ISSN 2278-3091

Volume 8, No.3, May - June 2019

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse86832019.pdf

https://doi.org/10.30534/ijatcse/2019/86832019

A survey on prediction approaches for epidemic disease outbreaks based on social media data

S. Ravi Kumar[1], Dr. M. Vamsi Krishna[2], Dr. Anurag[3]

[1]Research Scholar, Department of Computer Science and Engineering, Centurion University of Technology and Management, Paralakhemundi Odisha, India, *Email: ravikumarsuggala1113@gmail.com

[2]Professor & Principal, Department of Computer science, Chaitanya Institute of Science & Technology, Kakinada, India

[3]Professor and Principal, School of Computing Skills, Bhartiya Skill Development University, Mahindra World City, Jaipur, India

ABSTRACT

Currently, in order to recognize the Epidemic diseases in the earlier stage, social media and search queries are the most reliable source of data to mine and obtain new information about current activities. These interactive groups play a significant role where users give data of their likings and acquaintances. This data can be utilized to quantify the impact of thoughts and the general public conclusions continuously, being helpful on a few fields and research regions, for example, advertising efforts, budgetary expectation or open Medicare applications among others. Recently there are many techniques and approaches were handled by the researchers for monitoring the web data sources in order to identify the health measures of the public has developed as another important imperative so-called Epidemic Intellect. For early discovery of disease outbreaks in order to decrease the effect of epidemics by monitoring, several intelligence systems were generally used by the health care organizations. In this paper, the literature review is based on the different prediction methods and their approaches to extract web data from social medias like Facebook and twitter which are all gained much attention on early disease detection for public wellbeing. Among 60 papers related to prediction of epidemic outbreaks using social media data are reviewed and analyzed based on various diseases and prediction approaches.

Key words: Disease outbreak, Epidemic diseases, healthcare applications, social media.

1. INTRODUCTION

In the most recent couple of decades the world wherein we live has changed quickly in the course of Dangers of bioterrorism, flu pandemics, and rising irresistible ailments combined with phenomenal populace versatility prompted the advancement of observation frameworks for public wellbeing [1]. Research studies demonstrate that social media might be profitable devices in the illness reconnaissance toolbox utilized for improving general wellbeing experts' capacity to recognize disease occurrences quicker than conventional techniques and to upgrade flare-up reaction [2]. Conventional perception frameworks generally rely upon case reports, for example, flu like disease chart have the delay in time for informing the details of the patients [3]. To empower the previous recognition of irresistible infection flare-ups, a syndromic reconnaissance framework ought to use continuous or close constant information, i.e., school or work non-attendance, over the counter prescription deals, or the volume of Internet based on the health inquiries [4]. Among various elective information sources, social media, search queries, and sites are visited, and have demonstrated possible for computerized observation frameworks [5].

The Control and Prevention (CDC) and the WHO are the Center for Diseases and have teamed up with in excess of 30 nations to reinforce wellbeing frameworks and address preparing requirements for disease recognition and reaction in a nation explicit, adaptable, and supportable way [6]. In order to provide genuine and well-timed data to decision makers an efficient public health observation system must be implemented. The utility of the information gathered can be seen as quick, yearly or documented, based on the moves that can be made [7]. Additionally, four-dimensional goals of the information gathered from full scale versus miniaturized scale territories might be yielded with the objective to improve sensibleness and protect assets [8]. Thus, it is neither constantly conceivable nor compelling to send complex investigation frameworks. The basic investigation of the health organization is to guarantee quality and adequacy of the investigation in disease spread out locations in most of the developing countries [9]. National-level projects and inspection framework supervisors may lose control of the quality and sensibleness of the information gathered and benefactors, seeing inadequacy in the national framework, may make parallel nongovernmental observation frameworks to assemble legitimately to gather the information they need [10]. These frameworks for the most part work for the time being, however over the long haul, hinder considerably more the general health observation [11].

All cooperative prediction is introduced with an alternate arbitrarily drawn example of unique information to prepare a relapse model and steadily includes increasingly various types of models to improve the capacity to conjecture the future direction of illness plague by changing the model parameters [12]. Likewise, that a gauge model ought to be enhanced under a few execution estimates at the same time so as to give increasingly vigorous forecasts of things to come unfurling direction of a nearby infection outbreak [13]. In order to accomplish the objective of goal, researchers make use of various approaches and multi-objective optimization algorithm to optimize the disease prediction approach. This paper reports a survey on various prediction approaches and the techniques used by the researchers in recent years. The construction of the remainder of this proposed method was composed as pursues: section 2 depicts the inspiration behind this survey paper on epidemic disease outbreak. Section 3 illustrates about the preliminary phases and various techniques and approaches used for prediction of epidemic disease outbreak were delineated in section 4. Section 5 depicts the various social media applications used for monitoring and predicting epidemic outbreaks. Section 6 presents the discussion and concludes the paper in section 7.

2. INSPIRATION AND CHALLENGES BEHIND THIS SURVEY

Irresistible infections place an unsuitable and unbalanced social and financial problem on low-pay nations. National illness control projects have the troublesome undertaking of assigning constrained spending plans for intercessions crosswise over areas of their nations, in view of regularly different datasets of shifting quality from a scope of sources including centers, emergency clinics, town wellbeing specialists, the private division and non-government organization (NGOs). Each phase of the information gathering and investigation for reconnaissance frameworks might be influenced by an absence of limit just as by inclinations and skewed motivating forces for revealing and controlling information. Tending to these issues will be basic for successful decrease in the burden of endemic irresistible sicknesses universally just as to getting ready for developing epidemic dangers. In the meantime, scholarly experts are growing progressively refined strategies to gather and analyze information to improve spatial appraisals of infection burden by utilizing new Big Data sources, portable Health or m-Health approaches or robotic and measurable displaying methods. While these advances jump ahead, in any case, many stay most valuable for evaluating worldwide sickness dispersion, instead of for national control program prioritization. Making an interpretation of these new strategies to educate approach in endemic settings stays difficult. The articulated disengage between health frameworks and the scholarly world may restrict the utility of new methodologies. The high burden of work put on medicinal services specialists in low-salary settings further restrains their degree and time accessible for commitment with methodological improvements.

Regardless of continuous difficulties to execution, in any case, there are no promising logical methodologies that can use even inconsistent and low-quality information. Therefore we need an assorted new information streams that can be gainfully saddled to fortify procedures for asset designation when coordinated with existing reconnaissance frameworks. The information and examination difficulties looked by national disease control programs, diagram potential arrangements offered by investigative methodologies and new information streams and finish up by laying out obstructions to execution. By and large, epidemiological information about patients are accounted for by medicinal services professionals by means of inactive reconnaissance frameworks to a focal database, which is utilized to decide slants after some time in and map the geographic circulation of burden of infection in various districts just as the dynamic observation by means of sentinel locales may likewise inform these activities. This provincial information thus serves as a significant reason for resource allocation decisions.

Figure 1 delineates the progression of information by national control programs. Information moves through health frameworks and real difficulties looked by control programs. A subset of Reporting incentives, which regularly speak to just a subset of absolute contaminations both asymptomatic and clinical, are first identified by neighborhood health blue-collar worker, most normally in health offices and medical clinics. Neighborhood health specialists are additionally in charge of lining up people with endless diseases requiring numerous medicines over months or years. Some part of clinical cases are lab affirmed, contingent upon limit, and answered to provincial or locale focuses, which thusly report to national control programs. Information is frequently accumulated before being accounted for regional officials. NGOs and the private organization may likewise create a lot of epidemiological information. National control projects total and break down information to delineate circulation of ailment trouble, mediation adequacy, etc. New immediate mHealth methodologies and latently gathered facts from cell phones by means of Call Data Records (CDRs); and satellites might be utilized legitimately by control projects to delineate dangers and populace dispersions. At each dimension, there are huge issues for routine observation, and preparing for new methodologies will challenge for most control programs. At various dimensions of the health framework, responsible persons for reporting precisely might be skewed, and practicality of revealing might be especially hazardous for developing disease threats on non-administrative associations. The above crises and lack of information of epidemic disease outbreak on timing due to various reasons motivated us to do the survey on this aspect.

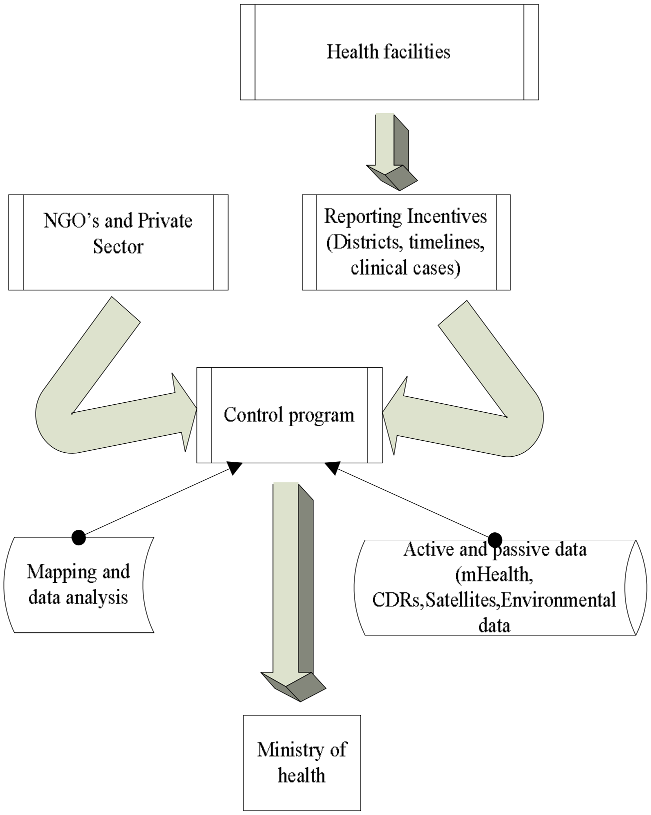
Health facilities
NGO's and Private Sector
Reporting Incentives (Districts, timelines, clinical cases)
Control program
Mapping and data analysis
Active and passive data (mHealth, CDRs,Satellites,Environmental data
Ministry of health

Figure 1: Data progression by national control programs

3. PRELIMINARY PHASES

3.1 Epidemic potential evaluation

In last 2001, the epidemic is alluded to that the occurrence in a system or area of examples of a disease, unequivocal prosperity related direct or other prosperity related events evidently in wealth of common place expectation. The society and the period wherein the cases happen are resolved exactly. The amount of cases demonstrating the closeness of a epidemic changes as shown by the kind, specialist, and size of mass revealed; previous experience or nonattendance of intro to the ailment; and time and spot of occasion. Additionally in 2001 "Outbreak" is alluded to as a plague constrained to confined increment in the rate of a sickness, for example in a town, village or locked organization. Mostly, a sickness that has enormous yearly changeability can be considered as epidemic. The transmission of numerous irresistible ailments differs notably via season. For instance, most of flu outbreaks in the northern half of the globe happen in mid to pre-spring [14, 15] while, even in moderately stable trans-mission zones, top intestinal sickness transmission for the most part pursues times of overwhelming rainfall [16]. Where sickness is available in a region, variances in its rate are viewed as epidemics just if the quantity of cases surpasses a specific limit. The result in general were given for the patients within the limit of 1.96 increasing deviation for the rate in 2 weeks [17]. In all the epidemic cases, is characterized by best looking at constant lengthy haul datasets, hence set up reconnaissance centers is a significant starter prerequisite.

3.2 Locating geographical epidemic areas

Regardless of whether an irresistible illness is far reaching all through a nation or whole locale, geologically the peril of inter alia isn't ascend to at all territories and will reflect, bury alia, the assignment and direct of vectors in disease and hosts. Geographical assortment in threat of pestilences is commonly perceived anyway pandemic slanted regions are just once in a while described officially. This is required most of the way to the difficulties in describing scourges, somewhat to absence of lengthy haul reconnaissance information and change the study of disease transmission of infections after some time. For instance, intestinal sickness transmission in numerous swamp regions of Africa frequently is portrayed as holo-endemic, with all year transmission, while neighboring districts at higher height are viewed as epidemic prone. In these territories, natural conditions are by and large less ideal, and transmission happens as epidemics possibly on events when changes in ecological conditions or potentially populace resistance make tolerant conditions. Be that as it may, the challenges in portrayal are appeared by an ongoing report by [18]. While testing research speculations it is basic to apply steady definitions in order to perceive pandemic domains. Then again, to improve general prosperity this may be less huge than idea of whether the case of sending in a particular locale is sufficiently phenomenal to require an emotionally specific kind in operational response.

3.3 Climatic and non- climatic disease threat aspects identification

A wide number of researches have been attempted to recognize normal danger components, including air. There are two standard methodologies: biological and statistical modeling. Statistical models are used to perceive the direct quantifiable connections between marker (for example atmosphere) factors and the result of intrigue (for example ailment occurrence). Biological models contain total re-introductions of atmosphere's consequences for the populace elements of pathogens and vectors. Most of past investigations have utilized factual displaying of region explicit verifiable sickness measures as well as vector conveyances. Biological models conceivably offer more prominent bits of knowledge into the mechanisms driving variety in disease occurrence however require progressively broad comprehension of climatic impacts on all parts of pathogen and vector elements. They along these lines have been connected on very rare events [19]. Therefore both approaches are utilized, it is critical to assess non-climatic elements. These incorporate pointers of the defenselessness of populaces to ailment outbreaks, for example, low insusceptibility, high pervasiveness of HIV, lack of healthy sustenance, medication and bug spray obstruction [20]. Inability to assess such impacts can prompt either variety in infection rate being inaccurately ascribed to atmosphere impacts and additionally poor prescient exactness.

