DATA SCIENCE AND INTERDISCIPLINARY RESEARCH: RECENT TRENDS AND APPLICATIONS

Editor:
Brojo Kishore Mishra

Bentham Books

Advances in Computing Communications and Informatics

(Volume 5)

Data Science and Interdisciplinary Research: Recent Trends and Applications

Edited by

Brojo Kishore Mishra
Department of Computer Science & Engineering
NIST Institute of Science and Technology (Autonomous)
Institute Park, Pallur Hills, Golanthara
Berhampur-761008, Odisha, India

Advances in Computing Communications and Informatics

(Volume 5)

Data Science and Interdisciplinary Research: Recent Trends and Applications

Series Editors: Pradeep Kumar Singh, Bharat Bhargava & Wei-Chiang Hong

Volume Editor: Brojo Kishore Mishra

ISSN (Online): 2737-5730

ISSN (Print): 2737-5722

ISBN (Online): 978-981-5079-00-5

ISBN (Print): 978-981-5079-01-2

ISBN (Paperback): 978-981-5079-02-9

First published in 2023.

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
80 Robinson Road #02-00
Singapore 068898
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

PREFACE

Data science has recently gained much attention for a number of reasons, Big Data is the most significant among them. Scientists (from almost all disciplines including physics, chemistry, biology, and sociology, among others) and engineers (from all fields including civil, environmental, chemical, and mechanical, among others) are faced with challenges posed by data volume, variety, and velocity, or Big Data.

The book contains quantitative research, case studies, conceptual papers, and model papers, review papers, theoretical backing, *etc*. This book will cover data science and its application to interdisciplinary science.

This book will prove valuable for graduate students, researchers, academicians, and professionals in information science, business, health, planning, manufacturing, and other areas who are interested in exploring the ever-expanding research on Data Science.

Chapter-01 provides a detailed survey and comparative analysis of various methodologies in the prediction of rainfall over multiple countries.

Chapter -02 focuses on applying clustering for gaining the benefits of evolutionary computation to process large-scale data and based on optimality, the performance of the datasets can be measured.

Chapter-05 presents an investigation of the data obtained from IoT sensors and observed that a huge amount of work can be done in the reliability analysis of the data from the sensors deployed in the agricultural fields.

Chapter-06 says that - Smart devices have rapidly started intruding our lifestyles with the technological promotion of the Internet of Things. One of the most used smart devices is the electric meter. Urban areas witness power theft as well as un-proportionate billing, both incurring tremendous losses to the respective exchequers. We thought that if a system may be designed which can predict power utilization and also classify the current usage, it would be beneficial to both the service providers as well as the consumers. Equipped with such thoughts, thorough research was conducted to monitor electric consumption and fault detection in the devices.

Chapter-08 focused on undertaking a quick analysis of socio-economic conditions. Information on the aforementioned parameters was gathered in order to get insight into the research area's socio-economic profile.

In the end, we thank the contributory authors, reviewers and my family members for their support. Special thanks to Prof. (Dr.) Pradeep Kumar Singh for his best support as a Book Series Editor. The editors are also thankful to all members of Bentham Science Publication house.

Brojo Kishore Mishra
Department of Computer Science & Engineering
NIST Institute of Science and Technology (Autonomous)
Institute Park, Pallur Hills, Golanthara
Berhampur-761008, Odisha, India

List of Contributors

Adrija Dasgupta	Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India
A. Kavitha	Department of Computer Science, Kongunadu Arts and Science College, Coimbatore, Tamil Nadu, India
C. Kishor Kumar Reddy	Stanley College of Engineering and Technology for Women, Hyderabad, India
Diganta Sengupta	Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India
D. Alamelu	KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India
Debasish Swapnesh Kumar Nayak	Department of Computer Science & Engineering, Institute of Technical Education and Research, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India
Das Jayashankar	Avior Genomics, Mumbai, India
Fahmina Taranum	Computer Science and Engineering Department, Muffakham Jah College of Engineering and Technology, Hyderabad, India
Gouri C. Khadabadi	Department of CSE, KLS Gogte Institute of Technolgy, Belagavi, India
Jyoti Prakash Mishra	Gandhi Institute for Education and Technology, Baniatangi, hubaneswar, Affiliated to BijuPatnaik University of Technology, Rourkela, Odisha, India
Jayashankar Das	Avior Genomics, Mumbai, India
K. Manimekalai	Department of Computer Applications, Sri GVG Visalakshi College for Women, Udumalpet, Tamil Nadu, India
K. Sridevi	Computer Science and Engineering Department, Muffakham Jah College of Engineering and Technology, Hyderabad, India
Manjunath Managuli	Department of ECE, KLS Gogte Institute of Technology, Belagavi, India
Nguyen Gia Nhu	Dean, Graduate School, Duy Tan University, Da Nang, Vietnam
N. Venugopal	Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India
P.R. Anisha	Stanley College of Engineering and Technology for Women, Hyderabad, India
Pankaja S. Kadalgi	Department of CSE, KLS Gogte Institute of Technolgy, Belagavi, India
Sambit Kumar Mishra	Gandhi Institute for Education and Technology, Baniatangi, hubaneswar, Affiliated to BijuPatnaik University of Technology, Rourkela, Odisha, India
Sudha Slake	Department of CSE, KLS Gogte Institute of Technolgy, Belagavi, India
Subhash Mondal	Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India
Samrat Podder	Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India
Suharta Banerjee	Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India

CHAPTER 1

A Comprehensive Study and Analysis on Prediction of Rainfall Across Multiple Countries using Machine Learning

C. Kishor Kumar Reddy[1,*], P.R. Anisha[1] and **Nguyen Gia Nhu[2]**

[1] *Stanley College of Engineering and Technology for Women, Hyderabad, India*

[2] *Dean, Graduate School, Duy Tan University, Da Nang, Vietnam*

Abstract: Rainfall is one of the most considerable natural occurrences, which is important for both human beings and living beings. Since the environment is changing and there is a huge change in weather, it is noted that the rainfall cycles are also varying and the earth's temperature is increasing day-by-day. The changes in weather conditions like humidity, pressure, wind speed, dew point and temperature affect the agriculture, industry, production, and construction and also lead to floods and land-slides. Hence it is one of the important factors to be noted for human beings to keep track of the natural occurrences in order to survive. In order to overcome these issues, a system is required which is able to forecast and predict the rainfall using statistical techniques which is the most popular tool in modern technology. This paper provides a detailed survey and comparative analysis of various methodologies used in the prediction of rainfall over multiple countries. Comparison is made in terms of various performance measures: accuracy, precision, recall, RMSE, specificity, sensitivity, MAE, F-Measure, ROC and RAE. Further, the drawbacks with existing approaches applied so far in the prediction are discussed.

Keywords: Artificial Neural Networks, Classification Techniques, Decision Trees, Naïve Bayes, Rainfall, Random Forest, SVM.

INTRODUCTION

Weather forecasting has become one of the most serious problems in the world and many researchers, governments, industries, risk management communities and scientific communities are looking into this issue. Weather is a natural and climatic characteristic that affects the daily routine of human activities such as farming in agriculture, production, construction, generation of electricity, forest and many more factors. Weather forecasting is one of the most important factors

[*] **Corresponding author C. Kishor Kumar Reddy:** Stanley College of Engineering and Technology for Women, Hyderabad, India; E-mail: kishoar23@gmail.com.

Brojo Kishore Mishra (Ed.)

because if the weather changes, then it may lead to natural calamities such as landslides, volcanoes, earthquakes, hurricanes, *in situ*., which lead to a lot of loss to society. Therefore, it is suggested to have a proper approach for the prediction of rainfall with which we can take preventive measures for these natural calamities. This forecasting helps in supervising human activities such as agriculture, production, construction, tourism, and many more. This forecasting also helps disaster prevention agencies with proper predictions regarding the weather with which they can take corrective measures and make a decision in order to prevent society from natural calamities.

Weather forecasting has the highest impact on human life and activities. Changes in the weather are rapidly increasing, hence it is crucial to conduct research on weather prediction and provide precipitation data for the prediction of weather timely and give early warning to avoid natural disasters. Rainfall prediction is one of the most crucial areas in the field of weather forecasting. Rainfall is one of the essential occurrences within the climatic system, whose disordered nature has the highest impact on water resources, agricultural and biological systems. Hence rainfall prediction is important for agricultural farming, tourism, navigation, and sailing. The collection of data related to weather has become very easy and sorted and moreover meteorological data can also be collected due to innovation and research in the fields of science and technology. A large amount of meteorological data should be collected for weather forecasting and it is very difficult to attain good accuracy. Science and development have improved in the field of information technology and computers with which researchers are able to analyse large amounts of data using big data analytics, which have found hidden relationships using machine learning techniques.

