GENERATING PREDICTABLE CRIME AND DISEASE OCCURRENCE RATE BY ANALYZING DIFFERENT NEWS SOURCES

[1]MD ANISUR RAHMAN AKASH, [2]PROJNA SAHA, [3]SAJIB HASAN

[1,2,3]Computer Science and Engineering American International University-Bangladesh Dhaka, Bangladesh

Abstract - There is so much potential and importance of geographical information retrieval systems in the field of health care and crime. This paper addresses the usage of geodata in analyzing crime and disease rates of any area for future prediction. It allows a spatial perspective on disease and crime analysis. Used to their optimum level, as tools for analysis and decision-making. It evaluates the utility of integrating disease and crime information with geographic information systems (GIS) mapping of environmental data in a pilot study. The main purpose of this study is to raise awareness about public health and crime risk factors and utilizing GIS can further enhance the accessibility of the combined crime analysis and health information.

Keywords - Geo data, Geographical Information System, Disease rate, Crime rate, Crime and Disease Prediction, String Comparator

I. INTRODUCTION

Now a day's crime and disease are one of the greatest problems of our world. The crime rate is increasing in the world constantly. It cannot be prevented in a rational way. Because criminals also use modern technology for committing the crime. Day by day crime occurrence rate is increasing so rapidly that law enforcement could not control the crime. Criminals are getting smarter for committing their crimes [33]. But if we use the previous data of crime occurrence we can predict the crime occurrence rate of any particular area.

Not only crime is increasing in our society as well as disease also has spread more than ever in our society. Though our medical technology has developed rapidly, the new type of disease also spread out in our society [34] [35]. Every day a lot of people died because of the different types of diseases in our world. Not every disease is spreading in every society. Some disease is spreading in a particular area. The problem is people are moving every society in our modern world. For that reason, if a disease is spread out in a society that disease can be carried out by visiting people and they can spread that disease in their society [36]. If we can predict area wise disease occurrence rate then a visiting people can easily know about the area's current situation and take the necessary step to prevent that disease before traveling that particular area.

In our last two decades technology has developed so rapidly that the world has more than enough resources for predicting area-based crime and disease rate. For example, if any crime occurred any particular area then a lot of newspaper publishes a report about that crime or television also report about that crime as well as radio.

Every one of those reports can easily be found on the web. For any kind of disease is spreading out in any particular area then the media publish a report about that disease. From those sources, we can easily build a crime or disease database. The technology made spatial data mining a practical solution for handling this type of data. So our main goal is developing an accurate, better and efficient pattern so that we can easily identify the crime and disease of any area.

Though we can easily get enough data from various sources, we have some challenged that we are facing:

1. The data has some limitations. Some data are incomplete or some data have some inconsistency. So, Analyzing that data is a bit difficult.
2. Though the web has enough information about crime and disease, we need a well-structured way of fetching those data.
3. The accuracy of the result directly depends on the training set.

The steps for cleaning the data:

1. Create the algorithm(Word count, unique word counts, character counts)
2. Visualize the data
3. Create clusters

The most challenging task was finding a way to detect a pattern by this method we can detect the crime. If we can detect the crime and disease we can easily fetch the location of occurrence these events and also the date. If we find this information's then we can easily predict or generate any report about any area. In the old days, law enforcement used previous case information. But most of this information was not available. For that reason, they could not predict the next probably in any particular area. But in the modern era, data mining is introduced. By using data mining we can get the most accurate within a shorter period of time. If classified the data in different classification we can get the information easily we can use cauterization for an area-based result.

There are some step for analyzing the crime and disease:

1. Collecting the data from various sources
2. Preparing those data for analyzing
3. Classify those data
4. Detecting the data by using pattern
5. Locating the area as well as date
6. Calculate the result and generating the predication

II. LITERATUREREVIEW

2.1 Geographical Information System:

Geographical Information system represents a technology design to achieve a particular objective. In recent years the variety of GIS products to assist the management and manipulation of spatial and non-spatial data have arrived on the market and users worldwide have begun to gain similarity with this new system. Experience service that there can be no doubt but that the application of GIS is making significant contributions in the facility the ability integration and presentation of information.

A GIS can be defined as the set of tools for collecting sorting, retrieving, analyzing and displaying spatial data. As a technology, AGS is not necessarily limited to the conference confine of one independent system. It may well have several components each with a particular objective.

Information is knowledge about a particular subject, issue, event or process. Information can be obtained from various sources. All processes produce information and it is the values of characters in the processes of output that are information. This captures most concepts of information in individual disciplines. The number of possible values in the output and their relative frequencies of occurrence may be used in measuring the amount of information present. An information retrieval system is described that dynamically prioritizes search request results before outputting to a user. When a database search yields multiple hits, the results are first categorized into a series of groups. Categories are determined from any number of different factors such as geographical locations of the search results, amenities, hours of operation, etc. [2]

2.2 GIS and Public Health:

Our emphasis will be on applications of GIS in public health and crime occurrence location wise which has proved itself to a GIS approach. To put it differently, the question here being addressed is what role GIS has and could have played so far in studies focusing on health issues.

