                          precision    recall  f1-score   support

                                    acne       0.80      0.92      0.86        13
acute-respiratory-distress-syndrome-ards       1.00      0.75      0.86        24
                         acute-sinusitis       0.67      0.50      0.57        20
                      alcohol-withdrawal       0.96      0.92      0.94        25
                 alcoholic-liver-disease       1.00      1.00      1.00        18
                                 allergy       0.92      1.00      0.96        24
                                  anemia       0.95      0.76      0.84        25
                            appendicitis       0.67      0.73      0.70        22
                              arrhythmia       0.57      0.76      0.65        17
                                  asthma       0.73      0.76      0.74        21
                             astigmatism       0.41      0.41      0.41        17
                     bladder-obstruction       1.00      1.00      1.00        16
                          cerebral-palsy       0.96      0.96      0.96        25
                              chickenpox       0.76      0.90      0.83        21
                               cirrhosis       0.89      1.00      0.94        25
                             common-cold       0.47      0.44      0.46        18
                          conjunctivitis       0.75      0.88      0.81        17
                      contact-dermatitis       0.83      0.74      0.78        27
                            dengue-fever       0.62      0.40      0.48        20
                                epilepsy       1.00      1.00      1.00        18
                                     flu       0.70      0.84      0.76        25
            fungal-infection-of-the-skin       0.85      0.85      0.85        20
                               gallstone       0.60      0.75      0.67        16
                               gastritis       0.69      0.48      0.56        23
                                  goiter       1.00      0.89      0.94        28
                             heat-stroke       0.96      1.00      0.98        26
                             indigestion       0.74      0.70      0.72        20
                  iron-deficiency-anemia       0.87      0.93      0.90        14
                           liver-disease       1.00      0.89      0.94        18
                                 malaria       0.84      0.95      0.89        22
                                   mumps       1.00      0.80      0.89        20
                                  myopia       0.33      0.31      0.32        16
                          osteoarthritis       1.00      0.96      0.98        24
                               pneumonia       0.79      0.61      0.69        18
                      protein-deficiency       0.75      0.95      0.84        19
                                  rabies       0.81      0.68      0.74        19
                                  scurvy       0.96      0.96      0.96        24
                      sickle-cell-anemia       0.88      0.93      0.90        15
                       sinus-bradycardia       0.70      0.70      0.70        20
                                  stroke       0.95      0.95      0.95        21
                         thyroid-disease       0.90      1.00      0.95        18
                            tuberculosis       0.95      0.90      0.93        21
                           typhoid-fever       0.50      0.81      0.62        16
                    vitamin-a-deficiency       0.95      0.90      0.92        20
                    vitamin-b-deficiency       1.00      0.85      0.92        27
                  vitamin-b12-deficiency       0.89      1.00      0.94        16
                    vitamin-d-deficiency       0.95      0.95      0.95        20
                          whooping-cough       0.70      0.80      0.74        20

                                accuracy                           0.82       979
                               macro avg       0.82      0.82      0.81       979
                            weighted avg       0.83      0.82      0.82       979
```

- PyGad
  - Train Accuracy – 81.75%
  - Test Accuracy – 78.33%

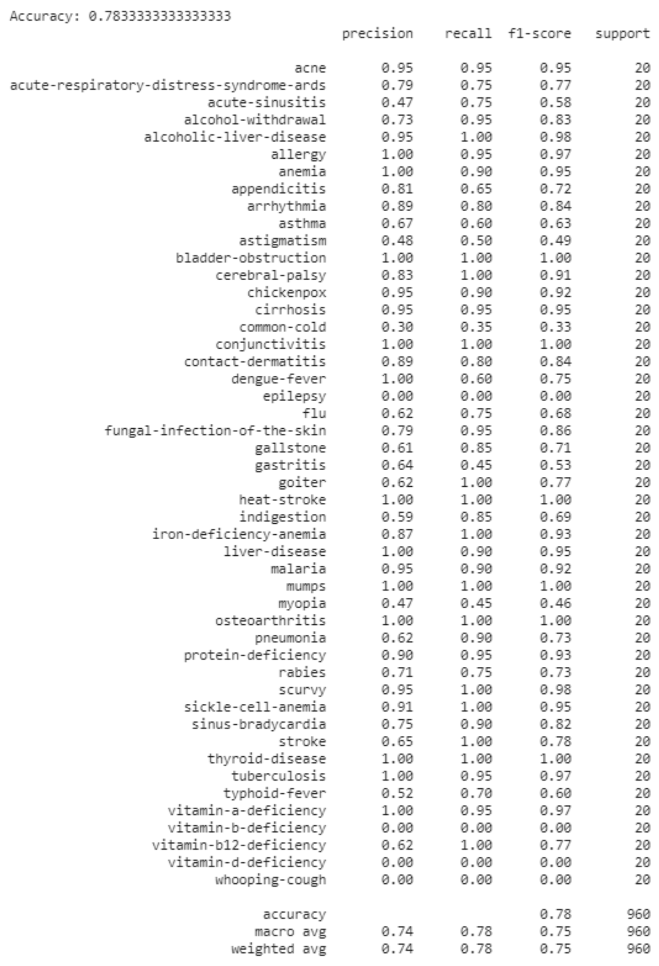


On Training Data



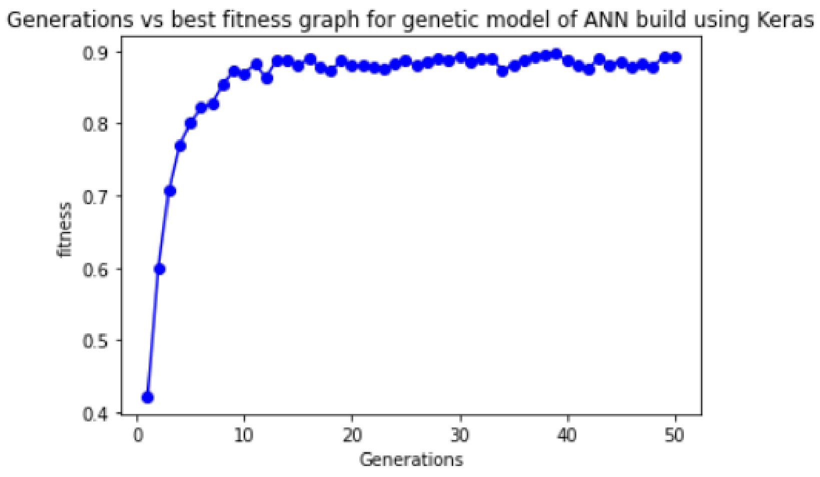
```
Accuracy: 0.8175951086956522
                                          precision    recall  f1-score   support

