

Disease Prediction Based on Identified Symptoms

In [44]:

```
import csv
import pandas as pd
import numpy as np
from collections import defaultdict
import seaborn as sns
import matplotlib.pyplot as plt
```

In [70]:

```
# Load Dataset
data= pd.read_excel("DSDataset.xlsx")
data
```

Out[70]:

	Disease	Count of Disease Occurrence	Symptom
0	hypertensive disease	3363.0	pain chest
1	hypertensive disease	NaN	shortness of breath
2	hypertensive disease	NaN	dizziness
3	hypertensive disease	NaN	asthenia
4	hypertensive disease	NaN	fall
...
1861	affect labile	NaN	bedridden
1862	affect labile	NaN	prostatism
1863	decubitus ulcer	42.0	systolic murmur
1864	decubitus ulcer	NaN	frail
1865	decubitus ulcer	NaN	fever

1866 rows × 3 columns

In [71]:

```
# Remove column 1
df = data.loc[:, data.columns!='Count of Disease Occurrence']
df
```

Out[71]:

	Disease	Symptom
0	hypertensive disease	pain chest
1	hypertensive disease	shortness of breath
2	hypertensive disease	dizziness
3	hypertensive disease	asthenia
4	hypertensive disease	fall

	Disease	Symptom
...
1861	affect labile	bedridden
1862	affect labile	prostatism
1863	decubitus ulcer	systolic murmur
1864	decubitus ulcer	frail
1865	decubitus ulcer	fever