GIS is gaining its popularity in terms of solving different problems of public health. An automated system can be created by using GIS which can be useful for capture, storage, retrieval, analysis and display of spatial data. A precise location of a particular area known as ‘pinpoint’ is essential to maximize the benefits of GIS for public health research. A point data could consist of individual patients' home or occupational addresses or where individuals came into contact with the disease is very important for health-related problems origins [10].

Although plots of data and other graphical displays are among the fundamental tools for analysts in general, for a spatial analyst, visualizing spatial data usually means using a map [11]. Disease mapping is one of the branches of geographical epidemiology fulfilling the need to create accurate maps of disease morbidity and mortality [12]. There are two main classes of disease maps for areal data: maps of standardized rates and maps of statistical significance of the difference between disease risk in each area and the overall risk averaged over the whole map [13] [14].

2.3 GIS and Crime Analysis:

Crime analysis is characterized as an arrangement of methodical, systematic procedures coordinated at giving auspicious and –relationships to help the operational and regulatory workforce in arranging the arrangement of assets for the counteractive action and concealment of criminal exercises, supporting the investigative procedure, and expanding worries and the leeway of cases. It underpins various division capacities including watch arrangement, uncommon activities, and strategic units, examinations, arranging and research, crime avoidance, what're more, authoritative administrations.

Criminological research has explained that crime can spread through local environments via a contagion-like process [16] For example, burglars will repeatedly attack clusters of nearby targets because local vulnerabilities are well known to the offenders (Bernasco and Nieuwbeerta 2005)[17]. A gang shooting may incite waves of retaliatory violence in the local set space (territory) of the rival gang (Tita and Ridgeway 2007; Cohen and Tita 1999) [18] [19]. The local, contagious spread of crime leads to the formation of crime clusters in space and time [15]. There is an association with criminality on offender characteristics such as the social status, peer group associations, biology and also place acknowledge the role of individual characteristics [20].

Computerized crime mapping has become a very important focus of innovation in policing. ‘Diffusion of innovations’ pioneered by Everett Rogers is a general approach to examine the diffusion of computerized crime mapping drawing upon the desired location [21]. Geographical studies reveal that crime is often concentrated in clusters, which in the literature are called hotspots [22]. Hotspots allow local law enforcers to determine the areas of greatest need. Hotspots have two different directions, one has more theoretical directions and another one is techniques for the detection of crime hotspots. At the stage, the techniques for detecting crime hotspots are

both definitive and easily applicable [23]. Existing hotspot mapping methods can be essentially divided into three main categories: point mapping, choropleth mapping, and kernel density estimation (KDE) [24] [25] [26].

Big data and predictive analysis have recently been using to impact social sciences and humanities in general [27] and criminology in particular [28]. In the context of crime analysis, the large amount of crime data available in police databases can be considered a valuable source of big data, which we can use to gain useable new insights and knowledge on current and emerging crime trends and patterns [29]. The application of advanced statistical methods to obtain this intelligence from big data is commonly referred to as predictive analysis. The use of predictive analysis in criminological applications is often referred to as predictive policing [30]. It can be defined as: "the use of historical data to create a spatiotemporal forecast of areas of criminality or crime hot spots that will be the basis for police resource allocation decisions with the expectation that having officers at the proposed place and time will deter or detect criminal activity" [31] [32].

III. RESEARCH METHODOLOGY

3.1 Collecting the data

3.1.1 Collecting the news as text file.

There were three ways of collecting news for this methodology. The very first of these three ones was to collect the news from different online news portals as a text format. The news could be in a different language. It could be in English and any other language. The news was collected in (.txt) format. This file can also be found from different types of websites. There was two way you could collect the text file. One was manually collected data from a different site and stored in the database. Another way of collecting text file was automatic. Give a URL in the system and then the system collected the data from that webpage if there was a text paragraph on that webpage. For this, we need to use jQuery. By using jQuery we could collect the text data very easily from an HTTP tag. The audio file generally stores in the text area or text field tag. So we collected all the collected txt file data from that site and try to extract our desired audio file from that page.

3.1.2 Collecting the news as audio file.

The next way of collecting the news was to retrieve information from the audio file format. The audio file can be collected from different TV news reports or any other source. This file can also be found from a different type of website. There was two way you could collect the audio file. One was manually collected data from a different site and stored in the database. Another way of collecting audio files was automatic. Give a URL in the system and then the system collected the data from that webpage if there was audio on that webpage. For this, we need to use jQuery. By using jQuery we could collect the audio data very easily from an HTTP tag. The audio file generally stores in the file tag. So we collected all the collected audio file data from that site and try to extract our desired audio file from that page.

3.1.3 Collecting the news as video file.

The next way of collecting the news was by collecting information from the video file format. Video reports can be found from different TV channel websites or different video sharing websites like YouTube. The video file URL is uploaded to the site and it could download the file and process the file for the next step. The system could also collect the video file HTTP tag.

3.1.4 Extracting the audio file from video file.

The video clip file could not be used for processing the data for the next step. For that, at first, the video must be converted to a text file. For converting the video file into a text file at first it must be converted into an audio file. For converting the video into audio, moviepy was used. Moviepy is a python module for video editing, which could be used for basic operations (like cuts, concatenations, title insertions), video compositing, video processing or to create advanced effects. It can read and write the most common video formats including GIF. In moviepy, at first the video file location was served and then it converted the video file into an audio file and it stored the audio file.

```
import moviepy.editor as mp
clip = mp.VideoFileClip("World News Today - BBC News Channel + BBC World News Aug 13.mp4")
clip.audio.write_audiofile("theaudio.mp3")
```

3.1.5 Converting the audio file into text file.

The text file was used for analyzing the data. So the audio file must be converted into a text file for analyzing the data. For converting the audio to a text file, the Speech Recognition module was used. SpeechRecognition is a python module for converting live audio to text and also audio file to text file.

```
import speech_recognition as sr

r = sr.Recognizer()

audio = 'BBC World News Business Headlines Sept 6 C, 2016 Subtitled.flac'

with sr.AudioFile(audio) as source:
    audio = r.record(source)
    print ('Done!')

try:
    text = r.recognize_google(audio)
    print (text)

    result = "{}".format(text)

    with open("Converted_audio_to_text.txt","w") as f:
            f.write(result)

except Exception as e:
    print (e)
```

3.1.6 Translating the text any language to English.

The text file could be in any language. But for analyzing the text file English language was used. For

that, the text file must be in the English language. For that, the text file must be translated into English. For translating the text file text blob module was used. Textblob is a python module which is used for processing the textual data. It provides a simple API for fibbing into common natural language processing (NLP) tasks such as part of speech tagging, noun phrase extraction, sentiment analysis, classification, translation and more. For translating the text file, the file must save the encoding mode as utf-8. Some language does not support the ANSI encoding. For that utf-8 encoding technology was used for saving the file. Textblob automatically detects the language of the text and then it was translated the text file into English. Like if the text file in the Bengali language the text blob at first detected the Bengali language and then it automatically translated the text into English.

```
from textblob import TextBlob
import io

file=io.open('bangla.txt','r',encoding="utf-8")
content=file.read()
file.close()
blob=TextBlob(content)
result=blob.translate(to='en')
print(result)
translate_copy=str(result)
text_file=io.open("bangla_en.txt","w",encoding="utf-8")
text_file.write(translate_copy)
text_file.close()
```

3.2 Analyzing the data

3.2.1 Extracting keywords from the file.

Keywords are needed for detecting the crime and disease in any particular area. For this step, we are using the "String Comparator". String Comparator has some different techniques. We are using "Approximate string matching". In this technique, if 80% of the string is matched with reference string then it counts that as match string. For reference string, we created some string dictionaries. Such as crime name dictionary, disease name dictionary, location dictionary, age dictionary, date pattern, day dictionary, etc. These dictionaries help us to fetch the necessary data from the text files. For generating victim name we use part of speech technique. Test results show that it works almost 80% for fetching the necessary data from the text files.

For example:" SYLHET: Two workers of Palli Bidyut were killed and two others sustained serious injuries as a passenger bus rammed a handcart in Osmaninagar Upazila of Sylhet on Saturday noon (Sep 22).The deceased were identified as Abdul Hai(35), Rashed Ahmed(20). The injured workers are Rajit Ali, 28 and Mamun Mia, 22. Eyewitnesses said the Palli Bidyut workers were heading towards Sylhet pushing a handcart loaded with electric poles while a Sherpur-bound bus (Shah Paribahan, Sylhet-J-110344) ran them over in Gadia area. Two workers died on the spot. Moulvibazar Sherpur highway zone sub-inspector (SI) Kamrul Islam told Banglanews that Fire Service members recovered the bodies and sent them to Osmani Medical College and Hospital. The injured men were also admitted to the hospital."

Location {SYLHET, Osmaninagar}

Crime {killed, injuries, deceased, pushing}

Date {Saturday, Sep 22}

Other {sub-inspector, Osmani Medical College}

Name {Abdul Hai, Rashed Ahmed, Rajit Ali, Mamun Mia, Shah Paribahan, Kamrul Islam}

Fig: Keywords from the text file

3.2.2 Finding the pattern among the keywords for finding the crime or disease of any particular location. For detecting the pattern of crime and disease we are using "Association rule mining". In association rule mining there is if some activity occurred then it can predict that other activity will also occur. Here we are using the keywords that are fetched from the text files for predicting the crime or disease. Here we make two sets of data for association rule. The first set contains the data from the previous step and the second step contain the data about crime occurred or not. For example:

{Sylhet, killed, Saturday, Osmani medical college, shah Paribahan} => {Crime occurred}

We set some rules for giving the result positive. The set of data should contain a location, crime or any criminal activity and also some other rules. If any data set can fulfil all the rule then the system can only tell that this crime has occurred in these areas.

3.2.3 Stored the data in the database.

If all the criteria fulfilled then this file marked as a crime occurred file. For generating result we need to store those data in the database. We are using a MySQL database for storing this information. We create some tables within the database and store those for future purposes.

3.3 Generating the result

3.3.1 Generates the result according to user inserted location. For generating the result the result user inserted his/her desired location and the system process the stored data and found out the location that was user's choice and then generates some reports for the user. For the detecting the location at first collect all the location that was stored in the database and the matched all the given location with the collective data and then found out the desired location. Then the system collected all the information for that location from the database and generated the following the results.

