International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 05 , May 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A Novel Data Mining Technique for Virus Generation and Impacts on Agriculture due to Weather Changes and Rainfall

Mr. Ramachandra Reddy Avula[1], Dr. A. Rama Mohan Reddy[2]

[1] *Research Scholar ,Department of CSE, Sri Venkateswara University, Tirupati,AP, India*

[2] *Professor, Department of CSE, University College of Engineering, SVU, Tirupati, AP, India.*

ABSTRACT

Virus or Bacteria generation, increase and spread through direct and indirect mediator from one place to another is very high in wet climate. Continues rainfall over 48 or 72 hours causes to wet weather. Temperature is also a big factor to generate Virus. In wet climate the temperature is low and is causes to increase the generation of bacteria/virus. The technique will be used to explain the future weather conditions like rain fall ratio and virus impacts on agriculture based on past data.

Disease classification

Broadly, infectious diseases may be classified into two categories based on the mode of transmission: those spread directly from crop/person to crop/person (through direct contact or droplet exposure) and those spread indirectly through an intervening vector organism (mosquito or tick) or a non-biological physical vehicle (soil or water). Infectious diseases also may be classified by their natural reservoir as anthroponoses (human reservoir) or zoonoses (animal reservoir).

Climate sensitivities of infectious diseases

Both the infectious agent (protozoa, bacteria, viruses, etc) and the associated vector organism (mosquitoes, ticks, sandflies, etc.) are very small and devoid of thermostatic mechanisms. Their temperature and fluid levels are therefore determined directly by the local climate. Hence, there is a limited range of climatic conditions—the climate envelope—within which each infective or vector species can survive and reproduce. It is particularly notable that the incubation time of a vector-borne infective agent within its vector organism is typically very sensitive to changes in temperature, usually displaying an exponential relationship.

Other climatic sensitivities for the agent, vector and host include level of precipitation, sea level elevation, wind and duration of sunlight.

Documented and predictive climate/infectious disease linkages

The seasonal patterns and climatic sensitivities of many infectious diseases are well known; the important contemporary concern is the extent to which changes in disease patterns will occur under the conditions of global climate change.

Over the past decade or so this question has stimulated research into three concentrations.

First, can the recent past reveal more about how climatic variations or trends affect the occurrence of infectious diseases?

Second, is there any evidence that infectious diseases have changed their prevalence in ways that are reasonably attributable to climate change?

Third, can existing knowledge and theory be used to construct predictive models capable of estimating how future scenarios of different climatic conditions will affect the transmissibility of particular infectious diseases?

Modifying influences

Climate is one of several important factors influencing the incidence of infectious diseases. Other important considerations include sociodemographic influences such as human migration and transportation; and drug resistance and nutrition; as well as environmental influences such as deforestation; agricultural development; water projects; and urbanization. In this era of global development and land-use changes, it is highly unlikely that climatic changes exert an isolated effect on disease; rather the effect is likely dependent on the extent to which humans cope with or counter the trends of other disease modifying influences. While recognizing the important independent role of these non-climatic factors, the focus of this section is to examine the extent to which they may compound the effects of climatic conditions on disease outcomes.

Disease classifications relevant to climate/health relationships

Climate variability's effect on infectious diseases is determined largely by the unique transmission cycle of each pathogen. Important environmental factors include temperature, precipitation and humidity (discussed in more detail in the following section). Several possible transmission

components include pathogen (viral, bacterial, etc.), vector (mosquito, snail, etc.), non-biological physical vehicle (water, soil, etc.), non-human reservoir (mice, deer, etc.) and human host. Epidemiologists classify infectious diseases broadly as anthroponoses or zoonoses, depending on the natural reservoir of the pathogen; and direct or indirect, depending on the mode of transmission of the pathogen.

The following is a description of each category of disease, discussed in order of probable increasing susceptibility to climatic factors.

1. Directly transmitted diseases

i. Anthroponoses

Directly transmitted anthroponoses include diseases in which the pathogen normally is transmitted directly between two crop/human hosts through physical contact or droplet exposure. The transmission cycle of these diseases comprises two elements: pathogen and human host. Generally, these diseases are least likely to be influenced by climatic factors since the agent spends little to no time outside the human host.

ii. Zoonoses

Directly transmitted zoonoses are similar to directly transmitted anthroponoses in that the pathogen is transmitted though physical contact or droplet exposure between reservoirs. However, these agents are spread naturally among animal reservoirs.

2. Indirectly transmitted diseases (anthroponoses & zoonoses)

Indirectly transmitted anthroponoses are a class of diseases defined by pathogen transmission between two human hosts by either a physical vehicle (soil) or a biological vector (tick). These diseases require three components for a complete transmission cycle: the pathogen, the physical vehicle or biological vector, and the human host. Indirectly transmitted water-borne anthroponoses are susceptible to climatic factors because the pathogens exist in the external environment during part of their life cycles. Flooding may result in the contamination of water supplies or the reproduction rate of the pathogen.

