INFLUENCE ASSESSMENT MODEL OF WATERSHED POPULATION ON WATER QUALITY IN A SRI LANKAN RIVER

スリランカの河川水質に対する流域人口の影響評価モデル

Dehiwala Liyanage Chamara Pramod Liyanage
Nagaoka University of Technology
Department of Information Science and Control Engineering

This dissertation is submitted in partial fulfilment of the requirements for the degree of Doctor of Engineering
December 2017
Supervised by
Professor Koichi Yamada

Graduate School of Engineering Nagaoka University of Technology, Japan

Dedication

I dedicate my dissertation work specially to my loving parents for always encouraging me to undertake studies for my professional development and complete my doctoral research. I also dedicate this dissertation to my loving wife for spending many days for proofreading this manuscript and providing financial support. Further, I dedicate this thesis to my fellow lab members for giving help to success my research. This dedication is also directed to my Sri Lankan friends who are studying in here for giving much advice and help to find the data sources on how to my research successful.

Special thanks are offered to all those people who helped me find the data. I appreciate the efforts of and feedback from the professionals and experts who applied their knowledge to solve practical issues involved in this thesis. Finally, I would like to thank the internet teachers tutorial and open source applications which enabled me to develop better skills to fulfil my research requirements.

ACKNOWLEDGEMENTS

One of the most influential advisors in my study career has been the supervisor of doctoral course, Professor Koichi Yamada, for giving me the freedom to pursue my interest in undertaking doctoral studies. His guidance enabled me to change my thinking, how to achieve the research goals and write journal papers. Further, I am especially grateful to my supervisor for giving me the financial support that allowed me to complete my thesis. I would like to thank the administrative staff at the Nagaoka University of Technology for the financial support and their help in my study journey at Nagaoka University of Technology, Japan. I would also like to express my sincere thanks to the examiners of the dissertation, Professor Yoshiki Mikami, Professor Zhidong Li, Associate Professor Shusaku Nomura and Associate Professor Tessai Hayama.

I also would like to show my gratitude to my former supervisor Associate Professor Asubhoda Marasinghe who encouraged me to undertake collaborative research on the subject of sustainable development in my home country, Sri Lanka. My gratitude also goes to the Central Environment Authority in Sri Lanka for access to the water quality data in Kelani River and the opportunity to discuss what is required of the monitoring system with specific reference to the quality of water in Kelani River.

DECLARATION

I, Dehiwala Liyanage Chamara Pramod Liyanage, declare that this thesis titled "Impact Analysis of Watershed population on Water Quality in a River Basin" and the work presented in it, are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this university
- Where any part of this thesis has previously been submitted for a degree or any other publication, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
- I have acknowledged all main sources of help

……………………… ………………

Student's signature Date

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the three year degree of Doctor of Engineering.

…………………… ……………..

Supervisor's signature Date

(Prof. Yamada Koichi)

ABSTRACT

Currently, critical issues affecting water quality in river basins are due to human activities such as urbanization, agricultural activities, and industries. Specially in the developing countries in Asia such as Sri Lanka, controlling the effects of these kinds of non-point source pollution is very difficult without assessing the influence of human activities on water quality and identifying the limitation of influence in terms of degradation of water quality. In this sense, the concept of watershed population is used to evaluate the quality of water in a river basin of Sri Lanka and is the key of this research. The influence of population on water quality is based on many human activities such as sewage disposal, land use for infrastructures, building houses, agricultural activities and the vehicles that they used for transportation. The approach taken here was to apply population to the two major water quality analysis processes that are parts of a water quality monitoring system. The contribution of this dissertation is incorporation of the watershed population in selecting the optimum water sampling site network and assessing the influence of non-point contamination sources on the water quality.

The research is expected to proposes a model to classify the quality of water in a river basin using the watershed population density in addition to the inclusions. The quantified influence of the population density on the water quality can be used to estimate the necessary facilities of waste water treatment to maintain the requirement of current and future population. Further, it can be involved to establish new urbanization in the water shed areas with maintaining optimum level of population. This might be a practical low-cost strategy for environment management especially for developing countries.

The proposed model is assumed to be implemented in water quality monitoring systems. Monitoring systems used for water quality in rivers usually have been simplified to functions to predict water quality, to find a point-source of contamination, and to control or mitigate the water pollution. The influence assessment model of watershed population on water quality in river basin can serve to enhance the water quality monitoring system while adding a function to assess the influence of the population as the major non-point source of water contamination in some developing countries.

Kelani River in Sri Lanka has been selected for this study because it is rich in biodiversity and many natural resources and plays a major role in the sustainable development of the country. More than 25% of the Sri Lankan population benefits from the river. Unfortunately, it is considered to be one of the most polluted rivers in Sri Lanka.

The first issue that we should solve is designing a water quality monitoring network in a

river basin. The population is used for that process as the factor of development pressure index (DPI) to supportively identify the polluted area due to urbanization. The proposed optimized selection of sampling sites network takes into account new possible sites identified by existing studies to examine pollution sources affecting the water quality. We used multi objective analysis method and genetic algorithm to find the optimized selection of sampling sites networks. In total 14 sites out of 29 were selected as constituents of the water quality monitoring network. The genetic algorithm is highly efficient in the design of the optimized sampling network compared to the brute-force approach.

The second objective of this research is to assess the influence of human activities on water quality. The study of the spatial correlation between urbanization and water quality parameters based on regional perspectives demonstrates that the human activities are positively correlated with degradation of water quality in Kelani River. The Total Coliform (TC), Dissolved Oxygen (DO), and Biochemical Oxygen Demand (BOD) were used to qualitatively define the population ranges using the Bayesian Network (BN) classification model. The results showed that the population density should be approximately less than 2375 to maintain water quality in the watershed for bathing and drinking purposes and less than 2672 for fish and other aquatic organisms. The population ranges proposed in the present study can be implemented by the relevant management authorities in Sri Lanka when they introduce new rules and regulations, set appropriate standards and improve waste water treatment facilities.

In summary, this research has quantitatively identified the ideal population density ranges for a watershed to maintain the quality of water in an appropriate level. This concept can be applied to predict water quality and it is a low cost method for proper environment management by evaluating the influence of human activities. It is suggested that the optimized selection methods to find monitoring network with 14 sampling sites with including new sites to enhance the current water quality monitoring of Kelani river in Sri Lanka.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

List of Abbreviations and Acronyms

CEA	Central Environment Authority
ICT	Information and Communication Technology
NWSDB	National Water Supply and Drainage Board
IWMI	International Water Management Institute
IUCN	International Union for the Conservation of Nature
DPI	Development Pressure Index
EPI	Environmental Pressure index
QGIS	Quantum Geographic Information System
WQCS	Water Quality Classification System
KRMP	Kelani River Basin Multi-Stakeholder Partnership
ANNs	Artificial Neural Networks
MLP	Multi-Layer perception
BN	Bayesian Network
TC	Total Coliform
BOD	Biochemical oxygen Demand
DO	Dissolved Oxygen

1 INTRODUCTION

1.1 Background

Maintaining good water in a river basin is necessary for any country to achieve the one of basic living requirements of citizens. Particularly for developing countries, a river basin must be a reliable natural source for drinking and work-related purposes [1], [2]. Kelani River has the most highly demanded river basins in the Sri Lanka being challenged by industrialization, urbanization and agricultural activities. It is the main source of drinking water for more than 25% of Sri Lankans and has the high and increasing pollution level due to human activities, industrial discharges and weakness of water quality monitoring system. Therefore, responsible authorities have proposed many projects to ensure the availability of quality water by controlling the effect of human activities such as Kelani River multi Stakeholder partnership project and National Pavithra ganga clean river project. In addition, the human activities are significantly highlighted as one of the major causes of surface water contamination in throughout the world[3]–[5]. Therefore, it is necessary to evaluate the influence of human activities as a non-point source pollution.

A major problem is that people do not realize the effects of their activities on water quality degradation [6]. Therefore, local authorities which are responsible for planning, implementation and controlling local initiatives on watershed areas in Kelani River need the proper estimation of influence of human activities. It will be beneficial to change the environmental protection policies and integrated domestic wastewater treatment facilities. The imposition of controls on contamination to protect the world's water ecosystems during the last few decades has shown the necessary to assess the influence of population on water quality in a river basin [7], [8]. Therefore, water quality monitoring systems are considered critical and must have the appropriate monitor features to control contamination in the river basin [9]. Based on these facts, the objective of this research is to develop an approach or methodology incorporating the watershed population to estimate water quality in the river basin.

Water quality monitoring systems have been employed for a few decades ago to check and assess the water quality. However, the sampling locations were often determined just for convenience or for certain reasons other than correct monitoring. After the provision of quality water has become a critical issue due to the pollution in rivers, studies about requirements for water quality monitoring systems have been popular [3]–[5]. In South Korea, short term and long-term water quality conditions are used to improve or lead to new policies that preserve the environment [10], [11]. The identified essential requirements of a water quality monitoring system according to the studies are listed in below[9], [12]–[14].

1. Water quality data analysis to obtain important information
    a. Analysis of changes in trends concerning water quality parameters
    b. Violation of water quality standards
    c. Calculating the pollution loads
2. Designing the water quality monitoring network
3. Identifying point and non-point sources affecting water quality changes
4. Real-time data gathering system from sampling sites
5. Communication and data sharing between responsible authorities
6. Real-time information visualization, and awareness programs that stakeholders can share effectively.

When considering the modern necessities of Sri Lanka as developing country, to keep the protection on the existing natural water resources is essential more than before. It is also necessary for sustainable development through the present industrialization economy and urbanization [15], [16]. Therefore, analysing external data rather than the usual water quality parameters is required to handle the current water issues occurring in Kelani River due to the urbanization, industrialization and agricultural activities [17], [18].

1.2 Problem Description

Sri Lanka uses natural water resources in assuaging people’s demand for water which is increasing as population also increases. Due to rapid urbanization, human activities have had wielded a significant impact on the ecological environment. The local pollution issues occur in the lower reaches and some part of upper reaches of the Kelani River and they can be explained as the result of activities by stakeholders such as households, restaurants and car service centres. Further, disposal of untreated or partially treated sewerage, solid waste and wastewater are the main pollution source of stakeholders. The National Water supply and Drainage Board (NWSDB) currently has faced many problems on the treatment process of water of Kelani River for providing drinking water due to water quality issues. Therefore, design the water quality monitoring network including the necessary locations to monitor the water quality for controlling the effect of human activities is one of the major requirement. Rather than that they are compelled to identify the new location for establishing the new water intake in a watershed which has not been exposed to pollution by human activities [16].

The limited amount of water quality data in river basin is one of the major problem to control the non-point source pollution in the developing countries. Human activities pose a significant threat to the water quality of rivers such as urban activities which represent a major cause of contamination in surface water bodies in Asian countries[3], [19], [20]. Evaluating and controlling the non-point source pollution of watershed areas is more complicated than point source pollution. Furthermore, contamination of watershed areas of a river has increased due to urbanization along many rivers [5], [8]. The human activities such as sewage disposal and other waste disposal, and unplanned land use for building houses, infrastructures developments and for farming mainly affect the water quality in river basin[21], [22]. The vehicles used in transportation also affect the water quality in river basin because the leaked fuel, the wash-down water, etc. drain to the river[23]. The forest degradation and deforestation based on unauthorized human activities affect the water quality in river basin [16]. Therefore, it should need proper evaluation method to identify the influence of human activates. As an example, identify the threshold limits of influence of human activities without having degradation of water quality in the river.

The challenge for developing countries is to undertake sustainable development without causing damage to the natural environment, e.g., avoiding crucial issues leading to the rapid deterioration and degradation of water quality in the water supply intake points [2], [4]. Further, cost-effective methods are more preferable for developing countries to protect their natural resources. The challenge for developing countries is sustainability development without causing damage to the natural environment, avoiding crucial issues leading to the rapid deterioration and degradation of the water ecosystem. Therefore, it can be expected to positive contribution to control the water pollution by developing water quality monitoring system based on this research.

1.3 Objectives

The aim of this research is to propose the model of influence assessment of watershed population on water quality in a river basin. The planning of human settlements in Sri Lanka is mainly executed under the guidance of the Urban Development Authority (UDA) with the involvement of local councils. The Zoning decisions in development plans are merely derived based on suitability and the capacity of the site itself, but water quality in the river basin is not taken into account. The proposed model can be employed by the planning agencies as a tool to assess the possible impact of zoning decisions on the natural environment. The scarcity of land for human settlement in Sri Lanka often leads to conflicts with the natural landscape which feature 103 river basins right throughout the country. Therefore, the assessment of impact of population distribution on water quality will facilitate to promote sustainable urban planning decisions in the long run.

To achieve this aim, we set three objectives to be solved shown below:

1. A method to find the optimal water sampling network (an essential set of sampling sites for water quality monitoring)
2. A model to classify water quality from water quality parameters.
3. Influence assessment of population growth on watershed

Therefore, assessment of the influence of the watershed population according to objectives mentioned in above can be implemented in the water quality monitoring system under the following two tasks, design the water quality monitoring network of sampling sites and evaluate the influence of external sources of contamination sources of river water respectively.

1.4 Research Framework

1.4.1 Framework

The research framework (Figure 1-1) consists of two approaches of water quality management system of the river basin based on objectives mentioned in above section. The Kelani River in Sri Lanka was selected as case study for this research work.

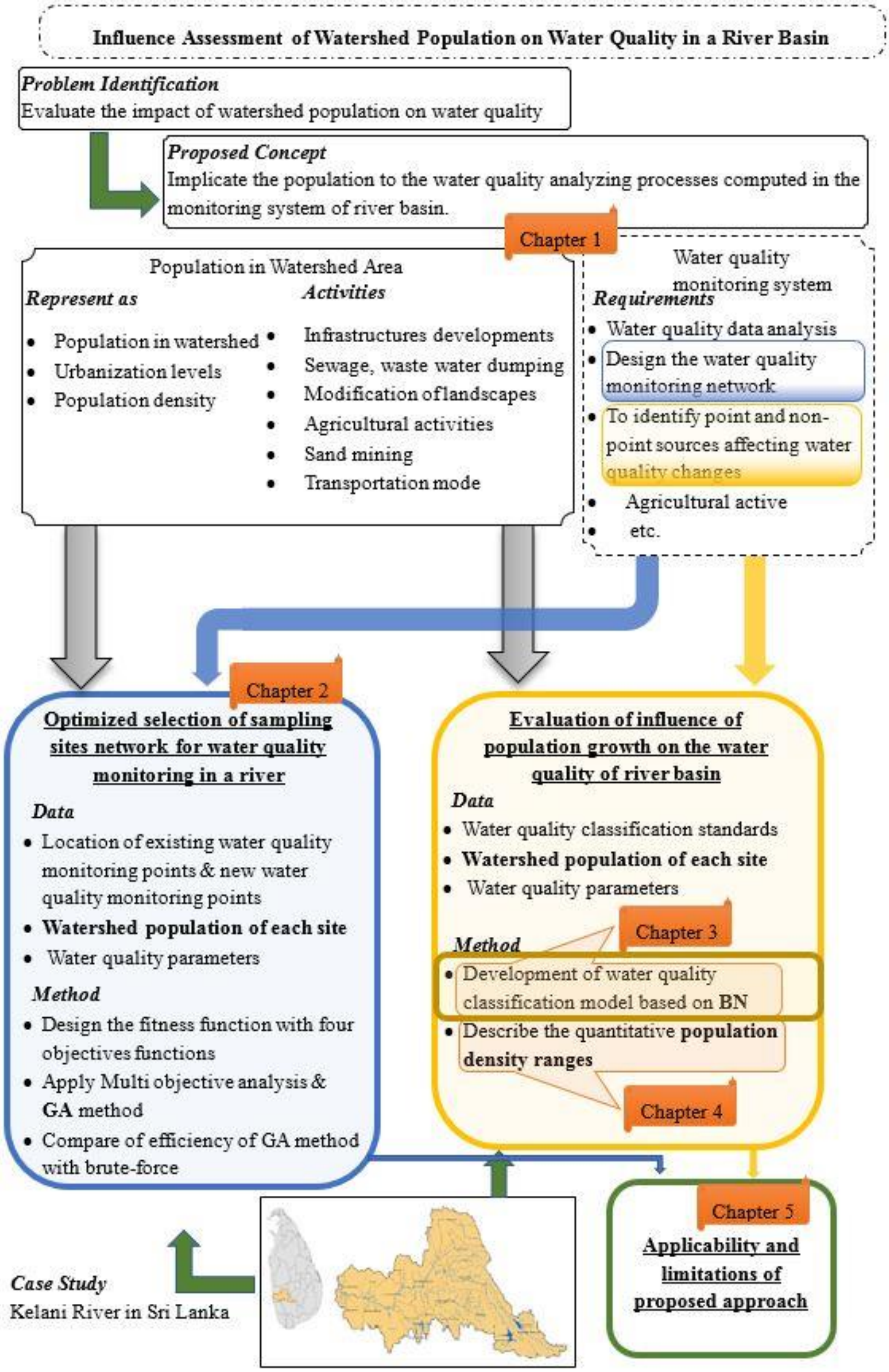


Figure 1-1 Framework of research

1.4.2 Target of the study

The study concerns the Kelani River (7º 10′- 6º 42′ N, 79º 12′ -80 º 33′ E), which is located in the Western Province of Sri Lanka (Figure 1-2). Kelani River is a 145km long river and 2292 km^2 of the river basin constitute the country's second largest watershed. Ranked as the fourth longest river in Sri Lanka, it stretches from the Seepada Mountain Range to Colombo. It flows through the Sri Lankan districts of Nuwara Eliya, Ratnapura, Kegalle, Gampaha, and Colombo. The Kelani River supplies approximately 80% of the water used in Colombo district, and it is a primary source of drinking water for the Colombo district.