                                    acne       0.95      0.98      0.97       184
acute-respiratory-distress-syndrome-ards       0.93      0.88      0.91       184
                        acute-sinusitis        0.63      0.82      0.71       184
                      alcohol-withdrawal       0.72      0.98      0.83       184
                 alcoholic-liver-disease       0.98      0.99      0.99       184
                                 allergy       0.90      0.98      0.94       184
                                  anemia       1.00      0.92      0.96       184
                            appendicitis       0.88      0.82      0.85       184
                              arrhythmia       0.96      0.84      0.90       184
                                  asthma       0.75      0.78      0.77       184
                             astigmatism       0.52      0.54      0.53       184
                     bladder-obstruction       1.00      1.00      1.00       184
                          cerebral-palsy       0.87      1.00      0.93       184
                              chickenpox       0.97      0.92      0.94       184
                               cirrhosis       0.99      0.98      0.98       184
                             common-cold       0.44      0.70      0.54       184
                          conjunctivitis       0.99      0.99      0.99       184
                      contact-dermatitis       0.83      0.83      0.83       184
                            dengue-fever       0.85      0.63      0.72       184
                                epilepsy       0.00      0.00      0.00       184
                                     flu       0.71      0.82      0.76       184
             fungal-infection-of-the-skin       0.88      0.89      0.88       184
                               gallstone       0.75      0.86      0.80       184
                               gastritis       0.68      0.66      0.67       184
                                  goiter       0.62      0.92      0.74       184
                             heat-stroke       0.99      1.00      1.00       184
                             indigestion       0.68      0.73      0.71       184
                    iron-deficiency-anemia       0.93      1.00      0.96       184
                           liver-disease       0.97      0.96      0.97       184
                                 malaria       0.94      0.99      0.97       184
                                   mumps       1.00      0.99      0.99       184
                                  myopia       0.52      0.49      0.51       184
                          osteoarthritis       1.00      0.99      1.00       184
                               pneumonia       0.73      0.92      0.81       184
                      protein-deficiency       0.98      1.00      0.99       184
                                  rabies       0.78      0.83      0.81       184
                                  scurvy       0.96      0.99      0.97       184
                      sickle-cell-anemia       0.94      1.00      0.97       184
                       sinus-bradycardia       0.83      0.96      0.89       184
                                  stroke       0.59      1.00      0.74       184
                         thyroid-disease       0.93      0.99      0.96       184
                            tuberculosis       1.00      1.00      1.00       184
                           typhoid-fever       0.58      0.69      0.63       184
                     vitamin-a-deficiency       0.99      0.96      0.97       184
                     vitamin-b-deficiency       0.00      0.00      0.00       184
                   vitamin-b12-deficiency       0.61      0.99      0.75       184
                     vitamin-d-deficiency       0.00      0.00      0.00       184
                          whooping-cough       0.00      0.00      0.00       184

                                accuracy                           0.82      8832
                               macro avg       0.77      0.82      0.79      8832
                            weighted avg       0.77      0.82      0.79      8832
```

On Test Data

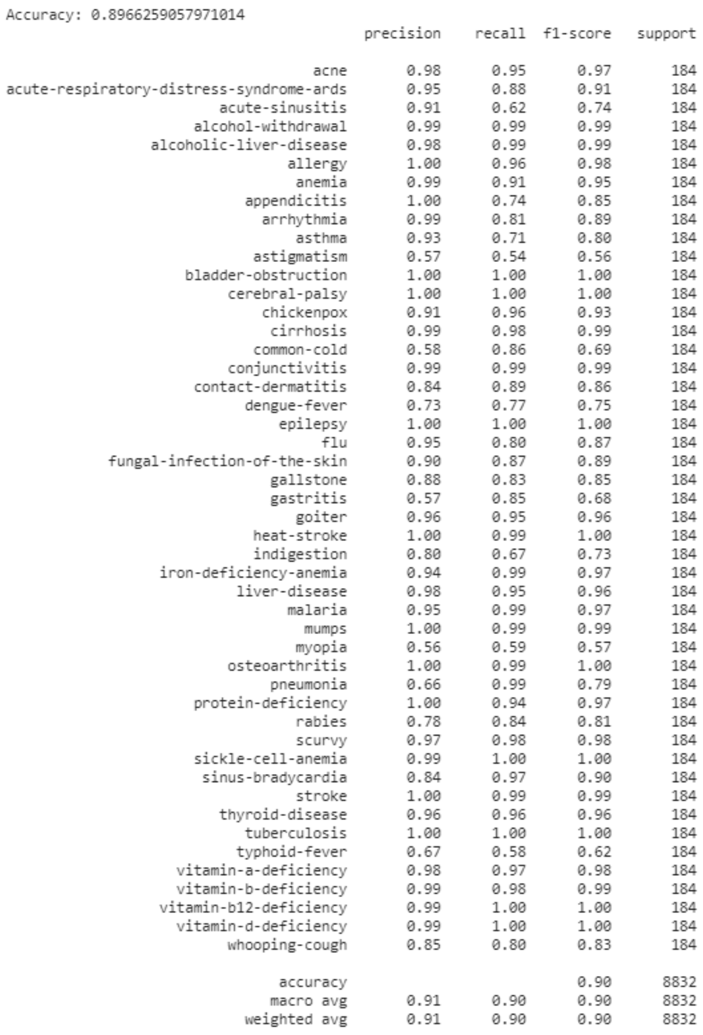


```
Accuracy: 0.7833333333333333
                                          precision    recall  f1-score   support

