**1. MOTIVATION**

Twitter is a large social media channel where users tweet about various topic which also includes health issues. Traditional disease surveillance was done manually by selecting some target population and collecting their view about any particular disease. Social media channels, like Twitter, provides continuous information of public opinion about any epidemic and other health issues which can help public health agencies in performing real time surveillance.

**2. GAP ANALYSIS**

There are some limitations which we noted during data set collection from twitter and analysis of that dataset.

1. Since our script for data collection from Twitter does not understand and filter out sarcasm, the dataset of tweets also includes sarcastic tweets and it is not possible for current sentiment analysis algorithms to accurately classify sarcasm into sentiment polarity.

2. Since our study is limited to very less number of languages so it certainly limits the accuracy of our results.

3.In collected dataset we noted that there were many tweets which do not make any sense in case of our study however it contains the keyword which we used to fetch the tweets from twitter.

**3.OBJECTIVE**

Research Objective 1 - To examine level of concern on 'Zika virus' by analyzing sentiment polarity of tweets .

Research Question 1 - To what extent shared contents on twitter give legitimate information ?

Research Objective 2 - To find out what number of people are twitting about prevention, transmission, treatment, symptom, mosquito, and pregnancy.

Research Question 2 - Are tweets on 'Zika virus' relevant ?

**4.GOAL**

We are going to use the user generated content which are available on twitter to perform disease surviellance. The disease we are interested in is caused by the Zika virus. In this paper, we will analyse how people reacted to Zika virus on twitter and what extent that information can be used for surveillance .