In [1]:

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import sklearn
```

In [2]:

```python
import glob
path="C:\S1_Dataset"
all_files=glob.glob(path+'/*')
print(all_files)
```

```
['C:\\S1_Dataset\\d1p01M', 'C:\\S1_Dataset\\d1p02M', 'C:\\S1_Dataset\\d1p03M', 'C:\\S1_Dataset\\d1p04M', 'C:\\S1_Dataset\\d1p05M', 'C:\\S1_Dataset\\d1p06M',
'C:\\S1_Dataset\\d1p07M', 'C:\\S1_Dataset\\d1p08F', 'C:\\S1_Dataset\\d1p09F', 'C:\\S1_Dataset\\d1p10F', 'C:\\S1_Dataset\\d1p11F', 'C:\\S1_Dataset\\d1p12F',
'C:\\S1_Dataset\\d1p13F', 'C:\\S1_Dataset\\d1p14F', 'C:\\S1_Dataset\\d1p15F', 'C:\\S1_Dataset\\d1p16F', 'C:\\S1_Dataset\\d1p17F', 'C:\\S1_Dataset\\d1p18F',
'C:\\S1_Dataset\\d1p19F', 'C:\\S1_Dataset\\d1p20F', 'C:\\S1_Dataset\\d1p21F', 'C:\\S1_Dataset\\d1p22F', 'C:\\S1_Dataset\\d1p23F', 'C:\\S1_Dataset\\d1p24F',
'C:\\S1_Dataset\\d1p25F', 'C:\\S1_Dataset\\d1p26F', 'C:\\S1_Dataset\\d1p27F', 'C:\\S1_Dataset\\d1p28F', 'C:\\S1_Dataset\\d1p29F', 'C:\\S1_Dataset\\d1p30F',
'C:\\S1_Dataset\\d1p31F', 'C:\\S1_Dataset\\d1p32F', 'C:\\S1_Dataset\\d1p33F', 'C:\\S1_Dataset\\d1p34F', 'C:\\S1_Dataset\\d1p35F', 'C:\\S1_Dataset\\d1p36M',
'C:\\S1_Dataset\\d1p37M', 'C:\\S1_Dataset\\d1p38M', 'C:\\S1_Dataset\\d1p39M', 'C:\\S1_Dataset\\d1p40M', 'C:\\S1_Dataset\\d1p41M', 'C:\\S1_Dataset\\d1p42M',
'C:\\S1_Dataset\\d1p43M', 'C:\\S1_Dataset\\d1p44M', 'C:\\S1_Dataset\\d1p45M', 'C:\\S1_Dataset\\d1p46M', 'C:\\S1_Dataset\\d1p47M', 'C:\\S1_Dataset\\d1p48M',
'C:\\S1_Dataset\\d1p49F', 'C:\\S1_Dataset\\d1p50F', 'C:\\S1_Dataset\\d1p51F', 'C:\\S1_Dataset\\d1p52F', 'C:\\S1_Dataset\\d1p53F', 'C:\\S1_Dataset\\d1p54F',
'C:\\S1_Dataset\\d1p55F', 'C:\\S1_Dataset\\d1p56F', 'C:\\S1_Dataset\\d1p57F', 'C:\\S1_Dataset\\d1p58F', 'C:\\S1_Dataset\\d1p59F', 'C:\\S1_Dataset\\d1p60F']
```

In [3]:

```python
li=[]
for file in all_files:
    #print(file)
    if file.endswith('.txt'):
        continue
    df=pd.read_csv(file,header=None,index_col=None)
    li.append(df)
healthy_op=pd.concat(li,axis=0,ignore_index=True)
```

In [4]:

```python
healthy_op.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52482 entries, 0 to 52481
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   0       52482 non-null  float64
 1   1       52482 non-null  float64
 2   2       52482 non-null  float64
 3   3       52482 non-null  float64
 4   4       52482 non-null  int64  
 5   5       52482 non-null  float64
 6   6       52482 non-null  float64
 7   7       52482 non-null  float64
 8   8       52482 non-null  int64  
dtypes: float64(7), int64(2)
memory usage: 3.6 MB
```

In [5]:

```python
healthy_op.columns=['Time','Acc.Front','Acc.vert','Acc.Lat','id','RSSI','Phase','Freq','Activity_Label']
healthy_op.head()
```

Out[5]:

	Time	Acc.Front	Acc.vert	Acc.Lat	id	RSSI	Phase	Freq	Activity_Label
0	0.00	0.27203	1.00820	-0.082102	1	-63.5	2.4252	924.25	1
1	0.50	0.27203	1.00820	-0.082102	1	-63.0	4.7369	921.75	1
2	1.50	0.44791	0.91636	-0.013684	1	-63.5	3.0311	923.75	1
3	1.75	0.44791	0.91636	-0.013684	1	-63.0	2.0371	921.25	1
4	2.50	0.34238	0.96229	-0.059296	1	-63.5	5.8920	920.25	1

In [6]:

```python
healthy_op.isnull().values.any()
```

Out[6]:

```
False
```

In [7]:

```python
cols=len(healthy_op.columns)-1
healthy_op1=healthy_op.values
X=healthy_op1[:, :8]
Y=healthy_op1[:,8]
```

In [8]:

```python
from sklearn.preprocessing import Normalizer
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split,GridSearchCV
from sklearn.metrics import accuracy_score
from sklearn.neighbors import KNeighborsClassifier
```

In [9]:

```python
normalize=Normalizer()
X=normalize.fit_transform(X)
print(X)
```

```
[[ 0.00000000e+00  2.93631563e-04  1.08825991e-03 ... -6.85424559e-02
   2.61778211e-03  9.97643541e-01]
 [ 5.41175984e-04  2.94432206e-04  1.09122725e-03 ... -6.81881739e-02
   5.12699303e-03  9.97657926e-01]
 [ 1.61998019e-03  4.83736884e-04  9.89656697e-04 ... -6.85791613e-02
   3.27354797e-03  9.97637799e-01]
 ...
 [ 4.64054952e-01  3.16054398e-04  9.08811718e-04 ... -5.78276511e-02
   4.78430619e-03  8.83902815e-01]
 [ 4.65829320e-01  1.14107909e-04  7.86629014e-04 ... -5.34284525e-02
   4.68329008e-03  8.83239105e-01]
 [ 4.67376010e-01  6.95187748e-05  1.26064817e-03 ... -5.40291449e-02
   9.91649969e-04  8.82396344e-01]]
```

In [10]:

```python
healthy_op.shape
```

Out[10]:

```
(52482, 9)
```

In [11]:

```python
train, test = train_test_split(healthy_op, test_size = 0.25)
print(train.shape)
print(test.shape)
```

```
(39361, 9)
(13121, 9)
```

In [12]:

```python
X_train = train[['Time','Acc.Front','Acc.vert','Acc.Lat','id','RSSI','Phase','Freq']]
Y_train =train.Activity_Label
X_test = test[['Time','Acc.Front','Acc.vert','Acc.Lat','id','RSSI','Phase','Freq']]
Y_test =test.Activity_Label
```

In [13]:

```python
X_train.head()
```

Out[13]:

	Time	Acc.Front	Acc.vert	Acc.Lat	id	RSSI	Phase	Freq
24690	110.50	1.16320	-0.059712	-0.093505	3	-59.0	4.16320	920.75
40747	343.45	0.40101	0.962290	-0.013684	1	-58.0	0.65348	924.25
502	82.25	1.10450	-0.266410	-0.150520	3	-56.5	1.07840	922.75
5953	348.00	1.24520	0.146980	0.054735	4	-58.0	4.88730	923.25
10519	451.75	1.26870	-0.151580	0.088944	3	-50.5	4.13560	922.75

In [14]:

```python
Y_train.head()
```

Out[14]:

```
24690    3
40747    1
502      3
5953     3
10519    3
Name: Activity_Label, dtype: int64
```

In [15]:

```python
#standardization
from sklearn.preprocessing import StandardScaler
X_train_stand = X_train.copy()
X_test_stand = X_test.copy()

num_cols = ['Time','Acc.Front','Acc.vert','Acc.Lat','id','RSSI','Phase','Freq']

for i in num_cols:
    scale = StandardScaler().fit(X_train_stand[[i]])
    X_train_stand[i] = scale.transform(X_train_stand[[i]])
    X_test_stand[i] = scale.transform(X_test_stand[[i]])
```

In [16]:

```python
# KNN
knn=KNeighborsClassifier(n_neighbors=3)
knn.fit(X_train_stand,Y_train)
y_pred_knn=knn.predict(X_test_stand)
print("accuracy KNN= ",accuracy_score(Y_test,y_pred_knn))
```