3.4 Enumerating the relation between disease outbreaks and climate variability

The path between sickness rate and the atmosphere parts can be assessed in a statistical or biological model that may in this way outline the reason of future estimates of disease episodes. Before this can be start, it is essential to ensure that both infection and legitimate data are open at fitting spatial and transient objectives and for a sufficient time distribution. The climatic data evaluated at standard atmosphere stations have the advantage of being exact, direct estimations of meteorological conditions – anyway this data will be specialist just of a storage region of the station itself. Then again, if the zone of interest does not contain meteorological stations, the use of this data depends upon appropriate extrapolation systems being connected to the data.

Determining the relation between climatic parameters and the event of irresistible infections and additionally their vectors so as to foresee geological and fleeting examples of sickness has been endeavored various occasions. In spite of the fact that these forecasts enable us to delineate and vector ranges, either in light of the fact that they intend to make spatial as opposed to transient expectations (for example forecast illness rates in areas that have not recently been overviewed), or in light of the fact that they are utilized to investigate potential impacts of long haul changes in climate over decades. In this way, prediction exactness can be spoken to by looking at the size of the surveyed and expected epidemic, utilizing the error of mean value, or as relationship amount among surveyed and anticipated cases [21-23]. In either case, model exactness ought to be evaluated against autonomous information to give a precise replication of an endeavor to foresee a future epidemic. Utilizing similar information to both form and test a model will in general overstate prediction exactness

3.5 Well-timed Warning System

A warning system are considered to originate from both model forecasts and infection observation (for example early discovery), and incorporate thought of operational conditions and reactions. Illness observation gives a methods for checking disease rate after some time and, contingent upon the idea of the framework, might be a proper instrument for identifying abnormal among occurred information. Carefully, infection reconnaissance does not comprise early cautioning, even where observation is completed inside an extraordinarily planned system of sentinel destinations. Observation gives methods for distinguishing as opposed to foreseeing the beginning of an epidemic. Be that as it may, an appropriately structured framework ought to present essentially the purpose of mediation, along these lines expanding the odds of intercession helping infection control. As a method for approving sickness forecasts created by climatic based models reconnaissance information comprise an indispensable piece of any completely fledged system. By and large, the presence of precise, approved prediction models relies upon the accessibility of reliable observed information.

Accordingly, proper illness reconnaissance frameworks must set up, in order to follow infection rate with the citation to the expected ordinary dimensions of occurrence can show the beginning of an epidemic will be observed the information of incorporate data on the region of cases and give data about its geological degree. Nonetheless, variations in observation information demonstrating strange dimensions of disease transmission ought to be examined before usage of huge scale mediations went for plague control. Such eruptions may comprise curios inside the reconnaissance framework (for example because of analytic practices changing and moves in the dimensions of use of seperate wellbeing offices by the overall population and so forth.) and it will not change the reflection in dimensions of sending the infection. It ought to likewise be borne at the top of the priority list that there is no single, standard methodology accessible for identifying variations (for example flare-ups) based on reconnaissance information. Various epidemic outbreak prediction approaches have been proposed and the affectability and particularity of each will change contingent upon the idea of the worldly dissemination of cases related with every disease type.

4. EPIDEMIC DISEASE OUTBREAK: TECHNIQUES AND APPROACHES

Techniques for gathering tolerant information, answering to health authorities and assembling reports are expensive and tedious. Given the fame of online networking, irresistible ailment observation frameworks that utilization information sharing advancements to precisely follow web-based social networking [24] information could possibly illuminate early alerting frameworks and outbreak reaction, and encourage correspondence between human services suppliers and nearby, national and universal health specialists. Some of the approaches using social media data for analysis were described in brief in this section.

An early warning decision based framework [25] is an instrument for conveying data about looming dangers to powerless individuals before a threat occasion happens, in this manner empowering moves to be made to relieve potential damage, and some of the time, giving a chance to keep the hazardous occasion from happening. The objective of a disease early warning framework is to give general wellbeing authorities and the overall population with however much notification ahead of time as could reasonably be expected about the probability of a disease outbreak in a specific area, in this way augmenting the scope of practical reaction alternatives. The most usually utilized bellwether of an approaching epidemic is the presence of early instances of the sickness in a populace. In certain examples, "sentinel" creatures are put in consecutively High-chance areas and checked for proof of disease, since contaminations among these creatures will ordinarily forecast human cases. These "observation and reaction" approaches give genuinely high prescient conviction of a looming infection outbreak however regularly leave general wellbeing specialists with minimal early notification for preparing activities to avert additionally spread of the illness operator. Conversely, natural perceptions and atmosphere conjectures can possibly be utilized in endeavors to anticipate the presence of a pathogen and accordingly enable chances to limit its transmission.

Fig. 2 shows the various segments important for an operational early warning framework. Atmosphere scales and data from continuous epidemiological reconnaissance and natural perceptions might be utilized as contribution for prescient models that produce perception or alerts about an approaching disease chance. . This data is then combined with vulnerability evaluations to figure out which fragments of a populace are destined to face damage from an approaching danger and hazard examination to decide the imaginable effect of the forthcoming threat on these gatherings. The activities required to lessen the effects of a looming danger are resolved through advancement of a feedback methodology. At long last, a communication with the public framework encourages the convenient scattering of data on looming dangers, hazard situations, and readiness methodologies to defenseless gatherings. The other focal parts of an early warning framework are talked about in more detail beneath.