Numerous natural disasters across the world are correlated to meteorological phenomena. In the most recent years, lots of machine intelligent learning techniques have been proposed to undertake, unlike issues in the world. In this paper, multiple intelligent learning models are discussed for precipitation forecasting across different parts of the earth. It is pragmatic that intelligent learning methods are able to make predictions with a smaller amount of error rate. An additional improvement in machine intelligent learning methods is computational performance, and it executes more rapidly with much lesser computer resources than methods based on differential equations that are presently used in operational centers. Henceforth, it is significant to consider intelligent machine learning models for precipitation forecasts in operational centres as a method to advance prediction quality and decrease computational costs. The utilization of machine intelligent learning methods coupled with meteorological science has encouraged scientists, academicians, and researchers to apply machine intelligent learning methods to classify and predict events in different climatic conditions.

Predictions can be made using several of methods such as machine learning techniques for classification and prediction, soft computing, and complex methodologies such as artificial neural networks (ANN). These are the most common techniques used for the prediction of weather. The functional correlation between data, and the correlation that is not known or difficult can be made using ANN's since they learn from examples and are based on self-adaptive mechanism. Problems that are complex can be solved using deep learning techniques. Deep learning refers to a series of several multilayers that are trained using unsupervised methodologies. Learning dense, valid and non-linear data using unsupervised methodologies results in knowledge development and we can be able to predict new data. The above methodology is used in the fields of natural language processing, bioinformatics, object recognition and computer vision. Deep learning has been promising for modelling time series data through methods such as Conditional RBM, Autoencoder, Restricted Boltzmann Machine (RBM), Recurrent neural networks, Convolution and pooling, and Hidden Markov Models. The main idea behind this paper is to present a wide survey of traditional statistical methodologies along with modern methodologies of Machine Learning in the prediction of rainfall accurately. Further, a comparison on rainfall predictions that used different approaches is made. Some reasonable solutions for predicting weather efficiently are also recommended.

RELEVANT WORK

Prediction of rainfall is a tedious work particularly when we expect accurate and exact values for predicting the rainfall. Rainfall prediction is trending research in the field of scientific research areas of technology and innovation in the modern world, since it has a large impact on the socio-economic life of human beings and all living beings.

Du *et al.* [1] researched and proposed deep belief networks methodology to be used in forecasting weather precipitation. A one year meteorological data is collected and used which is taken from Nanjing station. The author discussed the applications of big data processing methodologies in the field of meteorology for meteorological datasets. The proposed methodology is based on deep belief networks that develop a statistical model between precipitation characteristics and other meteorological information. Abhisheka *et al.* [2] conducted an experiment on the capability of ANN methodology by developing an efficient and consistent non-linear predictive model for analyzing weather. The authors have made comparison and assessed the performance of the model using diverse transfer functions, hidden layers, temperature of 365 days and neurons for maximal forecasting.

CHAPTER 2

A Novel Approach for Clustering Large-scale Cloud Data using Computational Mechanism

Zdzislaw Polkowski[1], Jyoti Prakash Mishra[2] and **Sambit Kumar Mishra[2,*]**

[1] *Department of Humanities and Social Sciences, The Karkonosze University of Applied Sciences in Jelenia Góra, Jelenia Góra, Poland*

[2] *Gandhi Institute for Education and Technology, Baniatangi, Bhubaneswar, Affiliated to BijuPatnaik University of Technology, Rourkela, Odisha, India*

Abstract: In the present situation, with the enhancement of virtualization techniques, it is very essential to keep track of accumulated large-scaled heterogeneous data in every respect. In addition to that, it is also necessary to prioritize the processing mechanisms when being linked with clustered data. Sometimes it has been observed that the large scaled datasets are too complex and therefore, the normal computation mechanisms are not sufficient or adequate for the specific applications. But it is highly required to observe the significance of each individual dataset and focus on the responses being accumulated from other aspects to make a suitable decision and generation of efficient analytical clustered data. The main aim of such applications is to apply the clustering gaining merits from evolutionary computation to process the large-scaled data and based on optimality, the performance of datasets can be measured.

Keywords: Clustered data, Entity integrity, Heterogeneity, Query term, Resource prioritization.

INTRODUCTION

Nowadays, the complexities linked to the large scaled heterogeneous data in general have been increased due to somehow incorporating virtual machines. In fact, it is very difficult to analyze either the complexities or the processing abilities using normal mining mechanisms or computational methods. As the application of these clustered datasets is quite relevant to the extraction of meaningful information as well as making decisions, specific tools and computation mechanisms may be deployed which in return may be more computationally efficient and adaptive to the situation. Of course, the integrity of

[*] **Corresponding author Sambit Kumar Mishra:** Gandhi Institute for Education and Technology, Baniatangi, Bhubaneswar, affiliated to BijuPatnaik University of Technology, Rourkela, Odisha, India; E-mail: sambitmishra@giet.edu.in

Brojo Kishore Mishra (Ed.)

data resembles complete accuracy along with its consistency. In a similar situation, the accumulation of processes can be maintained by having associated rules during the conceptualization stage. Also the suitable validation of data with check parameters can ensure the categorization of sensitive data while maintaining its integrity. Prioritizing both physical and logical integrity constrained, it has been observed that not only it will provide accuracy of data but also it will focus on unique parameters of data adopting entity integrity. Sometimes a group of similar processes intend to provide uniform storage with no provisioning commonalities. So, it is essential to adopt referential integrity rules to overcome the situation. Sometimes to preserve the accuracy, it is also required to prioritize the domain values and quantify the constraints. There are many mechanisms to search and obtain user-defined patterns as far as the large scaled data is concerned. In fact, specific classification mechanism can also be applied prioritizing the decomposition of fixed-length data. Somehow the primary purpose of this application is to obtain merits from evolutionary computation to process the large scaled data. Accordingly, the optimality linked to the datasets can be measured based on the performance.

REVIEW OF LITERATURE

Reiss *et al.* [1] in their work have focused on the criteria of cloud computing services specifically on scientific computations. They observed the importance of specific computations during the accumulation of resources from data centers.

Zhang *et al.* [2] in their work prioritized WSDF framework to observe the performance of transfer of data specifically linked with workflows on web services. Somehow it is provisioned with the involvement of data within services associated with web application. Also, it is associated with providing better performance with enhanced speed of transfer of data in the application.

Liu *et al.* [3] in their work have given more importance to clustered data and tried to obtain the machine response. Also they observed the performance of machines that have a storage capacity value of 0.5 In fact, the temporal locality of machine events is required to be considered in such situations.

Reiss *et al.* [4] in their work have discussed resource heterogeneity of the Google clusters. In fact, they observed considerable heterogeneity in different available resources along with resource allocation.

Yousif *et al.* [5] in their work have focused on clustered mechanisms linked to the cloud tasks which are intended to support the scheduler of the data center in the cloud. Probably this is directly linked to the replacement of virtual machines. In their work, they have prioritized k means as well as density-based clustering.

Reiss *et al.* [6] in their work tried to obtain the facts related to the utility of resources within the clusters and simultaneous requests on accompanying the tasks. In fact, the prediction of usage of resources during task execution can be judged with the actual usage of resources and probably the performance can be measured with not much difficulty. Accordingly, the authors tried to determine the actual implementation of tasks within the clusters.

Al-Dulaimy *et al.* [7] in their research focused on the enhancement of utilization of resources prioritizing the performance and minimizing the consumption of energy associated with data centers. In fact, the authors analyzed the large scaled tasks along with clustering.

Tayal *et al.* [8]. in their work focused on concepts of optimization linked to the scheduling of tasks toward cloud computing system using fuzzy genetic algorithms. They observed the appropriateness of the decision on scheduling of tasks during the entire process. In fact, these are quite modeled to imprecise scheduling parameters to attain the objectives.

Evangelinos *et al.* [9] in their research analyzed the implications associated with virtualized resources. As per their analysis, the performance achievement implied the complexities linked to specified applications on cloud resources. The enhancements in the process applications reflect mapping in the processors associated with clusters.

Raicu *et al.* [10] in their work prioritized the structure of managing data provisioned during the implementation of retrieval of data. They focused their work on grid computing towards handling the data. In fact, they observed that encapsulation as well as virtualization can be implemented quite efficiently in grid computing.

Agrawal *et al.* [11] in their work focused on complexities associated with the application of data science and business intelligence. They observed the importance of optimization criteria in specifically adopting the cross-platform systems and to some extent the machine learning models. The uniqueness lies in the encapsulation of the entire enumeration of plans. Initially the cost linked to transformation of queries can be avoided and then permitting to speed up the entire enumeration process by exploiting the processing elements for quantification.