The Kelani River is one the richest in terms of biodiversity in Sri Lanka. It passes through different landscapes such as lowland, sub-montane and mountains with montane forests. It gives life to many endemic plans and animals. The annual rainfall distribution of these regions varies from 2,001to 3,000 (mm) [16]. The Kelani River is used for hydropower hand it has two main tributaries in its upper reaches, these being Kehelgamu Oyaand Maskeli Oya. The most famous sub-stream that tourists can visit are We Oya at

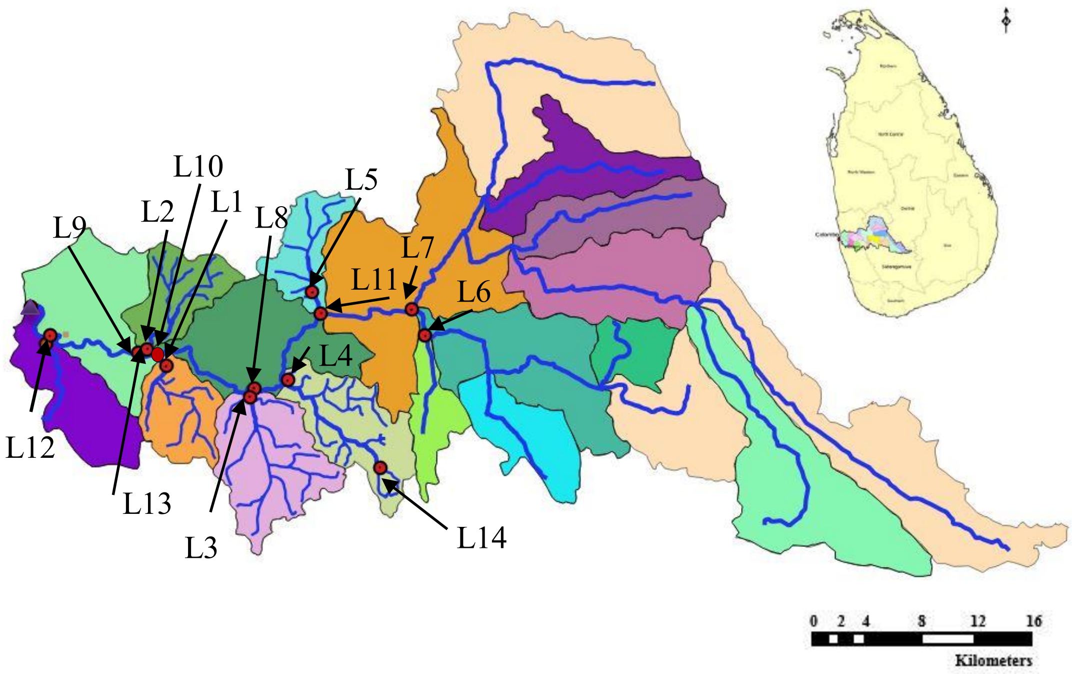


Figure 1-2 Kelani River with 20 sub-watersheds and 14 current water quality monitoring sites (red colored points)

Yatiyanthota, Gurugoda Oya and Kithulgala at Ruwanwella and Seethawaka Ganga at Avissawella. In addition, Kelani River is used for bathing, washing, transport, irrigation, fisheries and sand is extracted from its bed. In these ways, many people depend on the river for their daily livelihoods [16].

The water samples were collected according to the three grab samples technique, i.e., from two sides of the river and from the middle at the depth of 0–30 cm. The methods applied for the analysis were informed by the standard methods for the examination of water and wastewater by the Central Environmental Authority (CEA) in Sri Lanka. Finally, the quality control tests were conducted in line with American Public Health Association (APHA), American Water Works Association (AWWA) and Water Environment Federation (WEF) 2005 standards [24].

Kelani River is used for hydropower production in Sri Lanka. It has two main tributaries in its upper reaches as Kehelgamu Oya and Maskeli Oya which are contributed to part of hydro-electric production in the country by housing several major reservoirs, ponds and power stations. The most famous sub stream for tourism are We Oya at Yatiyanthota, Gurugoda Oya and kithulgala at Ruwanwella and Seethawake Ganga at Avissawella. In addition, the river is used for transport, irrigation, fisheries, and sewage disposal, and sand is extracted from its bed. In these ways, many people depend on the river for their daily life [15]. However, it is the most polluted river in Sri Lanka due to the rapid growth of industries located in close vicinity to it[17], [25].

1.5 Structure of the Dissertation

The rest of the dissertation is organized as follows:

Chapter 2 discusses the process of optimized selection of sampling sites network of a river basin for identifying the most necessary monitoring points should be monitored monthly. This study used four criteria including population as one urbanization factor which is in the development pressure index (DPI) rather than the usual factors of environmental pressure index (EPI) as the objectives to evaluate the monitoring samples. Proposed the optimized selection of 14 sampling sites network with new sampling sites should be monitored monthly.

Chapter 3 mainly discuss the development of classification models for water quality based on the standards required for drinking, bathing, industrial and agricultural activities. The advantage of the classification model than classification table is classifying records according probability than considering exact cut off levels. Therefore, application of classification model is more reasonable than classification table. In this case, the classification model is evaluating the correlations between the parameters. Further, the classification model is the most suitable mechanism for categorizing the different observation parameters based on classification standards of known parameters. Therefore, development of an accurate classification model in predicting water quality is a key point of this empirical study based on two classification models of ANNs and BN.

Chapter 4 concentrates on defining and describing the population by employing the natural environment whilst maintaining water quality based on classification standards with a BN classification model. The influence assessment of anthropogenic activities on the surface water environments has been studied around the world and their influence on the water ecosystem has been highlighted. Most of the water quality issues reported in Asian and other developing countries are due to the growth of population in watershed areas. In this study, we investigated the influence of the populations on the water quality of the Kelani River in Sri Lanka. The correlations between the water quality parameters and the populations of watersheds have been derived with the aim of quantitatively defining the populations corresponding to different water quality standards.

Chapter 5 consists with the conclusion of the research. This research also explained positive correlation between watershed population and water quality in the river basin. Further, this research assessed the influence of population on water quality in rivers based on the optimized selection of sampling sites networks and water quality classification model. Addition to that, here discussed the key finding of this research, contribution for protecting the natural environment and applicability of proposed concept based on two objectives. Further discussed the enhancement of proposed concept to overcome the limitation to offer more benefits for environment protection as the future works.

2 OPTIMIZED SELECTION METHODS OF SAMPLING SITE

2.1 Introduction

Designing an appropriate of water quality monitoring network for a river basin is most necessary when there are cost concerns of the process. Further, it should support to evaluate the external source of pollution such as the influence of human activities and agricultural activities. According our case study for Kelani River, 14 points along it are only monitored monthly. However, many water quality issues have emerged, so there are weaknesses in the current monitoring network. Many other research groups have identified several other locations should be monitored to obtain more accurate water quality data for Kelani River. Most of water quality issues have been occurred due to urbanization, agricultural activities, and industrialization. Therefore, find the essential water quality monitoring points should be monitored, population was applied as a selection factor. The objective of this study is to design an optimized selection of sampling sites network taking into account newly identified monitoring sites and not just those that are already available.

2.1.1 Background

To evaluate the water quality can consider many data rather than the water quality data such as population, and land use data. Of these inputs, the most important source is data concerning water quality parameters which is a costly process [26], [27]. Further, the evaluation of water quality of a particular region in a river basin requires obtaining data of water quality based on the physical, biological, and chemical properties of water but it is not only enough for evaluate and overcome current issues happen in particular site or sub watershed. The Environmental Sensitivity Index (ESI) data and Development Pressure Index (DPI) data such as water quality data and land use data should be used for analysis to find the reasons[9], [13], [27]. Because water quality monitoring system should evaluate and indicate the changes induced by anthropogenic activities, risk management and prediction. Identifying the contamination location and controlling the pollution is another major requirement of a water quality monitoring system specially in the catchment areas of a river basin [28]. Further, evaluation of the contamination and controlling the non-point source pollution is more difficult than those of point source pollution [5]. South Asian countries experience huge discharges of municipal wastewater and urban drainage into river basins [3]. The research conducted by L.J. Alvarez et al. in 2006 used coliform concentration to identify the critical points and control pollution caused by human activities; they employed an optimized solution for sampling site network[29]. Therefore, watershed population of each sampling site is considered as a criterion to realize the optimized sampling sites network for river basin.

When considering the expenditure on time, resources and efforts for monitoring the many sites in a the river basin, new concepts and methods have been devised to reduce the number of monitoring sites[30]. [31]. The monitoring system mainly considers the selection of the water quality parameters and measurement methods, selection of sampling sites and data analysis. When considering the cost of the process, it is possible to monitor essential points so that much accuracy is achieved [32],[26]. Many research studies have been done on developing optimal water quality monitoring networks and these methods depend on type of water resources [33]. Designing the sampling locations in the monitoring network is very important to obtain accurate estimates of water quality. Therefore, evaluating the existing sampling sites and proposing new sampling sites are very important processes [34].

To identify the sources of contamination is the most important requirement of a river basin mentioned. Huu Tuuan Do et al. used Geographic Information Systems (GIS) to obtain land use data to evaluate human activates and find the sampling locations at Chingtan Weir along Hsin-Dian River near Taipei, Taiwan [35]. The necessity of finding the best solution for a suitable sampling network is done so can we understand the situations like those discussed above. Much research has been conducted to find the best sampling network that considers the different criteria as mentioned a above[36]. Most studies looked at several objectives or criteria to find the best monitoring network [26], [27], [36]. Identifying the optimal solution mostly relies on their requirements, existing resources and data. The constraints are another important factor that we have concern when designing the sampling network [14], [26], [33].

2.1.2 Situation analysis

As mentioned in Chapter 1, Kelani River (Figure 1) was selected because it is the most polluted river in Sri Lanka due to the rapid growth of urbanization and industries located along the river banks. It flows through the most populated cities in the country. However, it is the most important surface water source in Sri Lanka as it supplies potable water to the commercial capital of the country - Colombo and its' suburbs. It supplies approximately 80% of the water used in Colombo and it is a key source of drinking water for the Colombo District. There is a water supply intake point at Ambatale, 14 km inland from the river mouth. In addition, the river is used for transportation, irrigation, fishing, sewage disposal, and sand is extracted from its bed. As stated above, many people depend on the river for their livelihoods [30].

The main sources of water pollution in Kelani River are the land-based ones such as spatially treated and untreated industrial effluents, agricultural runoff, domestic and municipal effluents. However, sewage from low-income settlements and industrial effluents (especially from tanning and metal finishing and processing industries) from many industrial sites are discharged conveniently into the Kelani River [17], [37]. A recent incident (2015) which occurred recently at the Coca-Cola Beverages Company in Biyagama near to Ambatale water intake serves as a case study [4]. The river has been contaminated with diesel for and this contaminated water had been distributed for drinking purposes for several days before it was brought to anyone's notice. This shows that the existing network is simply not working. There are increasingly serious water quality issue occurring downstream of Kelani River that are risking more salinity and anthropogenic activates. The Ambatale water intake has been facing many problems in providing water to meet demand. Therefore, the identification of water pollutant areas as well as non-pollutant areas are highly required because in order to locate a new area for water intake [37]. Furthermore, final report issued by KRMP in 2016 explains the proposed project for a long-term multi-stakeholder strategy and action plan to enhance the water quality monitoring system in Kelani River. This plan was also proposed by CEA and International Union for the Conservation of Nature (IUCN) Sri Lanka [15]. We conducted a questionnaire and held interviews on the issues and requirements of the Kelani River water quality monitoring system in March 2014 as part of research work of Master degree program. Water quality of Kelani River mainly depends on the data collected from samples obtained by the nodes in the network.

2.1.3 Water quality monitoring sites

2.1.3.1 Current water quality monitoring sites

The pollution status of Kelani River is monitored monthly by CEA in Sri Lanka. Simultaneously, the National Water Supply and Drainage Board (NWSDB) also monitors some of above locations and additionally two other points located in Ambatale and Labugama intakes. Considering the available resources, CEA can only monitor 12 points every month [24]. So, currently 14 sampling points are being monitored on a monthly basis (Figure 2-1). Six sampling locations along the Kelani River and others are located in sub tributaries. The existing system is based on the data of water quality parameters of sampling points and these points are already have been defined according to the anthropogenic activities based on point source pollution [24].

2.1.3.2 Other water quality monitoring sites

Many extensive studies have been carried out on many locations along the Kelani River which were already polluted and unfortunately these points were not monitored monthly because of a resource are lacking and the costs are large. A review of research papers referring to water management and quality in Sri Lanka were the main source for doing the situation analysis.

We identified many other water quality monitoring points utilized by different research groups. The data from Maskeliya, Mountain forest (Seepada), Kitulgala and Ginigathhena were obtained from the results of a review conducted by Weninger in 1972 while others are based on Cost and Starmuhlner (1980), Gunatilake (1983), Amarasinghe (1984) and Daniel (1986) [38]. Three other monitoring sites close to Ambatale intake well were also considered for this research [37].

The Pugoda Ela, Wak Oya, Kitulgala, Maha Oya, and Seetawake Oya are identified as the main sub-tributaries of Kelani River. Currently the points designated L1 to L12 are monitored monthly by CEA, points L13 and L14 are monthly monitored by NWSDB, All the other points (s0- s14) are not monthly monitored. All the sites are listed in Table 2-1 and shown in Figure 2-1.

Table 2-1 All Sampling locations including 14 monthly monitoring sampling sites and other identified 14 sampling sties

No.	Reference No.	Site location
1	L6	Seethawake Bridge
2	L7	Seethawake Ferry
3	L8	Hanwella Bridge
4	L9	Welivita Bridge
5	L10	At Kaduwela Bridge
6	L11	Pugoda Ferry
7	L12	Japanese Friendship Bridge
Tributaries connected to Kelani River		
8	L5	Pugoda Ela
9	L4	Wak Oya
10	L3	Pusseli Oya
11	L2	Maha Oya
12	L1	Raggahawatte Ela
Sampling points located near the main water intakes		
13	L13	Ambatale
14	L14	Labugama (Bulathkohupitiyaya)
Other used location		
15	s0	0.4 km from river mouth
16	s1	5 km from river mouth
17	s2	5.6 km from river mouth
18	s3	9 km from river mouth
19	s4	12 km from river mouth
20	s5	14.4 km from river mouth (Near to Ambatale)
21	s6	15.9 in upstream
22	s7	17.2 in upstream
23	s8	18.4 in upstream
24	s9	Near to Hanwella Bridge
25	s10	Norton Bridge
26	s11	Kitulgala
27	s12	Maskeliya (Gartemore Estate)
28	s13	Seepada
29	s14	Ginigathhena

2.1.4 Problem identification

The critical challenge of the current water quality monitoring system regarding Kelani River is to increase the accuracy of information using the existing resources. Currently, the efficiency in identifying the issues for water quality is low. As mentioned in situation analysis of 2.1.2 the influence assessment of anthropogenic activities is undermining the

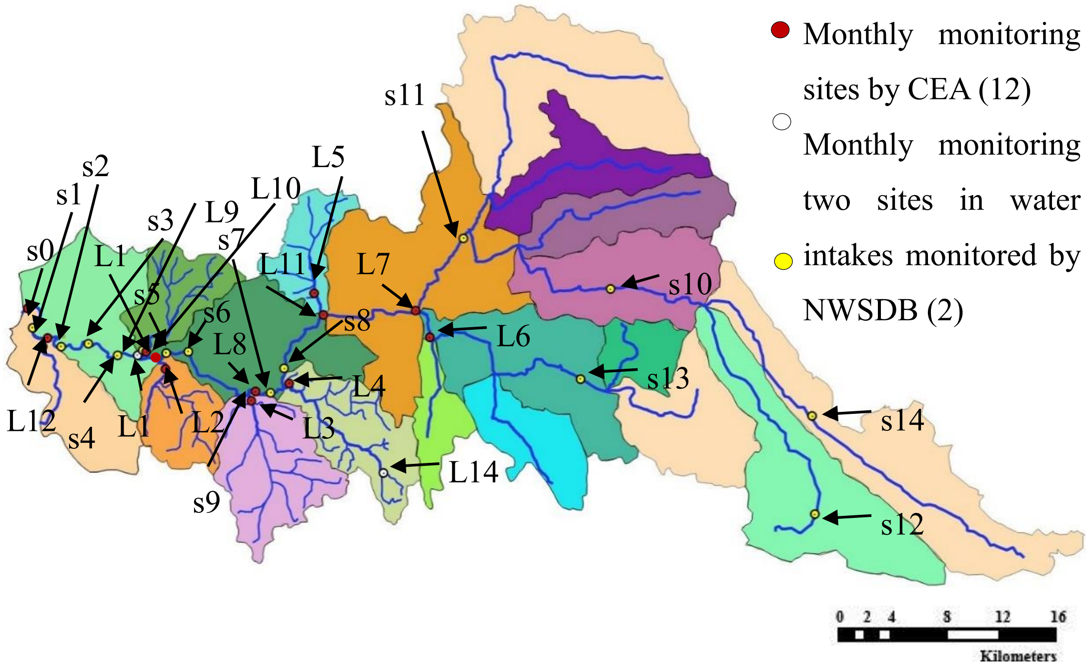


Figure 2-1 All water quality monitoring sites (total 29) in Kelani River

water quality. According to the current mechanism, the evaluation of a water quality of particular regions river basin mainly depends on capturing data from samples obtained at the site in the monitoring network, but it is not enough to the assess the influence of human activities. Considering the obtained information based on conducted interviews regarding current water quality monitoring system in Kelani River with director general of CEA , four other senior staff in environmental pollution control division of CEA, country representative of IUCN in Sri Lanka and two others senior lectures in department of town & country planning of Moratuwa university of Sri lanka while the had field visits in Sri Lanka, current monitoring sampling sites are mainly focused on the monitor of anthropogenic activities causing the point source contamination such as industries zones located in Biyagama and Awissawella [18], [24]. Most of contamination issues are caused by non-point source contamination of domestic and municipal effluents such as sewage from low-income settlements [24], [37]. As mentioned before, many extensive studies regarding water pollution caused by non-point source in river basin have been carried out on most of the locations along the Kelani River. Unfortunately, these points are not monitored monthly. Therefore, it needs to evaluate all points. A larger number of sampling sites enables retrieving more information than fewer ones, but it is a costly process. In the developed countries, in practice the number of sampling sites will usually be determined by the financial limitation and constraints of sampling and investigative facilities, and people involved[31]. Accordingly, examinations of Kelani River are currently limited to 14 sampling points even they propose new monitoring sites in KRMP report[16]. Considering that an optimized selection of sampling sites is crucially needed.

2.2 Literature review

Methods used to identify the most important sampling sites along a river basin were explained in sub-section 2.1.1. Most of these methods follow a set of evaluated criteria,the most important being water quality parameters, land use data, waste water loadings, and population. Hence the applications of multiple objectives analysis methods are more reasonable because it is possible to evaluate, explain and mitigate the sources of contaminations.

There are many applications that already attempt to build an optimal expansion strategy for a water quality monitoring system concerning river basins worldwide. Since the 1970s many research studies have focused on the effective design of monitoring networks using mathematical and statistical models combined with ICT based concepts. The ideal sampling locations as suggested by Sharp in 1971 [39], and Beckers and Chamberlain [32] developed a methodology to determine prioritized sections of a river. The application of genetic algorithms in 2005, integer programming and using GIS in 2008, graph theory, kringing theory and simulated annealing algorithm can also be listed as such ICT-based system [40]. Even though an optimized network had some limitations such as the inability to relocate assessment once the network is established, there are other strategies that have been developed for assessing water quality in river systems.

Many of the water quality monitoring networks in a river system have achieved specific objectives which are relevant in proving the efficiency of these networks [39], [41]. Other research studies used six criteria, namely representativeness of a river basin, compliance with water quality standards, surveillance of pollution sources, supervision of water use, examination of water quality changes and estimation of pollution load [42], [43]. The identification of the point source and nonpoint source pollution is one great benefit of water quality monitoring system. For this reason the present study used the population of each site to evaluate the nonpoint source pollution rather than the point source pollution [39].