1866 rows × 2 columns

```
In [73]: # Available data types
df.dtypes
```

```
Out[73]: Disease    object
Symptom    object
dtype: object
```

```
In [80]: # Preprocessing
import spacy

def preprocess(text):
    nlp = spacy.load("en_core_web_sm")
    doc = nlp(text)
    tokens = [word.lemma_ for word in doc if word.is_stop is False and word.is_alpha is True]
    return tokens

preprocess('shortness of breath')
```

```
Out[80]: ['shortness', 'breath']
```

```
In [87]: # Vectorize the data
from sklearn.feature_extraction.text import TfidfVectorizer

vectorizer = TfidfVectorizer(stop_words='english')
vector = vectorizer.fit_transform(df['Disease'])
```

```
In [89]: # Encode the Labels
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import LabelEncoder

label_encoder = LabelEncoder()
integer_encode = label_encoder.fit_transform(df['Symptom'])
print(integer_encode)
```

```
[252 323  84 ... 355 127 119]
```

```
In [94]: #Unique symptoms
```

```
unique_sym = np.asarray(df['Symptom'].unique())
unique_sym
```

```
Out[94]: array(['pain chest', 'shortness of breath', 'dizziness', 'asthenia',
      'fall', 'syncope', 'vertigo', 'sweating increased', 'palpitation',
      'nausea', 'angina pectoris', 'pressure chest', 'polyuria',
      'polydypsia', 'orthopnea', 'rale', 'unresponsiveness',
      'mental status changes', 'vomiting', 'labored breathing',
      'feeling suicidal', 'suicidal', 'hallucinations auditory',
      'feeling hopeless', 'weepiness', 'sleeplessness',
      'motor retardation', 'irritable mood', 'blackout',
      'mood depressed', 'hallucinations visual', 'worry', 'agitation',
      'tremor', 'intoxication', 'verbal auditory hallucinations',
      'energy increased', 'difficulty', 'nightmare',
      'unable to concentrate', 'homelessness', 'hypokinesia',
      'dyspnea on exertion', 'chest tightness', 'cough', 'fever',
      'decreased translucency', 'productive cough', 'pleuritic pain',
      'yellow sputum', 'breath sounds decreased', 'chill', 'rhonchus',
      'green sputum', 'non-productive cough', 'wheezing', 'haemoptysis',
      'distress respiratory', 'tachypnea', 'malaise', 'night sweat',
      'jugular venous distention', 'dyspnea', 'dysarthria',
      'speech slurred', 'facial paresis', 'hemiplegia', 'seizure',
      'numbness', 'symptom aggravating factors', 'st segment elevation',
      'st segment depression', 't wave inverted', 'presence of q wave',
      'chest discomfort', 'bradycardia', 'pain', 'nonsmoker', 'erythema',
      'hepatosplenomegaly', 'pruritus', 'diarrhea', 'abscess bacterial',
      'swelling', 'apyrexial', 'dysuria', 'hematuria',
      'renal angle tenderness', 'lethargy', 'hyponatremia',
      'hemodynamically stable', 'difficulty passing urine',
      'consciousness clear', 'guaiac positive', 'monoclonal',
      'ecchymosis', 'tumor cell invasion', 'haemorrhage', 'pallor',
      'fatigue', 'heme positive', 'pain back', 'orthostasis',
      'arthralgia', 'transaminitis', 'sputum purulent', 'hypoxemia',
      'hypercapnia', 'patient non compliance', 'unconscious state',
      'bedridden', 'abdominal tenderness', 'unsteady gait',
      'hyperkalemia', 'urgency of\x00micturition', 'ascites',
      'hypotension', 'enuresis', 'asterixis', 'muscle twitch', 'sleepy',
      'headache', 'lightheadedness', 'food intolerance',
      'numbness of hand', 'general discomfort', 'drowsiness',
      'stiffness', 'prostatism', 'weight gain', 'tired',
      'mass of body structure', 'has religious belief', 'nervousness',
      'formication', 'hot flush', 'lesion', 'cushingoid\x00habitus',
      'emphysematous change', 'decreased body weight', 'hoarseness',
      'thicken', 'spontaneous rupture of membranes', nan, 'hypotonic',
      'redness', 'hypesthesia', 'hyperacusis', 'scratch marks',
      'sore to touch', 'burning sensation', 'satiety early',
      'throbbing sensation quality', 'sensory discomfort',
      'constipation', 'pain abdominal', 'heartburn',
      'breech presentation', 'cyanosis', 'pain in lower limb',
      'cardiomegaly', 'clonus', 'unwell', 'anorexia',
      'history of - blackout', 'anosmia', 'metastatic lesion',
      'hemianopsia homonymous', 'hematocrit decreased', 'neck stiffness',
      'cicatrisation', 'hypometabolism', 'aura', 'myoclonus', 'gurgles',
      'wheelchair bound', 'left\x00atrial\x00hypertrophy', 'oliguria',
      'catatonia', 'unhappy', 'paresthesia', 'gravid 0', 'lung nodule',
      'distended abdomen', 'ache', 'macerated skin', 'heavy feeling',
      'rest pain', 'sinus rhythm', 'withdraw', 'behavior hyperactive',
      'terrify', 'photopsia', 'giddy mood', 'disturbed family',
      'hypersomnia', 'hyperhidrosis disorder', 'mydriasis',
      'extrapyramidal sign', 'loose associations', 'exhaustion', 'snore',
```

```

'r wave feature', 'overweight', 'systolic murmur', 'asymptomatic',
'splenomegaly', 'bleeding of vagina', 'macule', 'photophobia',
'painful swallowing', 'cachexia', 'hypocalcemia result',
'hypothermia, natural', 'atypia', 'general unsteadiness',
'throat sore', 'snuffle', 'hacking cough', 'stridor', 'paresis',
'aphagia', 'focal seizures', 'abnormal sensation', 'stupor',
'fremitus', "Stahli's line", 'stinging sensation', 'paralyse',
'hirsutism', 'sniffle', 'bradykinesia', 'out of breath',
'urge incontinence', 'vision blurred', 'room spinning',
'rambling speech', 'clumsiness', 'decreased stool caliber',
'hematochezia', 'egophony', 'scar tissue', 'neologism',
'decompensation', 'stool color yellow',
'rigor - temperature-associated observation', 'paraparesis',
'moody', 'fear of falling', 'spasm', 'hyperventilation',
'excruciating pain', 'gag', 'posturing', 'pulse absent',
'dysesthesia', 'polymyalgia', 'passed stones',
'qt interval prolonged', 'ataxia', "Heberden's node",
'hepatomegaly', 'sciatica', 'frothy sputum', 'mass in breast',
'retropulsion', 'estrogen use', 'hypersomnolence', 'underweight',
'dullness', 'red blotches', 'colic abdominal', 'hypokalemia',
'hunger', 'prostate tender', 'pain foot', 'urinary hesitation',
'disequilibrium', 'flushing', 'indifferent mood', 'urinoma',
'hypoalbuminemia', 'pustule', 'slowing of urinary stream',
'extreme exhaustion', 'no status change', 'breakthrough pain',
'pansystolic murmur', 'systolic ejection murmur', 'stuffy nose',
'barking cough', 'rapid shallow breathing', 'noisy respiration',
'nasal discharge present', 'frail', 'cystic lesion',
'projectile vomiting', 'heavy legs', 'titubation',
'dysdiadochokinesia', 'achalasia', 'side pain', 'monocytosis',
'posterior\rhinorrhea', 'incoherent', 'claudication',
'clammy skin', 'mediastinal shift', 'nausea and vomiting',
'awakening early', 'tenesmus', 'fecaluria', 'pneumatouria',
'todd paralysis', 'alcoholic withdrawal symptoms', 'myalgia',
'dyspareunia', 'poor dentition', 'floppy', 'inappropriate affect',
'poor feeding', 'moan', 'welt', 'tinnitus', 'hydropneumothorax',
'superimposition', 'feeling strange', 'uncoordination',
'absences finding', 'tonic seizures', 'debilitation',
'impaired cognition', 'drool', 'pin-point pupils',
'tremor resting', 'groggy', 'adverse effect', 'abdominal bloating',
'fatigability', 'para 2', 'abortion',
'intermenstrual heavy bleeding', 'previous pregnancies 2',
'primigravida', 'abnormally hard consistency', 'proteinemia',
'pain neck', 'dizzy spells', 'shooting pain', 'hyperemesis',
'milky', 'regurgitates after swallowing', 'lip smacking',
'phonophobia', 'rolling of eyes', 'ambidexterity',
'pulsus\paradoxus', 'gravida 10', 'bruit',
'breath-holding spell', 'scleral\rh\icterus', 'retch', 'blanch',
'elation', 'verbally abusive behavior', 'transsexual',
'behavior showing increased motor activity',
'coordination abnormal', 'choke', 'bowel sounds decreased',
'no known drug allergies', 'low back pain', 'charleyhorse',
'sedentary', 'feels hot/feverish', 'flare',
'pericardial friction rub', 'hoard', 'panic',
'cardiovascular event', 'soft tissue swelling', 'rhd positive',
'para 1', 'nasal flaring', 'sneeze', 'hypertonicity',
"Murphy's sign", 'flatulence', 'gasping for breath',
'feces in rectum', 'prodrome', 'hypoproteinemia',
'alcohol binge episode', 'abdomen acute', 'air fluid level',
'catching breath', 'large-for-dates fetus', 'immobile',
'homicidal thoughts'], dtype=object)

```

In []: