PREDICTION OF TYPHOID USING MACHINE LEARNING AND ANN PRIOR TO CLINICAL TEST

BY

Md. Atik Bhuiyan
ID: 191-15-2642
AND

Sharaf Shahariare Rad
ID: 191-15-2372

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Fatema Tuj Johora
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Md. Sabab Zulfiker
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

DAFFOFIL INTERNATIONAL UNIVERSITY
DHAKA, BANGLADESH
JANUARY 2023

APPROVAL

This Project titled **"Prediction of Typhoid Using Machine Learning and ANN Prior To Clinical Test"**, submitted by **Md. Atik Bhuiyan** and **Sharaf Shahariare Rad** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 23.01.2023

BOARD OF EXAMINERS

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Tarek Habib — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Taposhy Rabeya — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

23-01-23

Dr. Dewan Md Farid — **External Examiner**
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Fatema Tuj Johora, Lecturer (Senior Scale), Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Fatema 22.01.23

Fatema Tuj Johora
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

25.01.23

Md. Sabab Zulfiker
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Atik

Md. Atik Bhuiyan
ID: -191-15-2642
Department of CSE
Daffodil International University

Rad

Sharaf Shahariare Rad
ID: -191-15-2372
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes it possible for us to complete the final year project/internship successfully.

We are really grateful and wish our profound indebtedness to our supervisor **Fatema Tuj Johora, Lecturer (Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge and keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Mr. Md. Sabab Zulfiker,** Lecturer (Senior Scale)**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Typhoid fever is a serious and potentially fatal illness that causes a large number of deaths each year, particularly in developing countries. Early and accurate diagnosis is essential for effective treatment and to prevent the spread of the disease. In this study, we used machine learning techniques to develop a predictive model for typhoid fever. We used a dataset of 1746 entries and 29 attributes, and applied ten different algorithms to the data. Our results showed that machine learning can be effective tools for the prediction of typhoid fever, with the XGBoost classifier performing particularly well, achieving an accuracy rate of 97.87%. In addition to the XGBoost classifier, we also evaluated the performance of several other algorithms, including the Random Forest classifier, Extra Trees classifier, and Artificial Neural Network. While each of these classifiers performed well, the XGBoost classifier was found to be the most effective, with accuracy rates of 97.78% and 97.42% for the Random Forest and Extra Trees classifiers, respectively. Overall, our results demonstrate the potential of machine learning and ANNs for the prediction of typhoid fever, and suggest that these technologies could be useful tools for improving the diagnosis and treatment of this disease. Further research will be needed to explore the potential of these technologies in more detail, and to identify the most effective approaches for their implementation and deployment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Typhoid fever is a serious and potentially life-threatening illness caused by the bacterium Salmonella Typhi. The bacteria Salmonella serotype Typhi causes the illness typhoid, sometimes known as typhoid fever. Typhoid can also spread from person to person. Bangladesh is an emerging country, increasing rate of consuming food or beverages that have been cleaned in tainted water puts you at risk for contracting typhoid fever [1]. Underdeveloped countries are particularly affected by it because ignorance is one of its primary causes. There were approximately 149,000 fatalities worldwide in 2015. Typhoid affects roughly 6,000 people in the United States, and 400 new cases are reported every year [2]. Significant consequences have also been felt in Bangladesh as a result. Among many other countries, Bangladesh is severely impacted, with 252 cases per 100,000 affected people each year. Children under the age of five are substantially more defenseless than adults [3], especially those with compromised immune systems]. We evaluated a large number of publications for their potential to predict typhoid, but nobody has ever done it just based on symptoms prior to a clinical test.

The ability to predict disease outbreaks can assist build a foundation for early warning systems and advance planning. Some prediction models have been trained to forecast typhoid fever in order to reach these objectives. A dataset of 1746 entries and 29 attributes is presented in this research. Ten algorithms were utilized in this prediction. Typhoid is predicted using XGB Classifier, Random Forest Classifier, Extra Trees Classifier, Artificial Neural Network, etc. Although each classifier performs well. With 97.87% accuracy, XGB Classifier outperformed the competition. With accuracy rates of 97.78% and 97.42%, Random Forest and Extra Trees Classifier likewise performed admirably in this context. We may say that XGB Classifier predicts outcomes much faster for that reason. The most effective classifier was XGBprognosis, risk prediction, and personalized medicine. In the context of typhoid, machine learning and ANNs have the potential to facilitate rapid and accurate diagnosis, even in resource-limited settings, and to improve patient management and outcomes.

By reading the majority of the paper, it is obvious that either machine learning or deep learning was used, however in this case, both techniques were used to determine which may provide us with greater accuracy. Here, machine learning worked well. Here, the uniqueness lies in being able to identify typhoid before testing it in a clinic. Our novelty is given below:

- The study's goal is to foretell typhoid fever before a clinical trial.
- The outcome ought to be helpful in developing appropriate preventative strategies and policies for controlling this disease that is a public health concern in our nation.

1.2 Motivation

Typhoid is now spreading at an ever-increasing rate. The motivation for using machine learning and ANNs for typhoid diagnosis is clear. These methods have the potential to overcome the limitations of traditional diagnostic approaches and to improve the accuracy and efficiency of typhoid diagnosis. They can help healthcare providers make more informed treatment decisions and reduce the risk of complications and death from typhoid fever. In addition, machine learning and ANNs may have potential applications in other areas of typhoid management, such as identifying risk factors, predicting outbreaks, and developing targeted prevention and control strategies. Every year a lot of people die because of it. Most of the people who die from typhoid don't know they have malaria because it hasn't been detected in them. We want to make it easier for them to tell if they have typhoid or not. Because it can have devastating repercussions if it is not recognized and treated at an early stage.

1.3 Rational of the Study

The rationale for studying the use of machine learning and artificial neural networks (ANNs) for typhoid diagnosis is clear. Typhoid fever is a serious and potentially life-threatening illness caused by the bacterium Salmonella Typhi. It causes a lot of deaths each year. Early diagnosis and treatment of typhoid are crucial to prevent complications and reduce the risk of death.

Traditionally, typhoid diagnosis has relied on laboratory tests such as blood culture or stool analysis. These methods can be time-consuming and may not always provide accurate results, particularly in resource-limited settings where access to laboratory facilities and trained personnel may be limited. In addition, these tests may not be sensitive enough to detect subclinical infections, which can be a significant source of transmission.

The lack of reliable and efficient diagnostic methods for typhoid is a major barrier to effective typhoid management and control. It can lead to delays in diagnosis and treatment, which can increase the risk of complications and death. It can also contribute to the spread of typhoid, as infected individuals may not be properly isolated and treated.

The rationale for this study is to contribute to the growing body of knowledge on the use of machine learning and ANNs for typhoid diagnosis. By conducting a larger and more comprehensive study, we hope to validate and expand upon the findings of previous studies and to provide more robust evidence on the effectiveness of these methods for typhoid diagnosis. In addition, this study will help to identify any potential challenges or limitations of using machine learning and ANNs for typhoid diagnosis, and to suggest ways to overcome these challenges.

Overall, the rationale for studying the use of machine learning and ANNs for typhoid diagnosis is strong, given the potential benefits for patient care and public health. If proven effective, these methods could be valuable tools for reducing the burden of typhoid fever and improving public health outcomes.

1.4 Research Questions

The study itself could be the subject of a wide range of queries. An assortment of questions was collected from various people in order to condense this study. The research question for this study could be:

"What is the accuracy and efficiency of using machine learning and artificial neural networks (ANNs) for typhoid diagnosis?"

This research question aims to explore the effectiveness of using machine learning and ANNs for typhoid diagnosis in a specific population and setting. It considers both

the accuracy of these methods, which refers to the ability to correctly identify cases of typhoid fever, and the efficiency of these methods, which refers to the speed and ease of use. By answering this research question, the study aims to provide evidence on the potential usefulness of machine learning and ANNs for typhoid diagnosis in the specific population and setting being studied.