Climate sensitivity of infectious disease

i. Seasonality of infectious disease

Data patterns of winter mortality and infectious disease using the example of cyclic influenza outbreaks occurring in the late fall, winter and early spring in North America. This disease pattern may result from increased likelihood of transmission due to indirect social or behavioral adaptations to the cold weather such as crowding indoors. Another possibility is that it may be attributed directly to pathogen sensitivities to climatic factors such as humidity. Similarly, some vector-borne diseases also show significant seasonal patterns whereby transmission is highest in the months of heavy rainfall and humidity. Seasonal fluctuations of infectious disease occurrence imply an association with climatic factors. However, to prove a causal link to climate, non-climatic factors must be considered. Furthermore, in order to assess long-term climate influences on disease trends, data must span numerous seasons and utilize proper statistics to account for seasonal fluctuations.

ii. Vector-borne diseases

Important properties in the transmission of vector-borne diseases include:

• Survival and reproduction rate of the vector

• Time of year and level of vector activity, specifically the biting rate

• Rate of development and reproduction of the pathogen within the vector.

Vectors, pathogens, and hosts each survive and reproduce within certain optimal climatic conditions and changes in these conditions can modify greatly these properties of disease transmission. The most influential climatic factors for vector borne diseases include temperature and precipitation but sea level elevation, wind, and daylight duration are additional important considerations. Table 1.1
gives an overview of the impact of climatic change on each biological component of both vector and rodent-borne diseases.

Temperature sensitivity

Extreme temperatures often are lethal to the survival of disease-causing pathogens but incremental changes in temperature may exert varying effects. Where a vector lives in an environment where the mean temperature approaches the limit of physiological tolerance for the pathogen, a small increase in temperature may be lethal to the pathogen. Alternatively, where a vector lives in an environment of low mean temperature, a small increase in temperature may result in increased development, incubation and replication of the pathogen.

Precipitation sensitivity

Variability in precipitation may have direct consequences on infectious disease outbreaks. Increased precipitation may increase the presence of disease vectors by expanding the size of existent larval habitat and creating new breeding grounds. In addition, increased precipitation may support a growth in food supplies which in turn support a greater population of vertebrate reservoirs. Unseasonable heavy rainfalls may cause flooding and decrease vector populations by eliminating larval habitats and creating unsuitable environments for vertebrate reservoirs.

Humidity sensitivity

Humidity can greatly influence transmission of vector-borne diseases, particularly for insect vectors. Mosquitoes and ticks can desiccate easily and survival decreases under dry conditions. Saturation deficit (similar to relative humidity) has been found to be one of the most critical determinants in climate/disease models.

Sea level sensitivity

The projected rise in sea level associated with climate change is likely to decrease or eliminate breeding habitats for salt-marsh mosquitoes. Bird and mammalian hosts that occupy this ecological niche may be threatened by extinction, which would also aid the elimination of viruses endemic to this habitat. Alternatively, inland intrusion of salt water may turn former fresh water habitats into salt-marsh areas which could support vector and host species displaced from former salt-marsh habitats.

Water-borne diseases

Human exposure to water-borne infections can occur as a result of contact with contaminated drinking water, recreational water, coastal water, or food. Rainfall patterns can influence the transport and dissemination of infectious agents while temperature can affect their growth and survival. Table 1.2 outlines some of the direct and indirect weather effects on enteric viruses, bacteria and protozoa.

TABLE 1.1 Effects of weather and climate on vector and rodent-borne diseases*.

Vector-borne pathogens spend part of their life-cycle in cold-blooded arthropods that are subject to many environmental factors. Changes in weather and climate that can affect transmission of vector borne diseases include temperature, rainfall, wind, extreme flooding or drought, and sea level rise. Rodent-borne pathogens can be affected indirectly by ecological determinants of food sources affecting rodent population size, floods can displace and lead them to seek food and refuge.

Pathogen groups	Pathogenic agent	Food-borne agents	Water-borne agents	Indirect weather effect	Direct weather effect
Viruses	Enteric viruses (e.g. hepatitis A virus, Coxsackie B virus)	Shellfish	Groundwater	Storms can increase transport from faecal and waste water sources	Survival increases at reduced temperatures and sunlight (ultraviolet)[a]
Bacteria Cyanobacteria Dinoflagellates	Vibrio (e.g. *V. vulnificus*, *V. Parahaemolyticus*, *V. cholerae* non-01; Anabaena spp., Gymnodinium Pseydibutzschia spp.)	Shellfish	Recreational, Wound infections	Enhanced zooplankton blooms	Salinity and temperature associated with growth in marine environment
Protozoa	Enteric protozoa (e.g. Cyclospora, Crytosporidium)	Fruit and vegetables	Recreational and drinking water	Storms can increase transport from faecal and waste water sources.	Temperature associated with maturation and infectivity of Cyclospora

[a] Also applies to bacteria and protozoa.
Source: Reproduced from reference 20.

Temperature effects on selected vectors and vector-borne pathogens

Vector

• survival can decrease or increase depending on species;

• some vectors have higher survival at higher latitudes and altitudes with higher temperatures;

• changes in the susceptibility of vectors to some pathogens e.g. higher temperatures reduce size of some vectors but reduce activity of others;

• changes in the rate of vector population growth;

• changes in feeding rate and host contact (may alter survival rate);

• changes in seasonality of populations.

Pathogen

• decreased extrinsic incubation period of pathogen in vector at higher temperatures

• changes in transmission season

• changes in distribution

• decreased viral replication.

Effects of changes in precipitation on selected vector-borne pathogens

Vector

• increased rain may increase larval habitat and vector population size by creating new habitat

• excess rain or snowpack can eliminate habitat by flooding, decreasing vector population

• low rainfall can create habitat by causing rivers to dry into pools (dry season malaria)

• decreased rain can increase container-breeding mosquitoes by forcing increased water storage

• epic rainfall events can synchronize vector host-seeking and virus transmission

• increased humidity increases vector survival; decreased humidity decreases vector survival.