                                    acne       0.95      0.95      0.95        20
acute-respiratory-distress-syndrome-ards       0.79      0.75      0.77        20
                         acute-sinusitis       0.47      0.75      0.58        20
                      alcohol-withdrawal       0.73      0.95      0.83        20
                 alcoholic-liver-disease       0.95      1.00      0.98        20
                                 allergy       1.00      0.95      0.97        20
                                  anemia       1.00      0.90      0.95        20
                            appendicitis       0.81      0.65      0.72        20
                              arrhythmia       0.89      0.80      0.84        20
                                  asthma       0.67      0.60      0.63        20
                             astigmatism       0.48      0.50      0.49        20
                     bladder-obstruction       1.00      1.00      1.00        20
                          cerebral-palsy       0.83      1.00      0.91        20
                              chickenpox       0.95      0.90      0.92        20
                               cirrhosis       0.95      0.95      0.95        20
                             common-cold       0.30      0.35      0.33        20
                          conjunctivitis       1.00      1.00      1.00        20
                      contact-dermatitis       0.89      0.80      0.84        20
                            dengue-fever       1.00      0.60      0.75        20
                                epilepsy       0.00      0.00      0.00        20
                                     flu       0.62      0.75      0.68        20
             fungal-infection-of-the-skin       0.79      0.95      0.86        20
                               gallstone       0.61      0.85      0.71        20
                               gastritis       0.64      0.45      0.53        20
                                  goiter       0.62      1.00      0.77        20
                             heat-stroke       1.00      1.00      1.00        20
                             indigestion       0.59      0.85      0.69        20
                    iron-deficiency-anemia       0.87      1.00      0.93        20
                           liver-disease       1.00      0.90      0.95        20
                                 malaria       0.95      0.90      0.92        20
                                   mumps       1.00      1.00      1.00        20
                                  myopia       0.47      0.45      0.46        20
                          osteoarthritis       1.00      1.00      1.00        20
                               pneumonia       0.62      0.90      0.73        20
                      protein-deficiency       0.90      0.95      0.93        20
                                  rabies       0.71      0.75      0.73        20
                                  scurvy       0.95      1.00      0.98        20
                      sickle-cell-anemia       0.91      1.00      0.95        20
                       sinus-bradycardia       0.75      0.90      0.82        20
                                  stroke       0.65      1.00      0.78        20
                         thyroid-disease       1.00      1.00      1.00        20
                            tuberculosis       1.00      0.95      0.97        20
                           typhoid-fever       0.52      0.70      0.60        20
                     vitamin-a-deficiency       1.00      0.95      0.97        20
                     vitamin-b-deficiency       0.00      0.00      0.00        20
                   vitamin-b12-deficiency       0.62      1.00      0.77        20
                     vitamin-d-deficiency       0.00      0.00      0.00        20
                          whooping-cough       0.00      0.00      0.00        20

                                accuracy                           0.78       960
                               macro avg       0.74      0.78      0.75       960
                            weighted avg       0.74      0.78      0.75       960
```

- Keras with Steps of Evolutionary Algorithms
  - Training Time – 15 min 43 sec
  - Train Accuracy – 89.66%
  - Test Accuracy – 85.45%

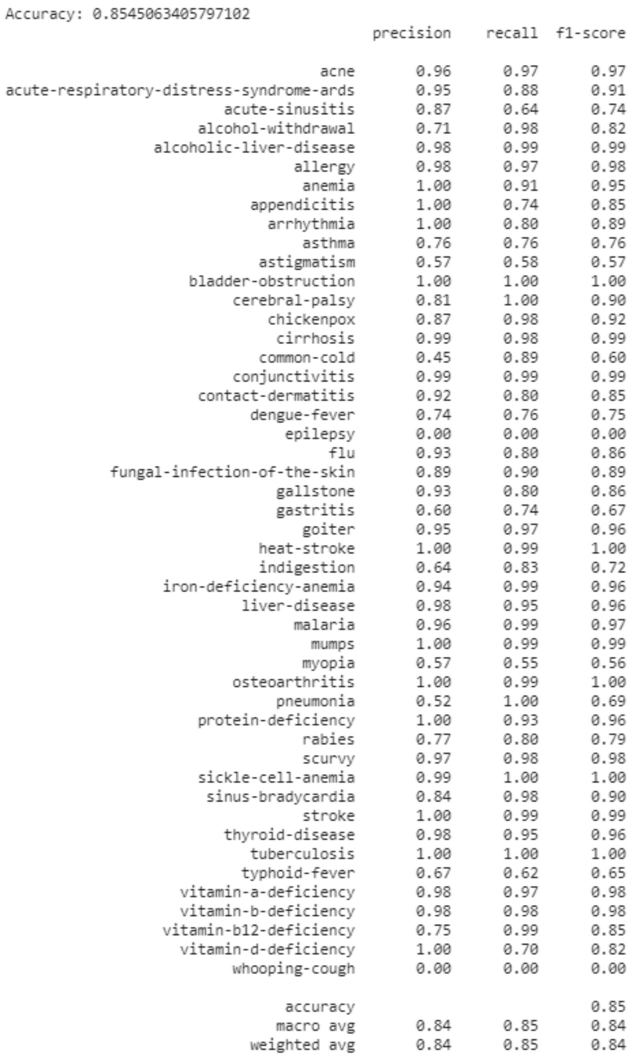


On Training Data



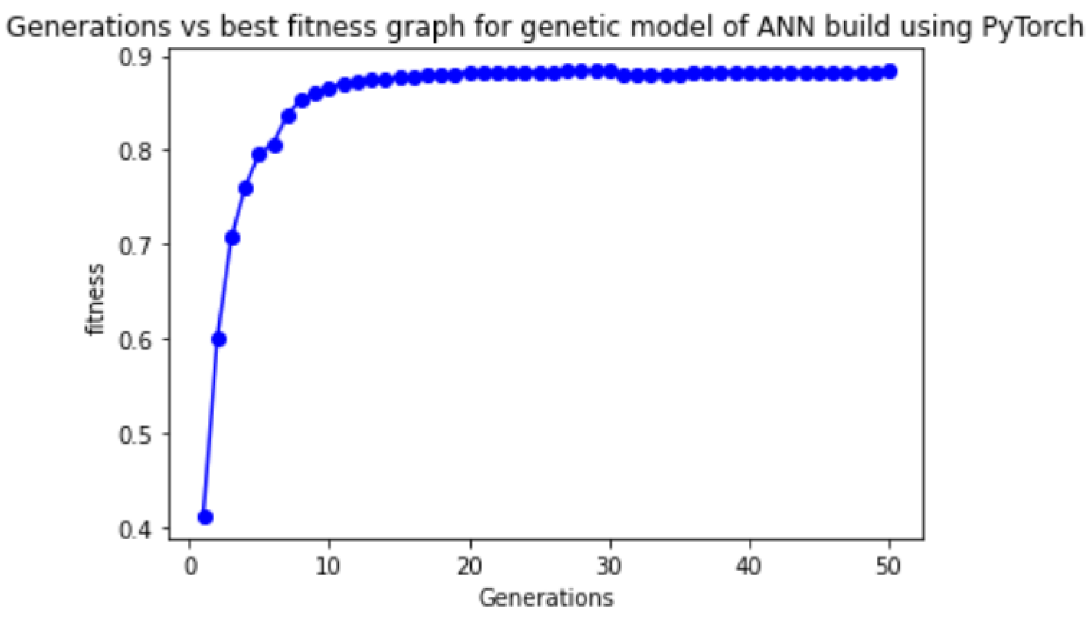
```
Accuracy: 0.8966259057971014
                                        precision    recall  f1-score   support

