Predicting Emergency Department Visits for Asthma-Related Issues using Online Networking

Priydarshini.M.R[1] Chandrakala.H.L[2]
[1]M.Tech. Student [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]HKBK College of Engineering Bangalore, India

Abstract— Asthma is a chronic incendiary issue of the aviation routes described by an airflow obstruction which cannot be completely cured, but appropriate measures could be taken to avoid or to control One such solution to treat the patients affected by asthma would be to send an Emergency Department. By doing so the patients could get treatment when needed and can be directed towards not so much excessive but rather more effective consideration locales. The main aim here is to predict the Emergency Department visits for asthma-related issues at constant or close ongoing using online networking such as Twitter information, Google search interests and Environment sensor data. Given these three information as input accurate results of the location where many people are affected can be obtained.

Key words: Asthma, Emergency Department, Twitter Information, Google Seek Interests, Environment Sensor Information

I. INTRODUCTION

Asthma, it is an interminable incendiary issue of the aviation routes which is characterized by airflow obstruction, which can affect people of all ages. The rate of asthma diagnosis is in the rise. Nations for example UK, USA, and Ireland are the worst asthma affected countries. People would require quick treatment to control the disease. Many actions have been taken so far to reach out people to help them. One such effort is Emergency Department. Statistics say that asthma represents for about half million hospitalization, about two to three million Emergency Departments. Still more that 3000 deaths occur annually in each country. There's loss of productivity due to missed school, office etc. The symptoms of asthma include inflammation of the bronchial airways resulting in wheezing, coughing, chest tightness and breathlessness. Recognition of the signs and symptoms of an asthma attack are vital to diminishing high bleakness and death rates as well as improving the quality of life for those diagnosed with asthma.

The problem with the existing system is response time. Surveys are conducted and the report is sent to Centers for Disease Control and Prevention (CDC), the report would take about two weeks for it to be delivered. To overcome this issue an application which could help us predict the Emergency Department visits at ongoing or close constant is required. This should also help us prioritize issues. One such solution could be social media. Twitter information, Google seek interests and Environment sensor information can provide accurate real-time result of the location where many people are affected by asthma. This could help the Emergency Department be prepared with appropriate equipments and staff.

The existing system includes past intense consideration use, drug use and sociodemoraphic attributes, basic information hotspots for these are Electronic Therapeutic records (EMR), population surveys and also medical insurance claims data.

Twitter is among the popular online social media which people use to post their messages called "tweets". These tweets could help in finding location where many people are affected as the tweets are regularly labeled by geographic area and time stamps. These data would be of great help for disease surveillance.

Google Search Trends, this is another way of obtaining required data. Google search trends helps to obtain results of how many people are making their search with the keyword "asthma" across a particular area.

Air pollution details of a particular area can be obtained from Environment Protection Agency (EPA) which would provide the details of the pollutants such as carbon monoxide, nitrogen oxide, etc.

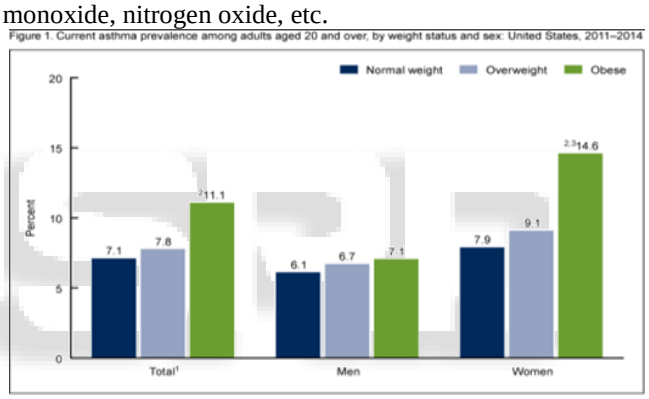

Fig. 1: Graph

The figure shows the asthma prevalence among adults aged 20 and over by weight status and sex.

II. ARCHITECTURE

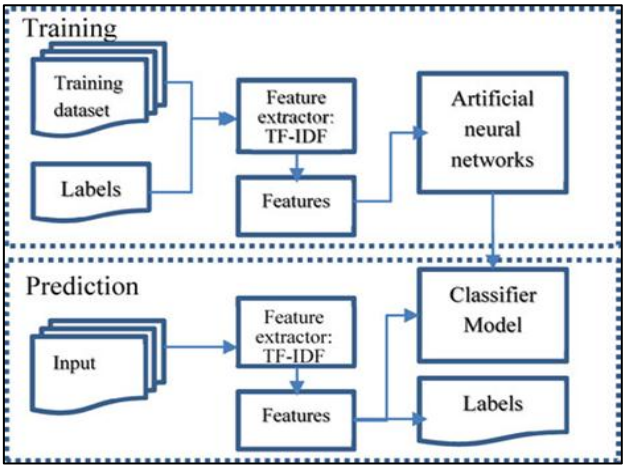

Fig. 1: Architecture

Two diverse Twitter datasets are gathered in this study. The general twitter stream: A basic collection of JSON snatched from the general Twitter stream. The general tweet numbers are utilized to gauge the Twitter populace and standardize

asthma tweet tallies and the asthma-related stream: to gather just tweets containing any of 19 related catchphrases that were proposed by the clinical associates from PCCI.

Once the twitter dataset is gathered, both the twitter dataset and the labels which include set of most common keywords related to asthma are sent to Feature extractor: TF-IDF. In Feature extractor we use Senti word algorithm to remove all non English tweets and also remove tweets containing special characters such as targets (@). After this, each tweet with the keyword is assigned tokens and this result is forwarded to Features, where the tweets are compared with the labels and separated into asthma relevant and asthma irrelevant treats. Now this result is sent to ANN to accurately identify relevant tweets.

In the view of the outcomes from relationship investigation, asthma tweets, CO, NO2 and PM2.5 are chosen as information into the expectation model. These inputs will be forwarded to the Feature extractor: TF-IDF where these inputs are assigned tokens and forwarded to Features where relation between the inputs are analyzed to find the most affected location by rating them as High, Medium and Low. This result is passed to both Labels and Classifier Model where Label stores this information for future purposes and Classifier model displays the most accurate result.

III. BACKGROUND

Some early studies suggest that Twitter information can be pertinent for early recognition of general wellbeing dangers. Examines demonstrate that Twitter can help wellbeing authorities track the spread of flu and online networking can battle interminable ailment. The eHealth activity report on online networking noticed that message sheets, sites, microblogs and interpersonal interaction locales enhance access to wellbeing data and give another channel to distributed correspondence among social insurance suppliers, purchasers and relatives. Air contamination constitutes a noteworthy boost of asthma intensifications. Studies show that increased convergences of ground-level ozone and particulates have been appeared to be connected with expanded rate of asthma intensifications. Aeroallergens, for example, dust and shape, may trigger hypersensitivity prompted asthma side effects. Seasons for different aeroallergens vary by topography. Higher temperatures may likewise be connected with asthma intensifications since they happen when there is more daylight, and daylight is important for outflows to be changed over to ground-level ozone.

An Environmental Public Health Tracking Program (EPHTP) was established by the region of Columbia Department of Health to show the connections between encompassing ozone and particulates and fleeting asthma wellbeing results furthermore to distinguish regions and populaces destined to be influenced by contamination. The objectives was to decide the extent to which asthma intensifications are connected with ozone and particulate focuses in the short-term, on the request of days after assumed presentation, and to recognize pediatric populaces that might be at expanded danger of these wellbeing impacts. Asthma intensifications were controlled by every day tallies of pediatric ED visits for asthma-related issues. Other spatial components were analyzed that show more huge provincial variety, including habitation postal district and financial status of the pediatric populace. Some intriguing results were gotten, particularly concerning financial status.

Google Trends, an openly available instrument that permits clients to associate with Internet seek information, which may give profound experiences into wellbeing related wonders and populace conduct. Social insurance writing was checked on utilizing Google Trends to order articles by point and study point; assess the procedure and acceptance of the apparatus; and location restrictions for its utilization in examination.

IV. METHODS

Social media is used to obtain accurate results about the location that is most affected. This is achieved by using Twitter dataset, where number of tweets from a particular location is analyzed. Also Google search interest where searches are made from the same location with the asthma keywords is analyzed. Prepared air quality information from sensors and historical emergency department room visits is included from the same location and same time period.

A. Information Collection and Processing:

1) ED Information:

Information based on the historic Emergency Department room visits are collected for a particular time period and are analyzed for a specific geographic location.

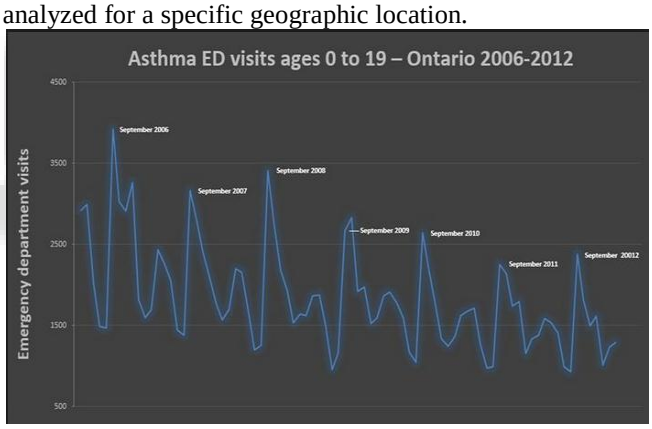

Fig. 1: This figure shows the number of Emergency Department visits for asthma related issue from year 2006 to 2012 for the location Ontario.

2) Twitter Data:

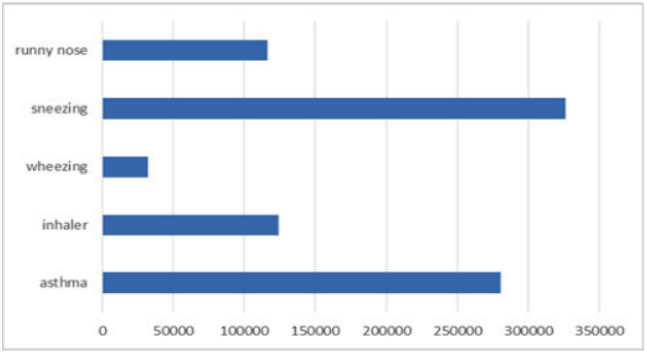

Fig. 2: This figure shows the keywords and frequency of tweets

Large volume of Tweets is downloaded from the Twitter API. Tweets with keyword "asthma" or any term related to asthma such as "inhaler" and "nebulizer"; and names of prescription drugs used to treat the condition, including

"albuterol" and "Singulair" are extracted for the noisy twitter dataset.

3) Google Data:

Google data is collected from Google search trends and is used to analyze the frequency of a particular searched term given a specific time and location. This specific searched term is based on the twitter keywords.

4) Sensor Data:

Air contamination information which contains six sorts of poisons is gathered. The six sorts of toxins are specific matter, ground-level ozone, carbon monoxide, sulfur oxides, lead and nitrogen oxides. The air quality records (AQI) connected with these contaminations is utilized. The higher the AQI esteem, the more conspicuous the level of air contamination and the more noteworthy the wellbeing concern. With AQI Site ID can be acquired i.e., the particular physical area and location.

B. Prediction Model:

In the first place the relationship between the asthma-related ED visits and information from Twitter, Google patterns, and Air Quality Sensors are investigated utilizing the Pearson connection coefficient. Likewise the relationship between asthma related tweet tallies and ED visit means stomach torment/stoppage patients is inspected. At that point, a forecast model is planned and actualized to gauge the occurrence of asthma ED visits utilizing a mix of free variables from the previously stated information sources. Since each dataset is from an alternate source and has diverse levels of granularity as for time and area, a few changes is performed on each dataset to make them good. A critical change is to standardize each dataset utilizing a standard standardization system, i.e., z-score.

$$z = \frac{x - u}{\sigma}$$

Where u is the mean and σ is the standard deviation.

In addition to this, Twitter information is standardized by figuring the proportion of asthma-related tweets to aggregate number of tweets in the general twitter stream gathered from the same geo-area in a given time. For the expectation model, four distinctive characterization techniques are utilized: Decision tree, Naïve Bayes, SVM and ANN, and their grouping precision are looked at. Additionally systems called versatile boosting and stacking are utilized to lessen arrangement mistakes. The ED visit checks are changed over from numerical to downright values in view of the z-values, where the perceptions are marked as "High", "Medium", or "Low". This model is utilized to characterize the anticipated variable, i.e., number of every day ED visits, into one of three reciprocal and totally unrelated classes- High, Medium or Low. The Naïve Bayes procedure requires ostensible information; consequently, another change was utilized to change over every single numerical data values into straight out qualities in view of the z values like the change utilized for ED visit checks.

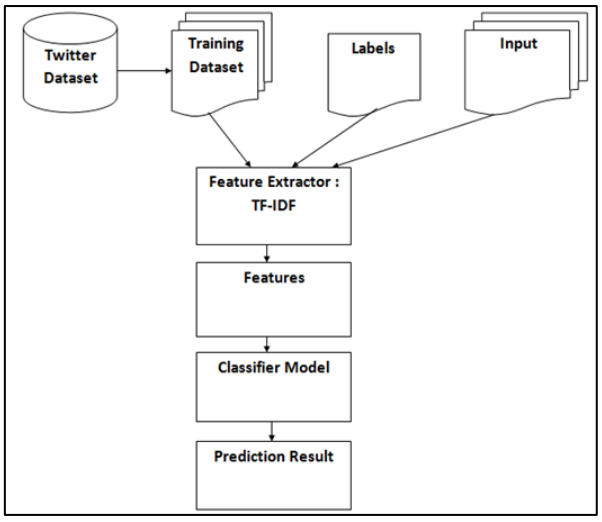

Fig. 3: Proposed Architecture

V. CONCLUSION

In this study we could see how online networking, historical ED data and environmental data can be used to obtain accurate result to predict ED visits for a particular location. The result obtained is real-time or near real-time therefore appropriate help can be provided when needed. And the results can be used for future purposes.

REFERENCES

[1] Big Data for Health Javier Andreu-Perez, Robert D.Merrifield, Carmen C. Y. Poon, Stephen T.C. Wong, and Guang-Zhong Yang, Fellow, IEEE

[2] L.J.Akinbami, X.Liu and J.E.Moorman, "Asthma prevalence, health care use, and mortality: United States, 2005-2009", National Center for Health Statistics, Hyattsville, MD,USA, National health statistics report no. 32, 2011.

[3] "Vital signs: Asthma prevalence, disease characteristics, and self-management education: United States, 2001-2009", Centers for Disease Control and Prevention Atlanta, GAm USA, mortality and morbidity weekly report, vol. 60, no. 17, pp. 547, 2011

[4] "Guidelines for the Diagnosis and Management of Asthma", National Institutes of Health, Bethesda, MD, USA, Expert Panel Report 3. Vol.2 , 1997.

[5] "Predicting Asthma-Related Emergency Department Visits Using Big Data", Sudha Ram, Member, IEEE, Wenli Zang, Max Williams, and Yolande Pengetnze.

[6] "Big Data for Precision Medicine" By Daniel Richard, Leff and Guang-Zhon Yang*

[7] How Twitter can help predict emergency room visits, 15, April 2015, by Alexis Blue.

[8] E.K.Kim, J.H.Seok, J.S.Oh, H.W.Lee and K.H.Kim, "Utilization of Hangeoul twitter to track and anticipate human flue contamination". Plos one, vol 8, no.7, p.E69305, 2013