Pathogen

Few direct effects but some data on humidity effects on malarial parasite development in the anopheline mosquito host.

Vertebrate host

• increased rain can increase vegetation, food availability, and population size

• increased rain can cause flooding: decreases population size but increases human contact.

Increased sea level effects on selected vector-borne pathogens

Alters estuary flow and changes existing salt marshes and associated mosquito species, decreasing or eliminating selected mosquito breeding-sites (e.g. reduced habitat for Culiseta melanura)

*The relationship between ambient weather conditions and vector ecology is complicated by the natural tendency for insect vectors to seek out the most suitable microclimates for their survival (e.g. resting under vegetation or pit latrines during dry or hot conditions or in culverts during cold conditions).

Table 1.2: Water and food-borne agents: connection to climate

Climate and weather can substantially influence the development and distribution of viral insects. Anthropogenically induced climatic change arising from increasing levels of atmospheric greenhouse gases would, therefore, be likely to have a significant effect on agricultural insect pests. Current best estimates of changes in climate indicate an increase in global mean annual temperatures of 1°C by 2025 and 3°C by the end of the next century. Such increases in temperature have a number of implications for temperature-dependent insect pests in mid-latitude regions.

Environmental changes	Example diseases	Pathway of effect
Dams, canals, irrigation	Schistosomiasis	↑ Snail host habitat, human contact
	Malaria	↑ Breeding sites for mosquitoes
	Helminthiasies	↑ Larval contact due to moist soil
	River blindness	↓ Blackfly breeding, ↓ disease
Agricultural intensification	Malaria	Crop insecticides and ↑ vector resistance
	Venezuelan haemorraghic fever	↑ rodent abundance, contact
Urbanization, urban crowding	Cholera	↓ sanitation, hygiene; ↑ water contamination
	Dengue	Water-collecting trash, ↑ *Aedes aegypti* mosquito breeding sites
	Cutaneous leishmaniasis	↑ proximity, sandfly vectors
Deforestation and new habitation	Malaria	↑ Breeding sites and vectors, immigration of susceptible people
	Oropouche	↑ contact, breeding of vectors
	Visceral leishmaniasis	↑ contact with sandfly vectors
Reforestation	Lyme disease	↑ tick hosts, outdoor exposure
Ocean warming	Red tide	↑ Toxic algal blooms
Elevated precipitation	Rift valley fever	↑ Pools for mosquito breeding
	Hantavirus pulmonary syndrome	↑ Rodent food, habitat, abundance

Source: reproduced from reference 3.

Table 1.3: Examples of environmental changes and possible effects on infectious diseases.

Changes in climate may result in changes in geographical distribution, increased overwintering, changes in population growth rates, increases in the number of generations, extension of the development season, changes in crop-pest synchrony, changes in interspecific interactions and increased risk of invasion by migrant pests. To illustrate some of these effects, results of a study investigating the impact of climatic change on the agriculture in West Godavari, East Godavari, Chithoor, and Visakhapatnam districts of Andhra Pradesh, India.

IJARSE
ISSN 2319 - 8354

Agriculture crop yield depends on the weather conditions of that region because weather can make huge impact on crop yield. Real time weather data can used to find the virus production and this causes to increase the diseases on crop. This research work is about studying the impact of weather disasters and viruses on agriculture crop using past data and applying data mining techniques to analyze, predict the useful patterns.

The availability of weather reports data during the last decades (observational records, radar and satellite maps, proxy data, etc.) makes it important to find an effective and accurate tools to analyze and extract hidden knowledge from this huge data. Meteorological data mining (process of finding the knowledge from agricultural data) is a form of Data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into useful knowledge. These insights can play important role in predicting the weather conditions, viral impacts and addressing the critical crop yield problem.

1. INTRODUCTION

Data mining is the process of extracting interesting (non-trivial, implicit, previously unknown and potentially useful) information or patterns from large data repositories such as relational database, transactional database, data warehouses, etc. Data mining is also a one of the core process of knowledge discovery from data (KDD). In this research work, the interesting weather patterns for wet and dry days causes the virus generation have been extracted using association rule mining approach. And this work also explained which type of virus generates and impacts on agriculture in future. The virus ratio with threshold value on the wet days has been estimated. Virus prediction is usually done for a region like individual gardens, fields and craft with vital installation. A methodology is using association rule mining of weather conditions for large agricultural fields, mango gardens, and rice craft stations for Andhra Pradesh (Chittoor district) region.

The objective of this research work is to construct relationship models between various weather and viruses variables and to develop specific data mining techniques and to interconnect these variables using the distribution patterns obtained out of the data. This work has been aimed to extract the behavioral aspects of weather data for wet and dry days prediction using association rule mining on micro station atmospheric dataset. The attention is focused to develop a method for virus generation and also to predict the minimum temperature during the winter season and maximum temperature during summer season in a similar fashion and to apply this technique for all types of stations having different topography. This forecasting model has been used to forecast the rainy days both during North East and South West monsoon seasons for specific stations, **Chennai Airport (**Latitude: 13.067439, DMS Lat: 13° 4' 2.7804" N/Longitude: 80.237617, DMS Long: 80° 14' 15.4212" E) and **Visakhapatnam Sea port: (**Latitude: 17.690474, DMS Lat : 17° 41' 25.7064" N/Longitude: 83.231049, DMS Long: 83° 13' 51.7764" E) and two typical coastal observatories, **Tirupati (**Latitude: 13.629065, DMS Lat: 13° 37' 44.6340" N/Longitude: 79.424446, DMS Long: 79° 25' 28.0056" E) and **Rajahmundry (**Latitude: 17.004393, DMS Lat: 17° 0' 15.8148" N/Longitude: 81.783325, DMS Long: 81° 46' 59.9700" E).