                                   acne       0.98      0.95      0.97       184
acute-respiratory-distress-syndrome-ards       0.95      0.88      0.91       184
                       acute-sinusitis       0.91      0.62      0.74       184
                     alcohol-withdrawal       0.99      0.99      0.99       184
                alcoholic-liver-disease       0.98      0.99      0.99       184
                                allergy       1.00      0.96      0.98       184
                                 anemia       0.99      0.91      0.95       184
                           appendicitis       1.00      0.74      0.85       184
                             arrhythmia       0.99      0.81      0.89       184
                                 asthma       0.93      0.71      0.80       184
                            astigmatism       0.57      0.54      0.56       184
                    bladder-obstruction       1.00      1.00      1.00       184
                         cerebral-palsy       1.00      1.00      1.00       184
                             chickenpox       0.91      0.96      0.93       184
                              cirrhosis       0.99      0.98      0.99       184
                            common-cold       0.58      0.86      0.69       184
                         conjunctivitis       0.99      0.99      0.99       184
                      contact-dermatitis       0.84      0.89      0.86       184
                           dengue-fever       0.73      0.77      0.75       184
                               epilepsy       1.00      1.00      1.00       184
                                    flu       0.95      0.80      0.87       184
             fungal-infection-of-the-skin       0.90      0.87      0.89       184
                              gallstone       0.88      0.83      0.85       184
                              gastritis       0.57      0.85      0.68       184
                                 goiter       0.96      0.95      0.96       184
                            heat-stroke       1.00      0.99      1.00       184
                            indigestion       0.80      0.67      0.73       184
                  iron-deficiency-anemia       0.94      0.99      0.97       184
                          liver-disease       0.98      0.95      0.96       184
                                malaria       0.95      0.99      0.97       184
                                  mumps       1.00      0.99      0.99       184
                                 myopia       0.56      0.59      0.57       184
                         osteoarthritis       1.00      0.99      1.00       184
                              pneumonia       0.66      0.99      0.79       184
                     protein-deficiency       1.00      0.94      0.97       184
                                 rabies       0.78      0.84      0.81       184
                                 scurvy       0.97      0.98      0.98       184
                     sickle-cell-anemia       0.99      1.00      1.00       184
                      sinus-bradycardia       0.84      0.97      0.90       184
                                 stroke       1.00      0.99      0.99       184
                        thyroid-disease       0.96      0.96      0.96       184
                           tuberculosis       1.00      1.00      1.00       184
                         typhoid-fever       0.67      0.58      0.62       184
                    vitamin-a-deficiency       0.98      0.97      0.98       184
                    vitamin-b-deficiency       0.99      0.98      0.99       184
                  vitamin-b12-deficiency       0.99      1.00      1.00       184
                    vitamin-d-deficiency       0.99      1.00      1.00       184
                         whooping-cough       0.85      0.80      0.83       184

                               accuracy                           0.90      8832
                              macro avg       0.91      0.90      0.90      8832
                           weighted avg       0.91      0.90      0.90      8832
```

On Test Data



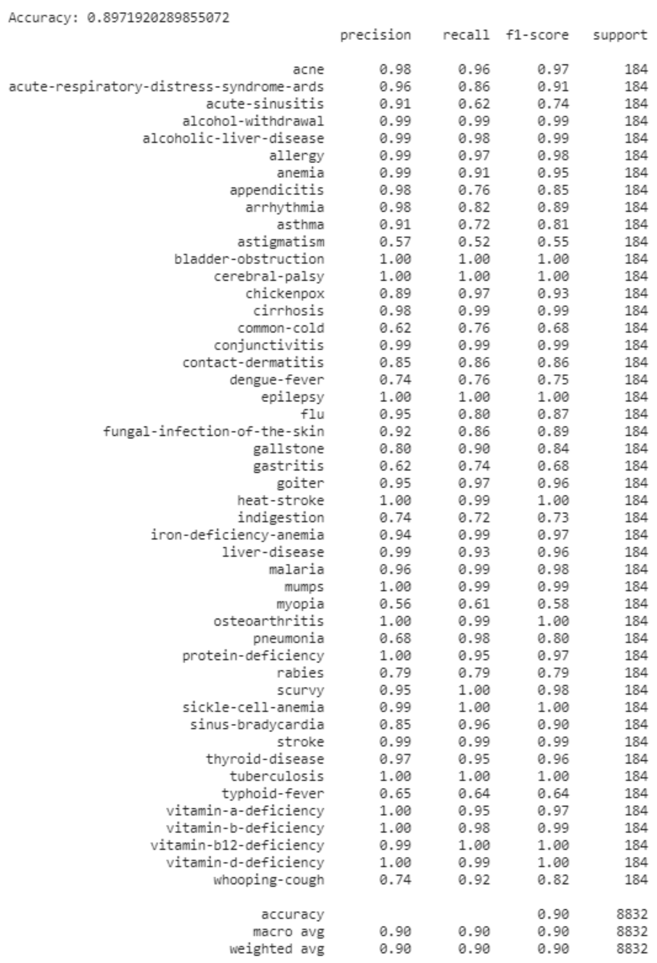
```
Accuracy: 0.8545063405797102
                                        precision    recall  f1-score