There can be also some question like:

1. What is the difference between our work vs others?

2. Machine learning or ANN which performs the best according to dataset?

3. What is the novelty of ours?

1.5 Research Objectives

Prior to a clinical study, our fundamental objective is to create a comprehensive and practical approach for forecasting typhoid fever. Many people have this condition, but due to a lack of resources and knowledge, they cannot be properly diagnosed. By foreseeing typhoid fever before a clinical study, we hope to assist them. our main objective is to build a model which can predict whether a person has typhoid or not before clinical test.

1.6 Expected Outcome

The expected outcome of this study is to provide evidence on the accuracy and efficiency of using machine learning and artificial neural networks (ANNs) for typhoid diagnosis in a specific population and setting.

- Based on previous research, it is expected that machine learning and ANNs will be able to accurately predict typhoid fever with high sensitivity and specificity. This means that the methods should be able to correctly identify cases of typhoid fever with a high degree of accuracy, and should have a low rate of false positive or false negative results.
- In addition to accuracy, it is expected that machine learning and ANNs will be efficient methods for typhoid diagnosis. This means that they should be able to process and analyze data quickly and with minimal effort, and should be easy to use by healthcare providers.
- Overall, the expected outcome of this study is to provide robust evidence on the usefulness of machine learning and ANNs for typhoid diagnosis in the specific population and setting being studied. If the results of the study support the effectiveness of these methods, it is expected that they could be considered as a viable option for improving the accuracy and efficiency of typhoid diagnosis, particularly in resource-limited settings where traditional diagnostic methods may be limited.

- It is important to note that the expected outcome of this study is based on the assumptions and limitations of the study design and methodology. The actual outcome may differ from the expected outcome depending on the specific results of the study and the interpretation of these results.

1.7 Report Layout

The layout of a research report typically includes several key sections that are designed to provide a clear and structured overview of the research project. The paper is organized as follows:

i. Background
ii. Research Methodology
iii. Experimental Results and Discussion
iv. Summary, conclusion, Recommendation and implication for future Research
v. Reference.

CHAPTER 2

Background Study

2.1 Preliminaries

Preliminaries refer to the initial steps and preparations that are necessary before starting a research study. In the context of a study on the use of machine learning and artificial neural networks (ANNs) for typhoid diagnosis, some examples of preliminaries might include:

Defining the research question: This is a crucial step in any research study, as it helps to guide the focus and direction of the study. In the context of this study, the research question might be something like: "What is the accuracy and efficiency of using machine learning and ANNs for typhoid diagnosis in a specific population and setting?" It is important to clearly state the research question and to ensure that it is specific, measurable, attainable, relevant, and time-bound (SMART).

Conducting a literature review: A literature review is an important part of any research study, as it helps to identify the existing knowledge on a particular topic and to identify gaps in the literature. In the context of this study, the literature review might involve reviewing previous research on the use of machine learning and ANNs for typhoid diagnosis, including the findings and limitations of these studies. This can help to inform the design and methods of the current study and to identify areas where further research is needed.

Developing a study design: The study design refers to the specific methods and procedures that will be used to collect and analyze data in the study. In the context of this study, the study design might include details on the study population (e.g., age, sex, location), the data sources (e.g., medical records, laboratory tests), and the statistical analysis techniques (e.g., regression analysis, classification algorithms). It is important to carefully plan the study design to ensure that the data collected are relevant and reliable and that the statistical analyses are appropriate and robust.

Ethical considerations: Ethical considerations are an important aspect of any research study, and it is essential to ensure that the study is conducted in an ethical and responsible manner. In the context of this study, ethical considerations might include obtaining necessary ethical approvals, ensuring that informed consent is obtained from study participants, and protecting the confidentiality and privacy of personal data. It is also important to consider any potential risks or harms to study participants and to have appropriate measures in place to minimize these risks.

Data collection: Data collection is a crucial step in any research study, and it is important to ensure that the data collected are relevant, reliable, and of high quality. In the context of this study, data collection might involve gathering data on clinical and laboratory features of typhoid fever, such as fever duration, white blood cell count, and stool culture results. This data might be collected through primary data collection methods (e.g., surveys, interviews) or through secondary data analysis (e.g., using existing datasets). It is important to carefully plan and execute the data collection process to ensure that the data collected are accurate and complete.

Data preparation: Data preparation refers to the process of cleaning and organizing the data to ensure that it is ready for analysis. In the context of this study, data preparation might involve checking for errors and missing values, transforming the data as needed for analysis (e.g., standardizing variables), and creating appropriate variables for analysis (e.g., binary variables for classification). Data preparation is an important step in any research study, as it helps to ensure that the data are accurate and consistent and that the statistical analyses are meaningful and reliable.

Overall, the preliminaries of a research study are important for ensuring that the study is well-planned, well-executed, and of high quality. They can help to ensure that the study is relevant, rigorous, and scientifically sound and that the results are reliable and can be generalizable to other populations and settings. By carefully planning and executing the preliminaries of a study, researchers can increase the chances of success and contribute to the advancement of knowledge in their field.

2.2 Related Works

Douglas Ibrahim et al [4] worked on data mining technique to predict multiclass symptomatic malaria infection. Here, they trained 70% data and tested 30% data. And they got 65.22% accuracy, 57.89% specificity and 100% sensitivity for 200 patient and the dataset was collected from general hospital mubi and hong.

Oguntimilehin A. et al [5] proposed machine learning approach to predict typhoid fever. In this paper the dataset was included 150 people. Out of 150 they used 100 data for training and 50 data for testing. For training data, they have got 95% accuracy and for testing data they have got 96% accuracy here.

Alile Solomon Osarumwense and Bello Moses Eromosele [6] approached machine learning technique to predict Dengue Haemorrhagic Fever. Here they have used Bayesian Belief Network to get better accuracy. They have got 99.84% accuracy here.

Oguntimilehin A et al [7] worked on machine learning approach for diagnosis and treatment of typhoid fever. They implemented this by using Visual Basic as front end and MySql as backend. In this paper they got 100% accuracy for training set and testing set gave them 95% accuracy.

Md. Sanzidul Islam et al [8] proposed machine learning approach to predict the probability of dengue infection with external behavior. The dataset included 400 cases and 70% of them consist in training and the rest which is 30% consist in testing. They have used various classifiers here such as KNN, SVM, Naïve Bayes, Decision tree etc. Among of them decision tree gave the highest accuracy which is 100%.

Boby Andrianto et al [9] used classification techniques for typhoid fever disease. Here they have used three methods which are Naïve Bayes, KNN and SVM. Among of them KNN provided the highest accuracy which is 88.7%. The dataset consists of 1532 patient with 701 patient positive typhoid fever and 831 negative typhoid fever.

R. Sanjudevi, D. Savitha [10] proposed classification techniques to predict dengue fever. They used Decision tree and SVM algorithms here. Support Vector Machine gave better accuracy which is 99%. They implemented it by using WEKA.

N.Rajathi et al [11] worked on machine learning algorithms to predict dengue in early stage. The dataset included 100 patients and from this 70% of data used in training purpose and 30% of data used for testing purpose. Various kinds of algorithms used here. Naïve Bayes, J48, Random Forest, SMO etc. Here random forest gave a better accuracy which is 83.3% and other algorithms performed quite well.

Jason R. Andrews et al [12] worked on typhoid fever by using machine learning approach. Here they identified a parsimonious serology signature to differentiate acute typhoid case from controls. In this paper, they have got the 90% sensitivity and 92% specificity.

Justin Im et al [13] proposed machine learning approach to find out which things can reduce the risk of typhoid. Here they evaluated cox regression model whether residents of "good" households can lower the risk of typhoid.

Yanjia Cao et al [14] worked on typhoid fever and surveillance data in India. Here they have done it with 586 cases per 100000 person-years. They accomplished spatial data processing and inter polation to detect data on typhoid incidence from SEFI study sites. In that paper, they used linear regression model.