The above-mentioned criteria are not applicable to the proposed new monitoring network or for a monitoring system with insufficient field data. A solution that comprises geospatial data such as catchment area, urbanization factors and pollutant loading is suggested where these are the criteria of an optimization algorithm known as simulated annealing[31], [44]. Another solution is to enumerate all the possible monitoring locations. Furthermore, it is feasible to identify the accidental contamination events in a drinking water distribution system used the brute-force approached to study all possible combinations of monitoring sites [45]. In the brute-force approach used here, the effort was made to calculate fitness function of all possible combinations of 29 locations. The most effective design of an optimized sampling network needs to identify the significant variations of each measuring criterion simultaneously. The research conducted by R.O.Strobl et al. used Water Quality Monitoring Sampling Analysis (WQMSA) model to find the critical sample monitoring points in small or large agricultural and forested areas the concentration of Total Phosphorus (TP) [13].

With the inclusion of AI for environmental management, water quality monitoring has had the ability to enhance the performance of monitoring systems [46]. The water quality components are the only sources used by the current system for Kelani River. Here the Development Pressure Index (DPI) and Environmental Sensitivity Index (ESI), are considered. There are many factors that can be used for measuring water quality such as remote sensing maps, populations, and land use data. Furthermore, some applications have offered new features such as tracking water quality distribution, variation, and ability to combine with other data sources such as GIS and remote sensing maps. Application of AI is more convenient for these purposes. Table 2-2 below summarizes the basic categories of AI with related water quality management applications [18].

Table 2-2 Basic categories of AI and applications related to water quality management

Technique	Application
Genetic algorithm	Optimization of selecting sampling sites
Fuzzy decision analysis	To identify the uncertainties in risk management
Knowledge-based system	Determination of various water quality modeling.
Artificial neural networks	Define relationships of different types of parameters. Optimization of sampling sites.

One of the most relevant techniques of AI is the genetic algorithm which was selected to assist with the optimization problem. The genetic algorithm (GA) also serves as a searching method based on the mechanics of natural selection and natural genetics [47]. GA has been successfully applied to achieve water quality systems' reliability, to create effective water quality models and to monitor networks. It has certain advantages over numerical optimization methods. In general GA is not based on objective functions or gradient estimations and in fact optimizes all objectives separately [42]. That requires only an estimate of the objective function value for each decision set. GA, is generally employed for optimization approaches. The major step in it defined as fitness function. It consists of a set of objective functions which represents different criteria and one objective function can consist more than one variable [39]. Then the algorithm generates random numbers within a set range which are initialized by the user. These numbers are called populations. Each generation calculates the fitness value. Then the operation terminates when the conditions given by the user are achieved. Therefore, GA has great potential for use in the optimization of water quality monitoring network at observation sites [48]. The computational complexity of the optimization problem is considered to be a hybrid scheme that offers a better solution. GA application with a filtering function helps reduce the seriousness of the problem and it leads to less computational time [44]. A solution can be obtained by applying the heuristic optimization algorithm. To evaluate the genetic algorithm, all equations and other given assumptions can be programmed using a program language such as Java, Python, C++, C#, and etc. The integrated

computation software can also be used to evaluate the GA in a convenient way, for instance Mathematica, Matlab, Weka, Goal, etc.
The application of a combined cluster and discriminate is also used for optimization rather than the above discussed methods. Their aims are recognized as seasonal characteristics that are reflected in the chemical parameters and seek to find spatially homogeneous sampling sites that are essential to the monitoring process. As a result, they reduce monitoring points from 14 to 11 and they suggests different frequencies for different seasons [49]. Even if we propose a solution there may be some barriers to overcome such as budget and practical issues. The constraint factors are mainly based on budget, detection sensitivity, and equality. The population data as an urbanization factor is introduced as a new criterion in this study. Furthermore, the distance to the nearest water intake along both sides of Kelani River were considered under the water use criterion instead of only distance to the nearest water intake in the upstream context [39].

2.3 Proposed method

In defining the optimality criteria based on data availability from a particular river basin, especially a developing country like Sri Lanka, this is difficult to achieve given the lack of data. A set of major objectives is considered to evaluate the sampling points as follows.

- Urbanization factor
  Population as a factor of DPI helps to evaluate the sampling points. This objective illustrates the effect of common geospatial factors that are related with particular monitoring sites. Identifying the point source and nonpoint source pollution is a great benefit of a water quality monitoring system. Thus we apply the population of each site for to evaluate the nonpoint source pollution rather than the point source pollution.
- Evaluating the violations of water quality standards by considering water quality parameters
- Coverage area of site in upstream of river
- Distance between intake and monitoring site
  Usually the distance between water intake and upper monitoring site only consider for the evaluation process. Here, the distance between water intake and

monitoring site in downstream was also applied. This because the Labugama water intake (L14) is located in the started area of Wak Oya and monitors the salinity effect happen in Ambathale water intake (L13).

Three constraints as motioned in below was built for by considering the practical feasibility.

- Budget constraint

  To ensure the cost effectiveness of the system we have limited the total number of monitoring stations. The current monitoring system considers only 14 points that are monitored monthly. According to the CEA; in has capability to monitor water quality in Kelani River downstream from Seethawake Bridge [24], [25].

- Distance constraint

  Effectiveness in monitoring the water quality in water intake location is very important.

- Equity constraint

  This ensures that each tributary in the river basin contains at least one monitoring station.

2.4 Methodology

Designing the optimal network of selected sampling locations efficiently is one major requirement of a water quality monitoring system for a river. To identify the criteria used to evaluate the points is one of the most typical aspects in this process. Two optimization methods were used to design and select the sampling sites networks. Finally, the efficiency of genetic algorithm versus the brute-force approach was evaluated.

2.4.1 Designing the fitness function

Fitness score, F can be estimated by using individual fitness functions. In this study four criteria were determined, namely compliance with water quality standards, supervision of water use, coverage area of site in upstream location, and population of each site. The fitness score can be written as in Equation 2-1.

$$\mathbf{F} = \mathbf{max} \sum_{i=1}^{n} \boldsymbol{w_i f_i} \qquad \textbf{(2-1)}$$

Where given n is the number of individual fitness functions, w_i is the weighting function for i^{th} individual fitness function, and w_i represents a weighting for each design criterion. In this case weight is defined according to importance of criteria utilized to measure water quality. During the field visit, we conducted the questionnaire to get answers from 21 people to obtain the ranking of criteria. These respondents were directly or indirectly involved with the water management sector in Sri Lanka and so were selected for this survey. The values of weighted are listed in Table 2-3. The questionnaire is used to assign the values and all criteria were ranked according to experts' preferences. The range of w_i is; $0< w_i <100$ and the total is 100 [3].

Table 2-3 Ranking values of criteria

Criteria	Ranking ratio
Water component analysis	0.28
Population based on selected location	0.22
Distance between the site and nearest downstream water intake	0.25
Coverage of site in upstream	0.25

Given f_i is the objective function of the i^{th} individual objective and all are defined in below.

2.4.1.1 Monitoring-violations of water quality standards specified for each watershed calculate using Equation 2-2

$$f_1 = \sum_{i=1}^{p}\sum_{j=1}^{q_i} L_{ij} \sum_{k=1}^{r} w_{ijk}\frac{|C_{ijk} - S_{ijk}|}{C_{ijk}}, \forall i, j, k \qquad (2\text{-}2)$$

Where C_{ijk} represents the mean concentration of the k^{th} pollutant of concern in the dry season at the j^{th} monitoring station in the i^{th} tributary (mgl^{-1}), S_{ijk} is the water quality standard of the k^{th} pollutant in the river j^{th} is the monitoring station in the i^{th} tributary is located(mgl^{-1}); The set of samples values are shown in Appendices A2.1. p is the total number of tributaries in the Kelani River Basin; q_i is the total number of candidate stations in the i^{th} tributary; r is the total number of pollutants of concern; and L_{ij} is a binary variable, in that it expresses whether the candidate site is currently included or not. According to available data S_{ijk} is common for all sections of the river. This study focuses only on dry season data to comply with water quality standards. Lastly, w_{ijk} is the weight for individual pollutants.

Water quality data from January to December 2010 with respect to pH, DO, COD, and BOD parameters are obtained from CEA. The water parameters of other points which are not monitored monthly were not extracted for the same period as stated above.

2.4.1.2 Coverage area of site in upstream location

$$\boldsymbol{f_2 = L_{ij}(U_{ij})^{-1}} \quad \textbf{(2-3)}$$

Where U_{ij} is the distance to next site located in the upstream (km) and L_{ij} is the binary variable in which 1 means the represents that the candidate location is present, and otherwise it is zero. Every ended monitoring site in each tributary is considered as a be a special site.

2.4.1.3 Distance between intake and monitoring site

Here all the monitoring sites are divided in two subsets according to location and specifically this refers to upstream (S_u) and downstream (S_d).

$$\boldsymbol{f_3 = \sum_{i=1}^{p} \sum_{j \in S} L_{ij}(U_{ij})^{-1}, \forall i, j \in S, \qquad \frac{f_3 = f_u \ when\ S = S_u}{f_3 = f_d \ when\ S = S_d}} \quad \textbf{(2-4)}$$

Where U_{ij} is the distance between the j^{th} monitoring site in the i^{th} tributary and the nearest upper stream and downstream with respect to f_u and f_d. This study area consists of with two intakes knows as Ambatale reservation and Labugama reservation. The Labugama reservation is located at the end of Wak Oya sub tributary of Kelani River while the former is located main stream. To do this calculation both intakes were considered as follows. The distance between the nearest downstream sampling point was considered instead of the distance from the nearest upstream sampling point.

2.4.1.4 Consider the urbanization factor as based on Equation 2-5

This objective illustrates the effect of urbanization because it helps to identify the impact of nonpoint source pollution rather than point source pollution [39]. As mentioned in the literature review there are many DPI factors which were included for this objective, but according to available data population density was used.

$$f_4 = \sum_{i=1}^{p} \sum_{j=1}^{q_i} L_{ij} P_{ij}, \forall i, j \qquad (2\text{-}5)$$

Pij is the population within a 10 km radius of the *jth* monitoring station in the *ith* tributary. Then the scaling function is applied to obtain the results. The buffer tool of Quantum Geographic Information System (QGIS) application, google classical map, Sri Lanka political context, and statistical data of the country's population in 2012 are used for this process. The number of buildings available in a particular area specially houses or factories can be used as a parameter.

2.4.2 Evaluation process

The multi objective analysis method and genetic algorithm were applied for find the optimal networks and three constraints are used to obtain a practical solution. The other main purpose of this study is to compare the efficiency of genetic algorithm and the brute-force approach.

2.4.2.1 Multi objective analysis method

Here the multi objective analysis (MOA) method was used to design the optimized selection of sampling sites network. The fitness score of each point was calculated using all fitness functions listed in the section above and highest values are chosen. Finally, three constraints were applied to achieve practical solutions such as budget constraint, distance constraint and equity constraint.

2.4.2.2 Genetic Algorithm

A genetic algorithm is a method for solving optimization problems even though they might be constrained or unconstrained. The most important thing is to generate a set of possible values based on population and repeatedly it modifies a population's individual criteria. It can be explained using fitness value of the individual variables or criteria, which in turn depends on a range of values of these variables. A designed water quality monitoring network includes optimized selection of sampling sites, but also needs to consider other points being generated and not only the ones. For this reason the fitness value has to be limited to some degree.

That means essential points have a minimum value for each criterion. There are many options which can be used to customize the searching mechanism, for example initial population fitness limit, parent selection rage, crossover, and mutation functions.

The approach regarding this section is to identify the most essential points in the existing network. The most applications of GA technique are used to find a best solution that but in this study needed most essential points, not only a one best point.

Accordingly, there are two options, where one is to terminate the process after several iterations considering the generated level of each variable and the other is to limit the fitness value in certain levels. According to the second option the fitness limit is changed from 10 to 100000 and executes the process while changing the number of generation with a fixed number of population. The initial population is kept as 100 because the range of criteria's values were 0 to 1 except in one criterion. Furthermore, the constraints were used to provide more practical the solutions rather than theoretical ones. Three constraints are used budget constraint, distance constraint and equity constraint.

2.4.2.3 Comparison of efficiency

The brute-force method was applied to compare its efficiency to that of the GA method. The brute-force method can identify all possible combinations and also verify the performance of the genetic algorithm. Let s be the number of all locations and p the possible number of locations able to be monitored. The brute-force solution shown as the number of combinations as C_p^s. The brute-force method give all possible combination of 14 sites from 29 sites and the fitness score for each combination is impossible to obtain in a realistic time.

The fitness score can theoretically be calculated for each combination but here it is not possible. Therefore, computation time of brute-force was calculated mathematically. As a first step the of computation time of one combination t is calculated by estimating time of fitness core function for one combination. Then total computation time T_{BF} can be obtained by multiplying the time of one combination (t) and the number of possible combinations according to the equation mentioned stated below. (according to budget constraint p =14 and total number of sites s = 29). The equation is written as in Equation 2-6.

$$\boldsymbol{T_{BF} = t * (C_p^s)} \qquad \textbf{(2-6)}$$

To perform the same process in the genetic algorithm, fitness score function is directly applied as mentioned in the previous section. Then average computation time of genetic algorithm T_{GA} was calculated employing Matlab R2015. All algorithms were executed in the same computational environment (CPU used i5 -3470, 3.20 GHz and RAM 8 GB).

2.5 Results and Discussion

2.5.1 Multi objective analysis

A summary of all results of individual criteria are shown in Table 2-4. The sites were listed at least in two criteria except the site s14.

Table 2-4 The obtained points respected to each individual objective function and proposed sites based on MOA

Ranking No.	Four criteria				Proposed sites
	Components f_1	Population f_4	Distance f_3	Covering f_2	
1	s14	s5	L9	L1	s0
2	s0	s6	L13	L3	L9
3	s1	L12	L4	L11	s4
4	s4	s3	L8	L5	s3
5	s11	s4	L14	L14	L13
6	s9	s1	s9	s13	s14
7	s12	s2	L3	s10	s2
8	s3	s0	L11	s12	L10
9	s5	L9	L5	L7	L4
10	s2	L13	L1	L6	L3
11	s10	s8	s5	s8	L8
12	L9	L10	L10	L10	s8
13	L8	L2	s8	L4	L14
14	L6	s7	L2	s11	s5
15	L13	L3	s7	L8	s11
16	L10	L1	s6	s14	s6
17	L4	L4	L7	s2	s9

According to this explanation proposed sites were not biased on one criterion and finally applied the three constrains.

2.5.2 Genetic Algorithm

The critical levels of each criterion at the termination point of the process are listed in Table 2-5 which is based on the application of Matlab R2015. The critical level cab be explained as the way of represent the involvement of each criterion but the coverage value of zero in Table 2-5 does not mean that criterion does not affect the result. Figure 2-2 displays the graph of best fitness and mean fitness.

Table 2-5 Critical levels of each criterion

Critical level of each criteria			
Component f_1	**Population f_4**	**Distance f_3**	**Coverage f_2**
0.46024345	0.10098865	0.335254	0

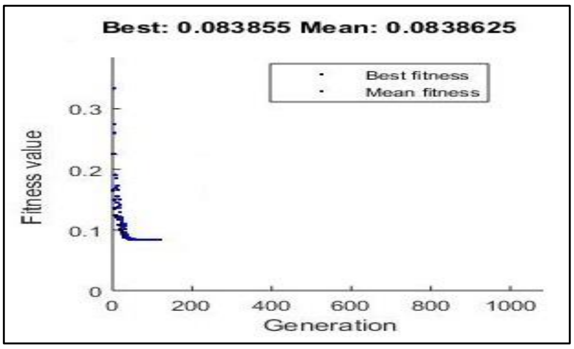


Figure 2-2 The graph of best fitness and mean fitness

The monitoring points which have higher value than critical value of each criterion as shown in Table 2-5 can be gave the priority to monitor in the Kelan River. 100 generation executed to get the optimization solution. To identify the more than one point as the essential points in here number of generation is limited in 100. All essential sites respected with respect to each individual objective function are summarized in Table 2-6 based on GA. These sites are scored higher than the critical levels explained in Table 2-5. Sites in each criterion are listed in descending order with a consideration of the essential points.

Table 2-6 The obtained sites of concerning each individual objective function criterion by GA

	Four criteria			
	Components f_1	Population f_4	Distance f_3	Covering f_2
1.	s14	s5	L9	s13
2.	s0	s6	L13	s12
3.	s1	L9		s10
4.	s4	L13		L14
5.	s11	s4		L1
6.		s3		L3
7.		L12		L11
8.				L7
9.				L6

Of these sites, L13 and L14 are linked to water quality in two intakes which are monitored by NWSDB. Consequently, set of 14 sites for optimized selection of sampling sites network were extracted by order of points in each criterion. Therefore, s14, s5, L9, s13, s0, s6, L13, s12, s1, s10, s4, L14, s11, and L1 are listed as essential points to monitor water quality in Kelani River. Then the constraints were applied to complete the process.

2.5.3 Comparison of efficiency

According to the brute-force approach, the combined of 14 sites from 29 sites is 77,558,760 and running time of one combination is 0.000077 seconds. The computation times (seconds) required to obtain the fitness function in both methods are approximately 6000 and 0.000012 for brute-force and genetic algorithm, respectively. In above case, the computational time is considered for the genetic algorithm. Even consider time of one combination as 0.000077 and population size is 100, the total time 0.77 only taken for the genetic algorithm. The brute-force method gives all possible combination of 14 sites from 29 sites and the fitness score for each combination is impossible to obtain in a realistic time.

2.5.4 Evaluation of the Constraints

2.5.4.1 Budget constraint

The number of sites that can be monitored monthly is 14 and so this number was extracted from the original 29.

2.5.4.2 Distance constraint

The effectiveness of documenting water quality in a water intake location is very important. Two main intakes are considered here L13 and L14. Monitoring points L9, s5 and L2 are located in near L13. Likewise, L4 is the nearest point to L14.

2.5.4.3 Equity constraint

This ensures that each tributary in the river basin contains at least one monitoring station. Wak Oya, Pugoda Oya and Sithawaka Oya are the main sub tributaries in the study area L4, L5, and L6 or s13 are monitoring locations that correspond to the above mentioned sub-streams.

Summarizing the constraints, budget constraint is considered the primary factor for deciding the sampling sites network. According to the capability of CEA, s10, s11, s12 and s13 are unable to be monitored on a monthly basis. For equity constraint, it is necessary to monitor L5 and L6. According to distance constraint L4 also should be monitored. The remaining position is filled by s3 according to Table2. Thus s10, s11, s12 and s13 are replaced by s3, L5, L6, and L4. After applying constraints, the final optimized selection of sampling sites is listed in Table 2-7. The brute-force method gives all possible set of 14 out of 29. Then, we can assume the obtained selection of sampling sites of both methods are as solutions of brute-force method.

