proceedings

Role of Geographical Information System and Principal Component Analysis for identifying vulnerable zone of surface water epidemiology

Chandramohan Karuppaiah[1], Vijaya Ramachandran[1] and Sangu Muthuraju[2]

[1] Department of Environmental Remote Sensing And Cartography, School of Earth and Atmospheric Sciences, Madurai Kamaraj University, Madurai-625021, India
[2] Department of pharmacology and pharmaceutical, 4849 Calhoun Rd, University of Houston, Texas, 77204, U.S.A.
* Correspondence: drcmresearchlab@gmail.com

Abstract : The maintaining of drinking water quality, access to clean urban surface water are the most challenge resulting in water-borne diseases. Although, the developing countries, the Geographical Information System (GIS) technology has not been yet systematically utilized. Therefore, the present study aimed to establish the parameters for assessing surface water quality and how they would cause health issues in the study area. This study reveals the surface water quality variations by using principal component analysis (PCA) techniques with the help of Landsat 8 Operational Land Imager (OLI) satellite imagery. Regular monitoring of environmental quality database produced by (GIS) could manage information from various sources such as, domestic, industrial, recreational activities, point and non-points sources etc. This monitoring activity could make spatial correlations with epidemiology data about time and space distribution of water-borne diseases. The spatial correlation of water contamination tank and it's surroundings of land use / land cover (LULC) activities were calculated by the technique of buffer analysis. Using Medical GIS, could easily detect the circulation and spread of disease across the geographic regions which could be used for planning, policy making, water resource protection and avoid contamination. The study area also included the industrial and residential areas of Madurai urban region, Tamil Nadu, India, to validate ground truth verification. The study would help to reveal the disease monitoring, surveillance systems, improving the distribution of health resources by predicting available health care accessibility, the source of pollution and it impacts on public health. Especially in Madurai need more attention to epidemiology. Because, the 11,132 epidemic disease affected cases were recorded in 2019[1]. In every year government of India spends more money for this type of epidemiological spread; especially in Tamil Nadu spend 8868 lakhs of money. WHO also revealed that the developing countries are highly affected by the epidemic diseases.

Keywords: Remote Sensing; Medical geographical information system (GIS), Surface water; epidemiology; vulnerability; NDWI; Principal Component Analysis (PCA), LULC; Buffer

1. Introduction

In 2016, according to the Global Health Observatory of World Health Organization (WHO) reported, 1,32,121 cholera cases and 2420 deaths were recorded worldwide. Health & family welfare Department reported the Tamil Nadu is endemic for Acute Diarrhoeal Diseases with sporadic outbreak of cholera in most of the districts throughout the year, and in epidemic proportions during the rainy seasons and peak summer periods. Epidemiology is the scientific study [3] of disease spread by particular environmental issues like waterborne and airborne diseases, etc. Remote Sensing (RS) and GIS are very effective tools for analyzing the spatio temporal variations [4] to control of diseases. Remote Sensing is an excellent technology to observe and detect the changes occurs on earth surface features [5, 6] which are associated with diseases.

Geographic information system (GIS) is a Database Management System (DBMS) tool for the collection of data from many sources by various methods [7], which is also accomplishing of organizing, storing, retrieving, analyzing and presenting the spatial data [8].The ERDAS Imagine 9.1v image processing software and the Arc GIS 10.1v spatial analysis software are used for mapping, spatial analysis and image processing of the both non-spatial and spatial data [9].

2. Study Area and Data

Study area of Madurai city, Tamil Nadu, India, lay between the aerial extensions of 78°2'50.733"E 9°59'38.597"N and 78°11'55.885"E 9°49'23.73"N.(Figure 1-A). According to the Survey of India (SOI) toposheet showing in and around the study area of Madurai having more than 100 tanks, from these tanks 70 tanks are having greater in size of 15 acre. Less than 15 acres of tank already encroached by various construction activities. Remote sensing techniques of satellite image represented the tank with /without water. The urban settlement occupied by 25020 acres. The population of the city was 1,734,000 (Population Census 2011). The domestic waste water directly joins to the tank and stream.