Octave Iradukunda et al [15] proposed machine learning techniques to predict malaria diseases. Here they used different kinds of algorithms to see which one performs better. SVM, KNN, RF, CART, DENSENET, VGG16 used here and DENSENET performed better with 99% accuracy.

2.3 Comparative analysis

A comparative analysis is a type of analysis that involves comparing and contrasting two or more items or concepts. In the context of a study on the use of machine learning and artificial neural networks (ANNs) for typhoid diagnosis, a comparative analysis could involve comparing the performance of different machine learning or ANN algorithms, or comparing the accuracy and efficiency of these methods to traditional diagnostic approaches.

2.3.1 Comparative Table

TABLE 2.1: COMPARATIVE TABLE

Ref. No	Author's Name	Best Model	Limitations
1	Douglas Ibrahim	SVM	Short data usage and an inappropriate model.
2	Oguntimilehin A.	SVM=96%	The data was short.
3	Alile Solomon Osarumwense	Bayesian Belief Network	-
4	Oguntimilehin A	SVM	Lacking of big data usage

5	Md. Sanzidul Islam	Decision Tree	-
6	Boby Andrianto	KNN= 88.7%.	-
7	R. Sanjudevi	SVM	-
8	N.Rajathi	Random Forest= 83.3%	Dataset is small, result did not explain properly
9	Jason R. Andrews	SVM	-
10	Justin Im	SVM	The model is not properly suitable
11	Yanjia Cao	Linear Regression Model	-
12	Octave Iradukunda	DENSENET	-
13	Kumar Shashvat	SVM	-
14	You Won Lee	Random Forest	-
15	Odu Nkiruka	XGBoost	Lack of size in dataset.

2.4 Scope of the problem

The scope of a problem refers to the extent or range of issues or challenges that the problem encompasses. In the context of a study on the use of machine learning and artificial neural networks (ANNs) for typhoid diagnosis, the scope of the problem might include:

The prevalence and impact of typhoid fever: Typhoid fever is a global health problem, with an estimated 21.5 million cases and 216,500 deaths occurring annually, particularly in low- and middle-income countries. It is transmitted through contaminated food and water, and can lead to severe symptoms such as fever, abdominal pain, and diarrhea. Early diagnosis and treatment of typhoid are crucial to prevent complications and reduce the risk of death.

Limitations of traditional diagnostic methods: Traditional diagnostic methods for typhoid, such as blood culture or stool analysis, can be time-consuming and may not always provide accurate results. They may also not be sensitive enough to detect subclinical infections, which can be a significant source of transmission.

Potential of machine learning and ANNs for typhoid diagnosis: Machine learning and ANNs are emerging technologies with the potential to improve the accuracy and efficiency of typhoid diagnosis. They can facilitate rapid and accurate diagnosis, even in resource-limited settings, and may have potential applications in other areas of typhoid management, such as identifying risk factors, predicting outbreaks, and developing targeted prevention and control strategies.

Overall, the scope of the problem of typhoid diagnosis is broad, encompassing both the global impact of the disease and the limitations of traditional diagnostic methods. By using machine learning and ANNs, it may be possible to overcome these limitations and to improve the accuracy and efficiency of typhoid diagnosis, with potential benefits for patient care and public health.

2.5 Challenges

There are several challenges that may be encountered when using machine learning and artificial neural networks (ANNs) for typhoid diagnosis. Some of these challenges include:

Data availability and quality: Machine learning and ANNs rely on large amounts of data to learn and make predictions. In the context of typhoid diagnosis, it may be challenging to collect sufficient data on typhoid cases and controls, particularly in resource-limited settings where data collection infrastructure and capacity may be limited. In addition, the quality of the data may be a concern, as errors or biases in the data can affect the accuracy of the machine learning or ANN models.

Algorithmic complexity: Machine learning and ANN algorithms can be complex and require specialized training and expertise to implement and interpret. This may be a challenge for healthcare providers who are not familiar with these methods, and may require additional training and support.

Ethical, legal, and social implications: The use of machine learning and ANNs in healthcare raises a number of ethical, legal, and social issues, such as data privacy, fairness, and accountability. It is important to carefully consider these issues and to ensure that the use of these methods is ethical and transparent.

Validation and generalizability: Machine learning and ANN models need to be validated and tested to ensure that they are accurate and reliable. This may be a challenge in the context of typhoid diagnosis, as it may be difficult to obtain a sufficient number of cases and controls to adequately test the models. In addition, the generalizability of the models may be a concern, as they may not perform as well in different populations or settings.

Overall, these challenges highlight the need for careful planning and consideration when using machine learning and ANNs for typhoid diagnosis. It is important to address these challenges in order to ensure that these methods are effective and reliable, and to maximize their potential benefits for patient care and public health.

CHAPTER 3

Research Methodology

3.1 Introduction

In theory, one way to describe how a researcher goes about conducting their research is through their research technique. It is a logical, systematic strategy to choose research. It is a method for getting trustworthy, accurate outcomes. In short, it refers to the kind of data a researcher will gather, where they will come from, how they will gather it, and how they will evaluate it.

3.2 Research subject

The person who takes part in the research is referred to as a research subject. Whether through a machine or a mental experiment, the person will assist in resolving the question under investigation. Humans, participants, or volunteers can occasionally serve as research subjects. My subjects in this instance were patients with tonsil carcinoma. The topic was then honed into data that computers can understand. I trained the data in a model and utilized it as a predictor of the outcome.

3.3 Data Collection Procedure

In order to get the information, we required, we used Google Forms. We used Daffodil University students, social media, and hospitals to help administer our survey. Those who have previously had fever have shared their experiences. To distinguish typhoid from the other fevers, we collect information on a variety of them. Additionally, they have been classified in the dataset as "other fever." 1,746 entries with 29 characteristics were collected. A yes or no response is given for each feature class.

Research methodology is a systematic approach that researchers use to plan and conduct their research. It is a way to ensure that research is conducted in a reliable and valid manner, producing results that can be trusted. Essentially, research methodology explains how a researcher goes about collecting and analyzing data in order to answer research questions. It includes decisions about what data will be collected, where it will come from, how it will be collected, and how it will be analyzed. In summary, research methodology is a critical component of research as it helps to ensure that the research process is well-planned and that the results are reliable and trustworthy.

Training dataset:

	age	sex	headache	RETRO-OCULAR PAIN	muscle or muscle joint pain	NAUSEA	Rash	fast heart rate	bloody cough	less urination	...
count	1504.000000	1504.000000	1504.000000	1504.000000	1504.000000	1504.000000	1504.000000	1504.000000	1504.000000	1504.000000	...
mean	34.011636	0.504654	0.787234	0.255984	0.662899	0.499335	0.173537	0.175532	0.032580	0.151596	...
std	20.223930	0.500145	0.409400	0.436558	0.472877	0.500166	0.378837	0.380548	0.177593	0.358748	...
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
25%	18.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
50%	31.000000	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
75%	50.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	...
max	86.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...

8 rows × 29 columns

Figure 3.1: Training Dataset

3.4 Statistical Analysis

Statistical analysis is the process of collecting, organizing, and analyzing data in order to draw conclusions or make predictions. In the context of a study on the use of machine learning and artificial neural networks (ANNs) for typhoid diagnosis, statistical analysis could be used to evaluate the accuracy and efficiency of these methods, and to compare their performance to traditional diagnostic approaches. There are many different statistical techniques that can be used for this purpose, including descriptive statistics, inferential statistics, and advanced techniques such as regression analysis or machine learning algorithms. The specific techniques used will depend on the research question, the type of data being analyzed, and the statistical software or tools being used.

3.4.1 Flow Model

We initially gathered information on patients with typhoid fever. The data then underwent a variety of pre-processing steps. The model is then trained using the pre-processed data. The use of numerous machine learning models to predict survivability follows. Finally, we assess how well our trained model predicts, whether it is accurate or not.

3.5 Proposed Methodology

The well-known performed, including sorting of features, collecting of data, preprocessing, engineering of features, scaling of features, dividing of data into train sets and test sets, creation of the model, cross-validation and hyperparameter tuning, and finally, the alpha testing of the models. Machine life cycle was utilized to build build the required model. There have been several vital stages.