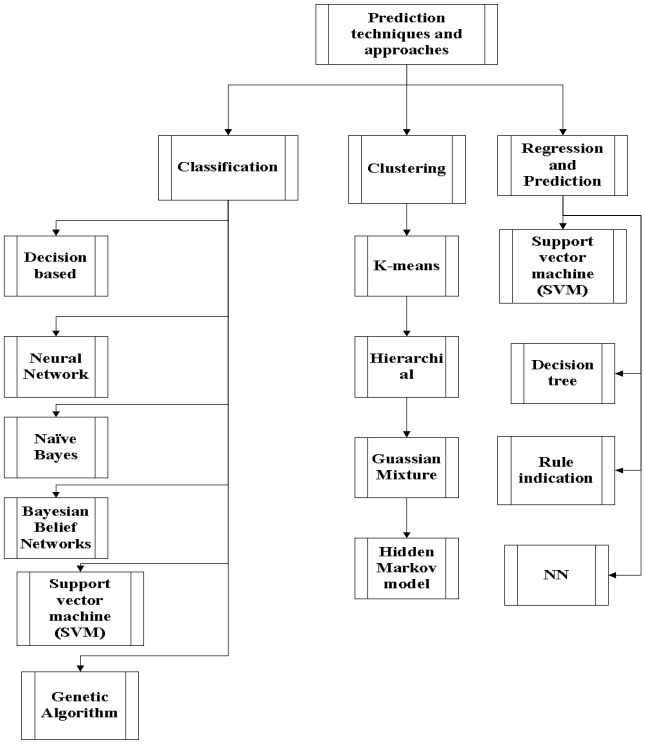

Figure 2: Techniques and approaches used for disease forecasting

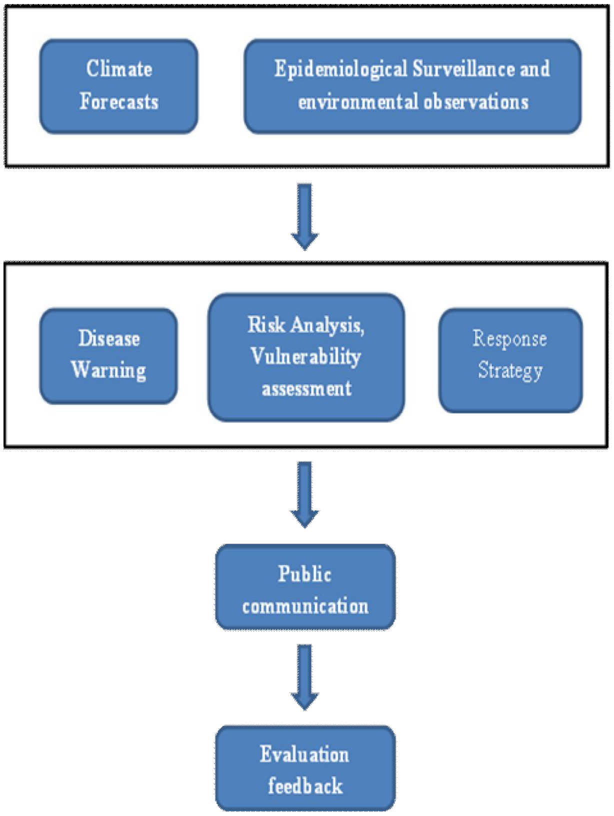
Climate Forecasts
Epidemiological Surveillance and environmental observations
Disease Warning
Risk Analysis, Vulnerability assessment
Response Strategy
Public communication
Evaluation feedback

Figure 3: Early Warning System

Epidemiological surveillance*:* Epidemiological surveillance frameworks that are continuous and efficient, that utilization institutionalized schedules for quality affirmation, and that accommodate examination and auspicious spread of data are basic to early warning frameworks.

Environmental observations*:* Orderly Environmental observations are a significant segment of an early warning framework, to some degree on the grounds that the effect of climate or atmosphere occasions regularly relies upon precursor conditions.

Vulnerability assessment*:* Vulnerability alludes to a populace's affectability to a threat, just as its capacity to adapt to the danger. Vulnerability evaluation gives a setting to translating observation information and for understanding the effects of interruption among any of the connections that associate food, cover, financial frameworks, and human wellbeing.

Risk analysis*:* Risk analysis is completed to dole out explicit probabilities to the reasonable effects of a forthcoming risk. Risk is a perplexing variable identified with risk types and examples of vulnerability, potential effects of the danger occasions, and the limit of networks to assimilate and recoup from these effects.

Preparation and response: Warning frameworks must be created working together with advancements in nearby, national-, or territorial dimension reaction abilities, especially in exceptionally vulnerable zones. Frequently, enhancements in logical prescient capacities are not joined by proportionate upgrades in the capacity to really utilize this warning statistics.

Public communication: To assure the warning information and prescribed reaction procedures are paid attention to by the populaces in danger, powerful open correspondence methodologies must be created. As examined in [26-28], compelling wellbeing correspondence projects distinguish and organize crowd fragments; transfer exact, science-based messages from dependable sources; and contact spectators through commonplace stations.

Wang et.al [29] looked the ANN models for the forecast of the irresistible runs in the China, Shanghai the FFBPNN (feed forward back propagation neural system), GRNN (Generalized regression neural system, RBFNN (Radial basis function neural system). the three layers are comprise, the FFBPNN information of specific in the 6 neurons, that was the layer of the single neuron. In the experimental section the amount of neurons layer are selected the hidden layer. The GRNN is the highly regression model depend on the kernel regression and concentrate on the creators. RBFNN is used to premise the work of the hidden layer, non-stop capacity and gauge for the precision. The combination of the three methods is brought the end of FFBPNN to execute the best least in the mistake for the error of the mean in blender. The analysis of aftereffects in the different direct relapse for the demonstrated anticipates and the qualities are lot near to real quality.

Human brucellosis (hb) as an experiment to distinguish significant ecological determinants of the infection and anticipate its outbreaks has examined by J. Wang et al. An epic counterfeit neural system (ANN) model was created, utilizing yearly region level quantities of HB cases and information on 37 natural factors, possibly connected with HB in Inner Mongolia, China. Data from 2006 to 2008 was used to get ready, support and test the model, while data for 2009–2010 was used to overview the model's presentation. The Enhanced Vegetation Index was perceived as the hugest marker of HB rate, trailed by means of land surface temperature and other temperature-and precipitation-related components. The sensible common claim to specialty of HB was shown reliant on these pointers. Model assessments were seen to be in incredible simultaneousness with declared amounts of HB cases in both the model improvement and evaluation stages. The examination suggests that HB case may be modeled, with a reasonable degree of accuracy, using the ANN model and environmental components acquired from satellite data. The examination expanded the understanding of environmental determinants of HB and moved the system for forecast of air that delicate powerful sickness flare-ups.

Wang et.al [30] has considered two well-known data mining arrangement calculations Support Vector Machine (SVM) and Artificial Neural Network (ANN) are used for Malaria figure using an immense dataset of Maharashtra state. Data of each one is 35 region of Maharashtra, from 2011 to 2014 has been considered. Parameters used are Average month to month precipitation, Temperature, Humidity, Total number of positive cases, Total number of Plasmodium Falciparum (pF) cases and flare-up occur in twofold characteristics Yes or No. An enormous amount of tests was accumulated from different sources. Root Mean Square Error (RMSE) and Receiver Operating Characteristic (ROC) are used to check the display of the models. It is seen that introduction of the model made using SVM is more exact than ANN. The SVM model can anticipate the flare-up 15 - 20 days early.

Y. Zhang et.al [31] proposes a Poisson-relapse based model first with the intra-common and between nearby factors included and the Empirical evaluations are coordinated subject to a certified enlightening list which records the 16-days-declared cases in the Yunnan zone of China for quite a while, from 2005 to 2011. To get familiar with the structure of the dissemination lattice, he proposes two methodologies –

using somewhere in the range of from the earlier learning and assessing it sans preparation by means of a scanty structure suspicion. The ensuing improvement issue of the most extraordinary a back game plan is a bended one and can be capably grasped by the substituting course technique for multipliers (ADMM). Besides, they combine furthermore the outside factor, i.e., the imported cases. With one reality that the scattering of the amount of corrupted cases over a year is (generally) for most torment and one supposition that the getting rate has a little variance consistently, we can derived the effect of the outside factor with a parametric limit (e.g., a quadratic limit) after some time. The consequent streamlining issue is so far bended and can be in like manner comprehended by the ADMM calculation.