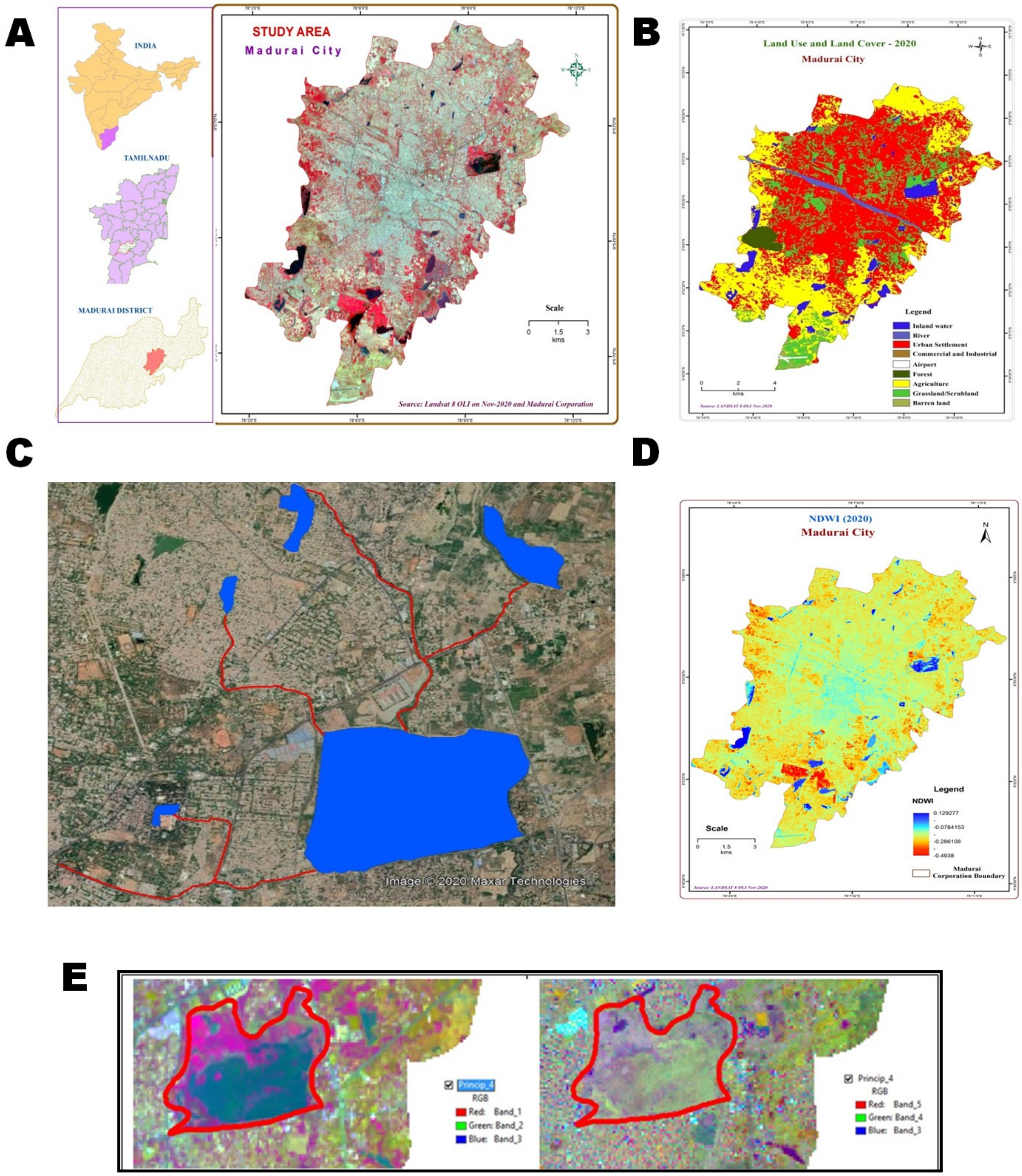

Figure.1 (A) Location of study area derived from India, Tamil Nadu, Madurai District and Madurai city boundary also viewed through satellite imagery with 30m resolution.**(B)** Land Use / Land Cover map using Digital Image Processing (DIP) of unsupervised classification techniques for area calculations of features.**(C)** Drainage network of Urban region. Blue colour shows the tank and red colour shows the canal connective to the tanks as well as open canal travel through the dense residential area. *Source: Google Earth pro – 2020. (***D)** Showing Normalized Differentiate Water Index using Spatial Analyst Tools of Raster Calculation techniques for differentiate the water density and other features. (**E)** (left image) PCA having band combinations of 5, 4, and 3 of tank showing the availability of water (dark blue) and vegetation (pink). € (right image) The same image PCA with the band combinations of 5, 4, and 3 indicate the quality variations of surface water fully reflect the similar pattern of vegetation because of the growth of algal bloom and other water living plants.