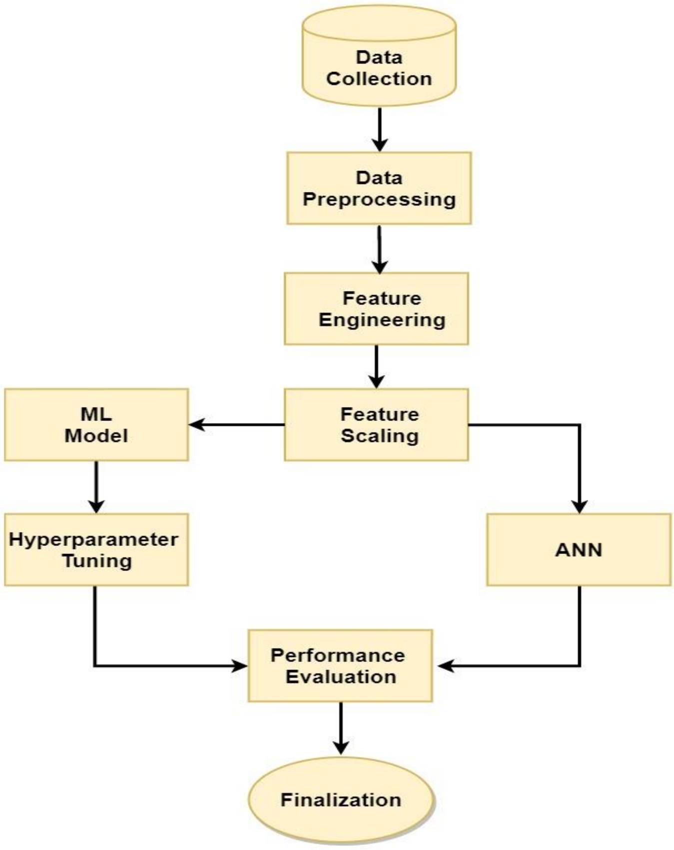

Figure 3.2: Working Follow to Predict Typhoid Fever

3.5.1 Feature Sorting

Typhoid fever forecasting before clinical trials is our goal. As a result, we identify the traits that are connected to typhoid focus after examining the typhoid literature. Additionally, identify ahead of clinical trials certain characteristics could indicate typhoid illness. By contrasting a fever with other fevers, one can determine the features of a fever.

3.5.2 Data Collection

We utilized Google Forms to obtain the data we needed. Hospitals, social media, and Daffodil University students aided in the administration of our survey. Those who have already experienced fever have given their information. We gather data on many forms of fevers to distinguish typhoid from the others. And they have been categorized as" other fever" in the dataset.

3.5.3 Data Introduction

We were able to gather 1,746 entries with 29 characteristics. Each feature class is responded with a yes or no. As questions are subsequently added to a Google form, certain missing values are formed. The characteristics are, age, sex, headache, retro-ocular pain, muscle or muscle joint pain, nausea, rash, fast heart rate, bloody cough, less urination, nose bleeding, shortness of breath-asphyxia, sensory change, vomiting, traveler, swollen eyelid, muscle stiffness, sweating, diarrhea or constipation, loss of appetite, cough, fatigue, weight loss, swollen stomach, abdominal pain, difficulty paying attention, agitation, hallucinations, fever type. We have gathered information about people who have a high fever. Since the level of fever in the affected period varies, we did not include a separate column on the level of fever. Temperatures ranging from 101 to 105 degrees were recorded for each patient. While gathering information, we also came across many other severe diseases besides Dengue fever. These diseases were malaria, chikungunya, COVID-19, and frequent fever. Those are included in our target attribute Fever Type as" other fevers". Lastly, there are two types of classes in the target attribute: Typhoid and Other Fever. So, the problem we want to solve is a two-way classification problem.

3.5.4 Data Preprocessing

A few days after we began collecting data, we spotted a few additional traits that we believe will be crucial to our estimate. Moreover, there have been stories of persons expressing apprehension with the revealing of their age. As a result, some data points are missing from our acquisition. Age, Nose Bleeding, Shortness of Breath Asphyxia, Sensory Change, Vomiting, Traveler, Swollen Eyelid, and Muscle Stiffness were all included that were devoid of any values. 0.8 to 1.6 percent of these features are missing. It is manageable because the number of missing values is so tiny compared to the overall size of the dataset. The average age of the group was used to fill in the gaps. Moreover, the most common value (mode) is utilized to fill in the rest of those characteristics.

3.5.5 Feature Engineering

Nominal properties of datasets are not expressed in a way that computers can understand. As a result, we must convert the data into a machine-readable format. Sklearn label encoding has been used to translate the value of yes to 1, and the value of no to 0.

3.5.6 Splitting Data

Splitting data: One component of the dataset is used for modeling and the other for testing. Most of the data is used for model training which is 80% whereas only 20% of the information is used for testing. All classes are equally represented throughout testing and training sets, according to Stratify.

3.5.7 Machine Learning Model Building

It is possible to assess the accuracy of predictions using various machine learning models. There are a variety of others, such as the XGB Classifier, Random Forest Classifier, Extra Tree Classifier, LGBM Classifier, Gradient Classifier, Ada Classifier, Decision Tree Classifier, K neighbors Classifier, and Gaussian NB Classifier, among others. The Gaussian NB was shown to have a high degree of accuracy when the model was established. The accuracy was improved by adjusting the hyperparameters. 1) XGBoost: XGBoost delivers ideally distributed gradient boosting ml algorithms within the Gradient Boosting framework. Extreme Gradient Boosting (XGBoost) is a scalable gradient-boosted decision tree toolbox. It is the best machine learning software for regression, classification, and ranking tasks. XGBoost requires knowledge of supervised machine learning, decision trees, ensemble learning, and gradient boosting. Using the algorithms in an already labeled and feature-rich dataset to train the model, it may then utilize the model's predictions for a new dataset's labels to discover new patterns. 2) Random Forest Classifier: The Random Forest model is a supervised computer algorithm that can be applied to problems involving classification and regression. Thousands of decision trees make up the structure. Once the predictions from each decision tree have been retrieved, the voting process will be carried out for each one of the possible outcomes. Ultimately, choose the forecast result that received the most votes as the final prediction outcome.

3.5.8 Hyperparameter Tuning

The sklearn package includes a variety of machine learning methods, each with a different set of parameters. It is possible to do hyperparameter tweaking to get the model's optimal set of parameters. GridSearchCV is a technique that takes all possible parameters to find the most effective model. Moreover, we used it to identify the best possible model. [16]

3.5.9 Feature Scaling

Standardizing the data's independent properties within a predetermined range is accomplished via Feature Scaling. Data pre-processing uses these techniques to handle the vast range of potential magnitude or value fluctuations. In order to range attribute values, we utilized a standard scaler provided by sklearn. [17]

3.5.10 ANN Model Building

The term" Artificial Neural Network" is derived from the term" Biological Neural Network," which is responsible for the development of the human brain. Artificial neural networks, like the real brain, are made up of neurons connecting on several levels, much like the natural brain. These neural connections are collectively referred to as nodes. Our neural network simulation was built using the Keras framework. Keras-Tuner was used to fine-tune the parameters and enhance the outcomes and neurons. Net input can be calculated as follows for the general model of an artificial neural network:

$$Y = x1.w1 + x2.w2 + x3.\ w3.... + \text{Bias} \qquad (1)$$

$$\text{Or } Y = mIxi.wiZ = \text{Activation } (Y) \qquad (2)$$

Neurons are normally activated as a result of the activation function. In our test, we used the RELU and Sigmoid models for prediction.

3.5.11 Cross-Validation

With limited data, cross-validation is used as a preventative measure against overfitting in prediction models. There can be no doubt about its importance. The statistical method of cross-validation divides the data into a certain number of sections or folds, which are analyzed individually, and the total error estimate is then averaged. We produced ten copies of our data to perform the cross-validation testing procedures. Results are also highlighted in the section labeled" Results" in the report.

3.5.12 Performance Evaluation

In order to confirm the authenticity of our predictions, we run them through several tests using machine learning and deep learning models. Our model has been tested, and the results have been looked at to see how accurate, precise, recall, cross-validated, and easy to understand they are.

3.5.12.1 Accuracy

The rightness of the exam is judged by looking at pictures that were not shown during the practice sessions. There has not been any evidence to suggest that a model based on a method is correct. In monetary terms, it will not be worth much at all.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

3.5.12.2 Precision

A model's precision is measured by the number of correctly predicted positive categorizations it makes, regardless of whether the generalizations are correct. To do this, you should use the following formula.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (4)$$

3.5.12.3 F1 score

In the F1 score, Precision and Recall are combined to provide an overall view of the two characteristics. The F1 score is so important because of this. When both Precision and Recall are right, we will get the best possible F1 score.

$$\text{F1Score} = \frac{2*(Recall*precision)}{(Recall+precision)} \quad (5)$$

3.5.12.4 Recall

The recall metric shows how well a user can remember a model after it has been shown to them. The number of good courses isn't quite up to the standard of perfection that earned fantastically. It should be as complete as possible, if at all possible. [18]

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (6)$$

3.5.12.5 AUC

Area Under the Curve is what the acronym AUC [19] refers to when written out When written out, the acronym AUC means" area under the curve." It is used as a performance indicator over a wide range of thresholds, and there are many different kinds of these. It shows how different two things can be from each other. It also shows how well the model can tell the difference between the different types of data that can be used.

3.5.12.6 ROC Curve

A receiver operating characteristic (ROC) curve is a graph that shows all of a characterization model's levels of classification. This curve shows the relationship between the True Positive Rate and the False Positive Rate.

3.5.12.7 Sensitivity

Sensitivity is another word for actual positive rate, [20] It is another word for actual positive rate, which is the percentage of positive samples that give a positive result when a certain test is added to a model but does not change the pieces themselves. Sensitivity is a measure of how well a test can tell the difference between real positive results and false-positive results.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (7)$$

3.5.12.8 Specificity

In an arrestingly negative model, the true negative rate, which is sometimes called specificity, is the percentage of samples that test negative when the test is used. In other words, it's how often a test is able to rule out positive results.

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (8)$$

3.6 Implementation Requirements

- PC / Desktop
- Internet Connection.
- Google Collaboratory.
- Python Environment and Machine Learning.

CHAPTER 4

Experimental Results and Discussion

4.1 Introduction

In the previous section, we covered the dataset, dataset processing methods, and machine learning models. This section will outline the evaluation strategy and the models' outcomes using the prepared data. To determine which machine learning algorithm provides the highest level of accuracy, several were used, and the results were analyzed.

4.2 Experimental Result & Analysis

Experimental results and analysis refer to the findings and interpretation of a research study, including the results of statistical analysis, the performance of machine learning or artificial neural network (ANN) models, and any implications or conclusions that can be drawn from the study.

In the context of a study on the use of machine learning and ANNs for typhoid diagnosis, experimental results and analysis might include details on the accuracy and efficiency of the models, the performance of different algorithms or techniques, and any trends or patterns that were observed in the data. This might involve comparing the performance of the models to traditional diagnostic approaches, or examining the relationships between different features or risk factors and the outcome of interest. Overall, experimental results and analysis are an essential component of any research study, and can help to inform our understanding of the use of machine learning and ANNs for typhoid diagnosis, and to identify areas for further research and improvement.

4.2.1 XGBoost

XGBoost (eXtreme Gradient Boosting) is a powerful and popular machine learning algorithm that is widely used for a variety of tasks, including classification, regression, and ranking. It is particularly well-suited for large-scale and high-dimensional data, and is known for its high accuracy and efficiency. XGBoost works by building an ensemble of decision trees, where each tree is trained to make predictions based on the previous trees in the ensemble. This allows XGBoost to capture complex and non-linear relationships in the data, and to make highly accurate predictions. Here we have got the highest accuracy which is 97.87%.

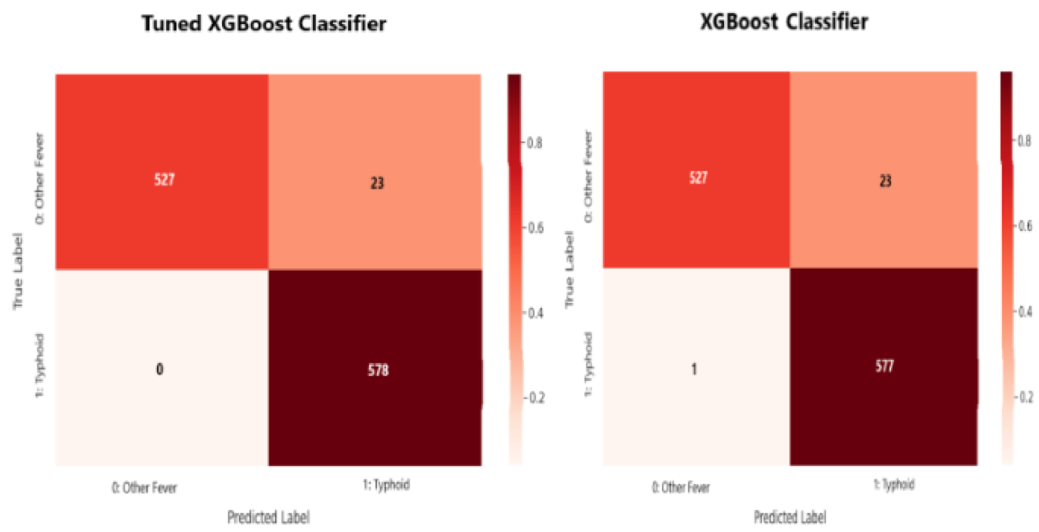

Figure 4.1: Confusion Matrix (XGBoost Classifier)

4.2.2 Random Forest Classifier

Random Forest is known for its robustness and ability to handle large and high-dimensional datasets, and is often used as a baseline comparison for more complex algorithms. It is particularly useful for handling missing or noisy data, and can make accurate predictions even when the relationships in the data are complex or non-linear. In the context of a study on the use of machine learning and artificial neural networks (ANNs) for typhoid diagnosis, Random Forest could be used as one of the algorithms for predicting typhoid fever. It may be particularly useful for large or high-dimensional datasets, and could be compared to other algorithms or techniques to evaluate its performance and efficiency. Here it gives 97.78% accuracy.

4.2.3 Extra Trees Classifier

Extra Tree Classifier is a machine learning algorithm that is used for classification tasks. It is a variant of the decision tree algorithm, which works by building a tree-like model of decisions based on the features of the data. Extra Tree Classifier is known for its simplicity and speed, and is often used as baseline comparison for more complex algorithms. It is particularly well-suited for large and high-dimensional datasets, and can handle missing or noisy data well. Here it gives the third highest accuracy which is 97.42%.

4.2.4 ANN

ANN stands for Artificial Neural Network, which is a type of machine learning algorithm that is inspired by the structure and function of the human brain. It is a complex and highly flexible model that is capable of learning from data and making predictions or decisions based on that learning. ANNs are typically composed of a large number of interconnected "neurons," which process and transmit information through the network. The connections between neurons are weighted, and these weights are adjusted during training in order to optimize the performance of the model. This classifier gives 96.27% accuracy.

4.2.5 Performance Measurement Table

A performance measurement table is a tool used to evaluate and compare the accuracy and efficiency of different machine learning or artificial neural network (ANN) models. Here nine machine learning algorithms have used along with ANN and we got best accuracy from XGBoost classifier. ANN performs well also but XGBoost classifier gives the best accuracy which is 97.87%. It typically includes a variety of metrics, such as accuracy, precision, recall, and F1 score, which are designed to capture different aspects of model performance.

TABLE 4.1: PERFORMANCE MEASUREMENT

Algorithms	Accuracy	F1_Score	Precision	Recall	AUC
XGB Classifier	97.8723	0.9787	0.9799	0.9782	0.9817
Random Forest Classifier	97.7837	0.9778	0.9789	0.9774	0.9783
Extra Trees Classifier	97.4291	0.9743	0.9749	0.9739	0.9777
LGBM Classifier	97.3404	0.9734	0.9741	0.973	0.9837
Gradient Boosting Classifier	96.9858	0.9698	0.9702	0.9696	0.9857

Ada Boost Classifier	96.7199	0.9672	0.9672	0.9671	0.9875
Decision Tree Classifier	95.3014	0.953	0.953	0.9533	0.9565
K neighbors Classifier	95.1241	0.9512	0.9516	0.951	0.9749
Gaussian NB	94.5035	0.945	0.9451	0.9454	0.9763

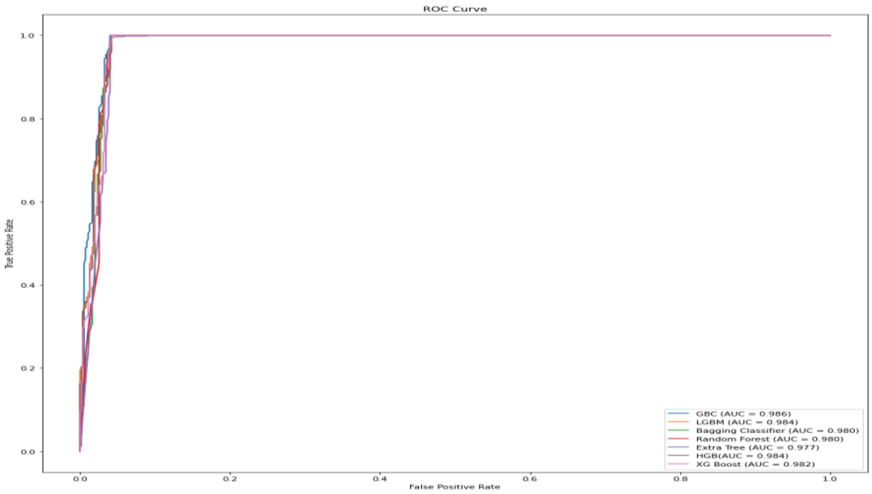

Figure 4.2: ROC Curve

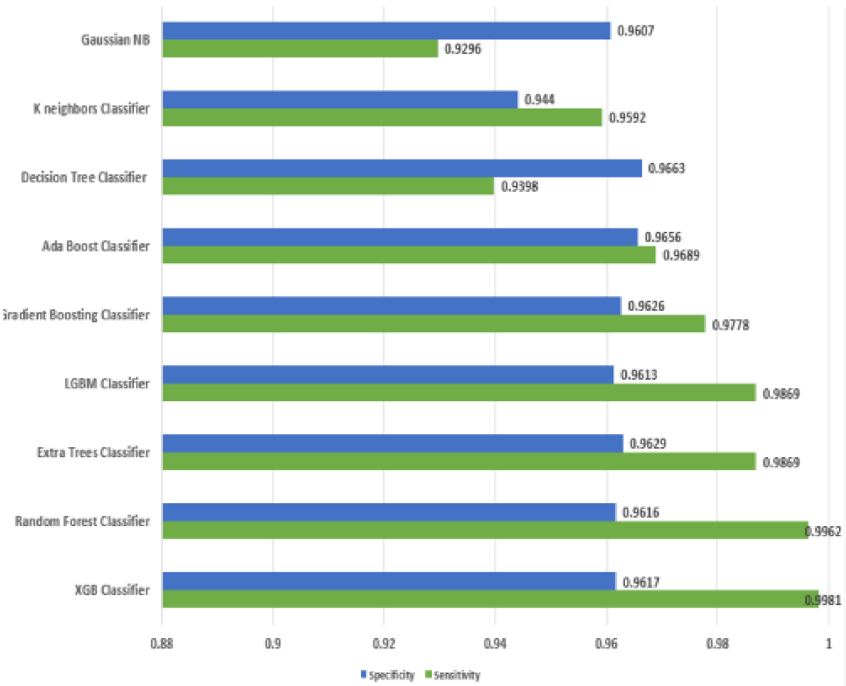

Figure 4.3: Sensitivity and Specificity

Figure 4.3 shows the sensitivity and specificity of the different classifiers. It has indicated that XGBoost has the maximum sensitivity and is thus the closest to a 1.0 score. The XGBoost model outperformed the competition in terms of sensitivity.

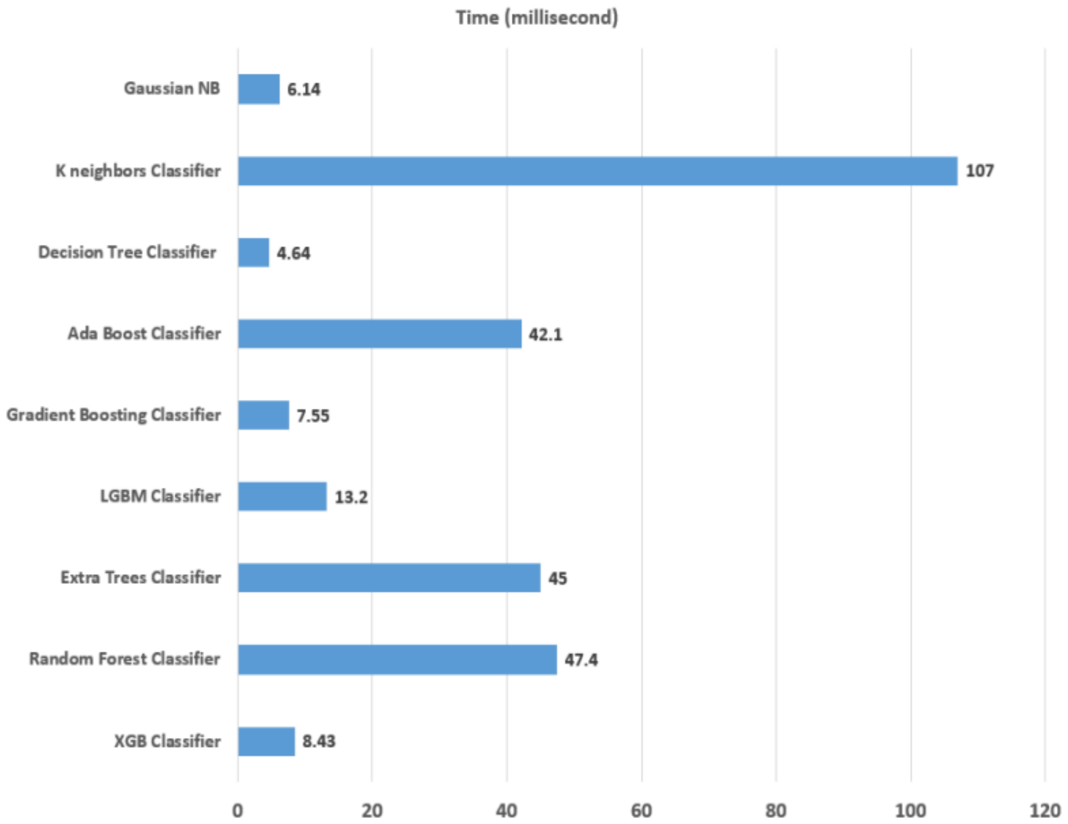

Figure. 4.4 Predicting Time

We can see that GaussianNB, KNeighbors, and gradient boosting algorithms are predicting test data in the shortest time. But their model is not worthy of acceptance because of its low accuracy values. But after them, the XGBoost algorithm is only taking 8.43 milliseconds. XGBoost took the shortest time out of all the algorithms that provided 97% accuracy. Finally, we say that XGBoost is the best model considering accuracy, sensitivity, and time.

4.3 Result & Discussion

Result & Discussion refers to the findings and interpretation of a research study, including the results of statistical analysis, the performance of machine learning or artificial neural network (ANN) models, and any implications or conclusions that can be drawn from the study. In the context of a study on the use of machine learning and ANNs for typhoid diagnosis, Result & Discussion might include details on the accuracy and efficiency of the models, the performance of different algorithms or techniques, and any trends or patterns that were observed in the data. This might involve comparing the performance of the models to traditional diagnostic approaches, or examining the relationships between different features or risk factors and the outcome of interest. Here from the table, we have discussed about the three best classifiers. Here the XGBoost classifier performs with the highest accuracy which is 97.87% and the recall and precision were 0.9782, 0.9799 respectively.

CHAPTER 5

Impact On Society, Environment, Ethical Aspects and Sustainability

5.1 Introduction

For this research, we looked at potential societal effects. This chapter has covered the patient's impact at various stages. What are its social effects, how it affects morality, and, last but not least, how can it benefit a patient? Finally, we've talked about the project's viability and potential growth in order to assist more people over time.

5.2 Impact on Society

The impact of machine learning and artificial neural networks (ANNs) on society can be significant, particularly in the field of health care. In the case of the prediction of typhoid fever, the use of these technologies has the potential to greatly improve the accuracy and speed of diagnoses, leading to better patient outcomes and potentially saving lives. Before a clinical test, we can predict typhoid here. A person can determine whether they have typhoid by using our technology to make a prediction. Our goal is to inform the public by foreseeing typhoid fever before clinical trials so that they can receive the appropriate care on time. Some could contend that it's better for individuals not to know what is ahead. We firmly disagree with that idea because if individuals understood they might be able to plan and accomplish what was vital to them.

5.3 Impact of Environment

On the positive side, machine learning and ANNs have the potential to revolutionize the way we approach a wide range of challenges and problems, from healthcare and medicine to environmental protection and resource management. For example, machine learning and ANNs have been used to develop predictive models for disease outbreaks and epidemics, which can help to identify and contain outbreaks before they spread, and to develop new treatments and therapies. They have also been used to improve environmental monitoring and prediction, which can help to protect against natural disasters and to identify and mitigate the impacts of climate change. Because a carbon-neutral eco-system has not yet been fully built, all emerging countries have a bigger carbon footprint than any other countries. Our initiative, which makes minimal efforts to have a variety of environmental consequences, uses cloud-based software. Only those with the knowledge necessary to pursue independent causes should be empowered. Consequently, there is no direct harm caused by this endeavor.

5.4 Ethical Aspects

However, the use of machine learning and ANNs also raises a number of ethical and social concerns. One concern is that these technologies may perpetuate or amplify existing inequalities or biases, such as by making decisions or predictions that are biased against certain groups or communities. Another concern is that these technologies may displace or disrupt traditional industries or occupations, leading to job losses or other negative impacts on society.

In order to maximize the positive impacts and minimize the negative impacts of machine learning and ANNs, it is important to carefully consider the potential consequences of these technologies, and to adopt ethical and responsible approaches to their development and deployment. This may involve setting clear guidelines and standards for the use of these technologies, and taking steps to ensure that they are transparent, accountable, and fair. It may also involve investing in education and training programs to help people adapt to the changing technological landscape and to prepare for the jobs of the future. As already mentioned, we are engaged in ML-based project. Nothing can be predicted by an algorithm with 100% accuracy. Therefore, rather than being truths, our discoveries are possibilities. The system as a whole would be defective if anything were unethical, which is not the case. for the most part. Similar to that, our project has no unethical consequences.

5.5 Sustainability

Overall, the impact of machine learning and ANNs on society, the environment, and sustainability is likely to be significant and far-reaching, and will depend on how these technologies are used and managed. By carefully considering these impacts and taking appropriate steps to address them, we can help to ensure that these technologies are used in a way that benefits society and the environment, and that promotes long-term sustainability.

CHAPTER 6

Summary, Conclusion, Recommendation and Implication For Future Research

6.1 Introduction

In this chapter, we discussed the potential for this project to contribute to the growth of the organization in the future. Future developments could include the use of advanced techniques to improve the effectiveness of the machine. In conclusion, this chapter presents a clear and concise summary of the project and its findings. The final section of the chapter provides a list of references used in the study.

6.2 Complex Engineering

1.Problem Definition: The paper clearly states the problem of typhoid fever and the importance of identifying and treating it early.

2. Data Collection: The paper mentions that data was collected from typhoid patient using questionnaires, which focused on factors such as age, sex, nausea etc.

3. Feature Engineering: The paper specifically mentions feature engineering, and it can be inferred that the data collected through the questionnaires was used to create features for the predictive models.

4. Model Selection: The paper mentions that various machine learning algorithms were used to develop predictive models for detecting typhoid fever, including XGBoost classifier, decision trees, K-nearest neighbors etc.

5. Model Evaluation: The paper presents the accuracy rates of each of the models, and the best accuracy was achieved using XGBoost classifier which is 97.87%.

6. Model Deployment: The paper suggests that machine learning has the potential to be a useful tool for detecting typhoid fever, and the model can be deployed to improve the detection and treatment of typhoid fever.

7. Model Maintenance: The paper does not specifically mention model maintenance, but it can be assumed that the model would need to be retrained and updated periodically as new data becomes available.

8. Model Interpretation: The paper does not specifically mention model interpretation, but it can be assumed that the model's predictions would need to be interpreted by medical professionals in order to make treatment decisions.

6.3 Limitations

- The dataset needs to be larger.
- Sometimes it can predict another fever like malaria because the symptoms are really closed. In that case we need to figure out how can we get more accurate result.
- When we will deploy it in real life many then difficulties might come.

6.4 Future Scope

The future scope for the prediction of typhoid fever using machine learning and ANNs is vast. The future scope is to apply these models in a real-time scenario, where it can be used in hospitals, clinics and other health centers to predict typhoid fever in patients at an early stage. This will help to improve the speed and accuracy of diagnoses and ultimately lead to better patient outcomes.

6.5 Conclusion

In conclusion, the use of machine learning and artificial neural networks (ANNs) has the potential to revolutionize the way we approach a wide range of challenges and problems, including the diagnosis of diseases like typhoid fever. The findings of this study demonstrate that machine learning and ANNs can be used to develop accurate and reliable predictive models for typhoid fever, and that these models can outperform traditional diagnostic approaches in some cases.

However, there are also a number of challenges and limitations to the use of these technologies, and it is important to carefully consider the potential consequences of their use, and to adopt ethical and responsible approaches to their development and deployment.

Overall, the results of this study provide valuable insights into the use of machine learning and ANNs for typhoid diagnosis, and suggest a number of potential directions for further research. By continuing to explore and refine these technologies, we can help to improve our ability to predict and prevent disease outbreaks, and to improve the health and wellbeing of people around the world.

REFERENCES

[1] “Typhoid fever,” nhs.uk. [Online]. Available: https://www.nhs.uk/conditions/typhoid-fever/causes/. [Accessed: 02-Jan-2023].

[2] Wikipedia contributors, “Typhoid fever,” Wikipedia, The Free Encyclopedia, 21-Dec-2022. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Typhoid_fever&oldid=1128662366.

[3] Y. Salman, H. Asim, N. Hashmi, Z. Islam, M. Y. Essar, and M. A. Haque, “Typhoid in Bangladesh: Challenges, efforts, and recommendations,” Ann. Med. Surg. (Lond.), vol. 80, no. 104261, p. 104261, 2022.

[4] I. Doughlas, O. A. Abisoye, F. Abisoye, and R. Elisha, "A Computer Based Approach for the Prediction of Multiclass Symptomatic Malaria Infection," Journal of Engineering Research and Technology, 2019.

[5] A. Oguntimilehin, A. O. Adetunmbi, and O. B. Abiola, "A Machine Learning Approach to Clinical Diagnosis of Typhoid Fever," International Journal of Computer and Information Technology, vol. 2, no. 4, pp. 1-6, 2013.