3.3.2 Crime rate of that area. The first result was about generating a report of crime of that particular area. First the system was found the location from the database and clustered the location from rest of the location and then found out all the occurred crime from the database. It collected all type of crime from the database. Then it calculated the total crime occurred from the database for all the location. Then

it calculated the percentage by (desired location crime occurrence/total crime occurrence) and then it generated a crime rate of that area by calculating those data.

3.3.3 Disease rate of that area.

The first result was about generating a report of disease of that particular area. First the system was found the location from the database and clustered the location from rest of the location and then found out all the affected disease from the database. It collected all type of disease from the database. Then it calculated the total disease affected from the database for all the location. Then it calculated the percentage by (desired location disease affection/total disease affection) and then it generated a disease rate of that area by calculating those data.

3.3.4 Specific crime or disease percentage list of that area. This system can also give some report about specific crime or disease. User gave information about the area. Then the system detected the location and collected all the information about that location from the database. After getting all the information then the system generated a list of all the crime occurred of that particular area. Such as if a user gave input the Dhaka as location input then the system tried to find all the crime list of Farmgate, Dhaka and then the system generate list of all crimes of Farmgate, Dhaka and give a list of all crime with the percentage.

3.3.5 Prediction of crime occurrence.

From the previous stored data, this system also could give prediction of crime occurrence. This system searched the location in the stored data and then it checked all the crime date and time. Then it could visualize how rapidly crime occurred on that particular area.

3.3.6 Prediction of affecting disease.

From the previous stored data, this system also could give prediction of disease occurrence. This system searched the location in the stored data and then it checked all the time and date when the disease was spread of that particular area. Then it could visualize how rapidly disease affected on that particular area.

IV. RESULTS AND DISCUSSIONS

4.1 Crime rate of that area

Here we visualized a graphical view of the result. From this result we could see that the crime percentage of Farmgate, Dhaka city is 75.2%. In this result it extracted only the rate of crime of user's desired location.

They can input a location in the system then the system is started to generate the rate of crime of that area. Then it collected the data from the database and calculate the result. After calculating the result it will create a pie chart and insert the data. After that the result will visualize in the pie chart.

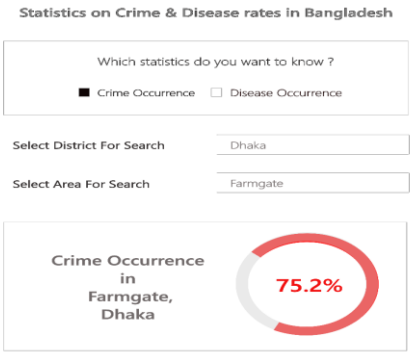


Figure 1: Rate of crime

4.2 Disease rate of that area

Here we visualized a graphical view of the result. From this result we could see that the disease occurrence percentage of Dhaka city is 34%. In this result it extracted only the rate of disease of user's desired location. They can input a location in the system then the system is started to generate the rate of disease of that area. Then it collected the data from the database and calculate the result. After calculating the result it will create a pie chart and insert the data. After that the result will visualize in the pie chart.

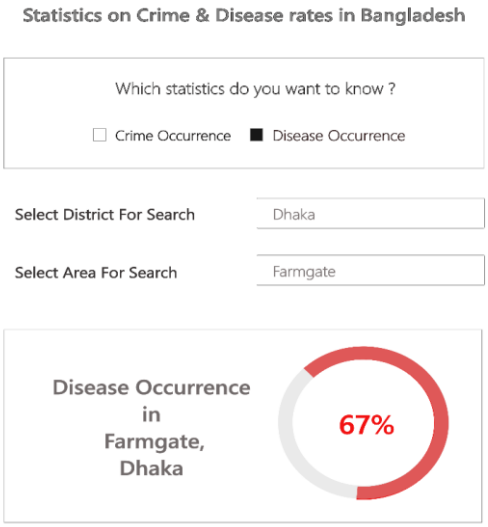


Figure 3: Rate of disease

4.3 Specific crime percentage list of that particular area

From the previous result we only know about crime rate of desired location. But we did not know anything about which crimes were occurring in that particular area. In this result it generates the list of crime for that location. It generates the result with the crime list with the percentage of the crime. For creating this result the system at first search the desired location in the database. Then it collected all the crime occurred in that location. Then it generates the crime list of the desired location. For creating the

list wise percentage of the crime, at first it calculates the total crime and also the number of specific crime occurrence number and then it calculates the percentage of the crime. After calculating the result it populates the data and give us graphical view of this result. From this result we can see that it generated a list of all the crime occurred in the Farmgate, Dhaka city and gave the percentage of all the crime in the Farmgate, Dhaka city.