Satellite Sensor and Image Selection:

Collection of four-band multispectral Landsat 8 OLI imagery acquired on 26 July 2020 was downloaded from the Earth explorer, USGS (United States of Geological Survey) which are free of cost. The image acquisition date was selected to coincide for the post monsoon period of November collecting more amount water through rainfall. In each band or combinations of bands having unique characters of earth features based on it's observing and reflecting capacity of Ultra Violet (UV) visible, Near Infrared (NIR), and Infrared (IR) spectrum. Blue, Green, and Red band combination were used to classifying the LULC classification, band 3 and band 5 used for calculating NDWI, band 1 to band 5 used for calculating the PCA. Table.A shows clear explanations of the spatial resolutions and spectral characters of the image.

A

Image Type	Band	Wavelength (um)	Spatial Resolution (m)
Landsat 8 OLI	Band 1- Coastal aerosol	0.43 – 0.45	30
	Band 2- Blue	0.45 – 0.51	30
	Band 3-Green	0.53 – 0.59	30
	Band 4- Red	0.64 – 0.67	30
	Band 5- Near Infrared (NIR)	0.85 – 0.88	30
	Band 6- SWIR 1	1.57 – 1.65	30
	Band 7- SWIR 2	2.11 – 2.29	30
	Band 8- Panchromatic	0.50 – 0.68	15
	Band 9- Cirrus	1.36 – 1.38	30
	Band 10- Thermal Infrared1	10.60 – 11.19	100
	Band 11-Thermal Infrared 2	11.50 – 12.51	100
Source:USGS, Science for a changing world			

B

S_No	Classified features	Area in Hectare
1	Forest	240.66
2	Agriculture	4439.16
3	Grassland/Scrubland	2027.52
4	Barren land	587.52
5	Urban Settlement	7080.57
6	Commercial and Industrial	194.31
7	Inland Water	585.18
8	River Water	203.04
9	Airport	21.69
	Total	15379.65

C

S.No	Physiography	Area in Sqkm	Depth from the surface
1	Soil		
	Clay loam	58	Very deep >150 cm
	Clay	2	Moderately deep 75-100 cm
	Water body	1	Water body
	Clay	15	Moderately deep 75-100 cm
	Sandy clay loam	12	Deep 100-150 cm
	Clay	24	Moderately deep 75-100 cm
	Water body	17	Water body
	Clay loam	12	Very deep >150 cm
	Loamy sand	10	Moderately shallow 50-75 cm
	Sandy loam	3	Very deep >150 cm
	Total	153	
2	Geomorphology		Not available
	Pediplain	105	
	Denudational Hills	1	
	Alluvial Plain	38	
	Flood Plain	5	
	Structural Hills	1	
	Upland	4	
	Total	153	
3	Geology		Not available
	Intrusive Rocks (Archaean - Pre-cambrian)	101	
	Crystalline Rocks (Archaean - Pre-cambrian)	52	
	Total	153	

Table.1 Table: 1 (A) shows the details of sensor and bands, **(B)** shows the feature classification and its area calculation, **(3).** Physiographical characters and its aerial extension cover.

3. Objectives

To trace out the tanks / surface water bodies near by the settlement using satellite imagery

To identify the water quality index by NDWI

To identify the vulnerable zone of epidemiology of waterborne diseases

4. MATERIALS AND METHODS

Data Sets and Methodology

The availability of various data of primary and secondary data sets sources such as the primary data collected were the SOI toposheets of 1:50,000 scale for the consistent region, Landsat- 8 consistent path -154, row-053 LOR for year November 2020 from Earth Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. NIR band of Landsat 8 as water index was found more satisfactory in extracting water bodies compared to the multi-band water indexes. For this study, the satellite imagery was used with remote sensing techniques to identify the settlement density distribution and water spread area, NDWI method used to calculate the water index for surface quality variations within the tank, buffer analysis studied to identify the settlement or urban buildup arrangements in surrounding of tank/waterbody area.