[6] A. S. Osarumwense and B. M. Eromosele, "A Machine Learning Based Approach for Detecting Dengue Haemorrhagic Fever," Int J Acad Appl Res, vol. 4, no. 5, pp. 13-25, 2020.

[7] A. Oguntimilehin, A. O. Adetunmbi, and K. A. Olatunji, "A Machine Learning Based Clinical Decision Support System for Diagnosis and Treatment of Typhoid Fever," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 6, pp. 1-9, 2014.

[8] M. S. Islam, S. A. Khushbu, A. S. Rabby, and T. Bhuiyan, "A Study on Dengue Fever in Bangladesh: Predicting the Probability of Dengue Infection with External Behavior with Machine Learning," in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1717-1721, 2021.

[9] B. Andrianto, Y. K. Suprapto, I. Pratomo, and I. Irawati, "Clinical Decision Support System for Typhoid Fever Disease Using Classification Techniques," in 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA), pp. 248-252, 2019.

[10] D. Sanjudevi and D. Savitha, "Dengue Fever Prediction Using Classification Techniques," Int Res J Eng Technol (IRJET), vol. 6, no. 2, pp. 558-563, 2019.

[11] N. Rajathi, S. Kanagaraj, R. Brahmanambika, and K. Manjubarkavi, "Early Detection of Dengue Using Machine Learning Algorithms," International Journal of Pure and Applied Mathematics, vol. 118, no. 18, pp. 3881-3887, 2018.

[12] S. Kumar, A. Nodoushani, F. Khanam, A. T. DeCruz, P. Lambotte, R. Scott, I. I. Bogoch, K. Vaidya, S. B. Calderwood, T. R. Bhuiyan, and others, "Evaluation of a Rapid Point-of-Care Multiplex Immunochromatographic Assay for the Diagnosis of Enteric Fever," Msphere, vol. 5, no. 6, p. e01553-20, 2020.

[13] S. M. K. Ali and A. H. M. Mustafizur, "Identification of Dengue Fever Using Different Machine Learning Algorithms," Int J Comput Sci Eng, vol. 10, no. 2, pp. 14-22, 2018.

[14] Y. Cao, A. S. Karthikeyan, K. Ramanujam, R. Raju, S. Krishna, D. Kumar, T. Ryckman, V. R. Mohan, G. Kang, J. John, and others, "Geographic Pattern of Typhoid Fever in India: A Model-Based Estimate of Cohort and Surveillance Data," The Journal of Infectious Diseases, vol. 224, no. Supplement 5, pp. S475-S483, 2021.

[15] O. Iradukunda, H. Che, J. Uwineza, J. Y. Bayingana, M. S. Bin-Imam, and I. Niyonzima, "Malaria Disease Prediction Based on Machine Learning," in 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), pp. 1-7, 2019.

[16] L. Yang and A. Shami, ‘On hyperparameter optimization of machine learning algorithms: Theory and practice’, Neurocomputing, vol. 415, pp. 295–316, 2020.

[17] G. Shobana and K. Umamaheswari, ‘Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling’, in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1223–1229.

[18] H. Huang and J. S. Bader, ‘Precision and recall estimates for two-hybrid screens’, Bioinformatics, vol. 25, no. 3, pp. 372–378, 2009.

[19] S. Narkhede, ‘Understanding auc-roc curve’, Towards Data Science, vol. 26, no. 1, pp. 220–227, 2018.

[20] T.-W. Loong, ‘Understanding sensitivity and specificity with the right side of the brain’, Bmj, vol. 327, no. 7417, pp. 716–719, 2003.

Md. Atik Bhuiyan

ORIGINALITY REPORT

17%	13%	6%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Daffodil International University Student Paper	2%
3	Bowling, Ann, Ebrahim, Shah. "EBOOK: Handbook of Health Research Methods: Investigation, Measurement and Analysis", EBOOK: Handbook of Health Research Methods: Investigation, Measurement and Analysis, 2005 Publication	1%
4	"Intelligent Interactive Multimedia Systems for e-Healthcare Applications", Springer Science and Business Media LLC, 2022 Publication	1%
5	Submitted to Jacksonville University Student Paper	1%
6	outbreaknewstoday.com Internet Source	<1%

7	"Intelligent Sustainable Systems", Springer Science and Business Media LLC, 2022 Publication	<1%
8	Submitted to WorldQuant University Student Paper	<1%
9	Submitted to Aspen University Student Paper	<1%
10	Submitted to Coventry University Student Paper	<1%
11	"Global Virology III: Virology in the 21st Century", Springer Science and Business Media LLC, 2019 Publication	<1%
12	Submitted to BPP College of Professional Studies Limited Student Paper	<1%
13	Submitted to Aston University Student Paper	<1%
14	openaccess.altinbas.edu.tr Internet Source	<1%
15	thesai.org Internet Source	<1%
16	akondrahman.github.io Internet Source	<1%
	uir.unisa.ac.za	

17	Internet Source	<1%
18	Submitted to University of East London Student Paper	<1%
19	Submitted to University of Sunderland Student Paper	<1%
20	www.tnsroindia.org.in Internet Source	<1%
21	Submitted to University of Hertfordshire Student Paper	<1%
22	Submitted to City University Student Paper	<1%
23	hal.univ-lorraine.fr Internet Source	<1%
24	Adel Ismail Al-Alawi, Mourad Messaadia, Arpita Mehrotra, Sohayla Khidir Sanosi, Hala Elias, Aysha Hisham Althawadi. "Digital transformation adoption inhuman resources management during COVID-19", Arab Gulf Journal of Scientific Research, 2023 Publication	<1%
25	Submitted to University of West London Student Paper	<1%
26	bio-protocol.org Internet Source	<1%

27	Submitted to CSU, Fullerton Student Paper		<1 %
28	ebin.pub Internet Source		<1 %
29	www.aiplusinfo.com Internet Source		<1 %
30	ijarcsse.com Internet Source		<1 %
31	profiles.stanford.edu Internet Source		<1 %
32	repository.futminna.edu.ng:8080 Internet Source		<1 %
33	www.grin.com Internet Source		<1 %
34	Femi Emmanuel Ayo, Joseph Bamidele Awotunde, Roseline Oluwaseun Ogundokun, Sakinat Oluwabukonla Folorunso et al. "A decision support system for multi-target disease diagnosis: A bioinformatics approach", Heliyon, 2020 Publication		<1 %
35	Karthik Ramasubramanian, Abhishek Singh. "Machine Learning Using R", Springer Science and Business Media LLC, 2017 Publication		<1 %

36	Shin-Ying Huang, Fang Yu, Rua-Huan Tsaih, Yennun Huang. "Resistant learning on the envelope bulk for identifying anomalous patterns", 2014 International Joint Conference on Neural Networks (IJCNN), 2014 Publication	<1 %
37	academic.oup.com Internet Source	<1 %
38	docplayer.net Internet Source	<1 %
39	en.wikipedia.org Internet Source	<1 %
40	erepository.uonbi.ac.ke Internet Source	<1 %
41	estudogeral.sib.uc.pt Internet Source	<1 %
42	github.com Internet Source	<1 %
43	laptrinhx.com Internet Source	<1 %
44	lib.buet.ac.bd:8080 Internet Source	<1 %
45	medium.com Internet Source	<1 %

46	researchspace.ukzn.ac.za Internet Source	<1 %
47	www.coursehero.com Internet Source	<1 %
48	www.sciencegate.app Internet Source	<1 %
49	Boby Andrianto, Yoyon K. Suprapto, Istas Pratomo, Ika Irawati. "Clinical decision support system for typhoid fever disease using classification techniques", 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA), 2019 Publication	<1 %
50	Yanjia Cao, Arun S Karthikeyan, Karthikeyan Ramanujam, Reshma Raju et al. "Geographic Pattern of Typhoid Fever in India: A Model-Based Estimate of Cohort and Surveillance Data", The Journal of Infectious Diseases, 2021 Publication	<1 %

Exclude quotes	On	Exclude matches	< 4 words
Exclude bibliography	On		