IJARSE
ISSN 2319 - 8354

1.1 ASSOCIATION RULES

Association rule mining (Agrawal et al 1995) discovers the relationship between items from the set of transactions. These relationships can be expressed by association rules such as [$i_1 \Rightarrow i_2$, i_3 support, s = 2%, confidence, C = 60%]. This association rule means that 2% of the transactions under scrutiny show that items i_1, i_2 and i_3 appear jointly. A confidence of 60% means that 60% of the transactions carrying i_1 also carry i_2 and i_3. Associations may include any number of items on either side of the rule. Agrawal et al (1993) first introduced the problem of mining association rules and later it was enhanced by data mining scholars (Agrawal and Srikant 1994; Bayardo and Agrawal 1999; Sarawagi et al 2000).

Let I = {i_1, i_2, ..., i_m} denote a set of literals, namely, items. Moreover, let D represent a set of transactions, where each transaction T is a set of items such that T ⊆ I. A unique identifier, namely TID, is associated with each transaction. A transaction T is said to contain X, a set of some items in I, if X T. An association rule implies the form XY, where X I, Y⊂ I and X∩Y = ⊆⇒⊂φ. The rule X ⇒ Y holds in the transaction set D with confidence, c, where c% of transactions in D that contain X also contains Y. The rule has support, s, in the transaction set D if s% of transactions in D contain XY. An efficient algorithm is required that restricts the search space and checks only a subset of all association rule, yet does not miss important rules. The apriori algorithm is very much useful to retrieve the frequent patterns and its relationships. However, the interestingness (validity) of the rule is only based on support and confidence.

1.2. WEATHER PATTERN DERIVATION

The discovery of associations and correlations among items in large climate data sets can be made using frequent item set mining. With enormously huge amounts of data continuously being collected and stored, many industries are becoming interested in mining such patterns from their databases. In this proposed method, the discovery of interesting correlation relationships among huge amounts of surface daily data records which contains the weather information can help in the process of decision making, such as weather forecast relevant to agriculture operations, supply of analyzed data to end users for planning agricultural strategy and behavior analysis of virus generation during peak monsoons.

When the association rule is applied to the concept mentioned above for analysis, the various atmospheric observations like minimum temperature, maximum temperature, daily mean temperature, sea level pressure, cloud density, humidity etc taken at a definite time in definite area, the association rules could be seen like

Rule1: If the temperature is high, then there is no rain in the same area at the same time.

Although rule Rule1 reflects some relationships among the meteorological elements, its role in weather prediction is inadequate, as users are often more concerned about the weather along a time dimension like.

Rule2: If the humidity is medium and temperature is low, then it keeps warm for the next 24 hour.

The traditional association rules only capture associations among items within the same transactions. Therefore they are intra transactional rules. The notion of the transaction could be usage pattern of communication preference, the occurrence of the virus generation, the atmospheric events like hot day and cool day occurrence with similar weather patterns, and so on. However, an inter-transactional association rule can represent also the association of items among different transactions along certain dimensions like 24 hours advance, 48 hours advance, 72 hours advance, 1 week advance, etc.

For association rules mining from the filtered dataset, we use predictive Apriori algorithm for finding the hidden relationship between various atmospheric parameters. The basic property of Apriori is that all non-empty subsets of a frequent item set must be frequent. A frequent item set must be frequent in connection with the above Apriori algorithm searches with an increasing support threshold for the best 'N' rules concerning a support-based corrected confidence value. By using predictive Apriori algorithm the association rules are generated on weather data set with support and confidence threshold values.

1.3. BAYESIAN CLASSIFICATION FOR CLIMATE DATA

A classification is made by combining the impact that the different attributes have on the prediction to be made. Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Bayesian classification is based on Bayes' theorem. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases.

Given a data value Xi the probability that a related tuple, Ti, is in class Cj is described by P(Cj | Xi). Training data can be used to determine P(Xi), P(Xi|Cj), and P(Cj). From these values Bayes theorem allow us to estimate the posterior probability P(Cj | Xi) and then P(Cj | Ti). On classifying a target tuple, the conditional and prior probabilities generated from the training ser are used to make the prediction. This is done by combining the effects of the different attribute values from the tuple. Suppose that tuple Ti, has P independent attribute values {Xi1, Xi2,..............,Xip}. We then estimate P(Ti | Cj) by P(Ti | Cj) = P(Xik | Cj).

Classification is a form of data analysis that can be used to extract models describing important data classes and to predict future data trends. It will be a great value in weather forecasting with a better understanding of the data at large (Ancel et al 2004). The proposed Bayesian classification for meteorological data set is implemented with the help of machine learning tool Weka 3.6.6 that is available as open source. Many real time problems has been solved with the help of this open source tool (Weka 3). The learning process of the climate data classification has training data that are analyzed by a classification algorithm. Here the class label attribute is "precipitation" for virus generation prediction that is having two values; they are "yes" and "no". For the virus generation estimation, the class label attribute is "precipitation" which has two values such as "high" and "low". For the temperature prediction during winter monsoon, the class label is "post_minimum" which has two values such as "low" and "normal". For the maximum temperature prediction during summer months, the class label is "post_maximum" which has two values such as "high" and "normal". The learned model or classifier is represented in the form of classification rules. The classification process of the climate data set contains the test data that are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new data tuples. The figure 1.1 shows the proposed model for the learning procedure of the climate data classification process. The figure 1.2 shows the proposed model for the classification procedure of the climate data set.