```
accuracy KNN=  0.987653380077738
```

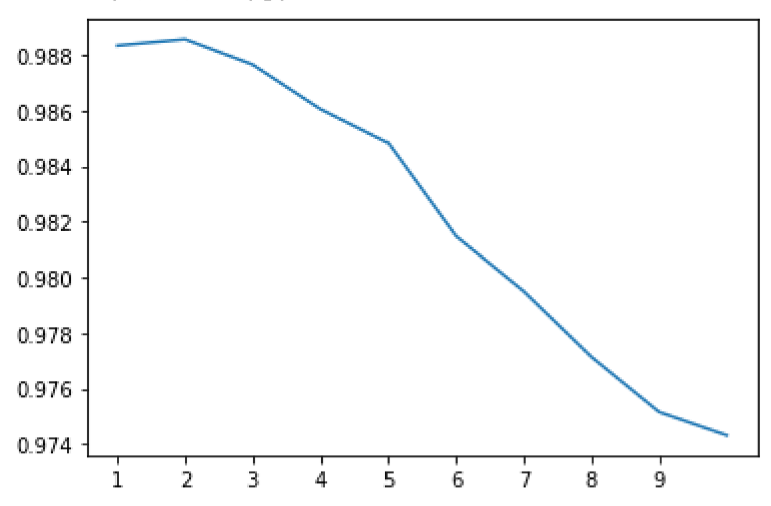
In [17]:

```python
a_index=list(range(1,11))
a=pd.Series()
x=[1,2,3,4,5,6,7,8,9]
for i in list(range(1,11)):
    model=KNeighborsClassifier(n_neighbors=i)
    model.fit(X_train_stand,Y_train)
    prediction=model.predict(X_test_stand)
    a=a.append(pd.Series(accuracy_score(prediction,Y_test)))
plt.plot(a_index, a)
plt.xticks(x)
```

```
C:\Users\Ganesh\AppData\Local\Temp/ipykernel_1568/4091055183.py:2: DeprecationWarning: The default dtype for empty Series will be 'object' instead of 'float6
4' in a future version. Specify a dtype explicitly to silence this warning.
  a=pd.Series()
```

Out[17]:

```
([<matplotlib.axis.XTick at 0x2c3cc4dcee0>,
  <matplotlib.axis.XTick at 0x2c3cc4dceb0>,
  <matplotlib.axis.XTick at 0x2c3cc4dc580>,
  <matplotlib.axis.XTick at 0x2c3cc760b50>,
  <matplotlib.axis.XTick at 0x2c3cc77a0d0>,
  <matplotlib.axis.XTick at 0x2c3cc77a760>,
  <matplotlib.axis.XTick at 0x2c3cc77aeb0>,
  <matplotlib.axis.XTick at 0x2c3cc781640>,
  <matplotlib.axis.XTick at 0x2c3cc781d90>],
 [Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, ''),
  Text(0, 0, '')])
```



In [18]:

```python
lr=LogisticRegression()
estimator={'solver':('newton-cg','liblinear','lbfgs','sag')}
gsc=GridSearchCV(lr,estimator)
gsc.fit(X_train_stand,Y_train)
y_gsc_pred=gsc.predict(X_test_stand)
print("accuracy gsc= ",accuracy_score(Y_test,y_gsc_pred))
print(gsc.best_estimator_)
```

```
C:\Users\Ganesh\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:763: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
C:\Users\Ganesh\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:763: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
C:\Users\Ganesh\anaconda3\lib\site-packages\sklearn\linear_model\_logistic.py:763: ConvergenceWarning: lbfgs failed to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression
  n_iter_i = _check_optimize_result(
accuracy gsc=  0.9019891776541422
LogisticRegression(solver='newton-cg')
```

In [19]:

```python
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
```

In [20]:

```python
model = svm.SVC() #select the algorithm
model.fit(X_train_stand,Y_train) # we train the algorithm with the training data and the training output
prediction=model.predict(X_test_stand) #now we pass the testing data to the trained algorithm
print('The accuracy of the SVM is:',accuracy_score(prediction,Y_test))#now we check the accuracy of the algorithm.
#we pass the predicted output by the model and the actual output
```

```
The accuracy of the SVM is: 0.9777455986586389
```

In [21]:

```python
rforest=RandomForestClassifier()
rforest.fit(X_train_stand,Y_train)
y_pred_rforest=rforest.predict(X_test_stand)
print("accuracy Random Forest= ",accuracy_score(Y_test,y_pred_rforest))
```

```
accuracy Random Forest=  0.991921347458273
```

In [ ]: