



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

## A Literature survey on Pulmonology Based Disorders Diseases Using Big Data Analysis

N.Sudha, P.Muthukumar

Associate Professor, Department of Computer Science, Bishop Appasamy College, Coimbatore, Tamilnadu, India

M.Phil, Research Scholar, Department of Computer Science Bishop Appasamy College, Coimbatore, Tamilnadu, India

**ABSTRACT:** The paper focus on comparison study of various researches based on lungs based disorders diseases. The base researcher focused on big data analysis of asthma based emergency visit. The previous research cannot cover all the lungs based disorders, so we focus on analyzing the researches made and proposing a survey on other lung based disorders like lung cancer, and pulmonology problems. A detailed big data analysis of various respiratory diseases existing in medical field. The researches shows that future big data analysis of lung based disorder has lot of scope towards global community.

**KEYWORDS:** Big data, image synthesis method, fuzzier synthetic images.

### I.INTRODUCTION

The comparison study of various researches based on lungs based disorders diseases namely asthma, cold, sneezing and so on. The researcher focused data analysis based emergency visit. The emergency visit data use on the researchers and the searching person to evaluate the number of affected person in only on the visited data. The proposed methodology deal with other lung related diseases in this paper, so we focal point on analyzing the research made and proposing a review on other lung based disorders similar to lung malignancy, and pulmonology troubles. The researches shows that potential large statistics analysis of lung based disorder has lot of capacity towards worldwide group of people.

### II.BACKGROUN WORK

The National ambulant medical aid Survey the National Hospital Discharge Survey, and also the National statistic System were wont to calculate national estimates . The systematic review covered a wide range of topics. Although the overarching framework for the review was based on the four essential components of asthma care, multiple subtopics were associated with each component .The research work mainly focused on knowing the major motive for commencement and indulging the youth in drugs. This was a prospective observational study that evaluated baseline predictors of asthma morbidity in adults in an urban, predominantly African American community in New York City .This paper summarizes our recently developed influenza infection detection algorithm that automatically distinguishes relevant tweets from other chatter, and we describe our current influenza surveillance system which was actively deployed during the full 2012-2013 influenza season. This study reviewed Twitter status updates mentioning "antibiotic(s)" to determine overarching categories and explore evidence of misunderstanding or misuse of antibiotics.

### III. LITERATURE REVIEW

1. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M.S., & Brilliant, L.

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza. One way to improve early detection is to monitor health-seeking behavior in the form of queries to online search engines, which are submitted by millions of users around the world



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

each day. Here we present a method of analyzing large numbers of Google search queries to track influenza-like illness in a population.

2. Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H.

Google Flu Trends (GFT) uses anonymized, aggregated internet search activity to provide near-real time estimates of influenza activity. GFT estimates have shown a strong correlation with official influenza surveillance data. The 2009 influenza virus A (H1N1) pandemic [pH1N1] provided the first opportunity to evaluate GFT during a non-seasonal influenza outbreak. In September 2009, an updated United States GFT model was developed using data from the beginning of pH1N1. Internet search behavior changed during pH1N1, particularly in the categories "influenza complications" and "term for influenza." The complications associated with pH1N1, the fact that pH1N1 began in the summer rather than winter, and changes in health-seeking behavior each may have played a part. Both GFT models performed well prior to and during pH1N1, although the updated model performed better during pH1N1, especially during the summer months.

3. Trasande, Leonardo, and George D. Thurston.

A growing body of research supports the role of outdoor air pollutants in acutely aggravating chronic diseases in children, and suggests that the pollutants may have a role in the development of these diseases. This article reviews the biologic basis of children's unique vulnerability to highly prevalent outdoor air pollutants, with a special focus on ozone, respirable particulate matter (PM 2.5 [ $<2.5$  microm in diameter] and PM 10 [ $<10$  microm in diameter]), lead, sulfur dioxide, carbon monoxide, and nitrogen oxides. We also summarize understanding regarding health effects and molecular mechanisms of action. Practitioners can significantly reduce morbidity in children and other vulnerable populations by advising families to minimize pollutant exposures to children with asthma, or at a broader level by educating policymakers about the need to act to reduce pollutant emissions. Management of children with asthma must expand beyond preventing exposures to agents that directly cause allergic reactions (and therefore can be diagnosed by means of skin tests) and must focus more attention on agents that cause a broad spectrum of nonspecific, generalized inflammation, such as air pollution.

4. Bates, David W., SuchiSaria, LucilaOhno-Machado, Anand Shah, and Gabriel Escobar.

The US health care system is rapidly adopting electronic health records, which will dramatically increase the quantity of clinical data that are available electronically. Simultaneously, rapid progress has been made in clinical analytics-techniques for analyzing large quantities of data and gleaning new insights from that analysis-which is part of what is known as big data. As a result, there are unprecedented opportunities to use big data to reduce the costs of health care in the United States. We present six use cases-that is, key examples-where some of the clearest opportunities exist to reduce costs through the use of big data: high-cost patients, readmissions, triage, decomposition (when a patient's condition worsens), adverse events, and treatment optimization for diseases affecting multiple organ systems. We discuss the types of insights that are likely to emerge from clinical analytics, the types of data needed to obtain such insights, and the infrastructure-analytics, algorithms, registries, assessment scores, monitoring devices, and so forth-that organizations will need to perform the necessary analyses and to implement changes that will improve care while reducing costs. Our findings have policy implications for regulatory oversight, ways to address privacy concerns, and the support of research on analytics

5. Culotta, Aron.

Rapid response to a health epidemic is critical to reduce loss of life. Existing methods mostly rely on expensive surveys of hospitals across the country, typically with lag times of one to two weeks for influenza reporting, and even longer for less common diseases. In response, there have been several recently proposed solutions to estimate a population's health from Internet activity, most notably Google's Flu Trends service, which correlates search term frequency with influenza statistics reported by the Centers for Disease Control and Prevention (CDC). In this paper, we analyze messages posted on the micro-blogging site Twitter.com to determine if a similar correlation can be uncovered. We propose several methods to identify influenza-related messages and compare a number of regression models to correlate these messages with CDC statistics. Using over 500,000 messages spanning 10 weeks, we find that our best model achieves a correlation of .78 with CDC statistics by leveraging a document classifier to identify relevant messages.

6. Paul, Michael J., and Mark Dredze.

Analyzing user messages in social media can measure different population characteristics, including public health measures. For example, recent work has correlated Twitter messages with influenza rates in the United States; but this has largely been the extent of mining Twitter for public health. In this work, we consider a broader range of public health applications for Twitter. We apply the recently introduced Ailment Topic Aspect Model to over one and a



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016

half million health related tweets and discover mentions of over a dozen ailments, including allergies, obesity and insomnia. We introduce extensions to incorporate prior knowledge into this model and apply it to several tasks: tracking illnesses over times (syndromic surveillance), measuring behavioral risk factors, localizing illnesses by geographic region, and analyzing symptoms and medication usage. We show quantitative correlations with public health data and qualitative evaluations of model output. Our results suggest that Twitter has broad applicability for public health research.

7. Collier, Nigel, Nguyen Truong Son, and Ngoc Mai Nguyen.

Micro-blogging services such as Twitter offer the potential to crowd source epidemics in real-time. However, Twitter posts ('tweets') are often ambiguous and reactive to media trends. In order to ground user messages in epidemic response we focused on tracking reports of self-protective behavior such as avoiding public gatherings or increased sanitation as the basis for further risk analysis. We created guidelines for tagging self protective behaviour based on Jones and Salathé (2009)'s behaviour response survey. Applying the guidelines to a corpus of 5283 Twitter messages related to influenza like illness showed a high level of inter-annotator agreement (kappa 0.86). We employed supervised learning using unigrams, bigrams and regular expressions as features with two supervised classifiers (SVM and Naive Bayes) to classify tweets into 4 self-reported protective behaviour categories plus a self-reported diagnosis. In addition to classification performance we report moderately strong Spearman's Rho correlation by comparing classifier output against WHO/NREVSS laboratory data for A(H1N1) in the USA during the 2009-2010 influenza season. The study adds to evidence supporting a high degree of correlation between pre-diagnostic social media signals and diagnostic influenza case data, pointing the way towards low cost sensor networks. We believe that the signals we have modelled may be applicable to a wide range of diseases.

## IV. EXPERIMENTAL RESULTS

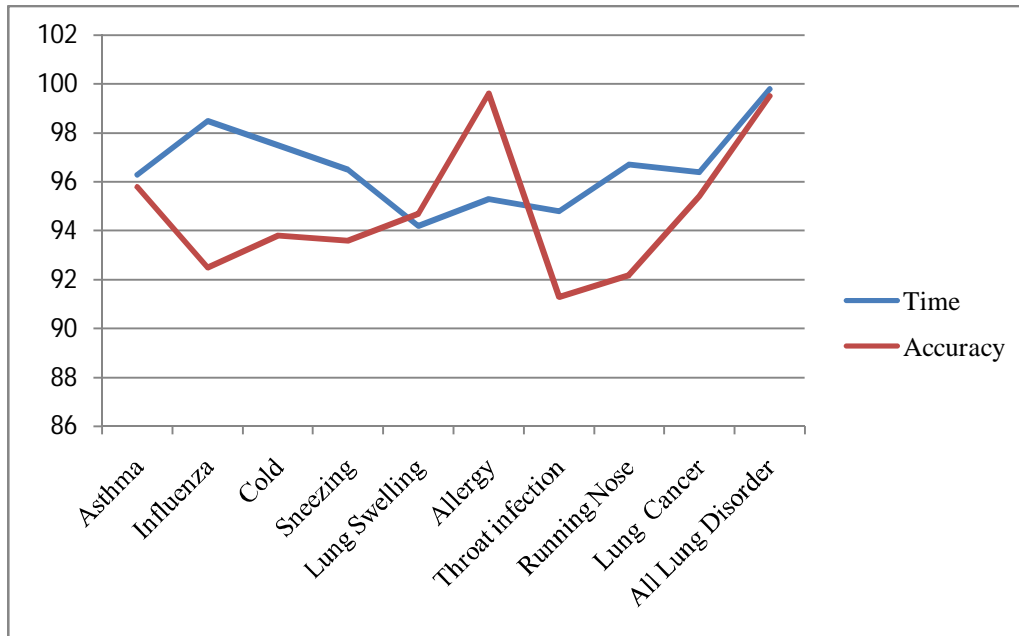
| Disorder                 | Time        | Accuracy    |
|--------------------------|-------------|-------------|
| Asthma                   | 96.3        | 90.5        |
| Influenza                | 98.5        | 91.5        |
| Cold                     | 97.5        | 92.5        |
| Sneezing                 | 96.5        | 93.5        |
| Lung Swelling            | 94.2        | 94.5        |
| Allergy                  | 95.3        | 95.5        |
| Throat infection         | 94.8        | 96.5        |
| Running Nose             | 96.7        | 97.5        |
| Lung Cancer              | 96.4        | 98.5        |
| <b>All Lung Disorder</b> | <b>99.8</b> | <b>99.5</b> |

In the experiments presented in this paper, the iterative nature of the proposed image synthesis method allows to reach higher accuracy values and synthesize qualitatively more realistic images. The optimal number of iterations seems to be about 2 or 3. A progressive increase of the feedback weight with the number of iterations, as suggested in leads to better synthesis results. Automatically setting the value for the feedback weight would require to consider the weight as an additional random variable together with an hyper prior which depends on the amount of spatial correlation of  $b$ . Several MR channels are simultaneously synthesized, which leads to a more constrained patch match (for iterations  $t > 1$ ), and ultimately to more consistent synthetic MRI.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 10, October 2016



However, using a higher number of MR sequences implies that patches lie in a space of higher dimension, which tends to homogenize patch distances. Indeed, for high dimensional data, every patch is more or less a close neighbor of others, and the notion of nearest-neighborhood becomes less relevant. This could result in fuzzier synthetic images. Conversely, less consistent, yet potentially more textured, results could arise when trying to synthesize a single image modality.

## VI. CONCLUSION

The paper concentrates on a comparison study of assorted researches supported by lung-based mostly disorders. The bottom investigator targeted on massive knowledge analysis of respiratory disorder based mostly on emergency visits. The previous analysis cannot cover all the lung-based mostly disorders, therefore we have a tendency to concentrate on analyzing the researches created and proposing a survey on alternative respiratory organ-based mostly disorders like carcinoma, and pulmonology issues. An in-depth massive knowledge analysis of assorted metabolic process diseases existing in the medical field. The researches show that future massive knowledge analysis of respiratory organ-based mostly disorder has a heap of scope towards world community.

## REFERENCES

- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M.S., & Brilliant, L. "Detecting influenza epidemics using search engine query data." *Nature* vol. 457, no.7232, pp. 1012-1014, 2008.
- Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. "Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic." *PloS one* vol.6, no.8, e23610, 2011.
- Trasande, Leonardo, and George D. Thurston. "The role of air pollution in asthma and other pediatric morbidities." *Journal of allergy and clinical immunology* vol. 115, no.4, pp. 689-699, 2005.
- Bates, David W., Suchi Saria, Lucila Ohno-Machado, Anand Shah, and Gabriel Escobar. "Big data in health care: using analytics to identify and manage high-risk and high-cost patients." *Health Affairs* vol. 33, no. 7 pp.1123-1131, 2014.
- Culotta, Aron. "Towards detecting influenza epidemics by analyzing Twitter messages." In *Proceedings of the first workshop on social media analytics*, pp. 115-122. ACM, 2010.
- Paul, Michael J., and Mark Dredze. "You are what you Tweet: Analyzing Twitter for public health." In *ICWSM*, pp. 265-272. 2011.
- Collier, Nigel, Nguyen Truong Son, and Ngoc Mai Nguyen. "OMG U got flu? Analysis of shared health messages for bio-surveillance." *Journal of Biomedical Semantics* 2, no.S-5, S9, 2011.
- Chew, Cynthia, and Gunther Eysenbach. "Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak." *PloS one* 5, no. 11, e14118, 2010.



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 10, October 2016**

9. Scanfled, Daniel, Vanessa Scanfled, and Elaine L. Larson."Dissemination of health information through social networks: Twitterand antibiotics." American journal of infection control vol. 38, no. 3, pp.182-188, 2010.
10. Myslín, Mark, Shu-Hong Zhu, Wendy Chapman, and Mike Conway."Using Twitter to examine smoking behavior and perceptions ofemerging tobacco products."Journal of medical Internet research vol. 15,no. 8, 2013.
11. Pesola, Gene R., FengXu, HabibulAhsan, Pamela Sternfels, Ilan H.Meyer, and Jean G. Ford. "Predicting asthma morbidity in Harlememergency department patients."Academic emergency medicine vol.11,no.9, pp. 944-950, 2004.
12. K. Arutchelvan\* Dr. PonPeriasamy, " Prognosis of Lung Cancer Using Data Mining Techniques" International Journal of Advanced Research in Computer Science and Software Engineering,Volume 6, Issue 3, March 2016