Agrawal *et al.* [12] in their application discussed the rules based as well as cost based optimizers. They observed that these may require minimum support from the system administrators to exhibit efficient execution plans.

CHAPTER 3

Secure Communication Over In-Vehicle Network Using Message Authentication

Manjunath Managuli[1,*], Sudha Slake[2], Pankaja S. Kadalgi[2] and **Gouri C. Khadabadi[2]**

[1] *Department of ECE, KLS Gogte Institute of Technology, Belagavi, India*

[2] *Department of CSE, KLS Gogte Institute of Technolgy, Belagavi, India*

Abstract: This chapter describes the initial working plan and the project carried out, along with their respective analyses. The project focused on implementing a cryptographic solution in the embedded system to ensure the authentication and integrity of CAN messages. With the rise of the autonomous driving concept in the automotive industry, there is a growing need for digital data processing and communication both within the vehicle and with the external world, creating a significant challenge to protect the data from cyber-attacks during communication. Hence, cybersecurity has become an important topic in the automotive industry. This project was carried out to ensure the security of data during communication on the vehicle Controller Area Network (CAN) using message authentication. The chapter concludes by highlighting the current technology's drawbacks and discussing potential future improvements.

Keywords: Automotive industry, Cyber-attacks, CAN, Cybersecurity, MAC, Message authentication, Secure communication.

INTRODUCTION

The latest key trends in the automotive industry involve transforming vehicles, especially cars, into computers on wheels, which requires the integration of numerous digital features into the vehicle system. These trends include autonomous driving, which grants greater control to the electronic system, as well as connectivity and digitization, electric vehicles, and shared mobility. However, these advancements also create new challenges as they make vehicles more vulnerable to cyber-attacks [1], thereby increasing the opportunities for attacks, broadening the threat landscape, and increasing the potential for attacks to occur.

[*] **Corresponding author Manjunath Managuli:** Department of ECE, KLS Gogte Institute of Technology, Belagavi, India; E-mail: manjunathm16@gmail.com

Brojo Kishore Mishra (Ed.)

Background for Vehicle Security

The latest trend in the automotive industry involves increasing the connectivity between vehicles, which is achieved by providing more connection ports to the external world, mainly through wireless communication technologies such as cloud connection to access the Internet [2]. As a result, communication between vehicles and other external entities, such as infrastructure, is becoming more common, known as Vehicle-to-Vehicle (V2V) and Vehicle-to-External (V2X) communication. However, this increased connectivity also brings with it an increase in threats to customer safety, security, and privacy, as depicted in Fig. (**1**).

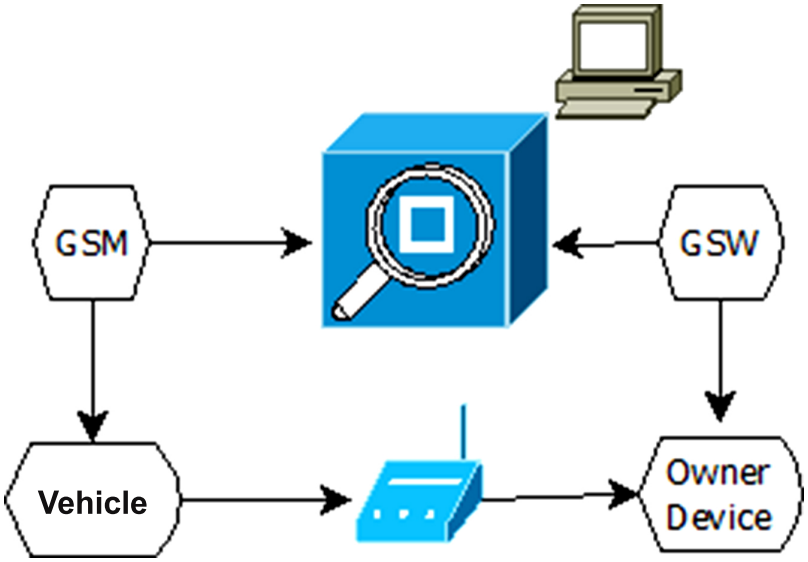

Fig. (1). Overview of connected car concept.

Hacking Incidents on Vehicles

The automotive industry is facing an increasing number of cyber-attacks, with hackers targeting vehicles that are loaded with new technologies [3]. These attacks involve stealing important data, which can later be used to demand money from either the vehicle owner or the manufacturer. The first instance of research on vehicle hacking began in 2010 to analyze the security of modern automobiles. In 2012, incidents of BMW and VW hacks were reported, where the immobilizer and OBD-II ports were used for hacking purposes. A Jeep hacking incident in 2015 became famous, as a researcher was able to hack into the infotainment system and take control of critical assets of the Jeep, leading to the recall of millions of Jeep vehicles [4]. Another incident in 2015 involved a Tesla vehicle being hacked, with the attacker gaining control over the vehicle's doors and windows remotely. The infotainment system, which has wireless connections with

the external world through Wi-Fi, Bluetooth, cellphones, and infrared, is often the first target for cyber-attacks on vehicles. Figs. (**2** & **3**) provide an overview of the most interesting and popular hacking incidents of recent years [5].

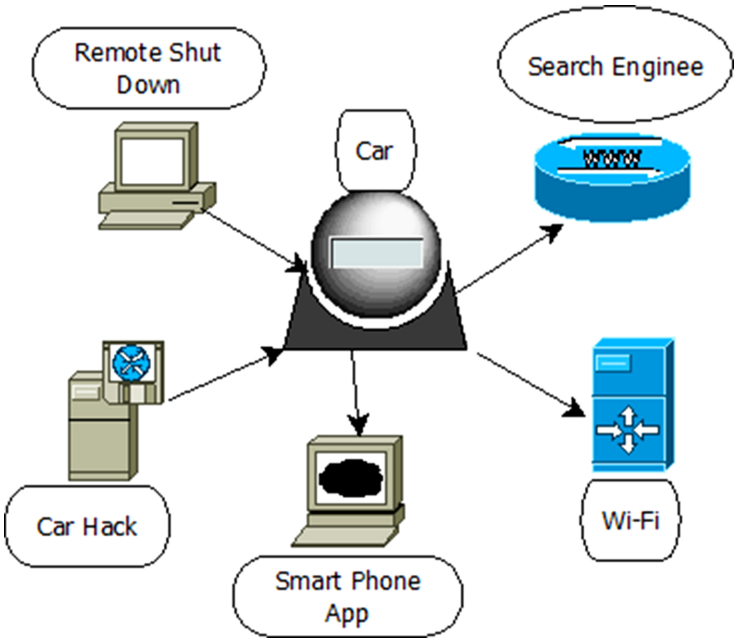

Fig. (2). Overview of connected car concept.

Economic Value at Risk Due to Poor Security Investments

In case of a vehicle hacking incident caused by inadequate security measures, the costs involved in recovering from the incident can be substantial and can have a negative impact on the brand name of the vehicle manufacturer. Naturally, implementing security measures in vehicles requires significant financial investment from both the vehicle manufacturer and product suppliers. The investment required for security implementation is compared to the consequences of vehicle hacking incidents in the Costs vs. Investments graph.

Security Goals

The major components of the cyber security goal are listed as Confidentiality, Integrity & Authentication and Availability in Fig. (**4**).

CHAPTER 4

A Decision Model for Reliability Analysis of Agricultural Sensor Data for Smart Irrigation 4.0

Subhash Mondal[1], Samrat Podder[1] and **Diganta Sengupta[1,*]**

[1] *Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India*

Abstract: Agriculture is the backbone of an Agro-based Country's Economic System as it employs the majority of the population. Internet-of-Things (IoT)-based intelligent systems help reduce losses and make efficient use of available resources. This paper aims to detect anomaly conditions that might occur in sensor nodes related to day-t--day smart irrigational activities in an agricultural field. IoT-based irrigation systems being prone to unauthorized intrusion can cause damage to smart farms in terms of crop damage and infertility of the soil. In this paper, we propose an intelligent decision-making system that can identify Anomalous Conditions and Suspicious Activities. The model discussed in this paper uses the idea of Gaussian distribution, which calculates the expected probability of a given state of an agricultural field and classifies anomalies based on what previous probabilities of an anomaly state looked like. The approach classifies the anomalies with an accuracy of 80.79%, a precision of 0.81, and a recall of 0.54 under test conditions.

Keywords: Smart Farming, Internet of Things, Anomaly Detection, Sensor Nodes, Smart Irrigation.