Table 2-7 Optimized selection of sampling sites based on two proposed two methods and sampling sites monitored by current monitoring network

Current monitoring network by CEA				**Multi objective analysis method**				**Genetic algorithm method**		
L12	L9	L2		s0	s3	s4		s0	s1	s3
L10	L1	L3		s5	s14	L9		s5	s6	s14
L8	L4	L11		L10	L3	L8		L9	L5	s4
L5	L7	L6		L4	L5	L6		L1	L6	L4
L13	L14			L13	L14			L13	L14	

2.5.5 Discussion

In this study, two optimal locations of sampling site networks were obtained. The individual comparisons of each criterion between listed in both methods are explained below.

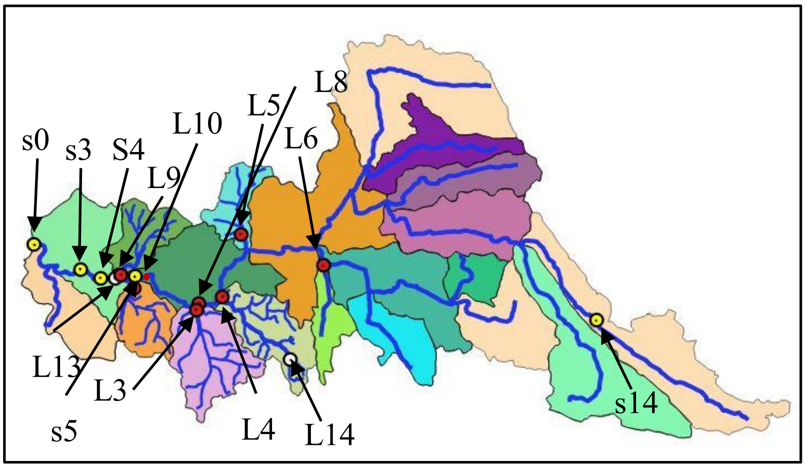
The essential sites of component criterion in GA are listed in result of MOA with same order. The set of 7 sites of population criterion in GA are already listed in result of MOA within top 10 sites. The set of 2 sites of distance criterion in GA are listed in same order in result of MOA. The set of 9 sites of covering criterion in GA are also listed in previous result within top 10. Since compare with criteria individually in both methods is shown more similarity of selected points.

The obtained sampling sites in Table 2-7 when consider the population objective function individually are s5, s6, L9, L13, s4, s3 and L12. Out of these seven, five sampling sites are listed in proposed optimized network by Multi objective analysis method and six sampling sites are listed in the proposed optimized network by GA. Due to that evaluate the impact of watershed population is very essential to find the critical points should be monitored. Further, we can prove there is strong relationship between of population and water quality.

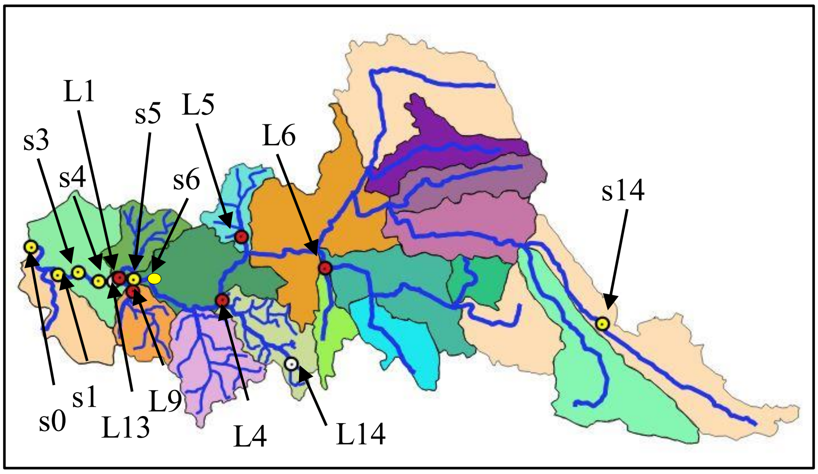
However, the order of set of points selected by coverage criterion is almost opposite in between top 10. The multi objective analysis method gave priority for points located in middle of the monitoring area but in the GA method, the selected points are located at the end of a particular sub tributary. According to the description of current monitoring points which has least concentration were not listed as high rank in the component criterion in both methods such as site ‘L6’ and ‘L4’[24].

Table 2-7 lists the proposed sampling sites both methods and current existing water quality monitoring network employed by CEA. The Figure 6 maps three river monitoring networks separately. As discussed in the situation analysis section, monitoring points which are identified as important and monitored by other researchers are also included in both proposed sampling sites networks. The points in the current network are mostly located according to the anthropogenic activities However, both methods used four different criteria including one urbanization factor. In general, both proposed networks identify more new points in near to Ambatale water intake and cover a larger area than the current monitoring network.

Point ‘L2’ in the current network has a high chemical oxygen demand (COD) variation level but considering variation differences in the components, this particular point is comparatively neutral [1]. Therefore, point ‘L2’ was not mapped in these new two



(a)



(b)

Figure 2-3 The proposed Kelani Kelani River water quality monitoring networks. (a) Proposed network using the multi objective analysis method (b) Proposed network using genetic algorithm method.

networks because these two systems could not evaluate the variation of each water quality components individually. The sites in the current monitoring system were selected with a regard for point source anthropogenic activities and sites are located at points L2, L12, L7, and L5 were deemed important to monitor. This is because many industries are located close to these points [24]. Of these only ‘L5’ is listed in both proposed networks. Locations ‘L2’ and ‘L12’ can very possibly be monitored by GA network rather than the multi objectives analysis method.

The brute-force method for a possible combination of all 14 sites and according to in this study it was more than 77 million. Then the fitness score for each combination is impossible in to find in a realistic time. Therefore, computation time is unreasonable. The results of computation time in the brute-force approach is above 500 million than the computation time of genetic algorithm. But when referring to the constraints, some points are necessary to monitor such as L5, L6, and L4. Yet some points are impossible to monitor and this means s10, s11, s12 and s13. Since if we apply some filtering rules as mentioned above, efficiency of brute-force can be increased. Furthermore, p = (14-3) =11, total number of sites s = (29-3-4) = 22 and the total number of combinations is 705,432. Therefore, the new brute-force time of filtering time is given below in seconds.

$$\boldsymbol{T_{BF-Filter} = t * \left(C_p^s\right) = 54} \qquad \textbf{(2-7)}$$

According to this equation the effectiveness of the brute-force approach is increased, but unlike the genetic algorism it is still unreasonable.

Considering practical issues such as budget constraint this research enhanced the existing monitoring network without expanding the number of monitoring sites. According to the KRMP report, around 25 new monitoring points were proposed except existing sampling locations[16]. A further four sites are introduced along the lower reaches of the kelani River. Therefore, the locations of all new suggested points in our experiment are almost the same as the proposed new locations cited in the KRMP report.

2.6 Summary

The watershed population was applied to assess the influence of human activates on water quality in the river basin. The obtained sampling five sites from seven in Table 2-7 when consider the population objective function individually are listed in the both proposed optimized networks. Due to that assess the influence of watershed population is very essential to find the critical points should be monitored. To realize an ideal water quality monitoring network for any river basin is very difficult the proposed. Therefore, both proposed optimized sectioned of sampling sites networks by MOA and GA methods are possible to apply for Kelani River to enhance the capability of current monitoring network. Even the brute-force method, while it can obtain the optimal solution, consumes the more time to calculate fitness functions. The genetic algorithm has the best efficiency in the design of an optimized sampling network.

According to one published project report in 2016 increasing the number of sampling sites up to 39 is suggested [16]. However, this research also studied about 29 total monitoring sites and proposed the number of new monitoring sites enhance the efficiency of the current network. Further, it was limited to 14 due to the possibility of CEA. Since proposed method can be used in the feature in Kelani River to select the essential points should be monitored based on the capability of CEA to achieve the goal of KRMP project proposal.

2.6.1 Achievement

Two proposed optimized selections of sampling sites networks can enhance the existing water quality monitor network at Kelani River. Population can be employed to evaluate the sampling sites. In this context the CEA has, applied the budget constraint to limit the number of monitoring sites to 14.

2.6.2 Benefits

According to Kelani river, this approach explains the how to find the 14 monitoring points rather than all 29 points. Since this network can use for the reduce the cost of analysing the water quality of samples. There is possibility change priority for any objective by changing the weight in Equation 2-1. A good feature of this method is that it can be used to consider different objective functions for any river and change the weight according to requirements as mentioned in Table 2-3.

2.6.3 Limitation and uncertainty

We can explain all four objectives and constraints as being the limitations of our proposed method. The accuracy is controlled by the number of objectives in the method. The accuracy of water quality component analysis is depending on the number of water quality parameters. According the commutation limits between responsible authorities in Sri Lanka, the dataset was limited. The data of water quality parameters used in both sets of sampling sites such as current monitoring sampling set and other samplings sites were not in same time period.

2.6.4 Suggestions for further studies

The water quality monitoring locations in a network cannot remain to be static forever. It is difficult to select most suitable monitoring network out of these two proposed monitoring networks. Further, investigations into the dynamic optimized selection of river sampling sites network should be attempted in the future. As a future work, to increase accuracy of this process better to use more suitable factor in DPI and ESI indexes which are very helpful to find source of contamination.

3 WATER QUALITY CLASSIFICATION

3.1 Introduction

The water quality standards are defined in the particular countries based on the quality of water that is required for the user requirement. Further, defined set of water quality standards are given in the table. These water quality standards are consisted with necessary parameters that should be evaluated if possible. The classification models are more convenient to analysis and prediction purposes in the particular water source. The classification model adept of imitating the basic relationship of the water quality parameters. Therefore, it has been widely adopted to the develop model for analysis and prediction based on both linear or nonlinear relationships on attributes.

Classification of the water quality in a particular river basin is necessary for evaluating the effects of external sources of pollution. The study is to discuss the possibility of development of the classification models from several water quality parameters. This chapter explains two water quality classification models using three-layer perceptron neural network and Bayesian Network (BN). The cross-validation technique was used to compare the results of the two models. Water quality data were obtained from 12 monitoring points over a 10-year period. Five water quality parameters were considered in defining four water quality standards of concerning Kelani River.

3.1.1 Background

Water quality classification of a particular river basin describes the water quality status of a specified locations. It is necessary to identify suitable locations to fulfil people's requirements, for instance drinking, bathing, agriculture and other activities. Additionally, the classification of water quality can be used to evaluate the human activities and the land use in the watershed around the location. The changes over certain period of time suggest those of human activities and the land use. Various kinds of mathematical assessment models have been used to study the physicochemical interrelationships between set of parameters: for an example, water quality index models, fuzzy synthetic evaluation approach and Bayesian network models, etc. [46], [50], [51]. However, water quality index models or classification models can obtain more interconnected results. Many water quality indices have been developed with a certain set of parameters in mind [52]. The CEA of Sri Lanka is, the over-arching organization responsible for monitoring the water quality in Kelani River. Its mission is to define the water quality classification system for Kelani River [9]. This study proposes a water quality classification considering five water quality parameters namely, Hydrogen Ion Concentration (pH), Dissolved Oxygen (DO), Chemical oxygen Demand (COD), Biochemical Oxygen Demand (BOD_5) and Nitrate (NO_3^-). The classification model based on BN can be applied to explain the correlation of each parameter by using conditional probability. Therefore, these kind of models offer more clarification about correlation of each parameters by explaining the ranges of parameters including dependent parameters based on the probability of the water classes.

3.1.2 Problem Identification

Kelani River is the most polluted river in Sri Lanka due to the rapid growth of industries located close to the river. Rather than classifying the water quality based on cut-off values in a rough guide of classification table such as Table 3-2, is it more reasonable to classify water quality using a classification model taking into account the interrelationships among water quality parameters. Classifying the river basin based on the classification results is essential to identify the best locations that satisfy users' requirements such as drinking, bathing, and agricultural activities. Therefore, to develop an accurate classification model in predicting water quality is a key point of this study.

3.2 Literature review

Artificial neural networks (ANNs) have the good potential to establish relationships based on well-defined data sets. Applications of ANNs to predict water quality variables in water systems have been increasing in the last decade [53]. BN models are able to learn patterns by evaluating the dependencies among parameters using Bayesian statistics [54].

Designing the most accurate and efficient model by comparing the two water quality classification models is the major goal of this chapter and in addition it is one of the major requirements of the CEA. The existing national environmental regulations are designed as tolerance limits for the discharge of industrial waste and are not enough to improve the quality of inland water streams. A few ambient water quality guidelines have been proposed but, not yet sanctioned in Sri Lanka, and specifically for inland water bodies. A recently published project proposal titled “Kelani River Basin Multi-Stakeholder Partnership 2016- 2020 (KRMP)” proposed inland water quality standards for Sri Lanka[16]. There are countries that have mandatory water quality indices based on ascertaining the status of water quality. For example, Kinta River in Malaysia has a water quality index water quality status [55], and so Dusit district canal in Bangkok [18]. A water quality modelling system of pollution index has been developed for Kaoping River in Taiwan [56].

To develop a predictive models rather than a traditional water quality assessment method, using machine learning algorithms for this research was greatly encouraged [20]. Particularly application of ANNs and BN, due to their ability to develop a model based on existing data, can predict the water quality parameters in water streams [57]. The application of the ANN model has been rapidly increasing in many scientific disciplines due to its capability of overcoming stifling arbitrary nonlinear relationships [58]. The BN also has the capability to acquire complex relationships among probabilistic variables.

The main difference in these two technologies is correlation or dependence between input variables. BN classifiers follow an assumption of stochastically independent inputs but in ANNs there may be an assumed correlation between input variables. In this study MLP neural network and BN were utilized in the design of a water classification model. The cross validation technique was used to evaluate the performance of the two models. The selection of input data or consideration of the number of input variables in the input layer also supports an to improvement in the efficiency by; firstly, eliminating unnecessary or redundant variables; and secondly, combining different set of parameters [59], [60]. The Kappa statistic and F-measure are often used to evaluate the performance of machine learning algorithms. The F-measure helps to evaluate the effectiveness of a model using the harmonic mean of recall and precision. The Kappa statistic was based on computing the observed accuracy and expected accuracy of the model.

The empirical approach planned in this study proposes a water quality classification system and compares the three-layer perceptron neural network model and BN model which can serve as the basis for an alternative for the proposed classification. The 3-fold cross-validation was used to analyse the impact of the number of hidden neurons and training time. The K2 local searching algorithm and A 0.5 estimator are used to evaluate the BN model. Finally, the compared the accuracy and computational time of both models the with cross-validation method are compared.

3.3 Methodology

3.3.1 Study materials

3.3.1.1 Water quality data

The study area is 20 km upstream basin from the Kelani River mouth (Figure 2). All monitoring points are monitored by the CEA, except two points which are located in water intakes. These two locations are monthly monitored by the NWSDB. Six sampling locations along the main stream of Kelani River and others are located in sub tributaries. The total dataset consists of monthly data from 2003 to 2013 except 2007. As mentioned in Chapter 1 they followed standards techniques to analyse the water quality parameters. In order to propose a water classification system for Kelani River, data of 5 water quality components during 12 months in 10 years were used. The data set consisted of with 1360 monthly samples of five water quality parameters collected over a period of 10 years. Ordinary mean and mode values are used to treat and remove missing and inconsistent observations. The sample data set is given in Appendix A2.3.

3.3.1.2 Water quality classification standards

The two proposed water quality classifications were studied. According to the CEA, two proposed categories of water quality guidelines for Kelani River were “water suitable for drinking with simple treatment” and “water suitable for bathing”. The Sri Lanka Standard Institute (SLSI) also proposed guidelines for selecting five conditions: irrigation and aquaculture, recreational use and one another drinking water category rather than the two in the CEA classification. Three selected water guidelines of SLSI are listed in Table 3-1. According to an paper published in 2014, Gemunu Herath stated that this classification was prepared in collaboration with CEA and SLSI [61]. However, both of these classifications have not yet been officially sanctioned.

Table 3-1 Proposed water quality guidelines in Sri Lanka

Parameter	Proposed by CEA		Proposed by SLSI		
	Water for drinking	Bathing water	Raw water for drinking	Bathing water	Agriculture Water
pH	6.0-8.5	6.0-9.0	6.0-9.0	6.0-9.0	6.0-8.5
EC dS/m	-	-	-	-	.7
TURB NTU	5	-			
TEMP ^{0}C	-	-			
DO mg/l at 25oC	6	5			
COD mg/l	15	20			
BOD_5 mg/l(5days at 20oC or 3 days at30oC)	3	4	5	4	5
CT mg/l	-	-			
NO_3^- as N mg/l	5	5	5	5	5
Colour Pt unit			100	-	-
Total coliform MPN/100ml (P=95%)			5000	1000	1000
Total phosphate mg/l			.7	.7	.7
Fecal coli form MPN/100ml (P=95%)			-	50	-
Aluminiu(µg/L)			200	-	5.0

The values of each category are similar except for some cases as follows; BOD value which is 3.0 in CEA and 5.0 in SLSI for drinking water category. According to the KRMP report, the proposed water quality standards by Western Region Megapolis Plaining(unpublished,2015) had same clarification levels as shown in Table 3-1 [16]. It lists standards of 39 parameters and mainly has three classes with 7 sub-categories. Further it gave more classification than Table 3-1 such as two drinking water categories and explained the DO level for Agriculture. The water quality standard of category 2 in class I of KRMP is almost related with category of water for drinking proposed by CEA in Table 3-1. When considered all water quality classification standards available in Sri Lanka, uncompleted categories were identified. Therefore, here some other classifications available in Asian countries were studied to propose the completed water quality classification standards.

There are many water quality standards for inland water streams available in the world but here we considered in South Asian countries. For example the Department of Environment in Malaysia has defined classes of water quality and water quality index [62]. Another example is water quality classification for Dusit district canal in Bangkok [63]. Both these studies classify water quality in rivers according to usages of river and some other classifications has suggested [56]. All values of water quality parameters are mentioned as ranges of values rather than a discrete value. The proposed local water quality guidelines defined the discrete values for parameters except pH.

After a thorough investigation, a water classification system was proposed to compute the classification models. Table 3-2 shows the proposed water classification table for Kelani River. These four categories are referred to as C1, C2, C3 and C4. Category C1 and C2 correspond to water for drinking and bathing water proposed in CEA classification respectively. Water quality of in C4 means it is polluted and needs special treatment for recreational use. The category of raw water for drinking C1 in Table 3-2 is related with drinking water with simple treatment category 2 in class I of KRMP standards[16]. The proposed water quality classification is mainly based on local water quality guidelines rather than the existing international classifications.