The proposed back propagation neural model for meteorological data classification learns by iteratively processing a weather data set of training tuples, comparing the network's prediction for each tuple with the actual known target value. The target value may be the known class label of the training tuple. For each training tuple, the weights are modified so as to minimize the mean squared error between the weather prediction and the actual target value that is stated in the figure 1.3.

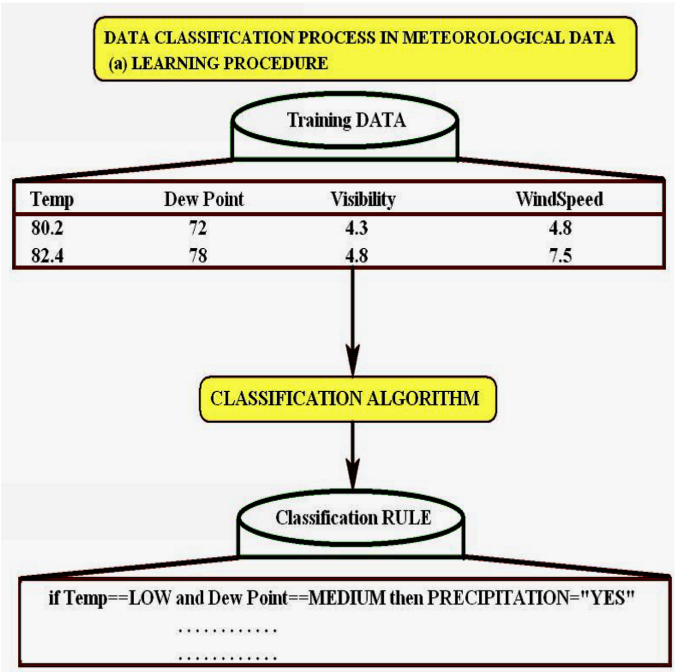

Figure 1.1. Learning procedure of meteorological data classification process.

The frequent weather patterns are extracted using predictive Apriori algorithm for virus generation prediction, virus generation estimation on Northeast and Southwest monsoon seasons for a location specific forecasting model. Summer maximum and winter minimum temperature patterns are also generated. By using the classifier model given weather data set can be classified. The significance of the present study would become clear if it is seen in broader perspective. While scholars have attempted to forecast tropical cyclone intensity change, evolution of association rules for droughts and floods, spatial association rule mining technique over the plateau, dissolved gas analysis etc. using the data mining technique, the application of this technique towards spot specific forecast of weather events is yet to be attempted, particularly to Indian sub continent. The earlier works have concentrated the data mining concept towards forecast of patterns of synoptic scale like tropical cyclone and droughts etc, but the efficacy of data mining towards virus generation forecast is yet to be established. The present work does an attempt towards that direction for peninsular India.