5. Result and Discussion

5.1. Water Bodies and Settlement Density using Image Classification Method:

The study area of Madurai urban region having 8 major tank and 20 small tanks of surface water bodies identified by the satellite imagery cross verification to SOI toposheet. There is high efficiency with the water index approaches when multispectral image classification of 30 m resolution, the water indexes is Normalized Difference Water Index (NDWI) [10], Modified NDWI (MNDWI) [11]. These two methods used to calculate the water index, distinguish between water and building, the optimal threshold to extract water is highly subjective, and also varies with region and time. But in that the every water index has its main problem is the NDWI was poor at distinguishing between water and buildings, mountain shadows [12]. Statistical methods of unsupervised classifications are used for identifying the water bodies. These methods are more accurate than other methods, because they are derived from the real satellite imagery. Prior knowledge is applied in the unsupervised classification makes classifications (figure 1-B) by learning from image interpretation key. Every feature's aerial extension calculates from unsupervised classification methods and its calculated values given table.B.

5.2. Calculation of NDWI

The image statistical calculation of NDWI (figure 1-D) useful to analyze the distinguishing of feature based on the pixel values. It was calculated by the band 3 and band 5 of the imagery using the formula of:

$$NDWI = (Green – NIR) / (Green + NIR)) \quad (1)$$

According to the table.A Landsat 8 OLI Band 3 (green) – Band 5 (NIR)

5.3. Identifying the contamination and it's source:

The study area settlement located nearby the major tanks which store rain water for ground water recharge as well as every home having at least one borehole for domestic water consumption. The septic tank leaching create the Giardiasis, typhoid, Toxigenic E.coli gastroenteritis, salmonellosis, etc [13 & 14]. Canal and nearest settlement, commercial, industrial, recreational activities are involving the major contamination sources. When the contamination increases it support profuse growth of weed plants on the surface water body [15 & 16]. The canal path and residential or buildup identified (figure 1-C) easily around the water source using satellite digital image processing (DIP) techniques (figure 1-B).

5.3.1. Principle Component Analysis (PCA)

Principle Component method usually capture all the spectral information from the bands, and compressed/merged it all into principle components. Technically two methods followed to firstly extract the future most of spectral information followed by secondly the rest of the spectral information which can't be stored in first calculation. PCA is widely used in Machine Learning task. PC is new variables that are constructed as linear combinations of the initial variables. The figure 1-E (left image) PCA1 having band combinations of 5,4, and 3 of tank showing the availability of water (dark blue) and vegetation (pink). The same image of figure 1-E (right image) with the band combinations of 5, 4, and 3 indicate the quality variations of surface water fully reflect the similar pattern of vegetation because of the growth of algal bloom and other water living plants. From this PCA generate eigenvalues follows.

5.4. Physiography of Study area:

Elevation, Geological (figure 2-A), geomorphological (figure 2-B) parameters and also physiological characters of precipitation, wind is important to study the water contamination as well as epidemiological intensity [17]. Some soils (figure 2-C), such as clays, absorb less water at a slower rate than sandy soils. Soils absorbing less water result in more runoff overland into streams. Generally hard rock does not have primary porosity. Secondary porosity in the form of fractures and weathering only allow groundwater occurrence [18]. The extent of weathering was identified based on VES which indicated weathering up to a depth of 40 m from the ground surface. High rank is assigned for meta-basalt rock (Table. C), which has high amount of porosity and good groundwater potential.