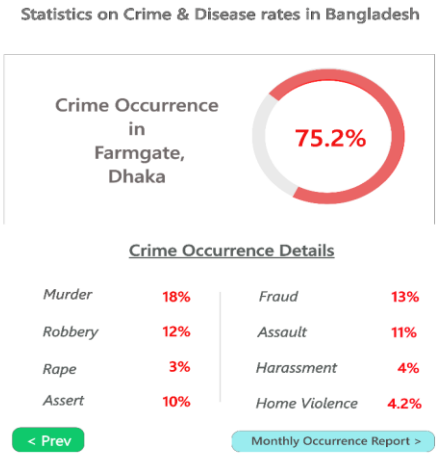


Figure 3: Area Based Crime Report

3.4 Specific disease category list of that particular area

From the previous result we only know about disease rate of desired location. But we did not know anything about which diseases were affecting in that particular area. In this result it generates the list of disease for that location. It generates the result with the disease list with the percentage of the disease. For creating this result the system at first search the desired location in the database. Then it collected all the disease affected in that location. Then it generates the disease list of the desired location. For creating the list wise percentage of the disease, at first it calculates the total disease and also the number of specific disease occurrence number and then it calculates the percentage of the disease. After calculating the result it populates the data and give us graphical view of this result. From this result we can see that it generated a list of all the disease affected in the Farmgate, Dhaka city and gave the percentage of all the disease in the Farmgate, Dhaka city

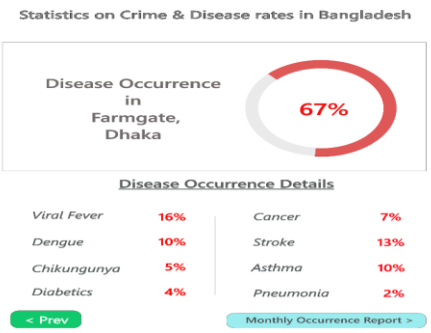


Figure 4: Area based disease report

4.5 Crime Prediction Report

For predicating the crime in a particular area we need to know about previous result. In the database we stored all the data of previous month. So that it can easily collects the previous data of any area what it needs. It calculates all the months' crime rate from the database. Then it populates the data in the list and give us a visualize view of the result .This result showed that how the crime percentage of a particular area on previous month. Here you can see the crime rate of Dhaka city for last 12 months. From this result a user can easily think about the crime prediction.

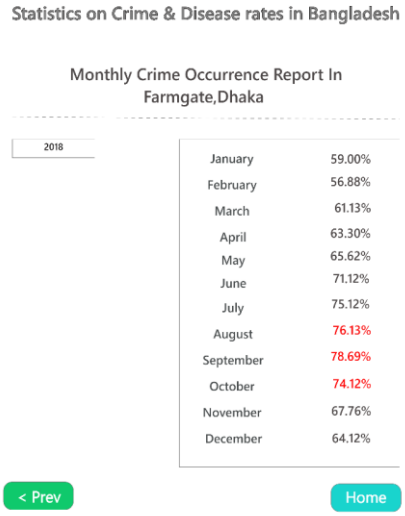


Figure 5: Crime prediction report of Dhaka

4.6 Disease Prediction Report

For predicating the disease in a particular area we need to know about previous result. In the database we stored all the data of previous month. So that it can easily collects the previous data of any area what it needs. It calculates all the months' disease rate from the database. Then it populates the data in the list and give us a visualize view of the result .This result showed that how the disease percentage of a particular area on previous month. Here you can see the disease rate of Farmgate, Dhaka city for last 12 months. From this result a user can easily think about the disease prediction.

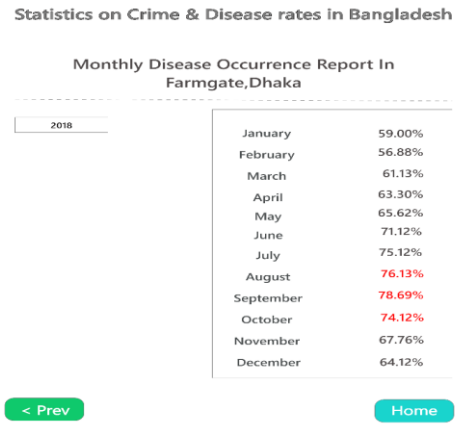


Figure 6: Disease prediction report of Dhaka.

V. CONCLUSIONS AND FUTUREWORK

This paper describes an approach on how can be geographically tag data used for analyzing the crime and disease of any area and predicting the future condition of crime and disease of that particular area. A prototype system has been created for understanding how this data can be used for the betterment of our society. This system provides a model of how can this data be collected from the website. The geographical content from web resources is currently an area of increased interest and research. Given that a large proportion of what people do is based around a location and the many web resources contain some kind of geographical context. In this paper, the web resource is the source of data. Different type of geographical data is being collected from a website. In this paper, three types of resources are being collected from the website. Text data, audio data, and video data are those resources that are the source of this paper result. These data are processed in different kinds of functionality and stores in the database for generating the result. In this data, the system tries to find any geographical tag data if it finds any type of data then it automatically stored that location and also try to find any information about crime or disease of that area.

The prototype system has not implemented in generating the result. The prototype system has tried to show how it can be done for generating the desired result. The prototype system has just shown that it can be possible to generate a prediction for crime or disease only from a different type of web resource that is stored on the website. From this paper, we also know that web resources can also be useful for decreasing crime in any area or spreading disease in any area.

In the future, there is a number of the new way we can show result and also try to find the easiest way to generate the result. We try to be a more precise report for any location and also we can build a prototype system for the whole model for showing the usefulness of the system. In the future, we will build a more resourceful dictionary for extracting the location more precisely and also detecting the crime or disease. Current the system is not that intelligent. In the future, we try to make this system more intelligent so that it can automatically generate our desired result. The system will be more users friendly and also more resourceful than this time. We try to bring the newest technology for detecting the geographical data more precisely.