INTRODUCTION

Anomaly detection, also known as Outlier Detection, is the technique of identifying events that deviate from a systems' normal behavior [1]. Such anomalous events can indicate situations such as technical glitches, a shift in monitored systems' behavior, fraudulent activities, or Cyber-Attacks. Therefore, it is crucial to detect such events for improvement in existing systems, to detect, and prevent cyber-attacks and fraudulent activities. To date, a huge amount of research has been done on Anomaly Detection, earlier by the Statistical Research communities and lately by computer science researchers. Some of the notable Anomaly Detection applications have been applied in detecting credit-card transaction frauds, malignant tumors, detecting a fault in manufactured

* **Corresponding author Diganta Sengupta:** Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India; Tel: 7044598807; E-mail: sg.diganta@ieee.org,msit.edu.in

Brojo Kishore Mishra (Ed.)

components like car engines, and for detecting unusual network traffic in a data center. Automated Irrigation Systems rely solely upon the data received from the Perceptron/Sensing Layer, which is made up of Physical Sensors and Actuators [2]. The data from sensing layer while being communicated to the middleware layer by the network layer is vulnerable. The sensing layer is vulnerable to node capturing, Sleep-Deprivation Attacks, False Data Injection Attack, and many more. The Network Layer is vulnerable to Phishing Attacks, Access attacks, DDoS/DoS attacks, Data Transmit attacks, Routing attacks. While the data is being recorded and sent for analysis, there remains a huge possibility of modification of data intentionally/unintentionally, especially in the case of False Data Injection Attacks. In addition, in some regions, manual damage to the agricultural fields by rival farmers is a possibility. Therefore, a pre-programmed routine Automated Irrigation System is not enough. In such cases, it is important to detect such outlier events and report the event to the System Administrator. The various layers discussed above are shown in Fig. (**1**). This chapter focuses on identifying such nonperforming or intruded nodes, and restricting transmission of altered data. The anomalies are detected based on historical data, and any deviation from the normal performance according to the learning curve of the algorithm using the historical data is cautioned. For analysis of the historical data, learning approach has been used which classifies the reliability in a binary class – True Node or False Node; True/False being the state of the node under attack or manipulation. Gaussian distribution function has been used as the learning algorithm. The results from our experiment are satisfactory and presented in the results section of this chapter.

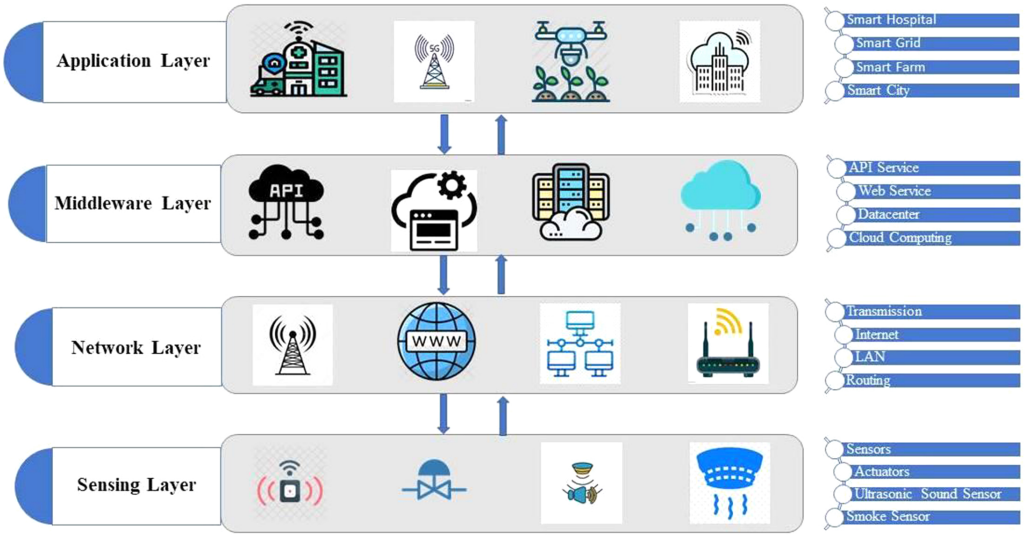

Fig. (1). Layer Structure of IoT.

Some crops are extremely sensitive to minor changes in the environment. In such cases [3], the availability of correct data about environmental conditions is very important for the IoT-based Smart Farming System. For example, if the transmission of soil moisture data is delayed or if False Data is sent to the system, then, it might lead to Over-Irrigation or under-Irrigation. Therefore, it is important for Decision Making Systems in Smart Farms to have an understanding of the data and then decide if the data received is reliable or not. To be applied to Smart Farming, Anomaly Detection Systems need to be lightweight because IoT systems are generally low-powered devices with limited memory and computational capabilities. In this paper, we propose a Lightweight Anomaly Detection Model which can detect and indicate anomalous conditions and give the System an indication of whether the data received by it, is reliable or not.

The rest of the paper is organized as follows. Section 2 presents the Literature Survey followed by the Proposed Methodology in Section 3. Section 4 presents the Results and Analysis of the model with the next section concluding with discussions.

LITERATURE SURVEY

Many academic and industrial researchers are working hard for implementing Secure and Reliable IoT Systems in Smart Farming and addressing open issues in the domain. Readers are directed to Table 2.1 of [4] for the taxonomy on methods for smart farming topics.

Strengthening the security of a system is essential to protect the data and prevent unauthorized access to resources. In this paper [5], a clustering-based algorithm, enhanced DBScan, was used to identify anomaly situations. It was concluded that enhanced DBScan had a higher recognition accuracy and as compared to other algorithms, had a special fast speed.

Agriculture has been one of the most important sectors in India's Economic Growth, yet this sector is lagging behind most of the world because of primitive agricultural methods. IoT and Automation Systems using Artificial Intelligence can handle the problem of crop monitoring and maintenance. An IoT device consisting of multiple sensors was designed [6]. This device ran multiple Artificially Intelligent workflows to ensure that water, fertilizers, and other important components reached the soil, as required.

Smart Farming environments implementing IoT Systems are extremely vulnerable to Cyber Attacks. Such attacks can be very harmful to the country's economy because such dynamic and distributed attacks can cause disruption in Food Supply Chain and limit the supply of raw materials to industries. Such cybersecurity

CHAPTER 5

Machine Learning based Smart Electricity Monitoring & Fault Detection for Smart City 4.0 Ecosystem

Subhash Mondal[1], Suharta Banerjee[1], Sugata Ghosh[1], Adrija Dasgupta[1] and **Diganta Sengupta[1,*]**

[1] *Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India*

Abstract: Growing electricity needs among the vast majority of the population seconded by a voluminous increase in electrical appliances have led to a huge surge in electric power demands. With thediminishing unit price of electric meters and increase of loading, it has been observed that a certain amount of electric meters generate faulty readings after exhaustive usage. This results in erroneous meter readings thereby affecting the billings. We propose a fault detecting learning algorithm that is trained by early meter readings and compares the actual meter reading (AMR) with the predicted meter reading (PMR). The decision matrix generates an alarm if |PMR-AMR|>T; where T equals the threshold limit. T itself is decided by the learning algorithm depending upon the meter variance. Moreover, our system also detects if there is any power theft as such an action would result in a sudden rise in AMR. The learning algorithm deploys six binary classifiers which reflect an accuracy of 98.24% for the detection module and an error rate of 1.26% for the prediction module.

Keywords: Smart Meters, Energy Prediction, Fault Detection, Machine Learning.

INTRODUCTION

Energy be it in any form is the driving force for the development and sustainability of large sectors of the world. The form of energy in which people's livelihood is largely dependent is electricity. It has become an indispensable resource in both commercial as well as residential sectors. Electricity can be transformed into other forms of energy and can be reserved, as well as can be re-transformed into its original form whenever needed [1]. This leads to the fact that it should be used judiciously in order to serve a large sector of people.

[*] **Corresponding author Diganta Sengupta:** Department of Computer Science and Engineering, Meghnad Saha Institute of Technology, Kolkata, India; Tel: 7044598807; E-mail: sg.diganta@ieee.org, msit.edu.in

Brojo Kishore Mishra (Ed.)

The growing population has led to a substantial rise in the demand for electricity. As a matter of fact, it has become necessary that we should use this resource wisely. Power consumption prediction can thus help to optimize usage of electricity [2]. Smart meters introduced for electricity management have the ability to increase the reliability and efficiency of power consumption [3, 4]. These meters contain a huge amount of data regarding electricity consumption [5] which allows us to invoke Machine Learning algorithms for future prediction. But these meters can also sometimes provide us with faulty data. Abnormal meter reading can be due to meter failure or a deliberate attempt to manipulating these smart meters. Fault Detection is an important aspect of identifying power theft and other electrical inefficiencies [6]. The researchers have been working from the time immemorial on developing an intelligent system capable of handling such issues.