Dependable forecasts of flu can help in the control of both regular and epidemic outbreaks. Nsoesie et.al [32] presents a simulation optimization (SIMOP) approach for determining the flu epidemic bends. The examination speaks to the last advance of a task went for utilizing a blend of reenactment, order, factual and streamlining systems to gauge the scourge bend and construe fundamental model parameters during a flu episode. The SIMOP system joins an individual-based model and the Nelder-Mead simplex enhancement strategy. The technique is utilized to estimate epidemics mimicked over manufactured informal organizations speaking to Montgomery County in Virginia, Miami, Seattle and encompassing metropolitan areas. Contingent upon the manufactured system, the pinnacle time could be anticipated inside a 95% as right on time as seven weeks before the real pinnacle. The pinnacle tainted and all out contaminated were additionally precisely determined for Montgomery County in Virginia inside the anticipating time frame. Anticipating of the epidemic bend for both occasional and pandemic flu outbreaks is an intricate issue, anyway this is a starter step and the outcomes recommend that more can be accomplished around there.

Numerous analysts are utilizing statistical and data mining procedures for the analysis of coronary illness. Bhatt et.al [33] utilized Naive Bayes a basic information mining system to show better outcome and precision. None of the system predicts heart disorders subject to risk factors, for instance, age, family heritage, hypertension, diabetes, alcohol confirmation, tobacco smoking, heftiness or physical inaction, raised cholesterol, etc. Framework dependent on such hazard elements would not just diminish the death rate of Heart Disease quiet in the country territories yet it would likewise give patients a notice and proposal about the plausible nearness of coronary illness even before he visits an emergency clinic or goes for exorbitant restorative checkups.

S. Vijayarani et.al [34] principally centers on foreseeing ailments from the hemogram blood test informational collection by utilizing information mining strategies. In the exploration, another grouping calculation which is named as weight based k-means calculation is created for distinguishing the leukemia, provocative, bacterial or viral contamination, HIV disease and malignant iron deficiency maladies from the hemogram blood test tests informational collection. The recently proposed weight based k-implies calculation proficiency is assessed with Fuzzy C-means and K-implies bunching calculations.

5. SOCIAL MEDIA

There are various research on websites and social media to collect datas about the epidemic disease outbreak [35-51]. Online data mining is a helpful technique for achieving tests for uncommon results or censored or hard to achieve populaces areas. Facebook is an online long range interpersonal communication website where clients make a profile, add different clients system, send messages to their system associations, and post messages to their profiles. Clients may likewise combine basic intrigue gatherings and speak with organizations. Like Face book, Twitter, might be utilized for dynamic or detached information accumulation. It has much of the time been utilized to select members however the data made by clients has additionally been utilized for research. Facebook clients have the choice of posting interests, for example, films, books, sports crews, or exercises, on their profile.

Land regions where a higher extent of clients embraced action related interests, for example, wellbeing and health or open air wellness exercises, and a lower extent of clients supported interests in inactive practices, especially TV viewing, would in general have less rates of heftiness. The anticipate wellbeing practices are in the post in endorsement of the clients “likes” by Face book. The clients are the postal division that ‘like’ is the data of classes in the program interface. The Behavioral Risk Factor Surveillance System announced the conditions that connect with future and wellbeing numeral information.

Information mining systems have been connected on the new wellspring of colossal information supplier for example internet based life. This will almost certainly give constant input to outbreak probability. Roughage et al [52] talk about different huge information approaches for sickness reconnaissance. They proposed utilization of cartographic procedure to produce a ceaseless information layer of illness hazard for different diseases. The procedure included infection record events to acquire focuses where disease have been accounted for and have characterized a complete degree of the ailment alongside epidemiologically important natural covariant. The authors likewise talked about carious difficulties of utilizing huge information as volume, frequency and decent variety for each phase of the procedure.

Xie et.al [53] examines different internet based life stage from which information can be gathered. The researchers examined four online networking datasets that can be utilized. The health related sites, are Facebook open dividers of restorative brands of associations, Facebook open posts referencing wellbeing media and terms transfers from Instagram (INST). Odlum and Yoon [54] have utilized Twitter to show the utilization of twitter as an ongoing strategy for Ebola outbreak identification. Patterns were performing on the tweets and substance examination was led to identify the similitudes and assemble groups. K-means calculation was utilized to assembly the clusters.

BioCaster [55] is an operational metaphysics based framework for observing on the web media information since 2006. This framework depends on content digging systems for recognizing and following the irresistible illness outbreaks through the inquiry of semantic sign. The framework persistently examinations archives announced from more than 1700 RSS channels, ProMED-mail, European Media Monitor, Google News, and the WHO, among others suppliers. The extricated content is grouped for topical importance and plots them onto a Google guide utilizing geo-data. The framework comprises of four fundamental stages: topic grouping, named element recognition (NER), disease/area identification and occasion acknowledgment. In the main stage, the writings are arranged into important or non-pertinent classifications utilizing to prepare a naïve Bayes classifier. At that point, for important record corpus are look substances of enthusiasm from 18 idea sorts dependent on the BioCaster cosmology identified with sicknesses, infections, microorganisms, areas and side effects

As of late, a few works have just showed up appeared capability of Twitter messages to follow and anticipate ailment outbreaks [55, 56]. Nellore and Fishman [57] work is centered on utilizing forecast market to show open conviction about the likelihood that H1N1 infection will turn into an epidemic. So as to conjecture the future costs of the probability broadcast, they chose to utilize the Support Vector Machine calculation to complete regression

A report classifier to distinguish pertinent messages is displayed in Culotta et al. paper [58]. In the work, Twitter messages identified with influenza were recollected during 10 weeks utilizing words, for example, influenza, hack, sore throat or migraine. At that point, a few characterization frameworks dependent on various regression models to connect these messages with CDC measurements were thought about, finding that the best model accomplishes a relationship of 0.78 (basic model regression).

6. ANALYSIS AND DISCUSSION

Every one of the frameworks and arrangements in table 1 exhibited for forecast have shown the effective and gainful utilization of strategies when connected to remove and get new information for open medicinal services purposes. The primary test of these frameworks is to translate the inquiry setting of a specific question or report, in light of the fact that a client can inquiry about a specific medication, indication or ailment for an assortment of reasons. This objective can be troublesome in light of the fact that a similar word can allude to an alternate thing relying on setting. Besides, a particular infection can have numerous names and side effects identified with it, which expands the unpredictability of the issue. In this way, to create techniques for diminishing false alerts and diminishing level of superfluous occasions distinguished by the epidemic frameworks can be a significant issue for future works and inquires about on the field. Also, to distinguish the time and area of messages is an esteem included for expanding the nature of recognizing conceivable new diseases outbreaks. Even so, practically speaking area names are regularly exceptionally confusing in light of the fact that geo-transient disambiguation is so troublesome, and on account of the assortment of manners by which cases are depicted crosswise over various writings. There are a few ongoing works demonstrate the capability of social Medias to follow and distinguish illness outbreaks. These works exhibit that there are health confirmations in web based life which can be recognized. In any case, there can be inconveniences in regards to the conceivable inaccurate forecasts on account of the immense measure of social information previously compared and the limited quantity of significant information identified with potential infections outbreaks. In this manner, it is important to test and approve cautiously every one of the models and strategies utilized.