REFERENCES

[1] Muhammad Ichsan Ali, Muhammad Rais, "Spatial Pattern of Crime with Geographic

[2] Information System (GIS) in Makassar, Indonesia" ISSN (Online): 2319-7064

[3] 2.Method and system for selectively presenting database results in an information retrieval system [2000-07-05].

[4] MILLER, H. J. (1991). Modelling accessibility using space-time prism concepts within geographical information systems. International Journal of Geographical Information Systems, 5(3), 287–301.

[5] Janssen R., Rietveld P. (1990) Multicriteria analysis and geographical information systems: an application to agricultural land use in the netherlands. In: Scholten H.J., Stillwell J.C.H. (eds) Geographical Information Systems for Urban and Regional Planning. The GeoJournal Library, vol 17. Springer, Dordrecht

[6] TROTTER, C. M. (1991). Remotely-sensed data as an information source for geographical information systems in natural resource management a review. International Journal of Geographical Information Systems, 5(2), 225–239.

[7] Bahaire, T., & Elliott-White, M. (1999). The Application of Geographical Information Systems (GIS) in Sustainable Tourism Planning: A Review. Journal of Sustainable Tourism, 7(2), 159–174.

[8] Church, R. L. (2002). Geographical information systems and location science. Computers & Operations Research, 29(6), 541–562

[9] Terry Holmes; Glynn Mangold; Fred Miller "Geographic Information Systems in Marketing"

[10] O'Dwyer, L. A., & Burton, D. L. (1998). Potential meets reality: GIS and public health research in Australia. Australian and New Zealand Journal of Public Health, 22(7), 819–823. doi:10.1111/j.1467-842x.1998.tb01500.x

[11] Bailey TC, Gatrell AC. Interactive spatial data analysis. Harlow: Longman, 1995

[12] Lawson AB, Bohning D, Biggeri A, et al, eds. Disease mapping and its uses. disease mapping and risk assessment for public health. Chichester: Wiley, 1999.

[13] Clayton D, Bernardinelli L. Bayesian methods for mapping disease risk. In: Elliott P, Cuzik J, English D, Stern R, eds. Geographical and environmental epidemiology–methods for small area studies. Oxford: Oxford University Press, 1996.

[14] Rezaeian, M., Dunn, G., St Leger, S., & Appleby, L. (2007). Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. Journal of Epidemiology & Community Health, 61(2), 98–102. doi:10.1136/jech.2005.043117

[15] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., & Tita, G. E.

(2011). Self-Exciting Point Process Modeling of Crime. Journal of the American Statistical Association, 106(493), 100–108. doi:10.1198/jasa.2011.ap09546

[16] Johnson, S. D. (2008), "Repeat Burglary Victimisation: A Tale of Two Theories," Journal of Experimental Criminology, 4, 215–240. [100,103,105]

[17] Bernasco, W., and Nieuwbeerta, P. (2005), "How Do Residential Burglars Select Target Areas? A New Approach to the Analysis of Criminal Location Choice," The British Journal of Criminology, 45, 296–315. [100,104]

[18] Tita, G., and Ridgeway, G. (2007), "The Impact of Gang Formation on Local Patterns of Crime," Journal of Research on Crime and Delinquency, 44 (2), 208–237. [100]

[19] Cohen, J., and Tita, G. (1999), "Spatial Diffusion in Homicide: Exploring a General Method of Detecting Spatial Diffusion Processes," Journal of Quantitative Criminology, 15 (4), 451–493. [100]

[20] SHANE D. JOHNSON (2010). A brief history of the analysis of crime concentration. European Journal of Applied Mathematics, 21, pp 349-370 doi:10.1017/S0956792510000082

[21] Weisburd, D., & Lum, C. (2005). The Diffusion of Computerized Crime Mapping in Policing: Linking Research and Practice. Police Practice and Research, 6(5), 419–434. doi:10.1080/15614260500433000

[22] Wang, D., Ding, W., Lo, H., Morabito, M., Chen, P., Salazar, J., & Stepinski, T. (2013). Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns. Computers, Environment and Urban Systems, 39, 93–106. doi:10.1016/j.compenvurbsys.2013.01.008

[23] Ratcliffe , Jerry H(2004). The Hotspot Matrix: A Framework for the Spatio- Temporal Targeting of Crime Reduction. Police Practice and Research (10.1080/1561426042000191305)

[24] Eck, J., Chainey, S., Cameron, J., Leitner, M., & Wilson, R. (2005). Mapping crime: Understanding hot spots . National Institute of Justice. ESRI (2011). Arcgis desktop: Release 10

[25] Williamson, D. McLafferty, S., McGuire, P., Ross, T., Mollenkopf, J., Goldsmith, V., et al. (2001). 9 tools in the spatial analysis of crime.Mapping and Analysing Crime Data: Lessons from Research and Practice , 187, CRC.

[26] Boba, R. (2005). Crime analysis and crime mapping . Sage Publications, Inc.

[27] Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. Big Data & Society, 1(1), 1e12.

[28] Chan, J., & Moses, L. B. (2015). Is big data challenging criminology? Theoretical Criminology. http://dx.doi.org/10.1177/1362480615586614.

[29] Rummens, A., Hardyns, W., & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. Applied Geography, 86, 255–261. doi:10.1016/j.apgeog.2017.06.011

[30] Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. RAND Research Reports. Santa Monica, CA: RAND Safety and Justice Program.

[31] Ratcliffe, J. (2010). Crime mapping: Spatial and temporal challenges. In A. R. Piquero, & D. Weisburd (Eds.), Handbook of quantitative criminology. New York: Springer Science.

[32] Rummens, A., Hardyns, W., & Pauwels, L. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. Applied Geography, 86, 255–261. doi:10.1016/j.apgeog.2017.06.011

[33] Uggen, C., & Larson, R. (2017). Is the Public Getting Smarter on Crime? Contexts, 16(4), 76–78. doi:10.1177/1536504217742400

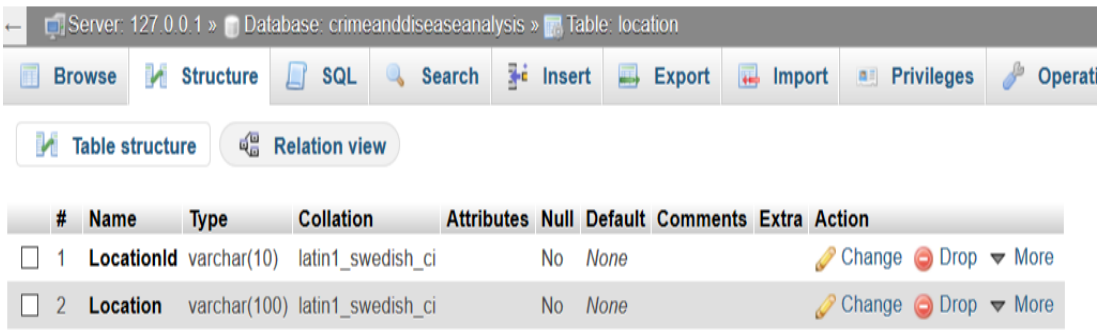
[34] Bacal, F., Silva, C. P., Pires, P. V., Mangini, S., Fiorelli, A. I., Stolf, N. G., & Bocchi, E. A. (2010). Transplantation for Chagas' disease: an overview of immunosuppression and reactivation in the last two decades. Clinical Transplantation, 24(2), E29–E34. doi:10.1111/j.1399-0012.2009.01202.x

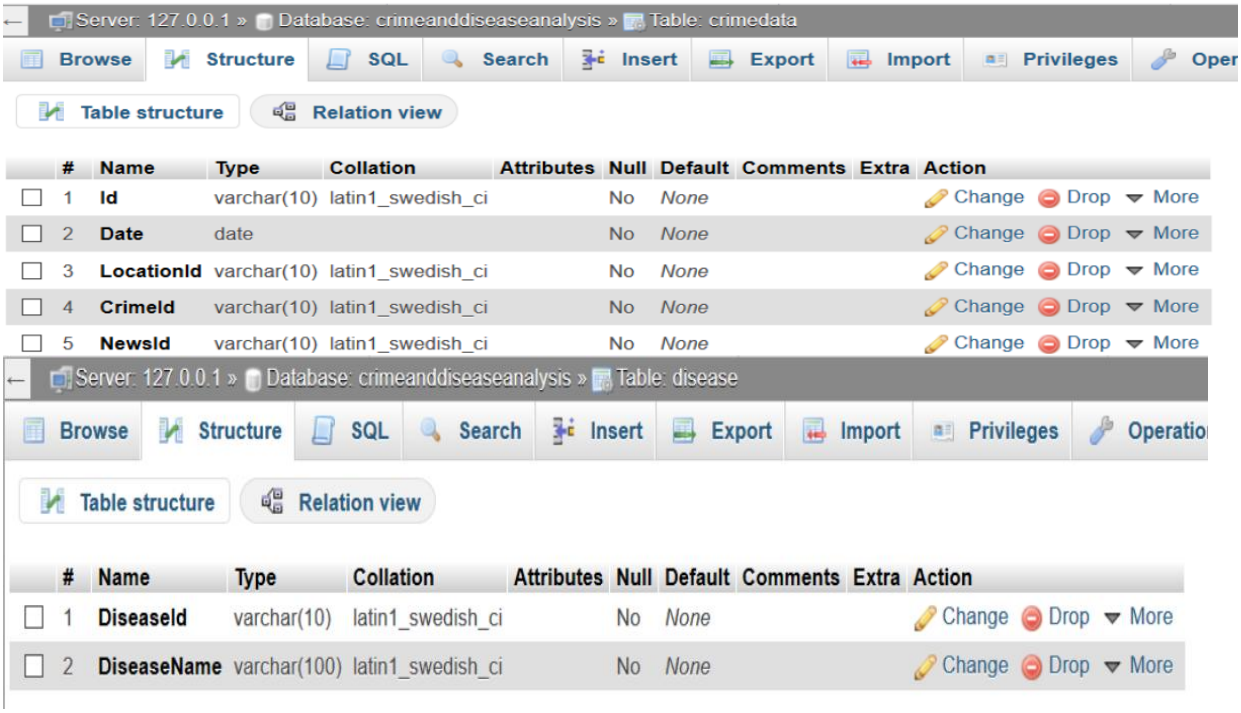
[35] Jess, T., Riis, L., Vind, I., Winther, K. V., Borg, S., Binder, V., … Munkholm, P. (2007). Changes in Clinical Characteristics, Course, and Prognosis of Inflammatory Bowel Disease during the Last 5 Decades: A Population-Based Study from Copenhagen, Denmark. Inflammatory Bowel Diseases, 13(4), 481–489. doi:10.1002/ibd.20036