Table 3-2 Proposed Water Quality Classification for Kelani River

Parameter	**Raw water for drinking C1**	**Bathing water C2**	**Agriculture Water C3**	**Polluted C4**
pH	6.5<x< 8.5	6<x<9	6<x<8.5	x<=6 Or 9>=x
NO_3^- as N mg/l	x < 5	x < 5	x < 5	x >= 5
DO mg/l	6 < x	5 < x	2 < x	x<=2
COD mg/l	x < 15	x < 20	x < 50	x >= 50
BOD_5 mg/l	x < 3	x < 4	x < 5	x>=5

3.3.2 Development of the ANNs Model

ANNs are a good machine learning technique to understand the relationships in a given data set, where the set is called the training data set. Subsequently another data set is used to verify the model, which is called the test data. The simplest form of ANNs is a perceptron. The multilayer perceptron neural network was used here to develop the model. It is a feed-forward ANNs that utilizes back propagation as its learning algorithm and is composed of an input layer, an output layer with actual results and hidden layers. The hidden layers contain neurons which serve to process the data and change the weights according to the backpropagation learning process. The 3-fold cross-validation technique and 1360 dataset was employed to develop the ANNs model development process and evaluation process. The accuracy was calculated by 1/3 of test dataset after training the model with 2/3 of the training sample.

The neural network structure and its parameter settings are crucial to improving the algorithm's performance. Deciding the number of neurons in input, hidden, and output layers, adjusting the learning rate, number of retrains, and training stopping criteria are also essential. The Figure 3-1 describes the ANNs structure at the beginning of training process. The five neurons in input layer including with values of pH, DO, COD, BOD_5, and NO_3^- water quality parameters. This set of water quality data were obtained from 12

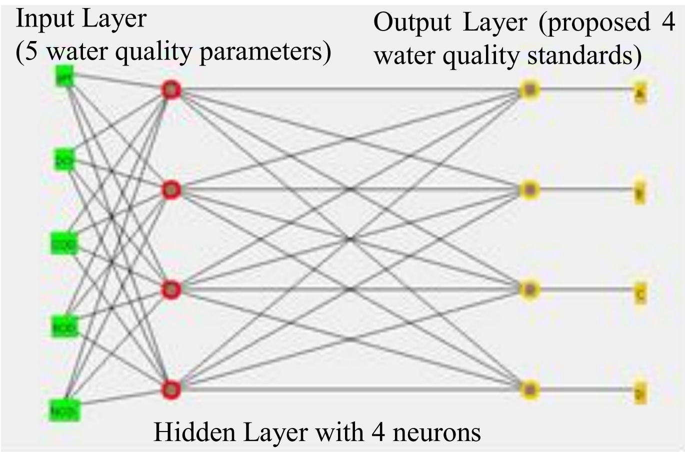


Figure 3-1 Architecture of the ANNs model for classification for water quality in Kelani River

monitoring points over a 10-year period by CEA in Sri Lanka. As mentioned in 1.4.2, they follow the international standard methods for the examination of water quality. The output layer consisted with four neurons of water quality classification classes of river basin such as raw water for drinking (C1), bathing water (C2), Agriculture Water (C3) and Polluted (C4) as explained in Table 3-2. The sample data of set of training data set is shown in Appendix A2.2.

3.3.3 Development of the BN Model

Bayesian network classification is a statistical supervised learning method. It uses explicit probabilities of hypothesis to provide an important perspective for studying many practical learning algorithms. The BN model is very easy to implement and update compared to the ANNs [64]. The BN structure interpret by the network with notes for each attribute and connected by directed edges according to a directed acyclic graph[65]. Further, the complexity of ANNs is greater than the BN model but has easier to learn algorithms for large data sets[53]. The learning a network structure and the probability tables are the two major steps when implementing the BN model. The K2, hill climbing add arcs with a fixed ordering of variables search algorithm to learn a network structure and simple estimator for to learn the probability tables. Figure 3-2 shown network of BN with one parent and default setting of all experimental values were applied.

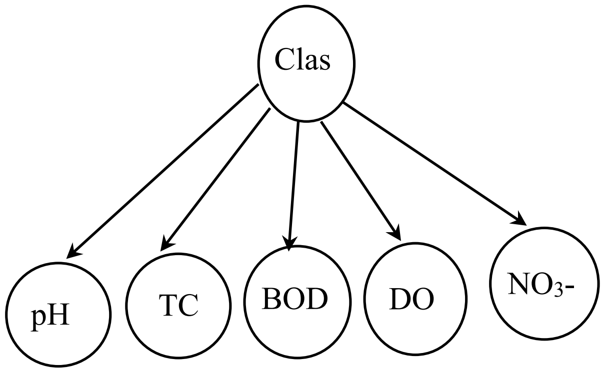


Figure 3-2 Structure of BN network with one parent

3.3.4 Optimization of model performance

The Weka workbench is used to understand the algorithm and evaluate its performance [65]. The Weka is a collection of state-of the-art machine learning algorithms and data preprocessing tools. The important prearrangements of the ANNs model are as follows: preparing the data sets, determining the neural network algorithm, defining model performance criteria and determining the evaluation or validation methods [62]. All algorithms were executed in the same computational environment (CPU used i5 -3470, 3.20 GHz and RAM 8 GB).

As mentioned in of the section on development of the ANNs model, multilayer perceptron neural network was used to establish the model and 3-fold cross-validation technique helped to evaluate the performance with a 1360 samples dataset. Then optimization of the performance was conducted in two main stages. In the first stage the impact of number of hidden neurons was assessed, followed by optimizing the training time on the network's test correctness in the second stage.

The multilayer perceptron classifier was evaluated for different training times by increasing the number of epoch in constant number of hidden neurons and other factors such as learning rate and momentum.

Secondly, the multilayer perceptron classifier was evaluated for several hidden neurons by constant number of epochs and other factors as in the first stage. Finally, considering these two scenarios the optimal point of the model's performance of model was obtained.

In the first step of implementation of the BN model, its accuracy was evaluated for 1,2 and 3 different number of parents of K2 (global score metric based algorithm) learning algorithm with default setting of simple estimator. Then the performance of both models was evaluated for different fold levels of cross-validation. The percentages of correct classification, Kappa statistic, F-measure and computational time are considered to evaluate the performance of models in each stage.

3.4 Results and discussion

In the proposed water quality classification system, drinking water category overlaps with that of bathing water and both these two overlap with the agriculture category. These water classifications are part of the existing local water classification guidelines in Sri Lanka but no other the international classification standards such as; Department of the Environment in Malaysia [62] or Kaoping River in Taiwan [56].

The results for test correctness evaluation by increasing training time are shown in Table 3-3. The constant value for hidden neurons was 4. After 1,000,000 epochs, the network's declined. The maximum percentage of correctly classified records reached 82.57% for 1,000,000 epochs. The maximum accuracy was found to be 82.57% for the highest F-measure value of 0.822 as shown in Table 3-3.

Table 3-3 Result of evaluation By based on increasing training time

Training time (epochs)	Learning Rate	Momentum	% Correctly classified	Kappa statistic	F-measure average
50	0.3	0.2	75.66	0.6481	0.748
500	0.3	0.2	79.26	0.7008	0.787
5000	0.3	0.2	81.18	0.7296	0.81
50000	0.3	0.2	82.13	0.7448	0.818
200000	0.3	0.2	82.28	0.7436	0.819
500000	0.3	0.2	82.13	0.7427	0.818
1000000	0.3	0.2	82.57	0.7492	0.822
5000000	0.3	0.2	82.21%	0.7436	0.819

The results derived from increasing the number of hidden neurons by constant number of epochs are shown in Table 3-4. The maximum percentage of correctly classified data records reached 93.01% for 45 hidden neurons.

Table 3-4 Result of evaluation by increasing hidden neurons

Hidden Layers	Learning Rate	Momentum	% Correctly classified	Kappa statistic	F-measure average
4	0.3	0.2	79.26	0.7008	0.787
15	0.3	0.2	92.21	0.8892	0.922
25	0.3	0.2	92.35	0.8913	0.923
35	0.3	0.2	92.72	0.8962	0.927
43	0.3	0.2	92.65	.8953	0.911
44	0.3	0.2	92.57	0.8942	0.925
45	0.3	0.2	93.01	0.9005	0.929
46	0.3	0.2	91.84	0.8837	0.917
50	0.3	0.2	91.84	0.8836	0.917
60	0.3	0.2	92.72	0.8963	0.927
65	0.3	0.2	92.43	0.8921	0.923

Then the training time was increased again for these 45 hidden neurons as shown in Table 3-5. This stage revealed that the optimal performance of the model is given in 45 hidden at about 50000 epochs training time.

Table 3-5 Results of evaluation by increasing training time with 15 hidden neurons

Training time (epochs)	Number of hidden Neurons	% Correctly classified	Kappa statistic	F-measure average
50	45	78.82	0.6946	0.78
500	45	93.01	0.9005	0.929
5000	45	96.76	0.9541	0.967
50000	45	97.50	0.9641	0.975
100000	45	97.50	0.9645	0.975

Even when using a proper water quality classification and a considerable number of datasets, some test records were not classified correctly. The reason is the inability to generate good estimates for extreme events which has been identified as one a the major weakness of artificial neural network models [63].

Increasing the number of parents of K2 learning algorithm, evaluated the performance of the BN model. At maximum number of parents 1 offered maximum accuracy for 15-fold cross-validation. The final results for both models concerning different fold levels of cross-validation are shown in Table 3-6.

Table 3-6 Result of evaluating of both models

Models	Folds level in Cross-validation	% Correctly classified	Kappa statistic	F-measure average	Computational time
ANN	3	97.50	0.9645	0.975	585.21
BN	3	98.52	0.9791	0.985	0.09
ANN	10	97.65	0.9666	0.976	575.97
BN	10	98.6	0.9801	0.986	0.02
ANN	15	97.35	0.9624	0.973	586
BN	15	98.75	0.9801	0.987	0.02

The ANNs obtained best accuracy at 10-fold cross-validation while the BN model obtained best accuracy at 15-Fold cross-validation. When comparing accuracy and computational time (the time taken to build the model) of both models, efficiency and accuracy of the BN model are much better than the ANN model. The number of total instances was 1360. Of these 1360, 1343 instances were correctly classified (98.75%) for the highest Kappa statistical value of 0.9801 and F-measure value of 0.987 by the BN model at 15-folds cross-validation. Both models achieved high accuracy levels, and the advantage of these kinds of alternatives is classifying records according to correlation of probability rather than considering the exact cut off levels as shown in Table 3-2. Further classify the records according to critical levels (cut off values) of important parameter and number of iteration should be great classification model for water quality in a river basin.

Both methods managed to attain considerably high accuracy. Accuracy of both models depends on the corrections between each parameter. The number of incorrectly classified instances is 31 and 17 which correspond to the ANNs and BN models, respectively. Out of these 3 and 1 records were selected to correspond to ANN and BN model which were not predicted closest neighboring category instead of given classification. Out of 3 in ANN model, 2 records were recommended with reasonable categories. In BN model couldn't provide relevant categories for the mentioned record. Therefore, the ANN model is well recommended for correcting instances rather than the BN model.

3.5 Summary

The developed ANNs model represented good mapping between the water quality data and defined water quality classification and had the highest accuracy. The BN model performed the best for one number of parent in K2 searching algorithm with simple estimator of probability table selecting algorithm. The accuracy of the multilayer perceptron neural network model is 97.64% (Kappa statistic =0.96 and F-measure-0.976). The most accurate percentage of accuracy 98 .75% (kappa statistic = 0.98 and F-measure =0.987) were obtained by BN model. Therefore, the model is more accurate than the neural network model. Nonetheless both models are powerful alternatives for the purposes of classification and to construct an effective water quality index in the Kelani River.

3.5.1 Achievement

Developed the two water quality classification models with high accuracy using ANNs and BN for classify the water quality according to standers of classification classes in river basin. The BN model has shown a high accurate level than the ANNs.

3.5.2 Benefits

The proposed water quality classification standards for inland water resources in Sri Lanka are shown in Table 3-2. This study successfully attempted to develop alternative water classification models for Kelani River. This model could be used as an expedient tool to evaluate the water quality without measuring many other ingredients exhaustively, for example, when we need to make a rough estimate of water quality under some conditions in the future.

In addition, the BN model gives us some suggestive information about the water quality classes and the ranges of input parameters. First, the conditional probabilities table in the class variable would show some water quality classes with probability values, meaning that we could know how certain or uncertain the water classes are. Then, the parameter rages in conditional probabilities, which are discretized by the supervised discretization filter in Weka tool, might suggest other thresholds than those in Table 3-2. Therefore, the BN model will be helpful to create a clear view about variation and tolerance of the water quality classification.

3.5.3 Suggestion for further studies

The results of this study encourage further study on the intelligent categorizing technique that employs filtering hidden relationships with critical levels (cut off levels in classification table, e.g. Table 3-2) of important parameters that refer to a river basin as scenario.

4 INFLUENCE ASSESSMENT OF HUMAN ACTIVITIES

4.1 Introduction

Human activities pose a significant threat to the water quality of rivers. Urban activities in particular are highlighted as one of the major causes of contamination in surface water bodies in Asian countries. Due to the significant threat to water quality of rivers by human activities the population was mentioned in Chapter 2 as the one of criterion to find the optimized selection of sampling sites. Evaluation of sustainable human population capacities in river watersheds is necessary to maintain better freshwater ecosystems in a country while achieving its development goals as a nation. Regarding the water quality classification classes in the river basin, we assess the influence of human activities. According to the efficiency and accuracy result obtained in Chapter 3, the BN model was selected to develop a water classification model. The Kelani River in Sri Lanka was selected and this research seeks to understand the direct influence of human populations on water quality in river basins.

4.1.1 Background

Owing to the huge discharge of municipal wastewater and urban drainage into river basins around the world, the effect is more pronounced areas. Other impacts as explained section 1.1 significant changes of landscapes of watershed based on human activities affects to the water quality. This in turn affects aquatic life, agriculture and hence the lives of people who reside in these areas. Cost-effective methods are critical for developing countries to protect their finite natural resources. In Chapter 2 proposed the optimized networks consists of new sampling site than current monitoring sites considering the water population as one criterion. That mean evaluation of watershed population is necessary to find the critical points to monitor the water quality in the river basin. It motivates to assess the influence of human activities on water quality in the river. Rapid population growth leading to urbanization is commonly observed in the flat regions of many countries, hence the flat beds of the river basins are more susceptible degradation and pollution.

The high average annual ranges in BOD, TC and low DO level reported in the rivers of some South Asian countries are mainly due to the huge discharge of municipal wastewater and urban drainage into river basins [3]. In another study, the fecal coliform analysis of a long-term stream water quality monitoring system in the city of Atlanta, Georgia, in the United States highlighted the effects of urbanization whereby contamination exceeded Georgia's water quality standard for all usage levels [66]. The study of the spatial correlation between urbanization and water quality parameters based on a regional perspective, indicated human activities are positively correlated with degradation of water quality in rivers. Urban population density is used to assess the influence of urbanization [67]. The urban overflow, forest disturbances, mining, sewage disposal mainly effect for changes of Turbidity, DO, BOD, total suspended solids and faecal bacteria in the river water [16].

4.1.2 Situation analysis

Kelani River in Sri Lanka is rich in biodiversity and has many natural resources, and plays an important role in the sustainable development of the country. More than 25% of the Sri Lankan population benefits from the river [15]. Unfortunately, it is considered to be one of the most polluted rivers in Sri Lanka [17]. Further, unplanned anthropogenic activities (towns) together with, industrial and agricultural activities have been highlighted as major threats [24]. The lower and middle regions of the Kelani River, which consists of flatbed areas, are endangered by human activities due to urbanization. In the Ma Oya tributary in the lower region of the Kelani River, the exceeded standards were BOD (60%) and DO (80%). Further, standard levels of COD, BOD and DO were exceeded in the Raggahawatte sub-stream in the middle section of the river basin, which includes flatbeds, sub-montane and dense forests [15]. The uncontrolled land use activities primary influence to erosion and sedimentation and over withdrawal of water, and sand are other issues happen on water quality in Kelani River[16]. Therefore, low pH level has been occurred in Ambatale plant which is used to distribute pipe-borne water for Colombo district.

Sri Lanka is a developing country with an increasing population but its government needs to show adequate concern for the proper management of existing water resources while achieving sustainable economic development. The responsible authorities in Sri Lanka have already proposed many sustainable utilization approaches for Kelani River [19]. Flood encroachment is another disaster which frequently occurs and a major contributory factor to the flooding has been the rapid illegal construction (e.g. buildings, filling in of marshlands for development) taking place in the lower reaches of the Kelani River (i.e. Colombo, Gampaha and Kegalle). This has increased the sediment loadings, and organic and inorganic loadings in the river as a consequence of frequent floods [68].

Here we propose an approach that can assess the influence of population growth on the water quality of natural water bodies and obtain the ideal population ranges that can be accommodated within the carrying capacity of the watershed. The specific context for the case study comprised the lower and middle regions of Kelani River.

4.1.3 Problem identification

The challenge is addressing this demand while achieving sustainable development through controlling and mitigating the impacts of urbanization, industrialization and agriculture on natural water sources. Responsible authorities in Sri Lanka have proposed a long-term strategy and action plan for managing Kelani River by taking the point and non-point source pollution into consideration [15]. Evaluating the contamination and controlling the non-point source pollution is more difficult to do than for point source pollution. As an example, one water intake point in Kelani River called Ambatale has faced problems many times recently. Maintaining proper drinking water standards is very difficult due to of the outcomes of monitoring human activities Due to those fact, it should need proper evaluation method for assess the influence of human activates such as identify the correlation between population growth and water quality parameters or pollution index and define the threshold limits of influence of human activities without having degradation of water quality in the river.

4.2 Literature review

To assess the influence of human activities on water quality we need identify and evaluate its relationship with water quality. There is much correlation with water quality parameters and population. Here considered how to evaluate and define the correlation between population and water quality parameters. Yangtze River in China was subjected to a pollution index to evaluate the impact of urbanization on the water quality. The patterns of urban, suburban and rural with the pollution index were also explained [7]. Hongmei Bu et al. did work on the Jinshui River in China, and they reported that the population is a significantly correlated to water quality parameters. The multivariate linear regression model was used to express the impact of population [8]. They identified a population capacity for the river basin, and estimated the possible rates of increase of population and what effect these will have.

Classification models can be used to analyse various influential factors in natural environmental processes and in this case, assess the influence of human activities using the water quality classification standards (WQCS) for a river basin [54]. Complexity and uncertainty are major problems when analysing spreading sources of pollution such as anthropogenic activities in or near natural water ecosystems [69]. Of the various classification models such as the Bayesian Network (BN), Artificial Neural Networks (ANNs), Decision Trees and Support Vector Machine we propose to use the BN model to find the optimum population ranges that can be sustained by the natural environment in a watershed [70]. In Chapter 2, the BN model was described as a network with nodes representing probabilistic variables and links representing probabilistic dependencies. The conditional probability distribution given to each variable represents the influence of the parent nodes. The BN model acted to derive sustainable population ranges for specific water uses such as drinking, bathing and fishing by investigating the probabilistic influence of water quality parameters on the population.

Artificial neural networks (ANNs) can help understand the relationship of well-defined data set. BN models can create patterns by evaluating the dependencies among parameters and multi-layer perceptron (MLP) is performed with Bayesian statistics [54]. For example, Kinta River in Malaysia has a water quality index based on the water quality status [62], and so too dose Dusit district canal in Bangkok [63].