Table 1: Comparative analysis

Author	Year	Disease	Contribution
Schmidt, Charles W et.al; Bauermeister, J, Jones, Altshuler et.al, Chaulket.al, Dal Moro et.al, Moreno et.al, Schumacher et.al,Sueki, H et.al,	2012, 2012, 2012, 2015, 2011, 2013, 2012, 2014, 2015		The social media helps to predict and track the disease outbreaks
Ginsberg et al. [2] Eysenbach G et.al; Imran et.al, Lampos, V et.al; Zhang et.al; Adler, A et.al; Camacho, C. FU et.al	2009, 2006, 2013, 2017, 2017, 2012, 2013, 2017		search engine query data on web used to detect the influenza epidemics
Cho S, et.al	2013		The South Korea in the

			Google trends are correlate with the national influenza surveillance
Belay et.al	2017		For enhancement of global health security emerged the infectious diseases in the Zoonotic disease
Althouse et.al, , Abdul Rahim	2011, 2008	Dengue	Prediction using search query surveillance
Xia et al., Kunpeng Zhang et.al,	2013, 2011		Sentiment elicitation system for social media data
Bonabeau et.al,	1998		The geographical spread of influenza
Macdonald et.al., Myers et.al; Hay et.al; Abeku et.al; Vijeta Sharma; E. Nsoesie et.al	1957, 2000, 2002, 2002, 2015, 2013	malaria	Forecasting disease risk, Epidemiology and control
Randolphpp et.al ,	2004	tick-borne	Evidence that climate change has cause diseases
Yongming et al., Wang, Junzhong et.al	2014, 2017, 2014	diarrhea	artificial neural networks to infectious forecasting
Y. Zhang et.al	2015		Regression approach based prediction
Sohana Saiyed et.al	2016	Heart disease	Naive Bayes Based Prediction
Vijayarani et.al	2015		Clustering based prediction
Harris et.al	2014	foodborne	The social media to identify illness of the health department
Thomas et. al.	2014	Tuberculosis and Lung Disease,	Outbreak identified by using the social media and genotype cluster analysis

7. CONCLUSION OF FUTURE DIRECTION

In this work we present a review on prediction approaches for epidemic disease outbreaks dependent via web-based social media information. Through the writing examined, it tends to be reasoned that different forecast models can be utilized to foresee the infection outbreaks; however these models depend just on the ailment information accessible in social network. The channel can be tapped in different mining methods to yield current input combined with quicker and increasingly exact recognition of the disease outbreaks. A few elements can be distinguished as significant segments in building up a decent forecast for danger of epidemic or infection outbreak. Essential among these are the precision of forecast, also its land scale and duration. Based on the survey, we have concluded that among the number of research works the machine learning approach performs in a better way for the prediction of epidemic outbreak rather than the traditional approaches. Therefore, these zones require further examination and planning.

REFERENCES

1. C. Schmidt. **Trending Now: Using Social Media to Predict and Track Disease Outbreaks**, *Environmental Health Perspectives,* Vol. 120, No. 1, pp. a30-a33, Jan. 2012. https://doi.org/10.1289/ehp.120-a30
2. R. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant. **Detecting influenza epidemics using search engine query**, data *Nature,* Vol. 457, No. 7232, pp. 1012-1014, Feb. 2009. https://doi.org/10.1038/nature07634
3. S. Triple Project. **Assessment of syndromic surveillance in Europe**, *The Lancet*, Vol. 378, No. 9806, pp. 1833-1834, Nov. 2011. https://doi.org/10.1016/S0140-6736(11)60834-9
4. G. Eysenbach. **Infodemiology: tracking flu-related searches on the web for syndromic surveillance**, in: *Proc. of AMIA Annual Symposium*, American Medical Informatics Association., USA, pp. 244–248, 2006.
5. S. Cho. **Correlation between National Influenza Surveillance Data and Google**, *Trends in South Korea*, *PLoS ONE*, Vol. 8, No. 12, p. e81422, Dec 2013. https://doi.org/10.1371/journal.pone.0081422
6. E. Belay Zoonotic. **Disease Programs for Enhancing Global Health Security**, *Emerging Infectious Diseases,* Vol. 23, No. 13, Dec. 2017. https://doi.org/10.3201/eid2313.170544
7. Y. Shin, C. Ryo, and J. Park. **Automatic extraction of persistent topics from social text streams**, *World Wide Web*, Vol. 17, No. 6, pp. 1395-1420, Nov. 2013. https://doi.org/10.1007/s11280-013-0251-3
8. S. Zhang, D. Cheng, R. Hu, and Z. Deng. **Supervised feature selection algorithm via discriminative ridge regression**, *World Wide Web* vol. 21, No. 6, pp. 1545-1562, Nov. 2017. https://doi.org/10.1007/s11280-017-0502-9
9. P. Guo. **Monitoring seasonal influenza epidemics by using internet search data with an ensemble penalized regression model**, *Scientific Reports,* Vol. 7, No. 1, Apr. 2017. https://doi.org/10.1038/srep46469
10. B. Althouse, Y. Ng, and D. Cummings. **Prediction of Dengue Incidence Using Search Query Surveillance**, *PLoS Neglected Tropical Diseases,* Vol. 5, No. 8, pp. e1258, Aug. 2011. https://doi.org/10.1371/journal.pntd.0001258
11. J. Tang, X. Hu, and H. Liu. **Social recommendation: a review**, *Social Network Analysis and Mining*, Vol. 3, No. 4, pp. 1113-1133, Dec. 2013. https://doi.org/10.1007/s13278-013-0141-9
12. Y. Xie, **MuSES: Multilingual Sentiment Elicitation System for Social Media Data**, *IEEE Intelligent Systems,* Vol. 29, No. 4, pp. 34-42, Jul. 2014. https://doi.org/10.1109/MIS.2013.52