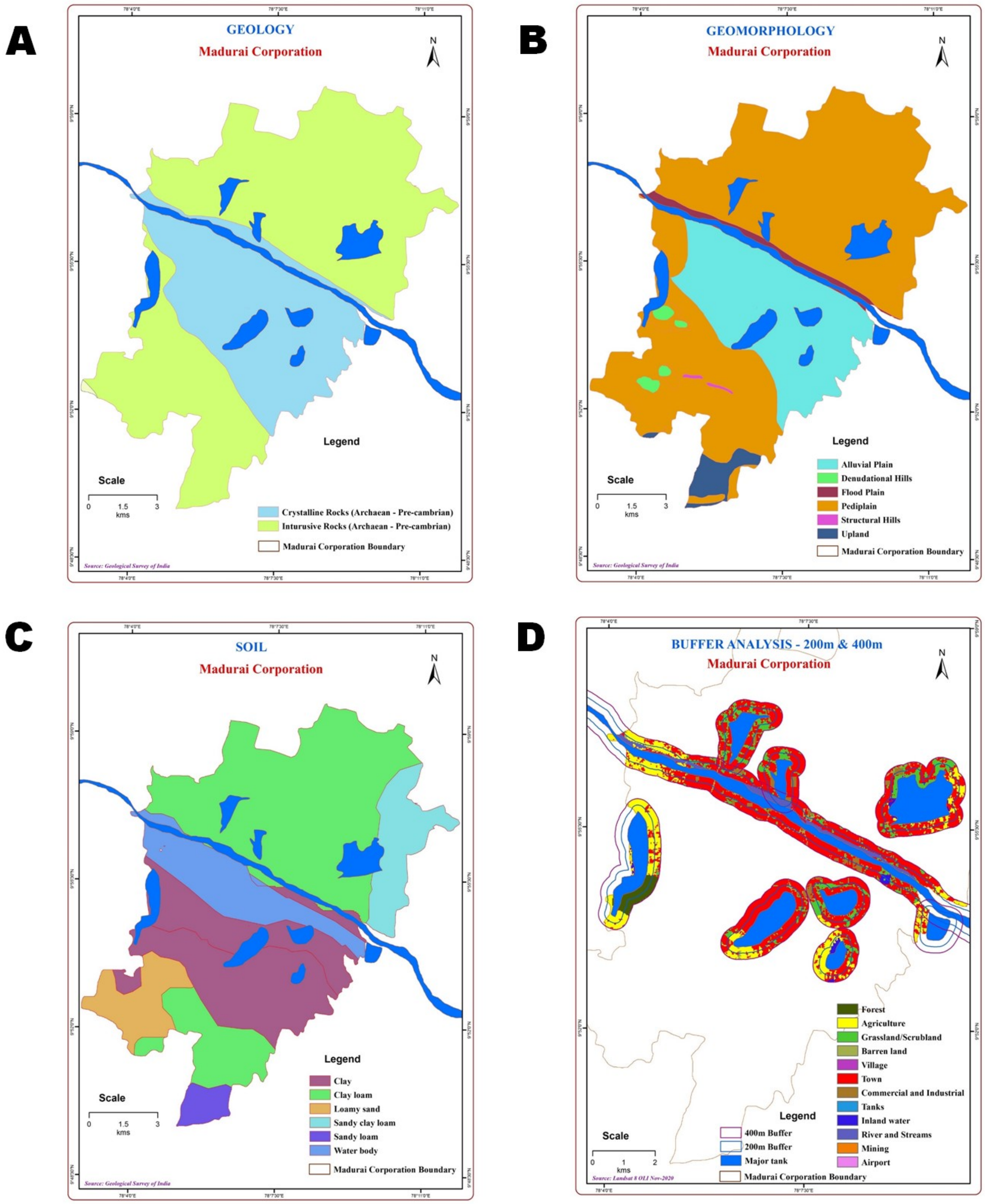

Figure 2. **Figure: (A)** Shows the type of geological characters available in the study area are Intrusive Rocks (Archaean - Pre-cambrian) and Crystalline Rocks (Archaean - Pre-cambrian). The aerial extension of these geological types intrusive is one of the two ways igneous rock can form. Igneous rocks tend to have low porosity and low permeability unless they are highly fractured by tectonic processes. Which cover maximum of 66% of study area. But the crystal or crystalline rock is a solid are arranged in a highly ordered microscopic structure, forming a crystal lattice that extends in all directions cover 34% of total study area. Crystalline rocks are having more capacity of water permeability rate when compare with the intrusive rock type as well as the Crystallinerocks spread major tank and river lying area. Which also increase the ground water contamination level. **(B)** shows the varied geomorphic characters in this study area Pediplain (68%), Denudational Hills (1%), Alluvial Plain (25%), Flood Plain (3%), Structural Hills (1%), and Upland (2%). From this observation alluvial plain is having more surface water permeability then the other geomorphic characters. Meanwhile al-