Given the work done on these system, this study likewise aims to propose a water quality classification standard applicable to for Kelani River. To develop predictive models rather than the traditional water quality assessment method, employing machine learning algorithms was emphasized [53]. The previous chapter compared the classification models explained the BN model functioned much better than the ANNs model.

This study focused on finding a level of population that can be sustained by the natural environment whilst maintaining water quality based on a BN classification model. The influence of anthropogenic activities on surface water environments has been studied around the world and their influence on the water ecosystem is highlighted [66], [71]. Most water quality issues reported in Asian and developing countries are due to the growth of population in watershed areas [3], [7], [8]. In this study, we assessed the influence of human communities on the water quality of Kelani River in Sri Lanka. The correlations between water quality parameters and watershed populations have been derived by quantitatively defining the populations corresponding to different water quality standards.

4.3 Proposed method

To assess the influence of human activities, here we classify population densities in watersheds by accounting for water classification classes. We can identify which part of a river basin is good for drinking, bathing or fishing by considering the population density of each watershed. We can also determine which part of the river is already contaminated due the population density. There are two main steps here:

- Identify the correlations between the water quality and population
- Qualitatively define the population ranges using the BN classification model

4.4 Methodology

4.4.1 Study Area

The study concerns the lower and middle regions of the Kelani River (6º 50′ - 7º 05′ N, 80º 12′ -79 º 10′ E), located in the Western Province of Sri Lanka (Figure 7). It stretches from the Sri Pada mountain range to Colombo. Colombo is one of the most highly urbanized areas in the Asian region [72]. Kelani River supplies approximately 80% of the water used to the Colombo district, and it is a primary source of drinking water. Some municipalities in the district are in L1 and L2 regions as shown in Figure 9. However, it is the most polluted river in Sri Lanka due to the rapid growth of industries located close to Kelani River [17], [25]. We used the data from five sampling sites located in five different watersheds, these are Raggahawathta Ela (Biyagama), Maha Ela (PallewelaOya), PusweliOya, WakOya, and PugodaOya. The points L1-L5 in Figure 4-1 respectively denote the above-mentioned sampling sites. The five water quality parameters namely, DO, COD, BOD_5 and NO_3^- served to examine the correlation between water quality and population.

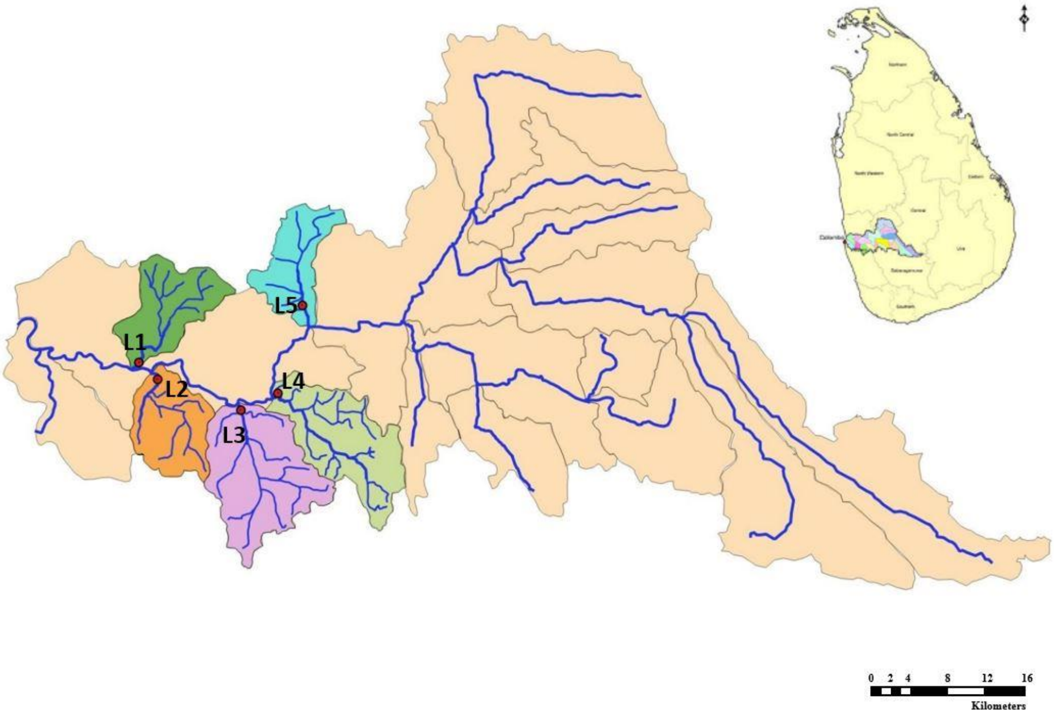


Figure 4-1 The five watersheds along Kelani River and their water sampling points

The dataset, includes five water quality parameters at five sampling points in different watersheds in the form of monthly data from 2003 to the end of 2013, with the exception of 2007 due to a lack of data for that year. We consider the data for all months in the dry and rainy seasons. There were some missing values and data errors due to both human and technical errors in the analysis, recording of results and failure to collect samples. Thus, we have 564 monthly data records of water quality in total.

Population data were obtained from the Department of Census and Statistics, Sri Lanka which are published on the city council's web site [73]. The percentages documented by local authority and secretariat divisions, in particular data from the report by the Kelani River Basin Multi-Stakeholder Partnership (KRMP) published in 2016, were used to calculate the population of each watershed [15]. The growth rate of the population, obtained from the 2001 and 2012 censuses, was applied to calculate the monthly population in each sub-watershed from 2003 to 2013. We assumed that the population of 2001 was the same as in 2003 and that of 2012 was the same as in 2013. The population densities in 2003 and 2013 of each watershed are given in Table 4-1. Estimating the population of L1 can be explained as follows: the population in January 2003 and in December in 2013 were assumed to be respectively 241,752 and 270,752 . Then a continuous growth rate was assumed at 221 per month.

Table 4-1. Population and density of watersheds

Watershed	Area (km^2)	Population density 2003	Population density 2013
L1	61.74	3916	4385
L2	61.46	3399	4150
L3	113.67	913	1133
L4	93.67	1791	2137
L5	52.14	2642	3099

Considering the existing data, the five selected water quality parameters of the river used in this analysis were TC, NO_3-, DO, COD and BOD_5 [20]. The WQCS given in Table 4-2 was mainly obtained from proposed water quality standards published by Western Region Megapolis Planning in 2015, which is attached in the KRMP report [16]. Considering the WQCS for the inland water source in Sri Lanka reported by the CEA, we assumed that the TC value of both Classes A and B are the same.

4.4.2 Assess the influence of population on water quality

We used the integration pollution index to yearly evaluate and visualize the influnce of human activities in each watershed area on the water quality of the river basin [7]. As mentioned in the above section, data of five water quality parameters of TC, NO_3^-, DO, COD and BOD_5 were used. The formula given in Equation 4-1 was used to calculate the integration pollution index of water quality.

$$\boldsymbol{PI = \frac{1}{m}\sum_{b=0}^{m}\frac{C_b}{C_0}(b = 1, 2, .. m)} \qquad \textbf{(4-1)}$$

Where *PI* is the integrated pollution index and C_b is the actual water quality parameter value of each sample. C_0 is the value of water quality standards; and *m* is the number of monitoring parameters. The five watersheds in Kelani River were classified into three categories, these being: urbanized level 1(UL1), which has highest population density with a threshold of 3,000, urbanized level 2 (UL2), with a threshold of 1,800; and urbanized level 3 (UL3), which has a low density of less than 1,800. The threshold urbanization level is obtained by considering the population density of each watershed given in Table 4-1. Next, we compared the *PI* of each category of urbanization for every year. Then the correlation coefficient of each water quality parameter with population was calculated separately to identify the greatest impact.

4.4.3 Development of Classification model

Development of the classification model based on a given data set is the typical role of data mining. Here, we develop the classification model used to classify the WQCS given in Table 4-2 based on water quality parameters and population. The water quality parameters selected based on their strength of correlation with the population in the previous step are shown in Table 4-2.

Table 4-2. Water Quality Classification Standards (WQCS)

Parameter	Drinking water with simple treatment Class - A	Bathing water Class - B	Fish and aquatic life Class - C
TC as MPN/100 ml	< 5000	< 5000	< 20,000
DO mg/l	> 6	> 5	> 3
BOD_5 mg/l	< 3	< 4	< 4

The values of water quality parameters which do not belong to any of the classes A, B or C were assumed to be polluted and deemed as class D. In Table 4-2, the WQCS are overlapping. When classifying each record, all five classification standards of class A were evaluated and if any were not satisfied, the classification standards of class B were checked. If any were not satisfied in B, then we moved to class C. If any records did not belong to any of the above three classes they were categorized under class D.

The problem domain of this classification is the set of variables *TC, BOD, DO* and population (*POP*) as well as the class variable $CV = \{A, B, C, D\}$. The BN classification model could be represented by the conditional probabilities of the unobserved class x_0 on the given observed data $x_1, x_2 \dots x_n$ in Equation 4-2.

$$\boldsymbol{P(x_0|x_1,x_2,..,x_n) = \frac{P(x_0,x_1,x_2,..,x_n)}{P(x_1,x_2,..,x_n)} \propto P(x_0,x_1,x_2,..,x_n)} \quad \textbf{(4-2)}$$

Where x_0 is a variable representing the unobserved class *CV* and $x_1, x_2, \dots, x_n$ are the set of variables of *TC, BOD, DO* and *POP*. The proportion ∝ holds because we assume the inputs are given. The BN implicitly encodes joint distributions and the probability of *n* attributes of x_i can be decomposed as a product of the joint probability distribution as shown in Equation 4-3.

$$\boldsymbol{P(x_0,x_1,,,x_n) = \prod_{i=0}^{n} P(x_i\,|x_i' s\ parents)} \quad \textbf{(4-3)}$$

The simple and very fast learning algorithm K2 and Tree Augmented Network (TAN) are two popular algorithms to training BNs (structure and probability distributions) from data [65]. Figure 4-2 shows two possible learning BN structures corresponding to the K2 and TAN algorithms. The K2 algorithm obtains the structure of the Naïve Bayesian (NB) network, which assumes conditional independency among $x_1, \dots, x_n$ given x_0 in Equation

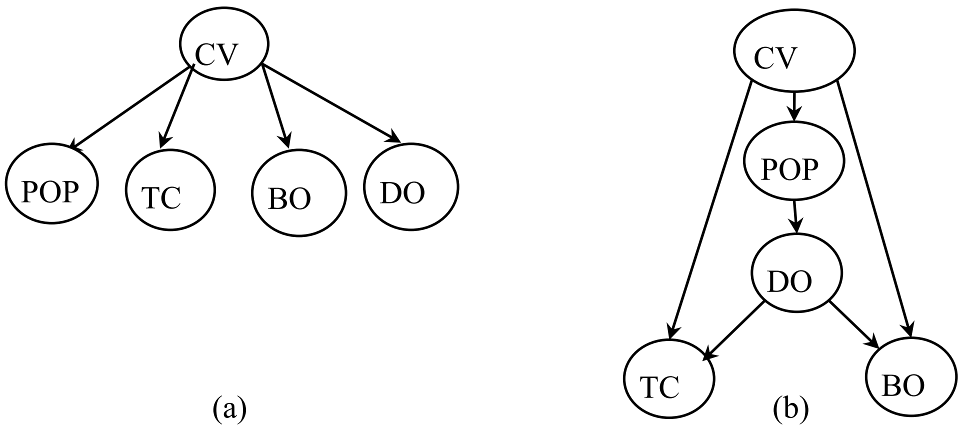


Figure 4-2. Structures of the Bayesian Network based on learning algorithms. (a) Structure of BN K2 learning algorithm. Structure of BN TAN learning algorithm

4-3. The Bayesian network with TAN learning algorithm indicates some dependencies in between variables other than the class variable.

The score for minimum description length (MDL) principle is employed to assess the models trained by K2 and TAN algorithms. It measures the quality of the network based on data by computing the log-likelihood of the resultant network while comprehending the network structure. This is achieved by searching possible sets of edges among each node and computing the conditional probabilities [64]. The equation computing the log-likelihood (*LL*) *l(V|D)* is shown in Equation 4-4. *V* is a set of random variables in the given data set (*D* = {*r1, r2,..ry.., rQ*}). *D* is the monthly data-set of water quality parameters containing DO, BOD and TC from the five sampling sites from 2003 to 2013 excepting 2007. The r_y is the y^{th} record of the data-set. *Q* is 564 which is total number of records in the data-set (*D*). Equation 4-5 explains the MDL score algorithm. *Y* is the number of variable; *LL* is the log-likelihood and q is the number of records in the data set (*D*). The value of *LL* is negative and the best structure should have the minimum score.

$$\boldsymbol{LL = l(V|D) = \sum_{y=1}^{Q} \log P(r_y|D)} \qquad \textbf{(4-4)}$$

$$\boldsymbol{MDL = -LL + \frac{Y}{2} logQ} \qquad \textbf{(4-5)}$$

To evaluate the accuracy of both BN models, a 10-fold cross-validation was applied for the given dataset. The Weka workbench was used to learn the algorithm and evaluate the performance of the algorithms [65]. Weka is a collection of state-of-the-art machine learning algorithms and data preprocessing tools. The parameters of the BN classifier were changed as follows in the Weka tool. The Simple Estimator-A 0.5 was used for both learning methods. It gives the direct estimates of the conditional probabilities. In the K2 learning algorithm, the random order was retained as false to use the order of nodes given by the dataset and the maximum number of parents was keep as 1 to obtain only the class node as parent. In the TAN algorithm, no specific options were applied. The given conditional probability table of the population node in the BN based on the TAN searching algorithm was analyzed to define the ranges of the population densities. The default given discretization filter of BN in Weka tool categorizes value range of each attribute in the nodes. It will help to identify the ranges of any attribute based on classification classes of parent node. The Fayyad and Irani's MDL method applied in supervised discretize filter given in default in Weka tool[74]. It divides the given range of an attribute based on information and decision theoretic notions.

4.4.3.1 Comparison of classification models

Using the same data-set, the accuracy and efficiency of the BN classification model was compared with the ANNs model. The three-layered ANNs with the backpropagation learning technique was developed for comparison. All algorithms ware executed in the same computational environment (CPU used i5 -3470, 3.20 GHz and RAM 8 GB).

4.5 Results and Discussion

4.5.1 Assess the influence of population on water quality

The population densities (per km^2) of each watershed in 2003 and 2013 are listed in Table 4-1. The average values of the above population densities were 4151, 3774, 1023, 1964 and 2871 which correspond to the watersheds L1–L5, considering that L1 and L2 are categorized as UL1, L4 and L5 as UL2, and L3 as UL3, as defined in subsection 2.2. These watersheds cover 2.65%, 2.63%, 4.87%, 4.02% and 2.23% of the entire watershed area of the Kelani River, respectively.

According to PI, the impact of the level of urbanization on surface water quality during the monitoring period is shown in Figure 4-3. PI was higher in UL1 followed by UL2 and UL3 for each year from 2003 to 2013. However, in 2005 and 2006 the differences in PI between UL1 and UL2 were not significantly higher compared to the other years. In the year 2006, both UL1 and UL2 scored a higher PI than UL3, with a significantly lower fall of 0.04 for UL1 compared to UL2 (UL1 obtained a PI value of 1.737 and UL 2 obtained a value of 1.726). In 2005, the PI value of UL1 was increased by 0.01 compared to UL2. Consequently, the contribution rate of pollution loads for each category showed the highest PI value in 2013 compared to other years.

The population growth rates from 2003 to 2013 in UL1, UL2 and UL3 are 16%, 18%, and 24%, respectively. When the 2003 scenario is compared with that of 2013 with reference to PI, the percentage increase in PI of the three urbanization levels were 50%, 109% and 141%, which also correspond to the population growth rates of each urbanization level. The PI values of all three categories in every year are above the standard value, except for UL3 in 2003. This indicates that the contamination of water occurs in all five watersheds.

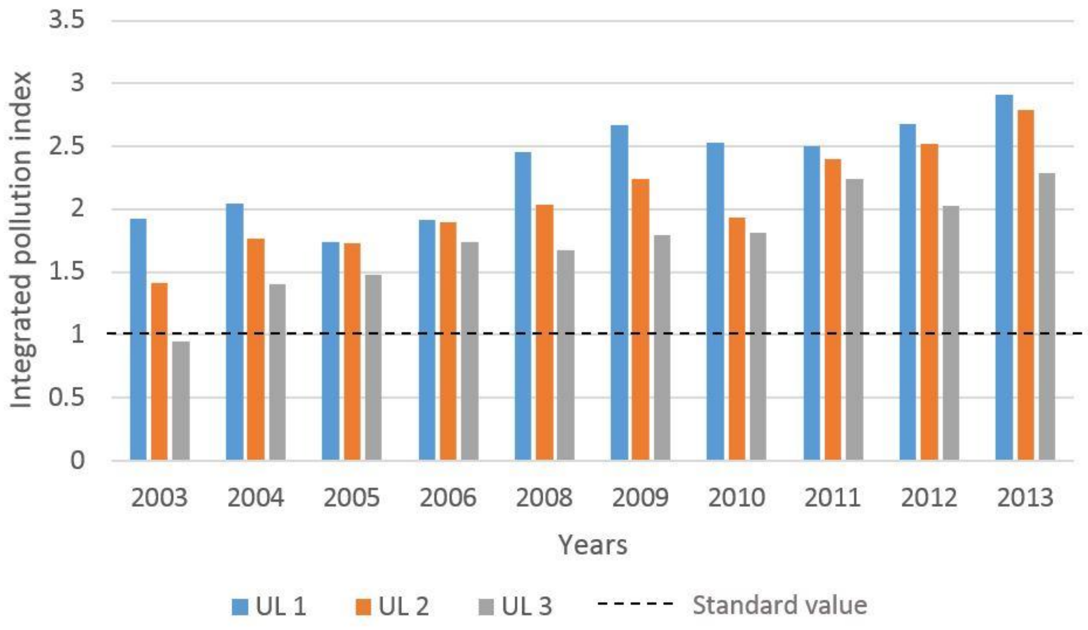


Figure 4-3 Spatial pattern of the integrated pollution index according to level of urbanization

The PI increases with rising level of urbanization in all five watersheds are shown in the three years 2003, 2008 and 2013 (Figure 4-4). The linear graph is used to represent the increment by considering the minimum and maximum PI values of the five watersheds and corresponding populations of a particular year. The calculation for 2003 can be explained as follows. The intensity of river pollution has clearly increased over time, as urbanization level also rose. The PI levels have increased from 1.14, 1.53 and 1.92 to 2.42, 2.66 and 2.91, respectively. The spatial pattern of the integrated pollution index given by the present study closely relates to similar existing results from a study conducted on the Yangtze River in China [7].