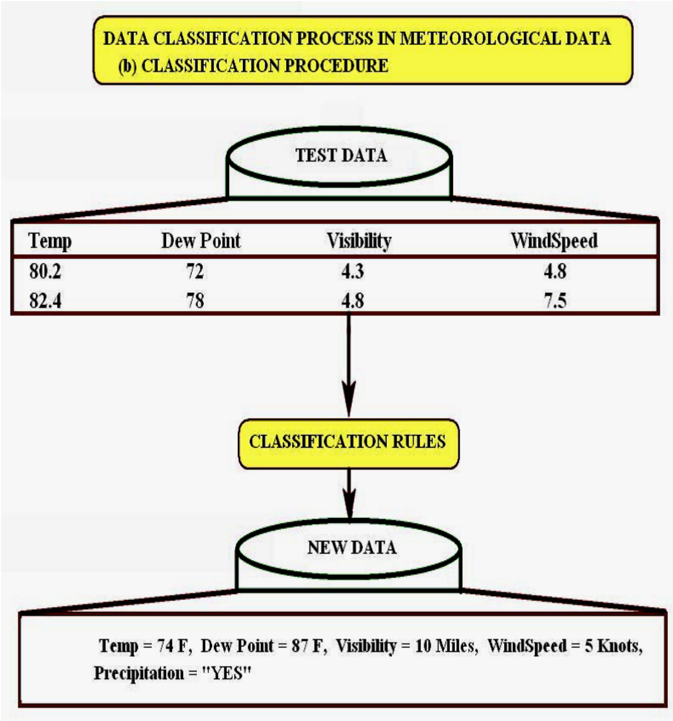

Figure 1.2. Classification process of meteorological data

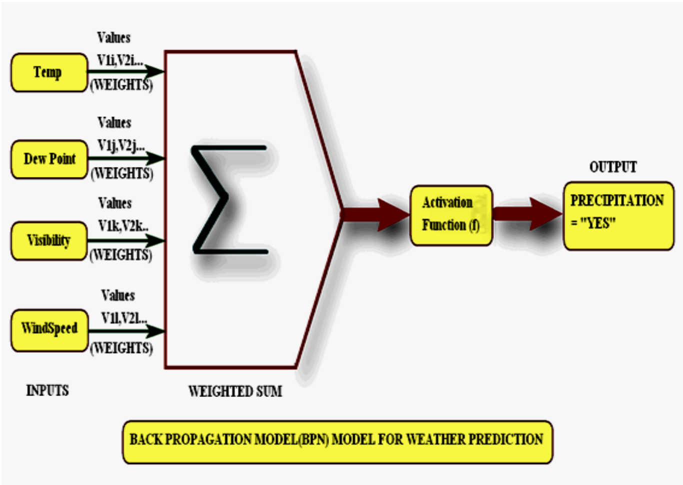

Figure 1.3. BPN model for meteorological data classification process

1.4. PROPOSED APPROACH

A few attempts on meteorological applications using data mining techniques have been made (Chen 2007; Cofina et al 2003; Dhanya and Nagesh Kumar 2009). However forecast for the monthly and seasonal virus generation for this Godavari delta region is needed. This has been attempted using the techniques of association rules in graphical mining, it is necessary to filter the raw data and convert into the format that is accepted by visualization tool, Gnuplot (Janert 2010). The proposed approach is described in Figure 3.3 that emphasizes the steps involved in the knowledge discovery using graphical mining of data mining technique

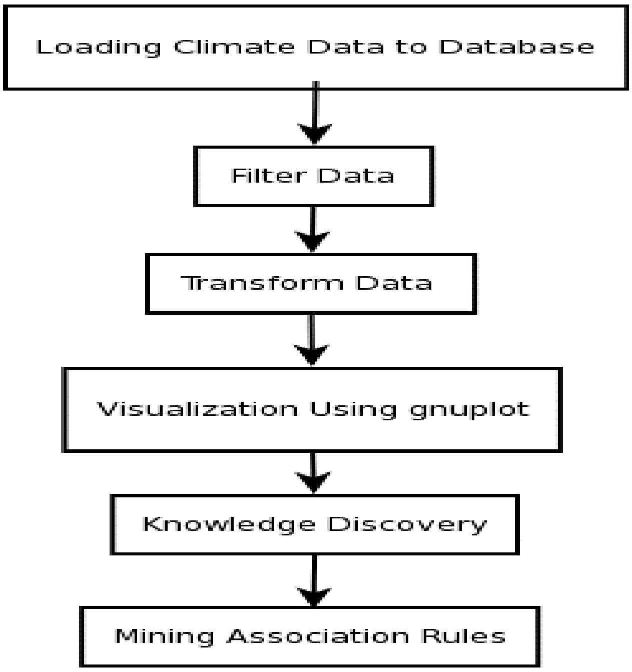

Figure 1.4. Steps involved in the knowledge discovery using graphical mining.

The procedure for the knowledge discovery using graphical mining is as follows:

Step 1 : Loading climatic data to SQL Database.
Step 2 : Filtering the noises and unwanted data from the dataset.
Step 3 : Transform data as required for the visualization software.
Step 4 : Generating graphs using the visualization tool.
Step 5 : Studying of patterns on the generated results.
Step 6 : Mining rules from those patterns.

1.5. RESULTS AND DISCUSSIONS

Using the graphical mining and data mining technique, virus generation due to rainfall and weather changes realizable in the North East monsoon was assessed from the virus generation due to rainfall and weather changes realized during Southwest monsoon months of June, July and August months with association rules. The running five-year mean rainfall is visualized using Gnuplot tool (Janert 2010) for different stations in Figures 1.5 to 1.10. In these figures the total monthly virus generation of June, July and August months of SW monsoon, September month of SW monsoon and October, November and December months of NE monsoon have been described. The total virus generation of the Southwest monsoon months and Northeast monsoon months were compared and the total monthly virus generation within the Northeast monsoon months were also compared.

International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 05 , May 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