luvial plain cover 38 sqkm of 25% total study area and also present in the major tank and river lying area. Due to this reason the contamination possibilities are very high. **(C)** showing the soil types of clay (26.62%), clay loam (51.95%), sandy clay loam (9.74%) and water body (18%) are present in the study area. Northern part of the study area fully occupied by clay loam and sandy clay lo.am which are highly favorable to water percolation and the southern part of the study area less percolation of water than the other area. When conclude this northern part of the study area is highly vulnerable than the other parts and settlements/build-ups are densely spread over in northern part. Table.3 gives an exact level of depth of the soil layer from the earth surface. **(D)** Shows the major tank and river present in the study area which they are closely relate with the residential area.

5.5. Buffering Analysis.

The study area tanks were extracted from the survey of India toposheet. Buffer analysis is one of the main techniques to identify the surrounding area calculation with specific distance from the center or outer of the features. Buffer builds a new object by identifying all areas that are within a certain specified distance of the original objects. In raster, buffers can be spread outwards from objects to create friction surface. The figure 2-D, shows the major water bodies selected for this study, figure 3-A is a villapuram tank more than 80% of the tank occupied by settlement, figure3-B of Vandiyur tank and it's surroundings for 200m buffer area generated from the center of the tank and for 400m buffer area generated. The main source of contamination identified as point source such as domestic solid waste and waste water collection through open urban sewage system (Canal) connectivity as well as direct waste dumping from the nearest residential and commercial. The ground water contamination process is more possible because the physiographical characters such as soil, geomorphology and geology are highly support the infiltration of contaminate surface water storage tank.