13. S. Moorhead, D. Hazlett, L. Harrison, J. Carroll, A. Irwin, and C. Hoving. **A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication**, *Journal of Medical Internet Research,* Vol. 15, No. 4, p. e85, Apr. 2013. https://doi.org/10.2196/jmir.1933
14. J. Martin. **Global institutions: the World Health Organization (WHO),** *Bulletin of the World Health Organization,* Vol. 87, No. 6, pp. 484-484, Jun. 2009. https://doi.org/10.2471/BLT.08.060814
15. E. Bonabeau, L. Toubiana, and A. Flahault**. The geographical spread of influenza,** *Proceedings of the Royal Society of London. Series B: Biological Sciences*, Vol. 265, No. 1413, pp. 2421-2425, Jan. 1998. https://doi.org/10.1098/rspb.1998.0593
16. G. Covell. **Epidemiology and Control of Malaria,** *bmj*, Vol. 2, No. 5059, pp. 1477-1477, May 1957. https://doi.org/10.1136/bmj.2.5059.1477
17. M. Fraser. **Viewpoint. Bioterrorism preparedness and local public health agencies: building response capacity**, *Public Health Reports*, Vol. 115, No. 4, pp. 326-330, Jul. 2000. https://doi.org/10.1093/phr/115.4.326
18. S. I. Hay, J. Cox, D. J. Rogers, S. E. Randolph, D. I. Stern, G. D. Shanks, M. F. Myers, R. W. Snow. **Climate change and the resurgence of malaria in the East African highlands,** *Nature*, Vol. 415, No. 6874, pp. 905-909, Feb. 2002. https://doi.org/10.1038/415905a
19. S. Randolph.**Evidence that climate change has caused 'emergence' of tick-borne diseases in Europe**, *International Journal of Medical Microbiology Supplements*, Vol. 293, pp. 5-15, Apr. 2004. https://doi.org/10.1016/S1433-1128(04)80004-4
20. Rogers. **Vulnerability, health and health care**, *Journal of Advanced Nursing***,** Vol. 26, No. 1, pp. 65-72, Jul. 1997. https://doi.org/10.1046/j.1365-2648.1997.1997026065.x
21. N. Tran Minh. **Zika virus: no cases in the Eastern Mediterranean Region but concerns remain**, *Eastern Mediterranean Health Journal*, Vol. 22, No. 5, pp. 350-353, May 2016. https://doi.org/10.26719/2016.22.5.350
22. T. Abeku. **Forecasting malaria incidence from historical morbidity patterns in epidemic-prone areas of Ethiopia: a simple seasonal adjustment method performs best,** *Tropical Medicine and International Health*, Vol. 7, No. 10, pp. 851-857, Oct. 2002. https://doi.org/10.1046/j.1365-3156.2002.00924.x
23. K. Mkutu. **Pastoralism, and Conflict in the Horn of Africa and the Sahel**, *Population and Development Review*, Nol. 44, No. 4, pp. 857-860, Dec 2018. https://doi.org/10.1111/padr.12211
24. M. Dhib, J. Pandey, K. Amine Al. **Majority Vote method for preferences detection: Application for Social Networks**, Vol. 8, No. 1, pp. 42-53, 2019. https://doi.org/10.30534/ijatcse/2019/08812019
25. V. Grasso, A. Singh, J. Pathak. **Early Warning Systems A State of the Art Analysis and Future Directions**, *Environmental Development*, Vol. 4, pp. 136-171, Jan. 2012. https://doi.org/10.1016/j.envdev.2012.09.004
26. M. Freimut. **Expanding the Scope of Dual Diagnosis and Co-Addictions: Behavioral Addictions**, *Journal of Groups in Addiction & Recovery*, Vol. 3, No. 3-4, pp. 137-160, Nov. 2008. https://doi.org/10.1080/15560350802424944
27. M. Hulwan. **Football Match Winning Team Prediction Using Machine Learning**, *International Journal of Advanced Research in Computer Science*, Vol. 9, No. 6, pp. 12-17, Nov. 2018. https://doi.org/10.26483/ijarcs.v9i6.6337
28. N. Sri Hari, P. Mishra, and K. Suvarna Vani. **Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems**, *International Journal of Advanced Trends in Computer Science & Engineering*, Vol. 8, No. 2, pp. 259-264, Feb. 2018. https://doi.org/10.30534/ijatcse/2019/26822019
29. Y. Wang, J. Li, J. Gu, Z. Zhou, and Z. Wang. **Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China),** *Applied Soft Computing*, Vol. 35, pp. 280-290, Oct. 2015. https://doi.org/10.1016/j.asoc.2015.05.047
30. J. Wang et al. **A Remote Sensing Data Based Artificial Neural Network Approach for Predicting Climate-Sensitive Infectious Disease Outbreaks: A Case Study of Human Brucellosis**, *Remote Sensing*, Vol. 9, No. 10, p. 1018, Sep. 2017. https://doi.org/10.3390/rs9101018
31. Y. Zhang, W. Cheung, and J. Liu. **A Unified Framework for Epidemic Prediction based on Poisson Regression**, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 27, No. 11, pp. 2878-2892, Nov. 2015. https://doi.org/10.1109/TKDE.2015.2436918
32. E. Nsoesie, R. Beckman, S. Shashaani, K. Nagaraj, and M. Marathe. **A Simulation Optimization Approach to Epidemic Forecasting**, *PLoS ONE*, Vol. 8, No. 6, p. e67164, Jun. 2013. https://doi.org/10.1371/journal.pone.0067164
33. N. Bhatt, A. Thakkar, A. Ganatra, and N. Bhatt. **Ranking of Classifiers based on Dataset Characteristics using Active Meta Learning,** *International Journal of Computer Applications*, Vol. 69, No. 20, pp. 31-36, Jan. 2013. https://doi.org/10.5120/12089-8269
34. S. Vijayarani, and S. Sudha. **An Efficient Clustering Algorithm for Predicting Diseases from Hemogram Blood Test Samples**, *Indian Journal of Science and Technology*, Vol. 8, No. 17, Jan. 2015.