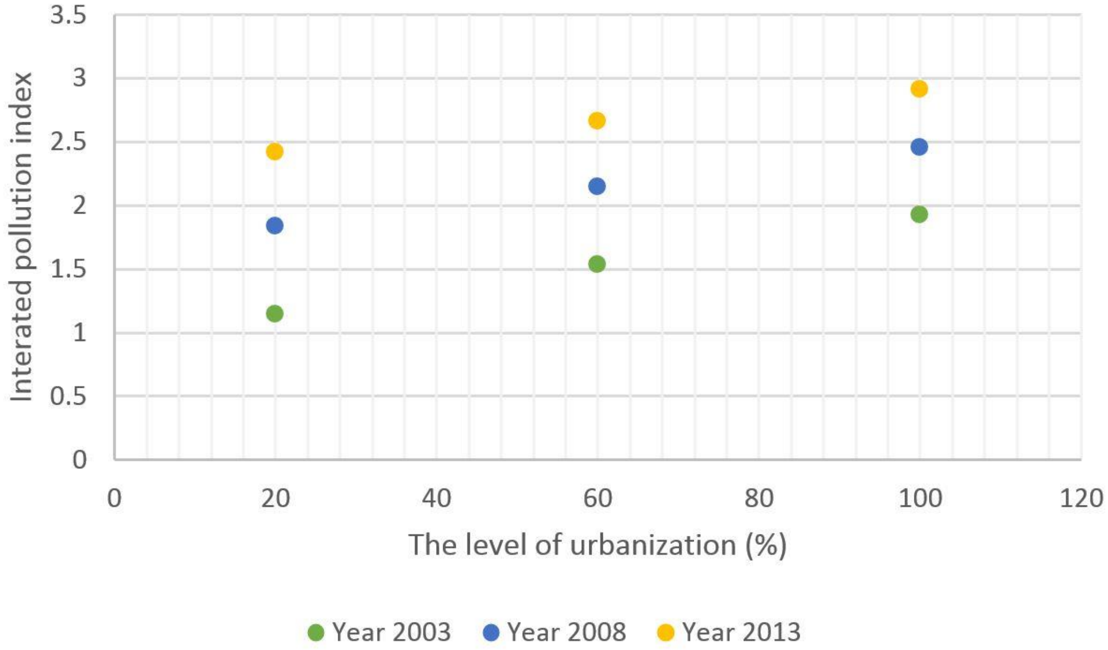


Figure 4-4 Relationship between the integrated pollution index and the level of urbanization considering the increment of population in 2003,2008 and 2013

4.5.2 Evaluation of correlation coefficients

We calculated the correlation coefficients between the five parameters and populations separately. The results are shown in Table 4-3. All the variables were positively correlated with population except DO ($p < 0.01$). TC, BOD and DO revealed a significantly high correlation with population. In general, the values of the correlation coefficients between 1.0 and 0.5 can be considered as indicating a strong relationship [75]. Further, to obtain a classification model of with higher accuracy, we selected water quality parameters for which the correlation coefficient with population density were greater than 0.5. Data used in the study suggested that the five parameters TC, BOD, COD, DO and NO_3^- (Table 4-3), can be used to develop the model. However, only three parameters (TC, BOD and DO) were selected, given that their correlation coefficient was greater than 0.5. The results for water quality in three urban areas in Nepal, India and Bangladesh also showed positive correlations of BOD and TC with population and a negative correlation of DO with population [3]. Less correlation between population density and NO_3^- has been shown in other research conducted in Sierra Nevada, California by Dylan S. Ahearn et al. [76]. The research conducted on the Jinshui River

Basin in the South Qinling Mountains, China, predicted the most correlated water quality parameters, which have a strong correlation with population, by defining the linear equations [8]. Comparison of the results for both rivers clearly illustrates the relationship between population and the water quality of the river basin.

Table 4-3 Correlation coefficients of water quality parameters and population

Water Quality Parameters	**Correlation Coefficients (p < 0.01)**
TC	0.687
BOD	0.745
COD	0.29
DO	−0.699
NO_3^-	0.400

4.5.3 The development of the Classification Model

The water quality classification standards given in Table 4-2 were used to determine the water quality classes of the training data. The performance of the BN classification model was discussed with two different learning algorithms. The result of the ANN model using a backpropagation algorithm served to compare the accuracy and efficiency of the BN model. The performance of the two classification models, ANN and BN are shown in Table 4-4. The TAN learning algorithm obtained −2280.95 as the minimum MDL score and had the highest accuracy level at 98.40% in the 10-fold cross-validation of the given dataset compared to the K2 learning algorithm. The least computational time also occurred in the BN model with the TAN learning algorithm. The accuracy of the ANN model was evaluated by changing some hidden layers and changing the training time (epochs). Finally, the best performance of the ANN model occurred at 2000 epochs with 15 hidden layers. In the previous study on water quality classification, the BN performed better than the ANN [55]. As shown in Table 4-4, the best performance was achieved by the TAN algorithm of the BN. Therefore, this research used the BN classification model with the TAN learning algorithm to quantitatively define the population ranges.

Table 4-4 The Summary of three classification models' performance

Method of model	Accuracy out of 564 (%)	Micro average of recall	Micro average of precision	Computational time (Seconds)
BN with K2 leaning algorithm	548 (97.16 %)	0.972	0.973	0.02
BN with TAN learning algorithm	555 (98.43%)	0.984	0.984	0.01
ANNs	551 (97.69 %)	0.977	0.977	3.62

The confusion matrix of the BN classification model is given in Table 4-5. Obtained values of recall and precision for each class based on the confusion matrix explain the performance of the classification model. Therefore, the total number of records of the test data set is 564. The recall values of A, B, C and D classes are 1.0, 1.0, 0.98 and 0.984 respectively. The precision values of the above classes are respectively 1.0, 0.97, 0.99 and 0.964 Further, the micro-averages of recall and precision are equal to 0.984.

Table 4-5 The confusion matrix of the BN classification model with TAN learning algorithm

		Forecasted quality			
		A	B	C	D
Actual quality	A	69	0	0	0
	B	0	36	0	0
	C	0	0	287	6
	D	0	1	2	163

4.5.4 Quantitative population range

The obtained probability distribution of the population density is shown in Table 4-6. The table represents the probability distribution of P ($POP|CV$) and helps us to understand the variation in the population density ranges according to the water quality classification classes.

Table 4-6 The population density classification ranges and probability of each classification class

Water Quality Classification Class	Probability of Class Variable in Each Density Range Defined by BN Models (... × 100%)		
	0 to 2375 (POP-1)	2375 to 2672 (POP-2)	Above 2672 (POP-3)
A	0.816	0.064	0.121
B	0.867	0.013	0.12
C	0.576	0.188	0.236
D	0.104	0.021	0.875

This study has identified three different population ranges according to the supervised discretize filter of Bayesian network given in Weka tool. The total probability of each class is equal to one. The water quality classification classes A and B demonstrate significantly higher probabilities compared to C in the lowest population density range of POP-1. The probabilities are lower in A and B than in C in the POP-2 range. The probability of class D denoted the highest probability (0.875) in the POP-3 range. We can consider a population density of between 2375 and 2672 as the critical point of population density for water contamination. Except for class C, the population range for POP-2 does not show high probability values. Therefore, the POP-2 range is suitable for maintaining water quality in class C. Taken together, a watershed with a population density of less than 2375, helps to maintain water quality at a higher level the classification classes A and B in the river basin. Conversely, a population density less than 2672 is helpful with regard to class C. This proposed population density can be taken into account when determining the watersheds'capacity to sustain a certain population density. Numerous studies conducted on water pollution in river basins have identified a high correlation between population and water quality parameters [3], [7], [8]. The present study further evaluated the effects and defined threshold values for population density with respect to the classification of water quality in river basins.

Research conducted along Jinshui River by Hongmei Bu et al. proposed a method to find the threshold value of population utilizing a quadratic equation of the pollution index and population [8]. The threshold values of population were based on the sample site's total factor score with the best water quality. In contrast, sustainable population density levels were derived from the WQCS of river water.

4.6 Summary

To assess the influence of human activates and water quality we selected three parameters which had the highest correlation coefficients of 0.7, 0.69, 0.69 ($p < 0.01$). These corresponded to biochemical oxygen demand (BOD), dissolved oxygen (DO) and total coliform (TC). Finally, we proposed a quantitative approach to estimate the population capacity of watersheds based on water quality classification standards (WQCS), employing the Bayesian network (BN) classification model. The optimum population ranges were obtained from the probability distribution table of the population node in the BN model. Results showed that the population density should be approximately less than 2375 to maintain the water quality in the watershed for bathing and drinking purposes and approximately less than 2672 for fish and other aquatic organisms.

4.6.1 Achievement

In this research, we have quantitatively identified the ideal range of population density for a watershed in order to maintain the quality of water at an appropriate level. We identified that water quality is worst in highly populated areas, average in medium populated areas and less serious in areas with small populations. Also identified population density as a major factor that should be well controlled to overcome the rapid deterioration and degradation of the water ecosystem. Finally, the proposed means can use to classify the water quality in the river basin based on population density of watershed.

4.6.2 Benefits

This proposed concept can be explained as a river classification method for different activities such as drinking, bathing and fishing according to the population density of a particular watershed. Therefore, any country can apply this model for the new establishment of a watershed management strategy. Apart from that, existing urbanized watersheds can identify the threshold value of population and offer many waste-water treatment facilities to remove pollution. After training, the classification model can be applied to different watersheds in the same river to identify threshold population densities by considering watershed population without using water quality parameters. For this reason, it is a superior and low-cost method for river water management in developing countries. The decision-making tools required for the above-mentioned processes can be augmented in a good practical sense by the threshold population densities derived in the present study.

Addressing the environmental management problems and making suggestions to change human behaviors that cause such problems are vital for protecting nature [77]. Studies have been carried out to predict acceptable population levels in river basins, i.e., Hongmei Bu et al. estimated the population capacity of a river basin [8]. However, limiting the population living near a watershed will not always be a practical solution and may be the last option when it comes to the most critical situations. The population density ranges proposed by the present study can be implemented by the relevant management authorities when introducing new rules and regulations and setting appropriate standards.

4.6.3 Limitation and uncertainty

The accuracy and ability of this method is based on the number of parameters showing a correlation with population and the strength of that correlation. This was based on our finding which revealed a positive correlation between the water quality parameters DO, BOD and TC and population. The applicability of the model will be compromised if there are not enough correlations and the correlation coefficient is unsatisfactorily low. Land use and environmental factors may sometimes affect the river's health more significantly compared to only the human population density of a river basin. For example, a river may go through heavily industrialized areas, large cultivated areas, or have a long course where soil erosion impacts are high [78].

4.6.4 Suggestion for further studies

The applicability of the present model is suitable for any country. Since relevant data are readily available in developed countries the model can be employed for existing data and will produce positive outcomes for multiple management objectives. The model can be further developed based on correlations between water quality and agricultural practices, industrialization, and infrastructure developments other than population density.

5 CONCLUSION

The study of the literature review regarding the degradation of water quality in the river basin showed the water quality is directly affected by the human influence. Thus, the dissertation has been able to highlight the importance of controlling the contamination due to human activities. The ultimate objectives of this research are to clarify the huge influence of human activities on water quality in a river, and to show the necessity to monitor the population as well as the pollution. Therefore, it has been evaluated the relationship between population growth, water quality parameters and the pollution index. Further, it has been estimated the limitation of the human activities without having degradation of water quality in the river and incorporated the population to design the water quality monitoring network. As the result, the dissertation proposed a new approach, which can be used to classify the water quality in the river basin according to watershed population density. The approach can be applied to practical environment management at low cost specially for developing countries. In addition, the optimal water sampling network consisting of the 14 points were proposed to enhance the current monitoring network in the Kelani River.

The Kelani River in Sri Lanka has been selected for the study. The Kelani River in Sri Lanka is rich with biodiversity and many natural resources, and plays an important role in the sustainable development of the country. More than 25% of the Sri Lankan population take benefits from the river. Unfortunately, it is considered to be one of the most polluted rivers in Sri Lanka.

5.1 Proposed model

The empirical study has proposed an influence assessment model of watershed population on water quality in the Kelani River in Sri Lanka. The model contains the following approaches introduced newly in this study.

- An approach to select the optimal water sampling site network for water quality
- A model to classify water quality from water quality parameters.
- A method to assess the influence of watershed population on water quality of the Sri Lankan river

5.2 Contribution

The overall contribution of this research is, to develop the assessment model on the water quality taking into account the influence of watershed population, which can be used for rivers in developing countries. The significant contribution can be listed as follows.

1. This empirical study discusses the relations between water quality and human activities. It has also been identified the population as a threat for the water resource and discussed possibility of the management. Furthermore, it showed the requirements that could be used to control the water quality degradation.
2. This study is going to propose a new method to select the optimal water sampling network, where the following actions are taken into account instead of the conventional approaches.
   - The watershed population is incorporated in the fitness function as the form of development pressure index (DPI).
   - To identify the most essential monitoring points to evaluate the water quality in the water intake, considered the distance between the intake and the monitoring site in nearest downstream sampling points rather than the nearest upper stream sampling points were mostly used in the conventional methods.
   - Due to the budget limitation common in developing countries, the proposed approach can be used to select the sampling site network with the given number of sites.
3. To develop the classification models for water quality, I proposed the water quality classification standards are given in Table 3-2.
4. According to the influence assessment of human activities on the water quality, this research has been able to generate an informative solution fulfilling its' third objective. In this case population densities are classified in watershed considering the water classification classes also. BN was applied to achieve this goal. The watershed was classified as follows using their population density. The population density should be approximately less than 2375 to keep the water quality in the watershed for bathing and drinking purposes and approximately less than 2672 for fish and other aquatic organisms.

The above mentioned information can be used by the Urban Development Authority (UDA), which is the monitoring body for urban planning and development in the entire Sri Lanka in collaboration with local councils. The Zoning decision would have been implemented based on the suitability and the capacity of the site itself regarding the development plans in the future. In present urban planning practice in Sri Lanka no such tools are employed to assess the potential impact of zoning decisions on the environment. The proposed model can be employed by the planning agencies as a tool to assess the potential impact of zoning decisions on the natural environment. Further, identification of the most relevant watershed for urbanization and capacity will be helpful to water resources management. Accordingly, the zoning regulations can be ratified without making harmful effects to the watershed for people who migrate to existing cities.

The scarcity of land for human settlement in Sri Lanka often leads to conflicts with the natural landscape which consists of 103 river basins right throughout the country. Therefore, the assessment of impact of population distribution on water quality will facilitate to promote sustainable urban planning decisions in the long run. As far as the existing population distribution pattern is concerned, information derived from the proposed model will be helpful to identify the minimum capacity of wastewater treatment facilities and change the environmental protection policies in existing human settlements. One of the best model to evaluate the efficiency of any solution that apply for controlling the influence of pollution sources in the watershed such as watershed population, land usages. This proposed low-cost approaches are most reliable for the water quality monitoring system to predict the water quality and while finding the contamination source to control the water pollution.

5.3 Limitations

The result of optimized selection of sampling sites network depends on the number of objective functions. According to existing data here used four fitness functions to evaluate the for fitness score. Also, only one DPI factor was used as objective function according the available data.

The accuracy of the means of population classification under the proposed method is limited for three water quality parameters. Lack of water quality parameters values has affected to find out robust data set. Due to that, the accuracy is depended on the correlation between population and water quality parameters.

5.4 Future works

The water quality monitoring locations in a network could not be static or unique forever. The development of the dynamic optimized selection of river sampling sites network can be suggested for future work. Further, to increase accuracy and efficiency of optimization process can use more suitable factor in DPI and ESI indexes.

The applicability of the present model of classifying the water quality based on watershed population is suitable for any country. Since relevant data are readily available in developed countries, the model can be trained on existing data, and it would produce positive implication s for multiple management objectives. The model can be further developed based on correlations between water quality and agricultural practices, industrialization, and infrastructure developments other than population density.

Finally, this research assessed the influence of population on water quality in rivers as follows. The developing optimized selection of sampling sites networks considering watershed population and defined the threshold value of population densities to keep quality water in river basin using quality classification model. Further, pointed out the possibility of controlling pollution by raising the awareness of stakeholders and relevant authorities of natural water bodies.

REFERENCES

[1] Sandra Fancieullo, Donna Somboonlakana, and Fred Suffian, “Protecting Clean Water for People and Nature | The Nature Conservancy,” *The Nature Conservancy*. [Online]. Available: https://www.nature.org/ourinitiatives/habitats/riverslakes/protecting-clean-water-for-people-and-nature.xml. [Accessed: 13-Apr-2017].

[2] Hari Srinivas, “We’re finally at the end of the UN Decade for Water 2005-2015 – It is time to say good-bye,” *International Decade for Action “Water for Life” 2005-2015*, 08-Sep-2015. [Online]. Available: http://www.un.org/waterforlifedecade/. [Accessed: 06-Apr-2017].

[3] S. K. Karn and H. Harada, “Surface Water Pollution in Three Urban Territories of Nepal, India, and Bangladesh,” *Environmental Management*, vol. 28, no. 4, pp. 483–496, Oct. 2001.

[4] Hafsa Sabry, “Thousands of Lives at Risk Kelani River Contamination,” *The sunday Leader*, 07-Sep-2015. [Online]. Available: http://www.thesundayleader.lk/2015/09/27/thousands-of-lives-at-risk-kelani-river-contamination/. [Accessed: 10-Jan-2017].

[5] H. Qin, Q. Su, S.-T. Khu, and N. Tang, “Water quality changes during rapid urbanization in the Shenzhen River Catchment: An integrated view of socio-economic and infrastructure development,” *Sustainability*, vol. 6, no. 10, pp. 7433–7451, 2014.

[6] J. . Wallace, “Increasing agricultural water use efficiency to meet future food production,” *Agriculture, Ecosystems & Environment*, vol. 82, no. 1–3, pp. 105–119, Dec. 2000.

[7] J. Wang, L. Da, K. Song, and B.-L. Li, “Temporal variations of surface water quality in urban, suburban and rural areas during rapid urbanization in Shanghai, China,” *Environmental Pollution*, vol. 152, no. 2, pp. 387–393, Mar. 2008.

[8] Hongmei Bu, Wenzhi Liu, Xianfang Song, and Quanfa Zhang, “Quantitative impacts of population on river water quality in the Jinshui River basin of the South Qinling Mts., China | SpringerLink,” *Enviromental Earth Sciences*, no. 75, p. 292, Feb. 2016.

[9] Chamara P. Liyanage and A. Marasinghe, “Requirements analysis for appropriate ICT based water management system for Sri Lanka,” presented at the 1st Int. Conf. on Energy, Environment and Human Engineering, Yangon, Myanmar, pp.256-271 Dec-2013.

[10] Yoon Seongkyu, “Ministry of Environment.” The environmental authority of the Republic of Korea, [Report], 2015.

[11] Wei Yang, Jun Nan, and Dezhi Sun, “An online water quality monitoring and management system developed for the Liming River basin in Daqing, China,” *PubMed Journals*, vol. 88, no. 2, pp. 318–325, Apr. 2007.

[12] Y. Miyata and T. Hata, “ICT-Based remediation with knowledge information management for contaminated groundwater,” *Geotechnical and Geological Engineering*, vol. 31, no. 3, pp. 911–926, 2013.