                                   acne       0.96      0.97      0.97
acute-respiratory-distress-syndrome-ards       0.95      0.88      0.91
                       acute-sinusitis       0.87      0.64      0.74
                     alcohol-withdrawal       0.71      0.98      0.82
                alcoholic-liver-disease       0.98      0.99      0.99
                                allergy       0.98      0.97      0.98
                                 anemia       1.00      0.91      0.95
                           appendicitis       1.00      0.74      0.85
                             arrhythmia       1.00      0.80      0.89
                                 asthma       0.76      0.76      0.76
                            astigmatism       0.57      0.58      0.57
                    bladder-obstruction       1.00      1.00      1.00
                         cerebral-palsy       0.81      1.00      0.90
                             chickenpox       0.87      0.98      0.92
                              cirrhosis       0.99      0.98      0.99
                            common-cold       0.45      0.89      0.60
                         conjunctivitis       0.99      0.99      0.99
                      contact-dermatitis       0.92      0.80      0.85
                           dengue-fever       0.74      0.76      0.75
                               epilepsy       0.00      0.00      0.00
                                    flu       0.93      0.80      0.86
             fungal-infection-of-the-skin       0.89      0.90      0.89
                              gallstone       0.93      0.80      0.86
                              gastritis       0.60      0.74      0.67
                                 goiter       0.95      0.97      0.96
                            heat-stroke       1.00      0.99      1.00
                            indigestion       0.64      0.83      0.72
                  iron-deficiency-anemia       0.94      0.99      0.96
                          liver-disease       0.98      0.95      0.96
                                malaria       0.96      0.99      0.97
                                  mumps       1.00      0.99      0.99
                                 myopia       0.57      0.55      0.56
                         osteoarthritis       1.00      0.99      1.00
                              pneumonia       0.52      1.00      0.69
                     protein-deficiency       1.00      0.93      0.96
                                 rabies       0.77      0.80      0.79
                                 scurvy       0.97      0.98      0.98
                     sickle-cell-anemia       0.99      1.00      1.00
                      sinus-bradycardia       0.84      0.98      0.90
                                 stroke       1.00      0.99      0.99
                        thyroid-disease       0.98      0.95      0.96
                           tuberculosis       1.00      1.00      1.00
                         typhoid-fever       0.67      0.62      0.65
                    vitamin-a-deficiency       0.98      0.97      0.98
                    vitamin-b-deficiency       0.98      0.98      0.98
                  vitamin-b12-deficiency       0.75      0.99      0.85
                    vitamin-d-deficiency       1.00      0.70      0.82
                         whooping-cough       0.00      0.00      0.00

                               accuracy                           0.85
                              macro avg       0.84      0.85      0.84
                           weighted avg       0.84      0.85      0.84
```

- PyTorch with Steps of Evolutionary Algorithms
    - Training Time – 2 min 7 sec (more than **7 times speedup when compared to Keras Model**. This is due to the fact that in PyTorch model, **we can transfer the data tensors along with model to the device side i.e. the GPU side and thus increasing the efficiency of GPU** and hence exponentially decreasing the training time)
    - Train Accuracy – 89.71%
    - Test Accuracy – 83.69%

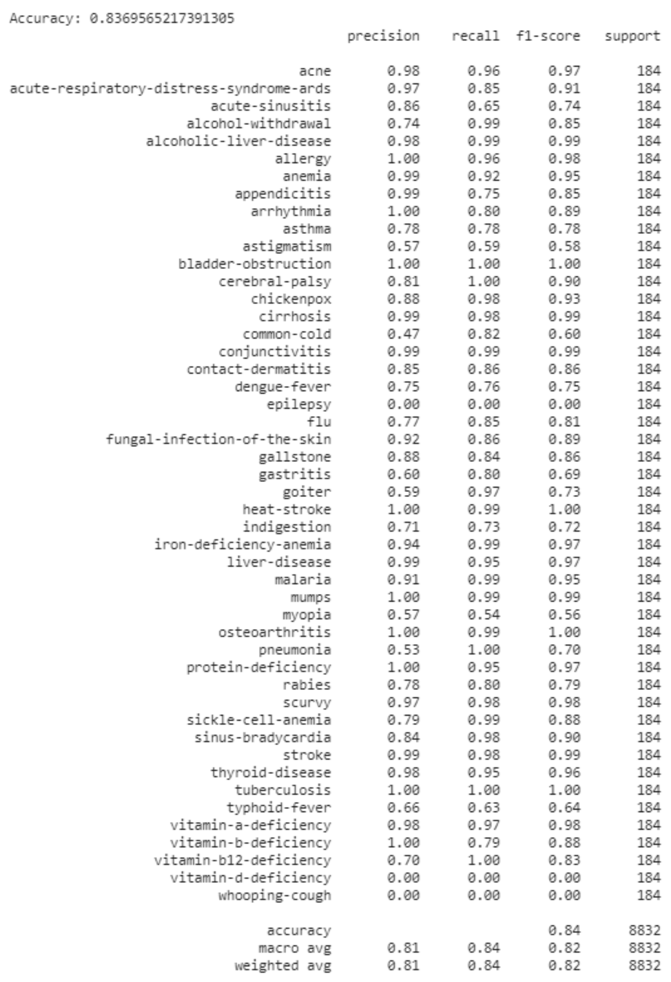


On Training Data

On Test Data



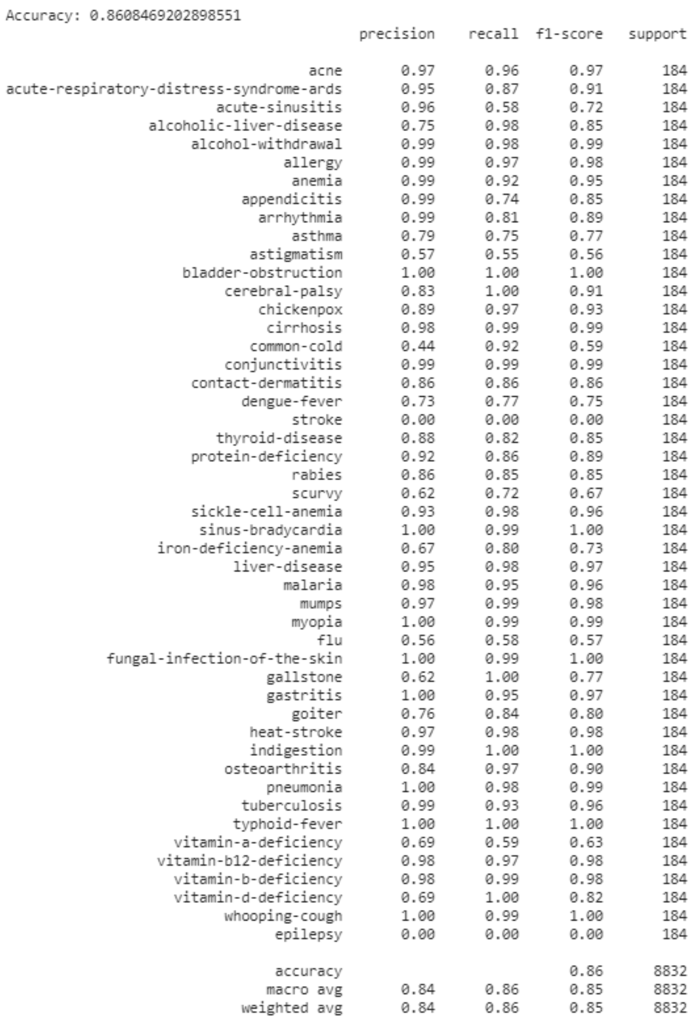
```
Accuracy: 0.8971920289855072
                                          precision    recall  f1-score   support