With increasing technological advancement, to combat the real-world issues, we need more accurate prediction techniques and an enhanced detection mechanism. The introduction of Machine Learning helped researchers in addressing difficulties with the least human interference [7]. In this paper, we proposed an architecture that uses a learning algorithm to predict the meter readings. A fault detection mechanism is then trained and AMR is compared with PMR. The decision module of the architecture generates an alarm when $|PMR-AMR| > T$, where T is calculated using a gradient descent. The proposed model is capable of generating monthly statistical reports which help in better analysis of each smart meter.With an increase in electric consumption data, it has become possible to design an algorithm for prediction of consumption levels, and detect any anomaly. The machine learning algorithms serve the purpose as they are trained on historical data, and can classify the recent consumption pattern accordingly.

The paper is arranged as follows: related works are discussed in Section 2, our proposed architecture is explained in Section 3, followed by Experimental Results & Analysis in Section 4 and finally the paper concludes in Section 5.

RELATED WORKS

Over the years, many researchers have been addressing the issue of electricity prediction and fault detection. They have used various approaches in order to come up with an efficient solution. In this section we have discussed some of the previous techniques that were made to combat these problems.

In a study [8], Himeur *et al.* reviewed the current trends and also new perspectives of anomaly detection in energy consumption and presented various works that have taken place in this field. They also discussed various problems and challenges that are still needed to be addressed by the researchers. Krzysztof *et al.*

[9], discussed the need for monitoring the rapidly rising electricity prices using smart meters. This model predicts the electrical energy consumption for the next 24 hours using Support Vector Machines and Neural Networks.

Jiang *et al.* proposed a hybrid Machine Learning model consisting of both unsupervised clustering and supervised classification, which has been used [10] to detect the electrical energy usage of consumers based on data relayed by smart meters. Cost-sensitive XGBoost Algorithm has been used for imbalanced data [11] to build a multi-class fault prediction model using Machine Learning Techniques. In another study [12] an LSTM based neural network has been proposed for predicting the electrical energy consumption as well as for detecting anomalous data relayed by the smart grids. This approach focuses on the seasonal and monthly trends in energy consumption and decreases the forecasting error.

Wang *et al*. [13] focused on improved electricity monitoring performance. It is comprised of two modules namely energy consumption habit classifier (ECHC) and an appliance pattern matching classifier (APMC). The clustering algorithms, and other algorithms based on Gaussian Mixture (EM-GMM) can be used to develop energy prediction models as proposed in a study [14]. The uniqueness of our proposed architecture with the existing state of the art architectures is discussed in Table **1**.

Table 1. Other proposed state-of-the-art architectures and their differences with our architecture.

Description	Difference	Refs.
This paper proposes the utilization of Computational Intelligence techniques in various sources of Building Energy Management Systems (BEMS) data to get relevant information.	The model of Building Energy Management Systems (BEMS) operations has been created using modified KNN and fuzzy logic rule extraction techniques while we have used the Machine Learning techniques of Regression and Classification to build our model.	[18]
This study proposes an autoencoder-based ensemble model for smart buildings to detect anomalies in analyzing complex, high-dimensional and large-scale data	Autoencoders are used for inconsistent detection while we are proposing the use of six different classifiers to detect the anomalous readings.	[19]
Atechnique called Symbolic Aggregate approXimation (SAX) has been proposed by Capozzoli *et al.* for efficient electricity management in buildings based on time series data.	The paper uses the SAX process for detecting unexpected patterns in energy values which is different from our anomaly detection layer in which we used the classification algorithms for the process.	[20]

CHAPTER 6

Investigating the Effectiveness of Mobile Learning in Higher Education

V. Kalaiarasi[1,*], **D. Alamelu**[2] and **N. Venugopal**[3]

[1] *PSG College of Arts and Science, Coimbatore, Tamil Nadu, India*

[2] *KGiSL Institute of Technology, Coimbatore, Tamil Nadu, India*

[3] *Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India*

Abstract: Technology is a fundamental part of the teaching-learning process which has brought plenty of benefits during the pandemic situation. Covid 19 has adversely affected the education system throughout the world. Due to this abrupt change, mobile learning has occupied a dominant place in helping students to handle this unavoidable crisis. The studies focused on online learning, e-learning and also m-learning. However, this study focuses on the effectiveness of m-learning by understanding the satisfaction and intention of the students towards m-learning. Mixed research method approach is used in this study. Structural Equation Modelling (SEM) is used to validate the proposed model. The findings revealed that the majority of the students are satisfied with m-learning. However, there are also some students who express dissatisfaction with learning practical courses through mobile phones, citing insufficient support from their parents and the institution. Additionally, the satisfaction and intention of students play a crucial role in determining the effectiveness of mobile learning (m-learning). The study presents a comprehensive framework for understanding, explaining, and predicting the factors that influence the effectiveness of mobile learning (m-learning) among higher education students. The study also supports the practitioners and educators with useful guidelines for designing a successful m-learning system, particularly in higher education. This will also enable government to frame its the digitized policies appropriately.

Keywords: Higher education, M-Learning, Technology acceptance, Teaching-learning, Triangulation.

INTRODUCTION

As a response to the global Covid-19 pandemic, numerous precautionary measures have been implemented to address the situation. Foremost among these

* **Corresponding author V. Kalaiarasi**: PSG College of Arts and Science, Coimbatore, Tamil Nadu, India; E-mail: kalaikavin@gmail.com

measures is the widespread adoption of online education. However, this sudden and unplanned shift to complete dependence on online learning has raised several concerns, despite the fact that online education was not widely prevalent prior to the pandemic. The adoption of online teaching in education has led to a distinctive revolution from traditional face-to-face learning to online learning. Several online tools are developed for assisting and improving learning to feel like a real classroom. Still, there is a lack, particularly in developing countries like India, where there is a substantial technical constraint related to the suitability of devices and availability of reliable internet connection, which poses a serious challenge [1].

The effectiveness and satisfaction levels of online learning are currently under scrutiny. With various modes available for accessing academic materials in online education, such as desktop computers, laptops, and tablets, many students prefer smartphones due to their ease of use. Over time, students' learning approaches have evolved from traditional techniques to e-learning and now to m-learning, with mobile devices playing a significant role in this shift [2]. Previously, mobile learning was not believed and recognized by many of them, even though it is one of the models of learning in higher education. Developments in technology have helped students to enhance their understanding through m-learning. Mobile learning means, “Any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of the learning opportunities offered by mobile technologies” (AICTE). M-learning is the up-gradation of e-learning.

In many developed countries, m-learning has become the favorite model of learning, due to the growth and utility of the internet and the World Wide Web. Mobile technology touched almost every aspect of human endeavor, including education. Due to this scenario, the advanced blended learning method also emerged.

Mobile learning provides a variety of benefits for learners especially in the education field. The major advantages of m-learning are portability, social interactivity, context sensitivity and connectivity, however, failures also exist. A large number of people who start m-learning are very enthusiastic at an earlier stage, but later, they lose their interest and ultimately stop using m-learning. Information system research clearly shows that user satisfaction is one of the most important factors in assessing the success of system implementation [3]. In mobile learning setting, numerous elements are involved in user satisfaction. The factors affecting m-learning effectiveness are revealed by previous researchers using various descriptive or analytical studies. These factors can be grouped into three groups: technology acceptance, system success, and environmental factors.

Therefore, understanding of m-learning will be helpful to engage students and to improve their learning outcomes.

The main objective of this study is to examine the influence of learner satisfaction and learner intention towards the effectiveness of m-learning in higher education.

The research design comprises both qualitative and quantitative methods to investigate the effectiveness of m-learning.

In India, education has shifted towards a learner-centered teaching model, and as such, this study aims to highlight the key factors that influence the effectiveness of mobile learning (m-learning), particularly in the context of practical-based subjects. Furthermore, it helps to identify and understand the students' perceptions and intentions toward mobile learning in higher education. It assists educators to frame an appropriate course plan to improve learning effectiveness. Further, it will enable to overcome difficulties and reduces the risk of failure during execution.