[13] R. O. Strobl, P. D. Robillard, R. L. Day, R. D. Shannon, and A. J. McDonnell, “A water quality monitoring network design methodology for the selection of critical sampling points: Part II,” *Environ Monit Assess*, vol. 122, no. 1–3, pp. 319–334, Nov.

2006.
[14] K. Mei, Y. Zhu, L. Liao, R. Dahlgren, X. Shang, and M. Zhang, "Optimizing water quality monitoring networks using continuous longitudinal monitoring data: a case study of Wen-Rui Tang River, Wenzhou, China," *Journal of Environmental Monitoring*, vol. 13, no. 10, pp. 2755–2762, 2011.
[15] Ananda Mallawatantri, Ajith Rodrigo, and Kushani De Silva, "Natural Resource profile of the Kelani River Basin," Central Environment Authority Sri lanka,[Report] 2016.
[16] Ananda Mallawatantri, Ajith Rodrigo, and Kushani De Silva, "Natural Resource Management and Monitoring Unit published final report of Medium to Long-term Multi-Stakeholder Strategy and Action Plan for Management and Conservation of the Kelanni River Basin,"[Report], Central Environment Authority Sri lanka, 2016.
[17] O A Ileperuma, "Environmental pollution in Sri Lanka: a review," *Journal of the National Science Foundation of Sri Lanka*, vol. 28, no. 4, pp. 301–325, Dec. 2010.
[18] Chamara P Liyanage, Marasinghe A, and Wijesinghe , "Applicability of AI for Water Quality Monitoring Network Design; a Case of Kelani River, Sri Lanka," presented at the Eleventh Annual Session, Artifical Intelligence, Colombo, pp. 86-92,Mar-2015.
[19] Ananda Mallawatantri, Ajith Rodrigo, and Kushani De Silva, "Preliminary perception survey on water safety: Kelani River Basin.",[Report] , Colombo, 2016.
[20] L. Y. Khuan, N. Hamzah, and R. Jailani, "Prediction of water quality index (WQI) based on artificial neural network (ANN)," *SCOReD 2002. Student Conference on*, 2002, pp. 157–161,2002.
[21] Rinkesh, "What is Waste Management and Methods of Waste Disposal?," *Conserve Energy Future*, pp. 23-28, Apr-2014. .
[22] Horst M Vogel, "Human Impacts to Water Quality - Limpopo River Awareness Kit." [Online]. Available: http://www.limpopo.riverawarenesskit.org/LIMPOPORAK_COM/EN/RIVER/WATER_QUALITY/HUMAN_IMPACTS.HTM. [Accessed: 30-Aug-2017].
[23] Fran Jamieson, "Let only rain down the storm drain," Viera, pp.364-369, 2002
[24] T W A W Wijesinghe, "Status of Water Quality of Kelani River," Central Environmental Authority , Sri Lanka,pp. 45-51, Colombo, 2010.
[25] C. P. Liyanage, A. Marasinghe, and K. Yamada, "Comparison of Optimized Selection Methods of Sampling Sites Network for Water Quality Monitoring in a River," *International Journal of Affective Engineering*, vol. 15, no. 2, pp. 195–204, 2016.
[26] J. H. Lee, "Determination of Optimal Water Quality Monitoring Points in Sewer Systems using Entropy Theory," *Entropy*, vol. 15, no. 9, pp. 3419–3434, Aug. 2013.
[27] N. B. Harmancioglu, Y. Icaga, and A. Gul, "The use of an optimization method in assessment of water quality sampling sites," *European Water*, vol. 5, no. 6, pp. 25–35, 2004.
[28] J.-J. Kao, P. H. Li, and W. S. Hu, "Optimization models for siting water quality monitoring stations in a catchment," *Environmental monitoring and assessment*, vol. 184, no. 1, pp. 43–52, 2012.
[29] L. J. Alvarez-Vázquez, A. Martínez, M. E. Vázquez-Méndez, and M. A. Vilar, "Optimal Location of Sampling Points for River Pollution Control," *Math. Comput. Simul.*, vol. 71, no. 2, pp. 149–160, Apr. 2006.
[30] C. P. Cetinkaya and N. B. Harmancioglu, "Assessment of water quality sampling sites by a dynamic programming approach," *Journal of Hydrologic Engineering*, vol.

17, no. 2, pp. 305–317, 2011.
[31] W. Dixon, G. K. Smyth, and B. Chiswell, "Optimized selection of river sampling sites," *Water Research*, vol. 33, no. 4, pp. 971–978, Mar. 1999.
[32] Q. Chen, W. Wu, K. Blanckaert, J. Ma, and G. Huang, "Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses," *Journal of environmental management*, vol. 110, pp. 116–124, 2012.
[33] I. T. Telci, K. Nam, J. Guan, and M. M. Aral, "Optimal water quality monitoring network design for river systems," *Journal of environmental management*, vol. 90, no. 10, pp. 2987–2998, 2009.
[34] Y. Ouyang, "Evaluation of river water quality monitoring stations by principal component analysis," *Water Research*, vol. 39, no. 12, pp. 2621–2635, Jul. 2005.
[35] H. T. Do, S.-L. Lo, P.-T. Chiueh, and L. A. P. Thi, "Design of sampling locations for mountainous river monitoring," *Environmental Modelling & Software*, vol. 27, pp. 62–70, 2012.
[36] A. W. Pollak, J. J. Peirce, L. J. Alvarez-Vázquez, and M. E. Vázquez-Méndez, "Methodology for identifying optimal locations of water quality sensors in river systems," *Environmental Modeling & Assessment*, vol. 18, no. 1, pp. 95–103, 2013.
[37] S. W. Athukorala *et al.*, "Ananlysis Of Water Quality Variation In Kelani River, Sri Lanka Using Principal Componet Analysis," in *SAITM Research Symposium on Engineering Advancements*, 2013, pp. 129–135.
[38] E silva, "Project report of Quality Assessment of Surface Water," Institute of Fundamental Studies, pp.269-275, 1996.
[39] S. Park, J. H. Choi, S. Wang, and S. S. Park, "Design of a water quality monitoring network in a large river system using the genetic algorithm," *Ecological Modelling*, vol. 199, no. 3, pp. 289–297, Dec. 2006.
[40] R. W. Meier and B. D. Barkdoll, "Sampling design for network model calibration using genetic algorithms," *Journal of Water Resources Planning and Management*, vol. 126, no. 4, pp. 245–250, 2000.
[41] P. F. Hudak and H. A. Loaiciga, "An optimization method for monitoring network design in multilayered groundwater flow systems," *Water Resources Research*, vol. 29, no. 8, pp. 2835–2845, 1993.
[42] D. P. Letternmaier, D. E. Anderson, and R. N. Brenner, "Consolidation of a Stream Quality Monitoring Network," *JAWRA Journal of the American Water Resources Association*, vol. 20, no. 4, pp. 473–481, 1984.
[43] N. B. Harmancioglu and N. Alpaslan, "Basic approaches in design of water quality monitoring networks," *Water Science and Technology*, vol. 30, no. 10, pp. 49–56, Nov. 1994.
[44] C. Lee, K. Paik, D. G. Yoo, and J. H. Kim, "Efficient method for optimal placing of water quality monitoring stations for an ungauged basin," *Journal of environmental management*, vol. 132, pp. 24–31, 2014.
[45] P. van Thienen, "Alternative strategies for optimal water quality sensor placement in drinking water distribution networks," 2014.
[46] R. O. Strobl and P. D. Robillard, "Artificial intelligence technologies in surface water quality monitoring," *Water international*, vol. 31, no. 2, pp. 198–209, 2006.
[47] J. A. Vasquez, H. R. Maier, B. J. Lence, B. A. Tolson, and R. O. Foschi, "Achieving water quality system reliability using genetic algorithms," *Journal of environmental engineering*, vol. 126, no. 10, pp. 954–962, 2000.
[48] Y. Icaga, "Genetic Algorithm Usage in Water Quality Monitoring Networks Optimization in Gediz (Turkey) River Basin," *Environ Monit Assess*, vol. 108, no.

1–3, pp. 261–277, Sep. 2005.
[49] P. Tanos, J. Kovács, S. Kovács, A. Anda, and I. G. Hatvani, "Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account," *Environmental monitoring and assessment*, vol. 187, no. 9, p. 575, 2015.
[50] Babaei Semiromi F, Hassani A H, Torabian A, and Karbassi A. R, "Water quality index development using fuzzy logic: A case study of the Karoon River of Iran," *African Journal of Biotechnology*, vol. 10, no. 50, 2011.
[51] M. E. Borsuk and C. A. Stow, "Bayesian parameter estimation in a mixed-order model of BOD decay," *Water Research*, vol. 34, no. 6, pp. 1830–1836, Apr. 2000.
[52] S. Tyagi, B. Sharma, P. Singh, and R. Dobhal, "Water Quality Assessment in Terms of Water Quality Index," *American Journal of Water Resources, American Journal of Water Resources*, vol. 1, no. 3, pp. 34–38, Jan. 2013.
[53] W. Wu, G. C. Dandy, and H. R. Maier, "Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling," *Environmental Modelling & Software*, vol. 54, pp. 108–127, 2014.
[54] K G Abistado, C N Arellano, and E A Maravillas, "Weather forecasting using artificial neural network and Bayesian network," *JACIII*, vol. 18, no. 5, pp. 812–817, Sep. 2014.
[55] C. P. Liyanage and K. Yamada, "Comparison of Water Quality Classification Models Using Artificial Neural Network and Bayesian Network," SCIS & ISIS 2016, Sapporo, Japan, pp.958-961,2016
[56] Y. C. Lai, Y. T. Tu, C. P. Yang, R. Y. Surampalli, and C. M. Kao, "Development of a water quality modeling system for river pollution index and suspended solid loading evaluation," *Journal of Hydrology*, vol. 478, pp. 89–101, Jan. 2013.
[57] Adam P Piotrowski, Maciej J Napiorkowski, and Jaroslaw J Napiorkowski, "Comparing various artificial neural network types for water temperature prediction in rivers," *ScienceDirect*, vol. 529, no. 1, pp. 302–315, Oct. 2015.
[58] S. Emamgholizadeh, H. Kashi, I. Marofpoor, and E. Zalaghi, "Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models," *International Journal of Environmental Science and Technology*, vol. 11, no. 3, pp. 645–656, 2014.
[59] G. G. Oliveira, O. C. Pedrollo, and N. M. R. Castro, "Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 47–61, Apr. 2015.
[60] M. J. Alizadeh and M. R. Kavianpour, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean," *Marine pollution bulletin*, vol. 98, no. 1, pp. 171–178, 2015.
[61] Gemunu Herath, "Water quality Management in Sri Lanka," The Asia Science & Technology seminar, Thailand,pp.489-495, 2014.
[62] N. M. Gazzaz, M. K. Yusoff, A. Z. Aris, H. Juahir, and M. F. Ramli, "Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors," *Marine Pollution Bulletin*, vol. 64, no. 11, pp. 2409–2420, 2012.
[63] Sirilak Areerachakul and Siripun Sanguansintukul, "Water Classification Using Neural Network: A Case Study of Canals in Bangkok, Thailand,", International

Conference for Internet Technology and Secured Transactions, pp.359-365, Bangkok, 2009.
[64] Richard E Neapolitan, *Introduction to Bayesian Network in Title of Learning Bayesian Networks*. Prentice Hall Upper Saddle River in US: Prentice Hall., 2004
[65] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*. Burlington, MA: Morgan Kaufmann, 2011.
[66] N. E. Peters, “Effects of urbanization on stream water quality in the city of Atlanta, Georgia, USA,” *Hydrol. Process.*, vol. 23, no. 20, pp. 2860–2878, Sep. 2009.
[67] H. Zhao, X. Duan, B. Stewart, B. You, and X. Jiang, “Spatial correlations between urbanization and river water pollution in the heavily polluted area of Taihu Lake Basin, China,” *J. Geogr. Sci.*, vol. 23, no. 4, pp. 735–752, Aug. 2013.
[68] I.P. Ajith Gunasekara, “Flood hazard mapping in lower reach of Kelani River, Sri Lanka,” *Journal of the Institution of Engineerings, Sri lanka*, vol. 41, no. 5, pp. 149–154, Jan. 2008.
[69] Day Sharon K, Thyne Geoff, Poeter Eileen, and MCCRAY Jone, “Effects of Population Growth on Water Quality in a Mountain Watershed,” Conference of Colorado, pp.l89-194, Oct-2002.
[70] R. Rosly, M. Makhtar, M. K. Awang, M. N. A. Rahman, and M. M. Deris, “The study on the accuracy of classifiers for water quality application,” *International Journal of u-and e-Service, Science and Technology*, vol. 8, no. 3, pp. 145–154, 2015.
[71] Paulo S pompeu and Carlos Bernardo M Alves, “The Effects of Urbanization on Biodiversity and Water Quality in the Rio das Velhas Basin, Brazil,” American Fisheries Society, 2005, vol. 47, pp. 11–22.
[72] M. Kawamura, S. Jayamanna, Y. Tsujiko, and A. Sugiyama, “Comparison of urbanization of four Asian cities using satellite data,” *Doboku Gakkai Ronbunshu*, vol. 1998, no. 608, pp. 97–105, 1998.
[73] Oskar Blakstad, “Sri Lanka: Provinces & Cities - Population Statistics in Maps and Charts,” *City Population*, 02-Jan-2017. [Online]. Available: http://www.citypopulation.de/SriLanka-Cities.html. [Accessed: 06-Jul-2017].
[74] Fayyad, Usama M, Irani, and Kebi B, “Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning,” *Proceedings of the International Joint Conference on Uncertainty in AI*, pp. 1022–1027, 1993.
[75] Oskar Blakstad, “Statistical Correlation - Strength of Relationship Between Variables.” [Online]. Available: https://explorable.com/statistical-correlation. [Accessed: 06-Jul-2017].
[76] D. S. Ahearn, R. W. Sheibley, R. A. Dahlgren, M. Anderson, J. Johnson, and K. W. Tate, “Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California,” *Journal of Hydrology*, vol. 313, no. 3, pp. 234–247, Nov. 2005.
[77] I Altman and J F Wohlwill, “Behavior and the Natural Environment - Springer Link.” Plenum Publishing Corporation, pp. 562-569, New York,1983.
[78] M. M. El-Sheekh, “Impact of Water Quality on Ecosystems of the Nile River,” in *The Nile River*, Springer, Cham, 2016, pp. 357–385.

APPENDICES

A1. Publications of the Research

A1.1 International Journal

C. P. Liyanage, A. Marasinghe, and K. Yamada, “Comparison of Optimized Selection Methods of Sampling Sites Network for Water Quality Monitoring in a River,” International Journal of Affective Engineering, vol. 15, no. 2, pp. 195–204, 2016.

C. P. Liyanage, and K. Yamada, “Impact of Population Growth on the Water Quality of Natural Water Bodies”, Sustainable, vol. 9, no. 8, 1405, 2017.

A1.2 Conference Publications

C. P. Liyanage and K. Yamada, “Comparison of Water Quality Classification Models Using Artificial Neural Network and Bayesian Network,” SCIS&ISIS2016, Sapporo, Japan, pp. 958-961, 2016 (Awards for *Best Poster Presentation*)

Chamara P Liyanage, Marasinghe A, and Wijesinghe, “Applicability of AI for Water Quality Monitoring Network Design; a Case of Kelani River, Sri Lanka,” Eleventh Annual Sessions of Sri Lankan Association for Artificial Intelligence, Colombo, pp.86-92, Mar-2015.

C. P. Liyanage, A. Marasinghe, and Bandunee Liyanage, “Assess the Applicability of Remote Sensing Mapping to Monitor the Water Quality in Kelani in Sri lanka”, IGCN2015, pp.56, Nagaoka, Japan, June 20, 2015.

C. P. Liyanage, A. Marasinghe, and T.W.A. Wijesinghe,” Effective ICT based water quality data visualization for the Kelani River in Sri Lanka”, Asiagraph 2015 conference, Tainan National University of the Arts, pp.269-275, Guant, April 2015

C. P. Liyanage, and A. Marasinghe, “Optimized Selection of Sampling Sites Network for Water Quality Monitoring of River in Sri Lanka”, ISASE, Tokyo, Japan. pp.220-226 March 2015.

C. P. Liyanage, A. Marasinghe, “ICT Based Water Quality Monitoring System of Kelani River in Sri Lanka”, First Symposium of Japan Sri Lanka Technology Promotion Association (JSTPA), Open University of Sri Lanka, Nawala, Sri Lanka. Pp.259-265, March 2015

A2 Water quality parameters values

A2.1Water quality standards of parameters

The values of S_{ijk} are shown in Table A2.1 as bellow.

Table A2.1 Standards values of water quality parameters

BOD5 mg/l(5days at 20oC or 3 days at30oC)	**COD mg/l**	**DO mg/l at 25 ^{0}C**	**pH**
5	25	5	7

The sample of actual values water quality parameters that used to calculate are shown in Table A2.2

Table A2.2 Water quality parameters of October in 2010

Ref of Site	**pH**	**DO**	**COD**	**BOD**
L9	7	7.8	2	0
L7	6.8	7.1	8	2
L6	6.8	7.1	5	1
L5	6.5	6.5	17	3
L4	6.7	7.4	5	1
L3	6.5	7.8	11	3
L2	6.5	4.9	7	2
L14	7.2	7.5	3	1
L13	7.3	6.7	3	0
L12	6.2	7.6	6	1
L8	6.8	6.6	10	3
L11	6.5	7.9	8	0
L10	6.6	7.9	6	1
L1	6.4	5.1	14	3

A2.3 Water quality data of training data set

The sample data set of water quality parameters which was used in Chapter 3 to develop the water quality classification model are given in Table A2.3.

Table A2.3 Sample data set of training dataset in classification model used in Chapter 3

pH	DO	COD	BOD	NO3-	Water quality class
6.65	6.35	13.5	3	0.335	C2
6.6	6.45	13.5	2.5	1.18	C1
6.6	4.05	18	2	1.08	C3
6.6	5.5	9	2.5	0.002	C2
6.8	5.2	31.5	10	0.66	C4
6.7	3.65	27	10	0.705	C4
6.5	5.45	15	2	0.76	C2
6.25	5.7	18	2.5	1.17	C2
6.55	5.95	24	2.5	0.615	C3
6.25	4.805	27	3.5	2.72	C3
6.15	3.305	13.5	6.5	2.15	C4
6.4	2.915	61.5	5.5	0.39	C4
6.9	5	21	3	0.11	C3
6.85	6.6	25.5	3	0.155	C3
6.7	6	19.5	2.5	0.08	C2
7.05	5.5	19	3.5	0.17	C2
6.95	4.9	111	7.5	1.285	C4
6.85	4.4	40.5	8.5	0.23	C4
6.4	6.45	27	3	0.525	C3
6.4	6.15	19.5	3	2.1	C2
6.7	6.25	25.5	3.5	0.39	C3
6.75	5.35	13.5	3	0.295	C2

The End…