[36] Mangili, A., & Gendreau, M. A. (2005). Transmission of infectious diseases during commercial air travel. The Lancet, 365(9463), 989–996. doi:10.1016/s0140-6736(05)71089-8

APPENDIX:

Database:





Analyzing the data:

```
import re
districts=['Dhaka','Barguna','Barisal','Barishal','Bhola','Jhalokati','patuakhali',
          'Pirojpur','Bandarban','Brahmanbaria','Chandpur','Chittagong','Chattogram','Comilla',
          'Feni','Khagrachhari','Lakshimpur','Noakhali','Rangamati','Faridpur',
          'Gazipur','Gopalganj','Kishoreganj','Madaripur','Manikganj','Munshiganj',
          'Narayanganj','narsingdi','Rajbari','Shariatpur','Sylhet','Sunamganj',
          'Moulvibazar','Habiganj','Thkurgaon','Rangpur','Panchagarth','Nilphamari',
          'Lalmonirhat','Kurigram','Gaibandha','Dinajpur','Sirajganj','Rajshahi',
          'Pabna','Chapai Nawabganj','Natore','Naogaon','Joypurhat',
          'Bogra','Sherpur','Netrokona','Mymemsingh','Jamalpur',
          'Satkhira','Narail','Meherpur','Magura','kustia',
          'Khulna','Jhinaidah','Jessore','Chuadanga','Bagerhat','Tangail',"Cox'sBazar","Cumilla"]

Crimes=['Murder','assault','Manslaughter','Cruelty','homicide',
'Larceny','Arson','Burglary','Fraud','Vandalism','Treason','Rape','Sexual','Kidnapping','Perjury',
       'Blackmail','Affray','Burglary','Bigamy','Consorting','Drunk','Forgery','Incest','Infanticide',
       'Manslaughter','Perjury','Riot','Robbery','Rout','Sabotage','stalking','Tax evasion','Damage','collapse ',
       'polluted ','pollutes','kills','killed','killing','rape','raped','win','won','won','lose','lost','Rohingya ',
       'lawless','bail','illness','drowned','victims'','bodies','body','dead','die','died','accident','accidents','police']
diseases=['AIDS','HIV','ALS','Alzheimer','Infection','Anthrax','Vaccination',
         'Antibiotic','Arthritis','Meningitis','Asthma','Autism','Bacterial',
         'Balantidium','Flu','Birdflu','Blood','Cancer','Diarrhea','Bronchitis',
         'virus','Syndrome','Chest','Chickenpox','pox','Chikungunya','Obesity',
         'Cold','Dengue','Diabetes','Ebola','Fever','stroke','brain','heart',
         'Surgical','Hepatitis','Herpes','Genital','Pressure','Hypertension',
         'Hyperthermia','Hypothermia','Infertility','Influenza','Malaria','Polio',
         'Pneumonia','Respiratory','Gonorrhea','Skin','STDs','TB','Typhoid']
for x in range(1,136):
    file=open('news/news{0}.txt'.format(x),'r')
    content=file.read()
    flag=0
    for c in content.split():
        for d in districts:
            c=c.replace('.', '')
            c=c.replace(',', '')
            c=c.replace(':', '')
            if c.lower()==d.lower():
                flag=1
                district=d
                break
```

```
        if(flag==1):
            check=0
            C_flag=0
            resultfile=open('crime.txt','a')
            for c in content.split():
                for k in Crimes:
                    if c.lower()==k.lower():
                        check=check+1
                        if(check==1):
                            resultfile.write(district)
                            C_flag=1
                            resultfile.write("     ")
                        Crime=k
                        resultfile.write(Crime)
                        resultfile.write("     ")
        if(flag==1 and C_flag==1):
            resultfile.write('\n')
            resultfile.close()

        if(flag==1):
            check1=0
            D_flag=0
            resultfile=open('diseases.txt','a')
            for c in content.split():
                for l in diseases:
                    if c.lower()==l.lower():
                        check1=check1+1
                        if(check1==1):
                            resultfile.write(district)
                            D_flag=1
                            resultfile.write("     ")
                        disease=l
                        resultfile.write(disease)
                        resultfile.write("     ")
        if(flag==1 and D_flag==1):
            resultfile.write('\n')
            resultfile.close()
area=input("please input any area:")

totalinput = len(open('crime.txt').readlines(  ))
count=0
file=open('crime.txt','r')
content=file.read()
for c in content.split():
    if c.lower()==area.lower():
        count=count+1
result=(count/totalinput)*100
result=round(result,2)
print("Crime rate of ", area," is : " ,result," %")
totalinput = len(open('diseases.txt').readlines(  ))
count=0
file=open('diseases.txt','r')
content=file.read()
for c in content.split():
    if c.lower()==area.lower():
        count=count+1
result=(count/totalinput)*100
result=round(result,2)
print("Disease rate of ", area," is : " ,result," %")
```

★★★