MODEL CONSTRUCTION AND DEVELOPMENT OF HYPOTHESIS

Technology Acceptance and Learner Satisfaction

The variables within the Technology Acceptance Model (TAM) have been found to exert a significant influence on learner satisfaction [4, 5]. The presumption is that a large number of learners perceive usefulness, and ease of use for course-delivering media, such as course websites, and file-transmitting software. The more positively influencing attitudes toward m-learning are subsequently improving learning experiences, satisfaction, and increasing intention towards using m-learning [6]. The perceived usefulness and ease of use are critical factors influencing learners' perceived satisfaction [7]. TAM consists of two main components namely; perceived ease of use and perceived usefulness in which, an individual perceives using a particular system that would improve performance. Perceived usefulness has a positive influence on students' satisfaction and retention [8]. It has a positive moderate linear relationship towards perceived e-learner satisfaction [9]. This leads to the hypothesis:

H1. Technology acceptances have a positive influence on Learner's satisfaction

System Success and Learner satisfaction

According to DeLone & McLean [10], a system can be evaluated in terms of information quality, system quality, and service quality. These characteristics affect the subsequent system use or intention to use and user satisfaction.

CHAPTER 7

Socio-Economy of Coastal Fishing Community of Southern Coast of Odisha: A Case Study

T. Padmavati[1,*]

[1] *Department of Marine Sciences, Berhampur University, Odisha, India*

Abstract: Many of the country's most important essential zones and urban regions are found on the outskirts of the coastal zone. The activity or processes of socio-economic growth among coastal areas on the usage of coastal resources are essential for understanding the socio-economic state of a region and its long-term management. Sustainable development shows two great notions for the protection of natural resources: (i) environmental protection and (ii) economic development. Development activities are allowed as long as they do not compromise people's quality of life or the viability of the natural systems on which development is based, according to the above two notions. Advanced scientific and technological knowledge is now helping to improve and prevent the unsustainable exploitation of natural resources along the Indian coast. River floods, cyclones, depressions, and, most crucially, coastal erosions are common occurrences in the studied region. This effect can potentially change the socio-economic situations of coastal residents and society. As a result of this background, the author decided to quickly analyze socio-economic conditions and look at the current state and impacts on the study area's socio-economic conditions. Information on the above parameters was gathered to get insight into the research area's socio-economic profile. The current project involves conducting a socio-economic survey on Orissa's southern coast.

Keywords: Fishing Community, Socio-Economy, Southern Coast.

INTRODUCTION

India is a peninsular country in the south Asian subcontinent, with a 7,500 km long coastline. This shoreline comprises more or less fragile ecosystems such as sand dunes, beaches, wetlands, mangroves, and estuaries associated with tidal flats, backwaters, lagoons, and coral reefs. Thus, it is imperative to state that the Indian coastline plays a significant role in the national economy and has a strong imprint on the international economy for its inherent ocean-derived resources.

[*] **Corresponding author T. Padmavati:** Department of Marine Sciences, Berhampur University, Odisha, India; E-mail: padmasmart@gmail.com

Brojo Kishore Mishra (Ed.)

Being one of the developing countries of Asia, significant parts of the Indian coast are in a state of stress due to the increase of population density, rapid urbanization, and resultant developmental activities like the construction of ports, harbours, jetties, sea walls, and revetment. Many of the critical economic zones and major urban areas are located within the fringes of the coastal zone. As a result, economic and social processes inevitably characterize the coastal systems [1]. Limited coastal land resources and excessive land demand can result in human-earth conflicts. In the urbanization process, land-use changes lead to a series of environmental constraints, especially in developing countries during economic transition or recession [2]. The coastal environment is unable to maintain the flow of goods (*e.g.*, resources for exploitation) and services (*e.g.*, natural defense systems) indefinitely [3]. The coastal areas of India encompass nine states and four Union Territories (UTs). The socio-economic development activity or processes among these major coastal states and UTs are not per capita income, poverty, infrastructure, and socio-economic development. Thus, the study on the utilization of coastal resources is a prerequisite for understanding of the socio-economic status of a region and its sustainable management.

The accessible environmental resources along the fringes of the Indian coastline have been exploited in an unregulated and unsustainable manner with our arena of knowledge. Thus there is an increasing stress over the environmental components. In this context, it is understandable that human and associated environments are vulnerable to natural threats. Before a few decades, coastal environmentalists realized the vulnerability to potential natural hazards if natural resources are consumed uncontrolled. Since then, sustainable development has been challenged in India and worldwide. In the 1980s, the interdependence between economic development, the natural environment, and people was described with the issues on development that meet the national interest without compromising the ability of the future generations to meet their needs and aspirations [4]. In general, sustainable development unveils two admirable concepts for protecting natural resources (i) environmental protection and (ii) economic development. Based on the above two concepts, the development activities are permitted as long as they do not jeopardize people's quality of life or the viability of the natural systems on which development is based.

It has been widely documented that organic and chemical pollution simultaneously puts pressure on fragile coastal ecosystems and that natural resource deterioration is sometimes irreversible [5]. Coastal zones are more vulnerable in different parts of the world, but environmental pollution is higher in developing countries, particularly India. Sustainable development constitutes an integrated and interactive approach that allows understanding the complex relationship between society and nature, simultaneously respecting human rights

and assuming that the environment is a vital dimension of the future generation [5]. Moreover, the complex and conflicting interactions of social equity, human security, and environmental sustainability within the social processes of shaping and building development for present and future generations are important issues. Nowadays, advanced scientific and technological knowledge substantially contributes to improving and preventing the unsustainable use of natural resources along the Indian coast. There is a need for research on how societal driving forces (social and demographic, political and institutional, economic and commercial, cultural and technological) influence the nature and distribution of human activities on coastal zones of India as well as the impacts on coastal ecosystems associated with the current and potential alternative patterns of human activity [6].

The coastline of India is 7500 km long, with 3202 marine fishing villages,1332 landing centers, and an Exclusive Economic Zone of 2.02 million Sq. Km, which is set to grow. This is equivalent to 2/3rd of our land area, necessitating significant investment and long-term, balanced growth planning. The marine fisher population in India is about 5 million, of which 3.5 million are currently residents of coastal fishing villages. Most fishermen in coastal fishing settlements lack legal title deeds and adequate houses with basic facilities. Despite several developmental programs providing homes and title deeds, 40% still live in huts and Kutcha houses.

With a 480-kilometer coastline, Odisha has the country's most extensive mineral resource base in terms of iron ore, thermal coal, coking coal, and other essential minerals such as manganese copper and heavy minerals in coastal beach sands. Govt. of Odisha liberalized for the investors (public and private sectors) and provided adequate and additional benefits to Indian and foreign entrepreneurs to establish industries and ports as part of a significant drive for Indian economic growth. Among the five coastal districts of Odisha, Ganjam is one of the most developing districts regarding trading. The extensive stretch of beach bestowed with favorable ecosystem attracts many tourists of the country. The World's most significant nesting beach of olive Ridley sea turtles is adjacent to the Rushikulya Estuary, known as Rushikulya Rookery. The study region experiences multiple recurrent disasters such as river floods, cyclones (notably super cyclone of 1999;Phailin of October 2013;Hudhud of October 2014), depressions, and coastal erosion. On the other hand, Govt. of India has identified Gopalpur as one of the country’s Special Economic Zones (SEZs). Govt. of Odisha provided around 2980 acres of land near Chatrapur to Tata Steel Limited (TSL) for their industrial activities (Reference). Indian Rare Earth Ltd (IREL), a sister-wing of the Department of Atomic energy, Govt. of India, is adjacent to Gopalpur Port. The primary activity of this plant is to process the beach sand to acquire the heavy minerals for the multidimensional requirement. Jayasree Chemicals’ plant stands

CHAPTER 8

Filtering Techniques for Removing Noise From ECG Signals

K. Manimekalai[1,*] and **A. Kavitha**[2]

[1] *Department of Computer Applications, Sri GVG Visalakshi College for Women, Udumalpet, Tamil Nadu, India*

[2] *Department of Computer Science, Kongunadu Arts and Science College, Coimbatore, Tamil Nadu, India*

Abstract: Electrocardiogram (ECG) records cardiac electrical signal to check for various heart problems. However, it can be impaired by noise. Therefore, ECG signal denoising is a significant pre-processing step that reduces noise and emphasizes the characteristic waves in the ECG data. The frequency range of a simple ECG is usually between 0.5 Hz and 100 Hz. When processing the ECG signal, artifact elimination is the most important resource since artifacts in ECG signal impede the diagnosis of disorders. This work uses MATLAB to reduce noise by applying low pass, high pass, and derivative pass filters. On the PTB database, the performance of these approaches is compared using benchmark measures such as mean-square error (MSE) and signal-to-noise ratio (SNR) to compare various ECG denoising algorithms. The combination of low pass + high pass + derivative pass filters produces low mean-square error (MSE) and signal-to-noise ratio (SNR) values of 0.052 db and 1.185 db when compared to the raw signal.