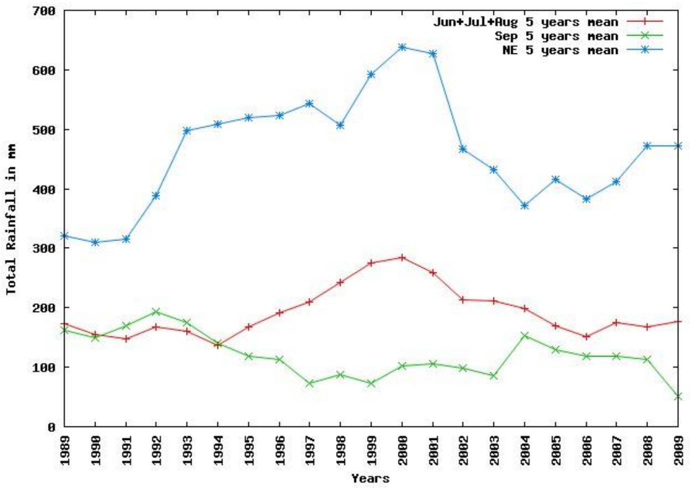

Figure 1.5. Five year mean rainfall of Bhadrachalam

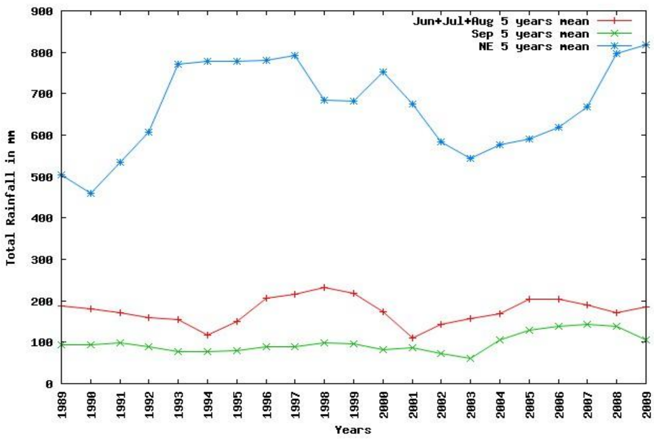

Figure 1.6. Five year mean rainfall of Polavaram

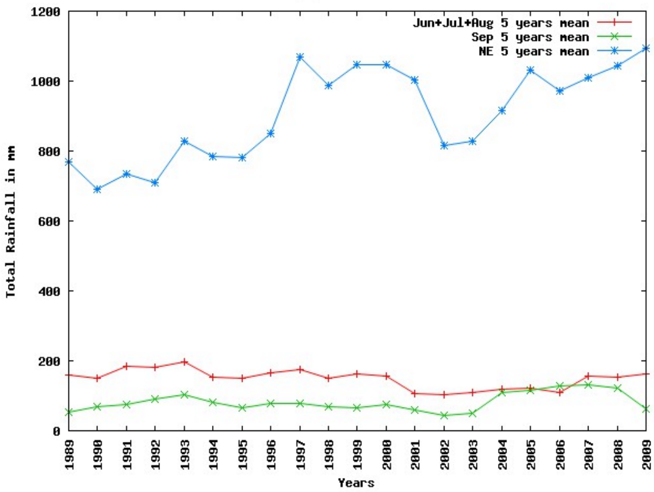

Figure 1.7. Five year mean rainfall of Kakinada

International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 05 , May 2016
www.ijarse.com

IJARSE
ISSN 2319 - 8354

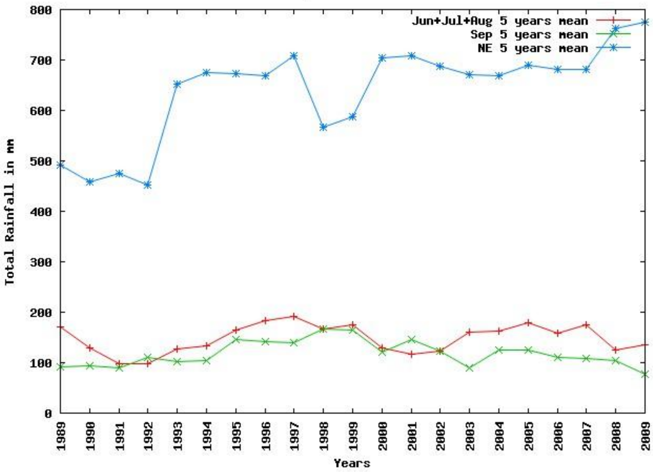

Figure 1.8. Five year mean rainfall of Pattiseema

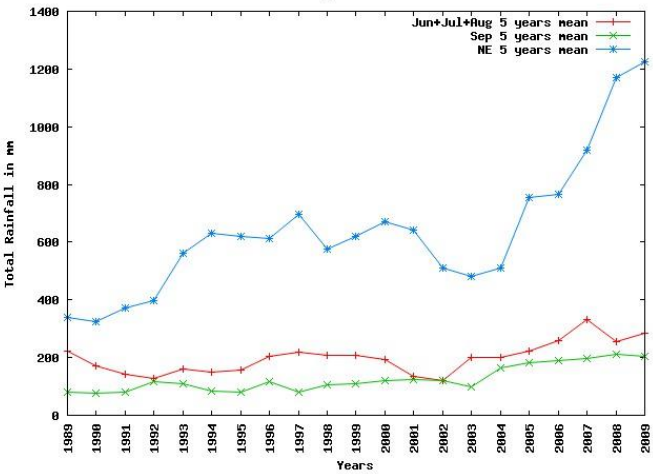

Figure 1.9. Five year mean rainfall of Rajahmundry

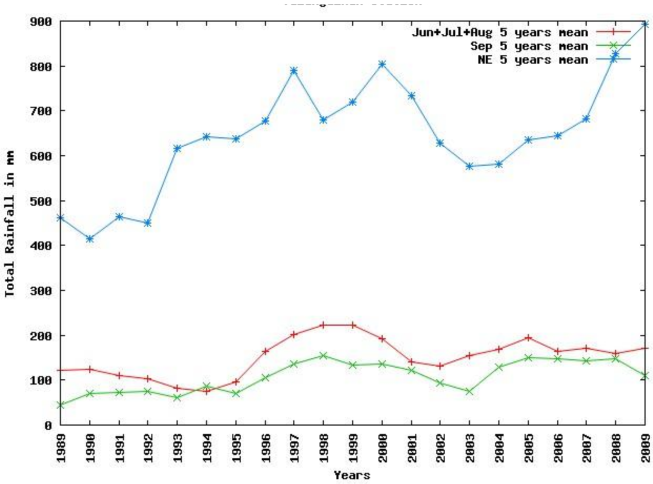

Figure 1.10. Five year mean rainfall of Yanam

After the analysis, the following rules are mined out from the above graphs:

$RF(X, "SW_C < SW_P") \Rightarrow NE_C(X, " > NE_P")$ (1.1)

RF(X, "OCT < 200mm") => NOV(X, "OCT R/F + 50mm") (1.2)

RF(X, "OCT > 200mm") => NOV(X ," > 400mm") (1.3)

The equation 1.1 describes the rule that when the total rainfall (RF) of current year of South West (SW_C) monsoon is less than the previous year of South West (SW_P) monsoon then the rainfall (RF) of current year of North East (NE_C) monsoon is greater than the previous year of the total rainfall (RF) of North East (NE_P) monsoon. The equation 1.2 describes the rule that when the total Rainfall (RF) of October month is less than 200mm then the total rainfall of the successive November month is most probably the actual October month rainfall with an excess of 50mm. The equation 1.3 describes the rule that if the total rainfall of October month is greater than 200mm then the total rainfall of the successive November month will have the precipitation of more than 400mm.