                                    acne       0.98      0.96      0.97       184
acute-respiratory-distress-syndrome-ards       0.96      0.86      0.91       184
                         acute-sinusitis       0.91      0.62      0.74       184
                      alcohol-withdrawal       0.99      0.99      0.99       184
                 alcoholic-liver-disease       0.99      0.98      0.99       184
                                 allergy       0.99      0.97      0.98       184
                                  anemia       0.99      0.91      0.95       184
                            appendicitis       0.98      0.76      0.85       184
                              arrhythmia       0.98      0.82      0.89       184
                                  asthma       0.91      0.72      0.81       184
                             astigmatism       0.57      0.52      0.55       184
                     bladder-obstruction       1.00      1.00      1.00       184
                          cerebral-palsy       1.00      1.00      1.00       184
                              chickenpox       0.89      0.97      0.93       184
                               cirrhosis       0.98      0.99      0.99       184
                             common-cold       0.62      0.76      0.68       184
                          conjunctivitis       0.99      0.99      0.99       184
                      contact-dermatitis       0.85      0.86      0.86       184
                            dengue-fever       0.74      0.76      0.75       184
                                epilepsy       1.00      1.00      1.00       184
                                     flu       0.95      0.80      0.87       184
               fungal-infection-of-the-skin       0.92      0.86      0.89       184
                               gallstone       0.80      0.90      0.84       184
                               gastritis       0.62      0.74      0.68       184
                                  goiter       0.95      0.97      0.96       184
                             heat-stroke       1.00      0.99      1.00       184
                             indigestion       0.74      0.72      0.73       184
                    iron-deficiency-anemia       0.94      0.99      0.97       184
                           liver-disease       0.99      0.93      0.96       184
                                 malaria       0.96      0.99      0.98       184
                                   mumps       1.00      0.99      0.99       184
                                  myopia       0.56      0.61      0.58       184
                          osteoarthritis       1.00      0.99      1.00       184
                               pneumonia       0.68      0.98      0.80       184
                      protein-deficiency       1.00      0.95      0.97       184
                                  rabies       0.79      0.79      0.79       184
                                  scurvy       0.95      1.00      0.98       184
                      sickle-cell-anemia       0.99      1.00      1.00       184
                       sinus-bradycardia       0.85      0.96      0.90       184
                                  stroke       0.99      0.99      0.99       184
                         thyroid-disease       0.97      0.95      0.96       184
                            tuberculosis       1.00      1.00      1.00       184
                           typhoid-fever       0.65      0.64      0.64       184
                     vitamin-a-deficiency       1.00      0.95      0.97       184
                     vitamin-b-deficiency       1.00      0.98      0.99       184
                   vitamin-b12-deficiency       0.99      1.00      1.00       184
                     vitamin-d-deficiency       1.00      0.99      1.00       184
                          whooping-cough       0.74      0.92      0.82       184

                                accuracy                           0.90      8832
                               macro avg       0.90      0.90      0.90      8832
                            weighted avg       0.90      0.90      0.90      8832
```



```
Accuracy: 0.8369565217391305
                                          precision    recall  f1-score   support