Keywords: Derivative Pass, ECG Signal, High Pass, Low Pass, MSE, SNR.

INTRODUCTION

Pre-processing techniques assist in the preparation of data for analysis. Before starting primary processing, the data should be pre-processed to remove any detector effects. The most important feature for executing data processing is the pre-processing step.

When data is obtained as a consequence of an experiment, the following step is to model the data so that the required information may be extracted. Globally, the data production will be either too much or too little, or it will be fractured. The

* **Corresponding author K. Manimekalai:** Department of Computer Applications, Sri GVG Visalakshi College for Women, Udumalpet, Tamil Nadu, India; E-mail: gvgmanimekalai@gmail.com

Brojo Kishore Mishra (Ed.)

term "pre-processing" refers to the act of classifying data into three categories and processing it accordingly. As a result, technologies such as Data Filters, Data Editing, Data Ordering, and Noise Modelling play an important role in any data pre-processing. The Pre-processing of ECG Signals is depicted in Fig. (**1**).

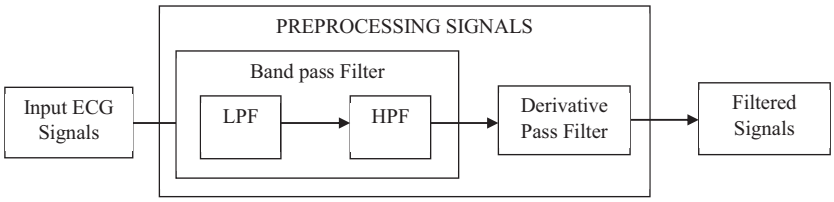

Fig. (1). Pre-processing of ECG Signals.

The signal processing method can be employed in a variety of different ECG analysis and interpretation systems. It is mostly utilised to get a few key characteristic parameters. Consequently, biomedical signal processing is applied to quantitative or primary analyses of physiological systems and is utilised to investigate the phenomena using signal analysis. The sector of biomedical signal analysis or processing methods has progressed to a higher-level characteristic for practical signal processing and pattern analysis approaches that have been proven to be effective. ECG signal processing is a broad topic that covers improving measurement accuracy and reproducibility. The focus of ECG analysis is on ECG interpretation, ambulatory monitoring (also known as Intensive Care Monitoring), and stress testing. Basically, these functions combine to form a basic set of algorithms that condition the data based on various sorts of noise and artifacts, resulting in basic ECG wave amplitude and duration measurements. It can also detect heartbeats and aggregate the information for storage or transmission. The basic ECG is made up of a frequency range ranging from 0.5Hz to 100Hz. As a result, artifact elimination is the most important resource in ECG signal processing. It is considered a critical task for the specialist to diagnose illnesses when artifacts are discovered in the ECG signal. The main contribution is to explore filters such as low pass, high pass, and derivative pass filters. When compared to the raw signal, the combination of Low Pass + High Pass + Derivative pass filters has the lowest MSE and SNR values (Fig. **2**).

ARTIFACTS

The electrocardiograph is a device that measures cardiac potentials on the body's surface, but it has nothing to do with the heart's electrical activity or function. It is critical to keep a focus on these types of artifacts, or else they will lead to unnecessary testing and therapeutic measures.

Monitoring electrodes on the skin surface are used to detect the electrical activity of the heart. It is denoted that the electrical signal is very tiny, often between 0.0001 and 0.003 volts. These signals are estimated in the 0.05 to 100 Hertz (Hz.) or cycles per second frequency range. Other artifact signals with comparable frequency and mostly bigger amplitude are inappropriately diffused across the skin surface and mixed with ECG signals.

An ECG artifact in electrocardiography is not related to the heart. Electrical interference from outside sources, electrical noise produced elsewhere in the body, or inadequate contact and machine malfunction are examples of these. Because artifacts are so ubiquitous, a thorough understanding of them is required to avoid cardiac rhythm misunderstanding.

Types of Artifact in ECG Signal

The goal of the ECG signal technique and signal processing system acquisition is to obtain a noise-free signal. Based on Rahul Kher's *et al.* [1] research, the key sources of noise are listed below.

The terminology and notations used in this study are listed in Table **1**.

Table 1. Terminologies and Notations.

Terminology	Description
ECG	Electrocardiogram
LPF	Low Pass Filter
HPF	High Pass Filter
MSE	Mean Square Error
SNR	Signal to Noise Ratio
DWT	Discrete Wavelet Transform
NI ELVIS	National Instruments Educational Laboratory Virtual Instrumentation Suite
MRA	Multi Resolution Analysis

CHAPTER 9

Deep Learning Techniques for Biomedical Research and Significant Gene Identification using Next Generation Sequencing (NGS) Data: - A Review

Debasish Swapnesh Kumar Nayak[1,*], Jayashankar Das[2] and **Tripti Swarnkar[3]**

[1] *Department of Computer Science & Engineering, Institute of Technical Education and Research, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India*

[2] *Avior Genomics, Mumbai, India*

[3] *Department of Computer Application, Institute of Technical Education and Research, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India*

Abstract: In the biomedical research areas of whole genome sequence (WGS) analysis, disease diagnosis, and medication discovery, Next Generation Sequencing (NGS) data are the most recent and popular trend. The use of NGS data has improved the analysis of infectious diseases, WGS, illness identification, and medication discovery. Although the amount of NGS data is massive, researchers have worked and are continuously working to improve its quality and precision. Modern computational techniques increase the biological value of NGS data processing, making it more accessible to biomedical researchers. Although the complexity of NGS and the required computational power to analyse the data pose a significant threat to researchers, the introduction of various branches of Artificial Intelligence (AI) such as Machine Learning (ML) and Deep Learning (DL) has given analysis, prediction, and diagnosis a new direction. Deep Learning's potential has been demonstrated in a variety of fields, including biomedical research, where it has outperformed traditional methods. The development of deep learning algorithms aids in the analysis of complicated datasets such as NGS by giving a variety of advanced computational methodologies. Different DL approaches are designed to manage enormous datasets and multiple jobs, and the genetic research business could be the next industry to benefit from DL. This paper discusses a variety of DL methods and tools for analysing NGS data in the fields of contagious diseases, WGS analysis, disease diagnosis, and drug design.

[*] **Corresponding author Debasish Swapnesh Kumar Nayak:** Department of Computer Science & Engineering, Institute of Technical Education and Research, Siksha 'O' Anusandhan Deemed to be University, Bhubaneswar, India; Tel: 8249718815; E-mail: swapnesh.nayak@gmail.com

Brojo Kishore Mishra (Ed.)

Keywords: Artificial Intelligence (AI), Deep Learning (DL), Infectious disease, Machine Learning (ML), Next Generation Sequencing (NGS), Whole Genome Sequencing (WGS).

INTRODUCTION

In the recent past, biomedical research has seen a heavy demand for high-throughput sequencing data. High-throughput sequencing in the current scenario is a primary approach for the research of genomics, molecular diagnosis, quantification and classification of metagenomics, genomic feature selection, prediction, and many more. A High-throughput sequence like Next Generation Sequencing (NGS) has a great impact on biomedical research due to its accuracy and time effectiveness [1]. Thousands of sequences can be obtained in very minimal time in the case of NGS. Most importantly, it comes with a very minimal cost compared to the other traditional sequencing techniques. The NGS in the computational view is rather a big omics data for analysis, classification, selection, and prediction; adding to this, the biological significance is a must for all the steps of analysis [2, 3].

The advent of NGS gives the ultimate flexibility to bring transformation in the research areas like bio-computational, biomedical, precision medicine, and biochemistry [4, 5]. It is also found that the huge amount of data in NGS helps to bring changes to the research area of proteomics, epigenetics, transcriptomics, and genomics [6]. In the areas of genomic research, NGS brings revolution by providing a huge amount of data with very less time. The NGS data can be obtained for nucleotides and gene expression, but the major limitation is the accuracy of the huge amount of data generated in a short period. However, it has been found that if the NGS data is processed extremely well before becoming input to any of the analytical phases, it might increase the potential and accuracy of the results [7 - 9].

The gene expression NGS data contains crucial information about different levels of differential activation of the genes which are involved in the evolution and development of any disease. The researchers and physicians have to apply an appropriate tool or technique which can efficiently identify the differentially expressed and significant genes. RNA-Seq is a high-throughput sequencing technique that contains the gene expression data and helps the researcher and physicians to go deeper into the diagnosis and prediction of the disease state of any specific disease especially cancer [10, 11]. The NGS data greatly influenced the research areas of biomedical and bio-computational science from 2010 to 2021. In a survey conducted in 2018, it is said that the global NGS market will grow more than 400 times by 2026 as shown in Fig. (**1**).