1.6. SUMMARY

The study reveals that in the Godavari delta basin, the rainfall pattern of many stations obey the association rule except Rajahmundry. Rajahmundry being coastal in nature may perhaps be the reason. The rule suggests a correlation between the Southwest and Northeast monsoon rains over this area when considered in relation to previous year's performance. It is also possible to assess the November month rain from the rains of October using the same rule. Frequent pattern mining leads to the discovery of associations and correlations among rainfall occurrence and virus impacts in large climate data set. This methodology can help to forecast rainfall and virus generation during the months of Northeast monsoon for the Godavari delta region of West and East Godavari districts of South India.

Conclusions and recommended future steps:

The purpose of this research work is to highlight the evidence linking climatic factors such as temperature, precipitation, and sea level rise, to the lifecycles of infectious diseases, including both direct and indirect associations via ecological processes. Many studies demonstrate seasonal fluctuations in infectious diseases but few have documented long-term trends in climate-disease associations. A variety of models has been developed to simulate the climatic changes and predict future disease outbreaks although few have controlled successfully for important sociodemographic and environmental influences. Gaps in knowledge indicate that future initiatives are required in the following areas: **Increase in active global disease surveillance:**

The lack of precise knowledge of current disease incidence rates makes it difficult to comment about whether incidence is changing as a result of climatic conditions. Incidence data are needed to provide a baseline for epidemiological studies. These data also will be useful for validating predictive models. As these data are difficult to gather, particularly in remote regions, a centralized computer database should be created to facilitate sharing of these data among researchers.

Continuation of epidemiological research into associations between climatic factors and infectious diseases:

In order to draw a causal relationship between climate change and patterns of infectious disease, research needs to prove consistent trends across diverse populations and geographical regions. This will best be accomplished by implementing rigorous study designs that adequately control for social and environmental confounders. International collaboration between researchers is important as well as interdisciplinary collaboration between specialists such as epidemiologists, climatologists and ecologists, in order to expand the breadth of information. A comprehensive study of mosquito-borne diseases, for example, requires a combination of entomologists, epidemiologists and climatologists to work together to examine the associations of changing vector habitats, disease patterns and climatic factors. Epidemiological data can be shared with policy-makers to make preventive policies.

Further development of comprehensive models: Models can be useful in forecasting likely health outcomes in relation to projected climatic conditions. Integrating the effects of social and environmental influences is difficult but necessary.

1.7. REFERENCES

1. Agnihotri, Dutta RK, Bhushan R, Somayajulu BLK. Evidence of solar forcing of southwest monsoon during last millennium. Earth Planetary Science Letters 2002.
2. Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C. 1993.
3. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules. Proc. of the 20th International Conference on Very Large Databases, Santiago, Chile, Sept. 1994; IBM Research Report RJ9839.
4. Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo AI. Fast Discovery of Association Rules. Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1995.
5. Akaike H. A new look at the statistical model identification. IEEE transaction on automatic control, 1977.
6. Alexander G, Kesavamurthy RN, De VS, Chellappa P. Fluctuation of monsoon activity. Indian Journal of Meteorology, Hydrology and Geo Physics,1978.
7. Ananthakrishnan R, Soman MK. Onset of southwest monsoon over Kerala 1901-1980. J. Climatology, 1988.
8. Basu BK. Diurnal variation of precipitation during summer monsoon. Monthly Weather Review, 2007.
9. Bayardo Jr RJ, Agrawal R. Mining the Most Interesting Rules. In Proc. of the 5th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, August 1999.
10. Cavozos T. Using self organization maps to investigate extreme climate event. Journal of climate. 2000.
11. Chen Mu. Ranking discovered rules from data mining with multiple criteria by data envelopment analysis. Expert systems with applications, 2007.

12. Das PK. The monsoons. National Book Trust of India, Third edition, 1995.
13. Dhanya CT, Nagesh Kumar. Data mining for evaluation of association rules for droughts and floods in India using climate inputs. Journal of Geophysical Research, 2009.
14. Dhar ON, Rakheja RR. Foreshadowing Northeast monsoon rainfall over Tamil Nadu, India. Monthly Weather Review, 1983.
15. Han J and Kamber M. Data Mining – Concepts and Techniques", Morgan Kaufmann Publications, 2008.
16. Climate change and infectious diseases. J. A. Patz,1 A. K. Githeko,2 J. P. McCarty,3 S. Hussein, 1 U. Confalonieri, 4 N. de Wet5

AUTHORS:

Mr. RAMACHANDRA REDDY AVULA,
Research scholar,
Department of Computer Science and Engineering,
University College of Engineering, Sri Venkateswara University,
Tirupati, Andhra Pradesh, India.

Dr. A. RAMAMOHAN REDDY,
Professor,
Department of Computer Science and Engineering,
University College of Engineering,
Sri Venkateswara University,
Tirupati, Andhra Pradesh, India.