                                    acne       0.98      0.96      0.97       184
acute-respiratory-distress-syndrome-ards       0.97      0.85      0.91       184
                         acute-sinusitis       0.86      0.65      0.74       184
                      alcohol-withdrawal       0.74      0.99      0.85       184
                 alcoholic-liver-disease       0.98      0.99      0.99       184
                                 allergy       1.00      0.96      0.98       184
                                  anemia       0.99      0.92      0.95       184
                            appendicitis       0.99      0.75      0.85       184
                              arrhythmia       1.00      0.80      0.89       184
                                  asthma       0.78      0.78      0.78       184
                             astigmatism       0.57      0.59      0.58       184
                     bladder-obstruction       1.00      1.00      1.00       184
                          cerebral-palsy       0.81      1.00      0.90       184
                              chickenpox       0.88      0.98      0.93       184
                               cirrhosis       0.99      0.98      0.99       184
                             common-cold       0.47      0.82      0.60       184
                          conjunctivitis       0.99      0.99      0.99       184
                      contact-dermatitis       0.85      0.86      0.86       184
                            dengue-fever       0.75      0.76      0.75       184
                                epilepsy       0.00      0.00      0.00       184
                                     flu       0.77      0.85      0.81       184
            fungal-infection-of-the-skin       0.92      0.86      0.89       184
                               gallstone       0.88      0.84      0.86       184
                               gastritis       0.60      0.80      0.69       184
                                  goiter       0.59      0.97      0.73       184
                             heat-stroke       1.00      0.99      1.00       184
                             indigestion       0.71      0.73      0.72       184
                  iron-deficiency-anemia       0.94      0.99      0.97       184
                           liver-disease       0.99      0.95      0.97       184
                                 malaria       0.91      0.99      0.95       184
                                   mumps       1.00      0.99      0.99       184
                                  myopia       0.57      0.54      0.56       184
                          osteoarthritis       1.00      0.99      1.00       184
                               pneumonia       0.53      1.00      0.70       184
                      protein-deficiency       1.00      0.95      0.97       184
                                  rabies       0.78      0.80      0.79       184
                                  scurvy       0.97      0.98      0.98       184
                      sickle-cell-anemia       0.79      0.99      0.88       184
                       sinus-bradycardia       0.84      0.98      0.90       184
                                  stroke       0.99      0.98      0.99       184
                         thyroid-disease       0.98      0.95      0.96       184
                            tuberculosis       1.00      1.00      1.00       184
                           typhoid-fever       0.66      0.63      0.64       184
                    vitamin-a-deficiency       0.98      0.97      0.98       184
                    vitamin-b-deficiency       1.00      0.79      0.88       184
                  vitamin-b12-deficiency       0.70      1.00      0.83       184
                    vitamin-d-deficiency       0.00      0.00      0.00       184
                          whooping-cough       0.00      0.00      0.00       184

                                accuracy                           0.84      8832
                               macro avg       0.81      0.84      0.82      8832
                            weighted avg       0.81      0.84      0.82      8832
```

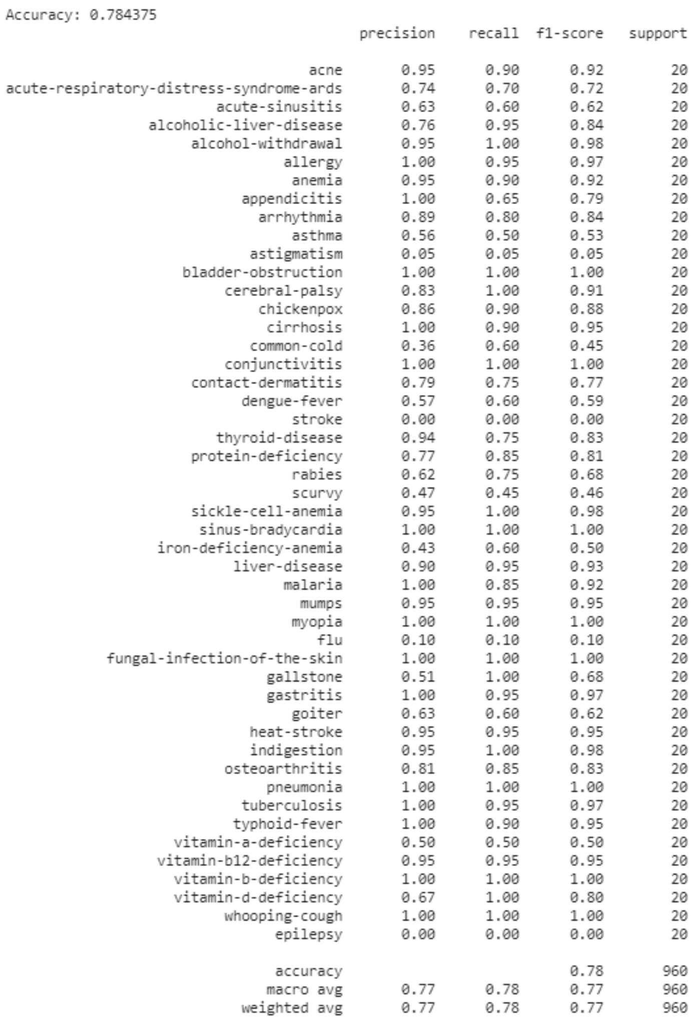
- Random Forest Classifier
    - Train Accuracy – 86.08%
    - Test Accuracy – 78.43%

On Training Data



```
Accuracy: 0.8608469202898551
                                          precision    recall  f1-score   support

                                    acne       0.97      0.96      0.97       184
acute-respiratory-distress-syndrome-ards       0.95      0.87      0.91       184
                         acute-sinusitis       0.96      0.58      0.72       184
                 alcoholic-liver-disease       0.75      0.98      0.85       184
                      alcohol-withdrawal       0.99      0.98      0.99       184
                                 allergy       0.99      0.97      0.98       184
                                  anemia       0.99      0.92      0.95       184
                            appendicitis       0.99      0.74      0.85       184
                              arrhythmia       0.99      0.81      0.89       184
                                  asthma       0.79      0.75      0.77       184
                             astigmatism       0.57      0.55      0.56       184
                     bladder-obstruction       1.00      1.00      1.00       184
                          cerebral-palsy       0.83      1.00      0.91       184
                              chickenpox       0.89      0.97      0.93       184
                               cirrhosis       0.98      0.99      0.99       184
                             common-cold       0.44      0.92      0.59       184
                          conjunctivitis       0.99      0.99      0.99       184
                      contact-dermatitis       0.86      0.86      0.86       184
                            dengue-fever       0.73      0.77      0.75       184
                                  stroke       0.00      0.00      0.00       184
                         thyroid-disease       0.88      0.82      0.85       184
                      protein-deficiency       0.92      0.86      0.89       184
                                  rabies       0.86      0.85      0.85       184
                                  scurvy       0.62      0.72      0.67       184
                      sickle-cell-anemia       0.93      0.98      0.96       184
                       sinus-bradycardia       1.00      0.99      1.00       184
                  iron-deficiency-anemia       0.67      0.80      0.73       184
                           liver-disease       0.95      0.98      0.97       184
                                 malaria       0.98      0.95      0.96       184
                                   mumps       0.97      0.99      0.98       184
                                  myopia       1.00      0.99      0.99       184
                                     flu       0.56      0.58      0.57       184
            fungal-infection-of-the-skin       1.00      0.99      1.00       184
                               gallstone       0.62      1.00      0.77       184
                               gastritis       1.00      0.95      0.97       184
                                  goiter       0.76      0.84      0.80       184
                             heat-stroke       0.97      0.98      0.98       184
                             indigestion       0.99      1.00      1.00       184
                          osteoarthritis       0.84      0.97      0.90       184
                               pneumonia       1.00      0.98      0.99       184
                            tuberculosis       0.99      0.93      0.96       184
                           typhoid-fever       1.00      1.00      1.00       184
                    vitamin-a-deficiency       0.69      0.59      0.63       184
                  vitamin-b12-deficiency       0.98      0.97      0.98       184
                    vitamin-b-deficiency       0.98      0.99      0.98       184
                    vitamin-d-deficiency       0.69      1.00      0.82       184
                          whooping-cough       1.00      0.99      1.00       184
                                epilepsy       0.00      0.00      0.00       184