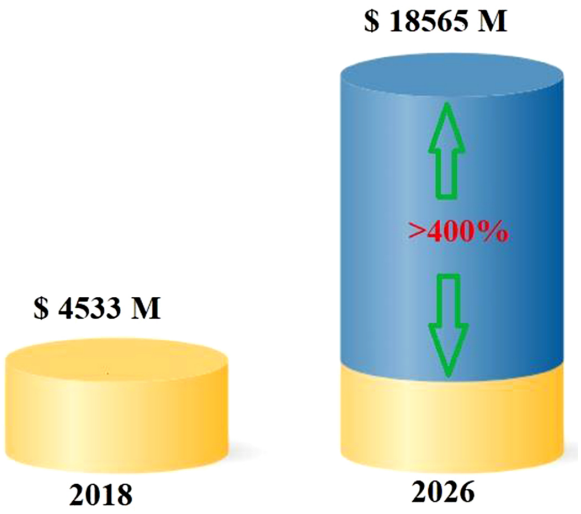

Fig. (1). Global next generation sequencing-market.

Researchers in biomedical, computational biology, and life science fields are nowadays in the search of rapid and accurate computational techniques which can be developed with a very minimal cost and produce more significant results. The advancement of computer science, particularly the newly created computational methods in Artificial Intelligence (AI), makes it possible to handle big data.

At the moment, the trend in infectious disease research is promising. Recently, the study has focused extensively on the infectious disease brought on by the transmission of an infectious agent from an animal or a human to a human. The infectious disease is categorized into four different parts based on the infection agent, Fungi, Viruses, Bacteria, and Parasites. Globally and especially in the Asian continent, infectious diseases like malaria, dengue, typhoid, Tuberculosis (TB), and Hepatitis are major threats. It is also found that there is a rapid growth in infectious diseases in India, and this is due to three major factors *i.e* huge population, lack of medical facilities, and most importantly lack of awareness among the people. The majority of deaths in India are caused by a variety of infectious diseases as shown in Fig. (**2**).

CHAPTER 10

Breast Cancer Detection Using Machine Learning Concepts

Fahmina Taranum[1,*] and **K. Sridevi**[1]

[1]*Computer Science and Engineering Department, Muffakham Jah College of Engineering and Technology, Hyderabad, India*

Abstract: Machine learning is applied in medical diagnosis to do early prediction of diseases, for increasing the possibility of recoverability around the globe. Cancer is a disease, which spreads quickly and would be difficult to control in advanced stages. The idea is to diagnose the disease at an early stage, so as to increase the chances of fast recovery. Breast cancer is common in women, and is a disease that causes the death of women in the age of fifty years or older. The purpose is to apply machine learning concepts to do early detection of disease. The system is fed with the images of all stages of cancer patients and the classification tools are used to train the system with the cases. This helps to predict the stage of cancer. After the prediction of the stage, the patient is prescribed with the medication or other appropriate treatment processes by the doctor. The right time diagnoses help to improve the prognosis and increase the chances of survival. The type of the tumour, size and its re-occurring nature need to be monitored from time to time to check it in control. The Data Mining algorithm in collaboration with Deep learning or Machine learning concepts can be used to design a system for early predictions. The proposal is to use the machine learning concepts to do performance comparison using different classifiers, such as Support Vector Machine (SVM), Decision Tree and K-Nearest Neighbour (KNN) on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [1]. The main aim of cancer detection is to classify tumours into malignant or benign, thus we use machine learning techniques to improve the accuracy of diagnosis.

The main objective is to assess the efficiency, effectiveness and correctness of the algorithm using performance metrics like Accuracy, Precision, F1 score and Recall Experimentation is done using Jupyter Notebook.

Keywords: Cancer, Classifiers, Machine Learning, WDBC.

[*] **Corresponding author Fahmina Taranum:** Computer Science and Engineering Department, Muffakham Jah College of Engineering and Technology, Hyderabad, India; E-mail: ftaranum@mjcollege.ac.in

INTRODUCTION

Background

Breast Cancer is a disease responsible for majority of women's deaths 50 years or older. It is the second most dangerous disease after lung cancer. The statistics provided by the World Cancer Research Fund (WCRF) in 2018, stated that a record of 626,679 deaths were reported of two million cancer cases. Breast cancer is 24.2% of all the other types of cancers found in women.

The prediction helps in recovering the disease more easily; hence either deep or machine learning approaches can be applied. The process of patient's consultation with the doctor can be arranged in an online mode for quick treatment. The recent technology supports online consultation with the doctors for senior or weak patients, who have difficulty reaching the hospital. The method for consultation fee payment is also made easier with services like PAYTM, GPAY, PhonePe and BHIM.

Undertaking Thorough Medical History

The admission process proceeds with the discussion of the patient's medical history collection process; the medical supervisor creates a record containing the medical history of the patient. The Physical examination along with the prescribed test would be performed to rule out the doubts and get a clear picture of the patient's health after diagnosis. The screening of the area is done using Magnetic resonance imaging (MRI), Ultrasound, Tissue biopsy, X-ray or manual examination by an expert. On confirming about the disease based on removal of the tissue from the infected area showing the results as cancerous, or by Sentinel node biopsy test, the patient may be forwarded through multiple examinations, if required. The hospitals must be well equipped with the facilities and machineries to conduct the diagnosis as the delay may spread the disease and deteriorate patients' condition. Avoiding all these processes by using machine learning algorithms for detecting the state as Benign or Malignant makes the job easier and would definitely increase the survival rate to some extent as the pre diagnosis has helped to control the disease on time. The screening or testing used is listed in this subsection.

Imaging Tests

Mammogram is an examination of the breast, which is a special medical testing of images extracted using a low-dose X-ray system to check the inside tissues and

nerves of the breast and note any abnormal changes. Screening can be done based on symptoms like a lump in the breast, irritation in the breast epidermis, thickened parts of the breast, redness or flaky skin. The type of screening based on symptoms is named Diagnostic mammogram, otherwise screening mammogram.

Advanced Test

The use of computers to store images is a part of advanced technology named Digital mammography. These records are portable and sharable. The quality of the image is good with respect to sharpness and contrast. The 3-dimensional (3-D) breast imaging or breast to tomosynthesis is another type of advanced technology. These are known as efficient techniques as they capture images from different angles in less time. The collected information is filtered by a computer to create manifold thin section of the breast, which helps the radiologist to do accurate analysis.

Classification Using the Techniques

The machine learning tools help to declare the severalty of the case in advance. The input with multiple cases trains the system perfectly to detect any abnormalities in the initial stage itself. The prognosis and prediction of the disease are done based on ML approaches for modelling the succession of melanoma, which recognize the attributes that help guide towards accurate results. The research helps to design an efficient and fast-tracking system to categorize the type of infection into either malignant and benign class.

Dataset

Further accurate classification of benign tumours can prevent patients from undergoing additional treatments. Thus, the proper diagnosis of BC and classification of patients into cancerous or non-cancerous tumours is the major aim of research. Because of its unique advantages in critical features and analysis of complex BC datasets, machine learning (ML) is majorly used as the methodology of choice in BC pattern classification and forecast modelling. Classification and data mining algorithms are effectively used for classifying data. Especially in the medical field, where these algorithms are widely used in prediction and prognosis to make decisions. Machine learning classifiers use the patterns in the images from the huge dataset to perform testing, training and validations for producing accurate results. These complex working principles and quick execution of the testing results are benefitted over a minimized cost of

SUBJECT INDEX

A

B

C

D

E

F

G

H

I

L

M

N

W

Brojo Kishore Mishra

Dr. Brojo Kishore Mishra was awarded Ph.D. in Computer Science from Berhampur University in 2012 for his excellent work in Web Mining. He has worked in several reputed private Engineering Colleges and State Universities at different levels for the last 17 years. Presently, Dr. Mishra is a Professor in the Department of Computer Science & Engineering at GIET University, Gunupur, India and he is also working as a Joint Secretary of IEEE Bhubaneswar Sub-Section. He has published more than 43 publications in reputed international conferences, 48 peer-reviewed journals and 49 book chapters, 15 edited books, 02 authored books, 02 patents, 02 copyrights, 01 trademark (applied), and 04 book series contributions. He has successfully guided 01 Ph.D. research scholar and currently 06 research scholars are continuing. He has received more than 07 national awards so far. He is also a life member of ISTE, CSI, and a senior member of IEEE.