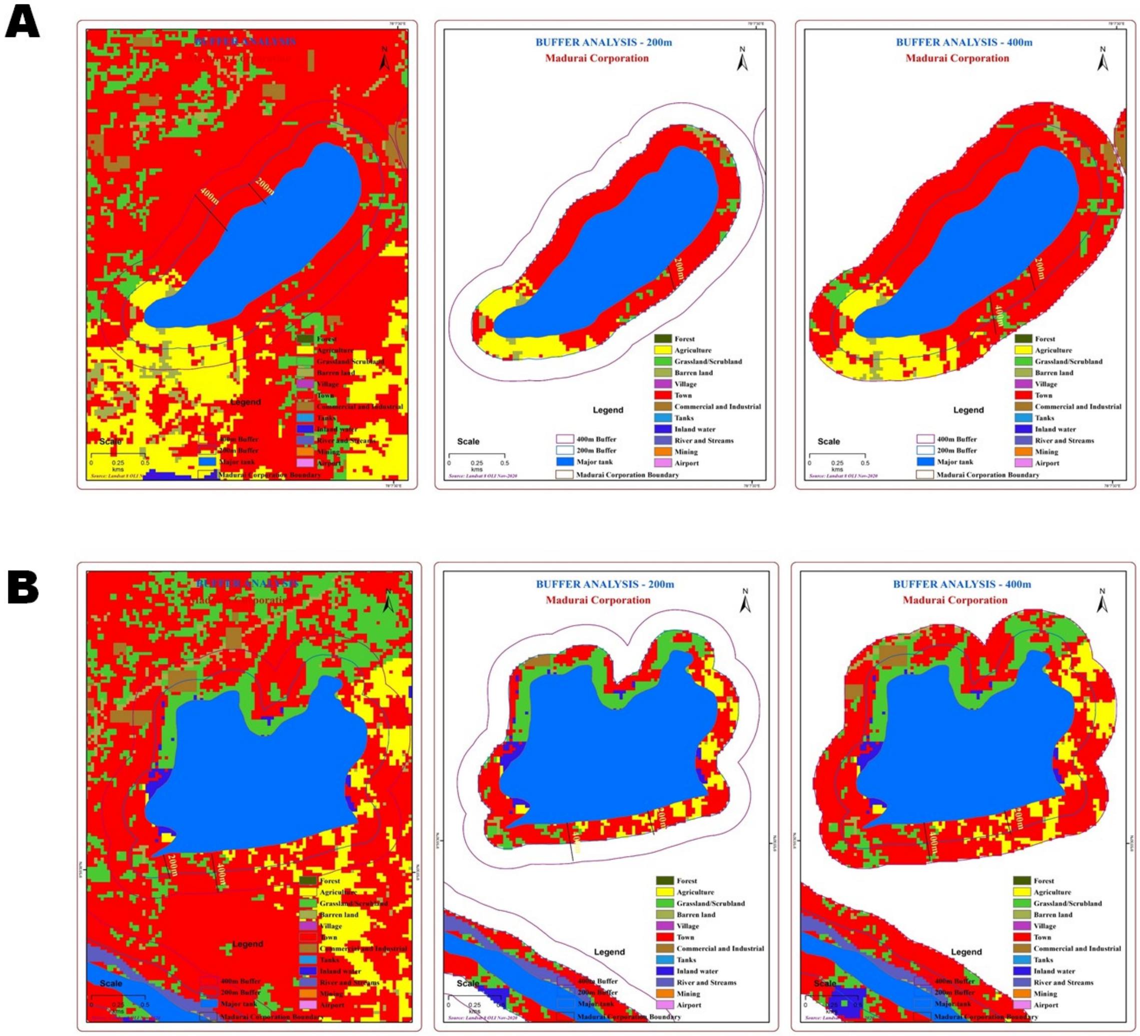

Figure 3. (A) Villapuram tank buffer analysis of 200m and 400m **(B)** Vandiyur tank buffer analysis of 200m and 400m.

5.6. Epidemiology of Water-Borne Diseases and their Impact

The contaminated water and their toxic exudates generate the pathogenic microorganisms, cause serious conditions such as cholera, diarrhea, typhoid, amebiasis, hepatitis, Diptheria, gastroenteritis, giardiasis, campylobacteriosis, scabies, and worm infections, etc [19]. Drinking or swimming in contaminated water could be dangerous to health [20]. The Landsat 8 OLI image of NDWI and PCA methods were indicate the clear identifications of surface water quality variation. Within the range of 200m to 400 m distance from the tank buffered zone living people are using contaminated ground water. Where, the buffered regions are considering as epidemiological vulnerable zone. Bore well water contaminants are associated with several illnesses including gastrointestinal illnesses, cancer, reproductive issues, and neurological disorders [21 & 22]. The collection of waste material and sewage water mixed to the tank in rainy season and due to the biological processes it will pollute and contaminate gradually.

5. Limitation of the study

At this COVID-19 pandemic situation is making difficult to connect the every house around the tank within the range of 200m and 400m because of the reason for collecting primary data. The remote sensing data of high resolution imagery (e.g. 1m) give more information of LULC than the low resolution of 30m of Landsat 8 OLI. But high resolution imagery is not available freely, cost is very high. Based on the time consumption the research might not carry out all major tanks in the study area.

6. Conclusion

This study reveals, the people are consuming contaminated water leads into more possibilities of epidemiology. The Remote Sensing techniques of PCA, NDWI, DIP and GIS techniques of buffer analysis, spatial anlyst tool, and mapping were highly effective to identifying the possible area of ground water contamination. As well as focusing the large spatial distribution of population of epidemiological spread vulnerable zone. Based on this study result we would prepare decision making and early warning system of epidemiological activities to the society. In future, this study extends further to reveal the disease type, epidemiological measurements of exposure and disease status, other sources of diseases, decision making of medical GIS, etc by primary data collection.