https://doi.org/10.17485/ijst/2015/v8i17/52123

35. J. Bauermeister, M. Zimmerman, M. Johns, and P. Glowacki, S. **Stoddard and E. Volz, Innovative Recruitment Using Online Networks: Lessons Learned From an Online Study of Alcohol and Other Drug Use Utilizing a Web-Based, Respondent-Driven Sampling (webRDS) Strategy**, *Journal of Studies on Alcohol and Drugs*, Vol. 73, No. 5, pp. 834-838, Sep. 2012. https://doi.org/10.15288/jsad.2012.73.834
36. L. Jones, B. Saksvig, M. Grieser, and D. Young. **Recruiting adolescent girls into a follow-up study: Benefits of using a social networking website**, *Contemporary Clinical Trials*, Vol. 33, No. 2, pp. 268-272, Mar. 2012. https://doi.org/10.1016/j.cct.2011.10.011
37. S.Adler, K. Eames, S. Funk, and W. Edmunds. **Incidence and risk factors for influenza-like-illness in the UK: online surveillance using Flusurvey,** *BMC Infectious Diseases*, Vol. 14, No. 1, Dec. 2014. https://doi.org/10.1186/1471-2334-14-232
38. P. Altshuler, H. Gerns Storey, and S. Prager. **Exploring abortion attitudes of US adolescents and young adults using social media,** *Contraception*, Vol. 91, No. 3, pp. 226-233, Mar. 2015. https://doi.org/10.1016/j.contraception.2014.11.009
39. J. Klein, R. Thomas, and E. Sutter. **Self-Reported Smoking in Online Surveys**, *Medical Care*, Vol. 45, No. 7, pp. 691-695, Jul. 2007. https://doi.org/10.1097/MLR.0b013e3180326145
40. M. Stein et al. **Online Respondent-Driven Sampling for Studying Contact Patterns Relevant for the Spread of Close-Contact Pathogens: A Pilot Study in Thailand**, *PLoS ONE*, Vol. 9, No. 1, p. e85256, Jan. 2014. https://doi.org/10.1371/journal.pone.0085256
41. M. Barratt et al. **Lessons from conducting trans-national Internet-mediated participatory research with hidden populations of cannabis cultivators**, *International Journal of Drug Policy*, Vol. 26, No. 3, pp. 238-249, Mar. 2015. https://doi.org/10.1016/j.drugpo.2014.12.004
42. M. Ben-Ezra et al. **Face it: Collecting mental health and disaster related data using Facebook vs. personal interview: The case of the 2011 Fukushima nuclear disaster,** *Psychiatry Research*, Vol. 208, No. 1, pp. 91-93, Jun. 2013. https://doi.org/10.1016/j.psychres.2012.11.006
43. Camacho, K. Eames, A. Adler, S. Funk, and J. Edmunds. **Estimation of the quality of life effect of seasonal influenza infection in the UK with the internet-based Flusurvey cohort: an observational cohort study,** *The Lancet*, Vol. 382, pp. S8, Nov. 2013. https://doi.org/10.1016/S0140-6736(13)62433-2
44. K. Chaulk, and T. Jones. **Online Obsessive Relational Intrusion: Further Concerns About Facebook,** *Journal of Family Violence*, Vol. 26, No. 4, pp. 245-254, May 2011. https://doi.org/10.1007/s10896-011-9360-x
45. F. Dal Moro. **Online Survey on Twitter: A Urological Experience,** *Journal of Medical Internet Research*, Vol. 15, No. 10, p. e238, Jun. 2013. https://doi.org/10.2196/jmir.2719
46. J. Janiec, A. Zielicka-Hardy, A. Polkowska, J. Rogalska, and M. Sadkowska-Todys. **Did public health travel advice reach EURO 2012 football fans? A social network survey,** *Eurosurveillance*, Vol. 17, No. 31, Aug. 2012. https://doi.org/10.2807/ese.17.31.20238-en
47. Kuehn. **Agencies Use Social Media to Track Foodborne Illness**, *JAMA*, Vol. 312, No. 2, p. 117, Jul. 2014. https://doi.org/10.1001/jama.2014.7731
48. M. Moreno, A. Grant, L. Kacvinsky, K. Egan, and M. Fleming. **College Students' Alcohol Displays on Facebook: Intervention Considerations,** *Journal of American College Health*, Vol. 60, No. 5, pp. 388-394, Jul. 2012. https://doi.org/10.1080/07448481.2012.663841
49. K.Schumacher et al. **Social Media Methods for Studying Rare Diseases**, *PEDIATRICS*, Vol. 133, No. 5, pp. e1345-e1353, May 2014. https://doi.org/10.1542/peds.2013-2966
50. H. Sueki. **The association of suicide-related Twitter use with suicidal behaviour: A cross-sectional study of young internet users in Japan**, *Journal of Affective Disorders*, Vol. 170, pp. 155-160, Jan. 2015. https://doi.org/10.1016/j.jad.2014.08.047
51. T. Thomas, S. Heysell, E. Houpt, J. Moore, and S. Keller. **Outbreak of pyrazinamide-monoresistant tuberculosis identified using genotype cluster and social media analysis**, *The International Journal of Tuberculosis and Lung Disease*, Vol. 18, No. 5, pp. 552-558, May 2014. https://doi.org/10.5588/ijtld.13.0663
52. S. Hay, D. George, C. Moyes, and J. Brownstein.**Big Data Opportunities for Global Infectious Disease Surveillance**, *PLoS Medicine*, Vol. 10, No. 4, pp. e1001413, Apr. 2013. https://doi.org/10.1371/journal.pmed.1001413
53. Y. Xie et al. **MuSES: Multilingual Sentiment Elicitation System for Social Media Data**, *IEEE Intelligent Systems*, Vol. 29, No. 4, pp. 34-42, Jul. 2014. https://doi.org/10.1109/MIS.2013.52
54. M. Odlum, and S. Yoon. **What can we learn about the Ebola outbreak from tweets?,** *American Journal of Infection Control*, Vol. 43, No. 6, pp. 563-571, Jun. 2015. https://doi.org/10.1016/j.ajic.2015.02.023
55. N. Collier et al. **BioCaster: detecting public health rumors with a Web-based text mining system**, *Bioinformatics*, Vol. 24, No. 24, pp. 2940-2941, Oct. 2008.

https://doi.org/10.1093/bioinformatics/btn534

56. N. Collier et al. **BioCaster: detecting public health rumors with a Web-based text mining system,** *Bioinformatics*, Vol. 24, No. 24, pp. 2940-2941, Oct. 2008. https://doi.org/10.1093/bioinformatics/btn534
57. A. Nellore, and J. Fishman. **Pandemic Swine Flu 2009**, *Xenotransplantation*, Vol. 16, No. 6, pp. 463-465, Jun. 2009. https://doi.org/10.1111/j.1399-3089.2009.00559.x
58. Culotta, A**.: "**Towards detecting influenza epidemics by analyzing twitter messages"**.** in: *Proc. of the First Workshop on Social Media Analytics*, Washington D.C., USA, pp. 115–122, 2010 https://doi.org/10.1145/1964858.1964874
59. Y. Wang, J. Li, J. Gu, Z. Zhou, and Z. Wang. **Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China),** *Applied Soft Computing*, Vol. 35, pp. 280-290, Oct. 2015. https://doi.org/10.1016/j.asoc.2015.05.047
60. M. Bates. **Tracking Disease: Digital Epidemiology Offers New Promise in Predicting Outbreaks**, *IEEE Pulse*, Vol. 8, No. 1, pp. 18-22, Jan. 2017. https://doi.org/10.1109/MPUL.2016.2627238
61. S. Sohana, B. Nikita, Amit P. Ganatra, **A Survey on Naive Bayes Based Prediction of Heart Disease Using Risk Factors**, *International Journal of Innovative and Emerging Research in Engineering* ,Vol. 3, No. 2, pp.228-2232, Apr. 2016
62. J. K. Harris, R. Mansour, B. Choucair, J. Olson, C. Nissen, J. Bhatt. **Health department use of social media to identify foodborne illness**, *Morbidity and Mortality Weekly Report*, Vol. 63, no. 32, pp. 681–685, Aug. 2014.