                                accuracy                           0.86      8832
                               macro avg       0.84      0.86      0.85      8832
                            weighted avg       0.84      0.86      0.85      8832
```

On Test Data



```
Accuracy: 0.784375
                                          precision    recall  f1-score   support

                                    acne       0.95      0.90      0.92        20
acute-respiratory-distress-syndrome-ards       0.74      0.70      0.72        20
                         acute-sinusitis       0.63      0.60      0.62        20
                 alcoholic-liver-disease       0.76      0.95      0.84        20
                      alcohol-withdrawal       0.95      1.00      0.98        20
                                 allergy       1.00      0.95      0.97        20
                                  anemia       0.95      0.90      0.92        20
                            appendicitis       1.00      0.65      0.79        20
                              arrhythmia       0.89      0.80      0.84        20
                                  asthma       0.56      0.50      0.53        20
                             astigmatism       0.05      0.05      0.05        20
                     bladder-obstruction       1.00      1.00      1.00        20
                          cerebral-palsy       0.83      1.00      0.91        20
                              chickenpox       0.86      0.90      0.88        20
                               cirrhosis       1.00      0.90      0.95        20
                             common-cold       0.36      0.60      0.45        20
                          conjunctivitis       1.00      1.00      1.00        20
                      contact-dermatitis       0.79      0.75      0.77        20
                            dengue-fever       0.57      0.60      0.59        20
                                  stroke       0.00      0.00      0.00        20
                         thyroid-disease       0.94      0.75      0.83        20
                      protein-deficiency       0.77      0.85      0.81        20
                                  rabies       0.62      0.75      0.68        20
                                  scurvy       0.47      0.45      0.46        20
                      sickle-cell-anemia       0.95      1.00      0.98        20
                       sinus-bradycardia       1.00      1.00      1.00        20
                  iron-deficiency-anemia       0.43      0.60      0.50        20
                           liver-disease       0.90      0.95      0.93        20
                                 malaria       1.00      0.85      0.92        20
                                   mumps       0.95      0.95      0.95        20
                                  myopia       1.00      1.00      1.00        20
                                     flu       0.10      0.10      0.10        20
            fungal-infection-of-the-skin       1.00      1.00      1.00        20
                               gallstone       0.51      1.00      0.68        20
                               gastritis       1.00      0.95      0.97        20
                                  goiter       0.63      0.60      0.62        20
                             heat-stroke       0.95      0.95      0.95        20
                             indigestion       0.95      1.00      0.98        20
                          osteoarthritis       0.81      0.85      0.83        20
                               pneumonia       1.00      1.00      1.00        20
                            tuberculosis       1.00      0.95      0.97        20
                           typhoid-fever       1.00      0.90      0.95        20
                    vitamin-a-deficiency       0.50      0.50      0.50        20
                  vitamin-b12-deficiency       0.95      0.95      0.95        20
                    vitamin-b-deficiency       1.00      1.00      1.00        20
                    vitamin-d-deficiency       0.67      1.00      0.80        20
                          whooping-cough       1.00      1.00      1.00        20
                                epilepsy       0.00      0.00      0.00        20

                                accuracy                           0.78       960
                               macro avg       0.77      0.78      0.77       960
                            weighted avg       0.77      0.78      0.77       960
```

Conclusion

All the above results can be summarised as -

S. No.	Apprach	Train Accuracy	Test Accuracy
1	Naïve Bayes Classifier	55.48%	53.75
2	Simple Artificial Neural Net Classifier using Keras	83.45%	82.12%
3	PyGad	81.75%	78.33%
4	ANN Model using Keras + manual evolutionary algorithm's steps	89.66%	85.45%
5	ANN Model using PyTorch + manual evolutionary algorithm's steps	89.71%	83.69%
6	Random Forest Classifier	86.08%	78.43%

From the above table we can see that –

- While a simple ANN can ooutperform naïve bayes by a large margin, it is still behind the Random Forest Classifier in terms of accuracy.
- On applying PyGad, a decrease in performance was observed and on further reasearch it was found that t is still a package still under development. Due to this they have yet not been able to provide proper benefit of Evolutionary Algorithms
- We can also see that while PyTorch and Keras Models give similar accuracies, there is a large difference in terms of their training time for same hyperparameters.
- We can see that on writing functions for selection, crossover and mutation from scratch and using them along with the ANN has lead to a significant increase in performance and the resulting model is now even better than the Random Forest Classifier.

So from all the above points we can safely conclude that **ANN Model using PyTorch along with manual evolutionary algorithm's steps** have outperformed all the other models.

References and Regards

- PyGAD documentation: https://pygad.readthedocs.io/en/latest/
- Application of GA in ANN: https://medium.com/swlh/genetic-algorithm-in-artificial-neural-network-5f5b9c9467d0
- Building models with Keras and PyTorch: https://medium.com/deep-learning-with-keras/which-activation-loss-functions-in-multi-class-clasification-4cd599e4e61f