Reference

1. Director General of Health Services (2019), The National Institute of Communicable Diseases, Government of India.
2. Avtar Singh Dua (2005. Programmes for the control of leprosy, tuberculosis and malaria.NCMH Background Papers·Burden of Disease in India. National Commission onMacroeconomics and Health Macroeconomics and Health
3. Ursula J. Blumenthal, Jay M. Fleisher, Steve A. Esrey and Anne Peasey (2001).Epidemiology: a tool for the assessment of risk. World Health Organization (WHO). Water Quality: Guidelines, Standards and Health. Edited by Lorna Fewtrell and Jamie Bartram.Published by IWA Publishing, London, UK. ISBN: 1 900222 28 0
4. NeclaUlugtekin, sevalalkoy, dursun z. seker&cigdemgokselUse of GIS in Epidemiology: A Case Study in Istanbul. https://doi.org/10.1080/10934520600780636
5. Wan Y., Liu Y., Peng Q., Jie F., Ming D. (2019) Remote Sensing Image Change Detection Algorithm Based on BM3D and PCANet. In: Wang Y., Huang Q., Peng Y. (eds) Image and Graphics Technologies and Applications. IGTA 2019.Communications in Computer and Information Science, vol 1043.Springer, Singapore. https://doi.org/10.1007/978-981-13-9917-6_50
6. Tamilenthi1.S, Punithavathi.J, Baskaran.R and ChandraMohan.K (2011). Dynamics of urban sprawl, changing direction and mapping: A case study of Salem city, Tamilnadu ,India. Scholars Research Library Volume 3 , Issue No1. SRL-AASR-2011-1252, CODEN (USA) , AASRC9 ISSN:0975-508X.
7. Osman Abdallah and Osman Akif (2001).The Development of a Database Management in GIS Applications in Oman. Sultan Qaboos University Journal for Science [SQUJS] 6(2), DOI: 10.24200/squjs.vol6iss2pp45-53, LicenseCC BY 4.0
8. Henk J. Scholten and John C. H. Stillwell (1990).Geographical Information Systems for Urban and Regional Planning.Springer, Dordrecht. Part of the TheGeoJournal Library book series (GEJL, volume 17). DOI https://doi.org/10.1007/978-94-017-1677-2
9. PalaniyandiMasimalai 2014. Remote Sensing and Geographic Information Systems (GIS) As The Applied Public Health & Environmental Epidemiology. International Journal of Medical Science and Public Health | 2014 | Vol 3 | Issue 12. DOI: 10.5455/ijmsph.2014.081020141
10. Mcfeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 1996, 17, 1425–1432.NICD (2019), Govt. of India
11. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 2006, 27, 3025–3033
12. Guojie Wang, Mengjuan Wu, Xikun Wei and HuihuiSong(2020). Water Identification from High-Resolution Remote Sensing Images Based on Multidimensional Densely Connected Convolutional Neural Networks. Remote Sens. 12, 795; doi:10.3390/rs12050795
13. Marylynn V. Yates (1985). Septic Tank Density and Ground Water Contamination.R.S.Ker Environmental Research Laboratory, P.O.Box 1198, Ada, Oklahoma 74820. Vol.23, No.5- Ground Water
14. Craun, G.F. (1984). Health aspects of groundwater pollution. In: Groundwater Pollution Microbiology. G.Bitton and C.P.Gerba, cds. John Wiley & Sons, New York
15. Lokeshwari.H, ChandrappaG. T. Effects of heavy metal contamination from anthropogenic sources on Dasarahalli tank, India.Lakes and Reservoirs, Science, Policy and Management for sustainable use. https://doi.org/10.1111/j.1440-1770.2007.00337.x
16. Ingole N. W. &Bhole A. G. (2000) Bio-accumulation of arsenic, mercury and lead by water hyacinth (Eichhorniacrassipes).J. of the IPHE, India, 22– 4.
17. Jones, R. Anne, and G. Fred Lee."Septic Tank Wastewater Disposal Systems as Phosphorus Sources for Surface Waters." Journal (Water Pollution Control Federation), vol. 51, no. 11, 1979, pp. 2764–2775. JSTOR, www.jstor.org/stable/25040491.Accessed 29 Nov. 2020.

18. Vittala SS, Govindaiah S, Gowda HH (2005) Evaluation of groundwater potential zones in the sub-watersheds of North Pennar river basin around Pavagada, Karnataka, India using remote sensing and GIS techniques. J Indian Soc Remote Sens 33(4):483–493
19. Health and Family welfare Department (2020), Department of Public Health & Preventive Medicine, Government of India (GOI).
20. Australian Government (2020), Department of Health. Environmental Health Practitioner Manual: A Resource Manual for Environmental Health Practitioners Working with Aboriginal and Torres Strait Islander Communities
21. Abraham Munene, Jocelyn Lockyer, Sylvia Checkley, David C. Hall. Exploring Well Water Testing Behaviour Through the Health Belief Model Show less. Environmental Health Insights. https://doi.org/10.1177/1178630220910143
22. Villanueva, CM, Kogevinas, M, Cordier, S, et al. Assessing exposure and health consequences of chemicals in drinking water: current state of knowledge and research needs. Environ Health Perspect. 2014;122:213-221. doi:10.1289/ehp.1206229.