**Prediction of Child Disease Based on Statistical Analysis of Survey Data Using Machine Learning**

**BY**

**Mohammad Shazzad Hasan Shipat**
**ID: 172-15-10149**

**Rajaul Islam**
**ID: 172-15-10159**

**Md. Safayet Hossain**
**ID: 172-15-10169**

This Report Presented in Partial Fulfillment of the Requirements for

The Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Warda Ruheen Bristi**

Lecturer

Department of CSE

Daffodil International University

**DAFFODIL INTERNATIONAL UNIVERSITY**

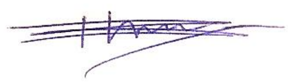
**DHAKA, BANGLADESH**

**MAY 2021**

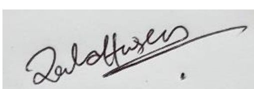
# APPROVAL

This Project titled "**Prediction of child disease based on statistical analysis of survey data using machine learning**", submitted by **Mohammad Shazzad Hasan Shipat**, **Rajaul Islam** and **Md. Safayet Hossain** 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 31 May 2021.

## BOARD OF EXAMINERS



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



**Zahid Hasan** **Internal Examiner**
**Assistant Professor**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

**Md. Riazur Rahman** **Internal Examiner**
**Assistant Professor**
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

**Dr. Md Arshad Ali**
**Associate Professor**
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

**External Examiner**

# DECLARATION

We hereby declare that, this thesis has been done by us under the supervision of **Warda Ruheen Bristi, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this thesis nor any part of this thesis has been submitted elsewhere for award of any degree or diploma.

**Supervised by:**

**Warda Ruheen Bristi**
Lecturer
Department of CSE
Daffodil International University

**Submitted by:**

**Mohammad Shazzad Hasan Shipat**
ID: 172-15-10149
Department of CSE
Daffodil International University

**Rajaul Islam**
ID: 172-15-10159
Department of CSE
Daffodil International University

**Md. Safayet Hossain**
ID: 172-15-10169
Department of CSE
Daffodil International University

# ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty God for His divine blessing makes us possible to complete the final year thesis successfully.

We really grateful and wish our profound our indebtedness to **Warda Ruheen Bristi, Lecturer,** Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Machine Learning Algorithms” to carry out this thesis. Her 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 thesis.

We would like to express our heartiest gratitude to **Professor Dr. Touhid Bhuiyan, Professor, and Head,** Department of CSE, for his kind help to finish our thesis 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 patients of our parents.

# ABSTRACT

This research aims to ascertain the prevalence and the factors affecting preventive infectious diseases among small children in Bangladesh. The students assumed that morbidity occurred in infants under 5 years of age was affected by various background features of children and their parents. In Bangladesh, the death rate of infants is much higher than natural deaths. Children are dying from several diseases every day. Malaria, diarrhea and Chickenpox are most common amongst them. There were 285731 confirmed malaria cases between January 1, 2008, and December 31, 2012. Though it will be decreased because the Bangladesh Government has taken many steps against it. But still, it's a big issue. Bangladesh had the greatest percentage of deaths due to acute bloody diarrhea in children from 1 to 4 years old: 27.8 percent (5/18). Chickenpox is also a massive problem behind child deaths. On the other hand, younger children were more likely than their older counterparts to have several health problems. Children from lower-income or middle-income families were at a higher risk of disease than those from higher-income families. The prevalence of childhood moral diseases was significantly influenced by changes in drinking water supply and care methods. The 85.8 percent more likely co-morbidity in children in homes with contaminated untreated sewage is compared to parents in families with piped water. Four standard ML algorithms were used. They are- Naive Bayes Classifier, SVM, Neural Network and Random Forest Support Machines. The highest accuracy of 96.17% has been forecast by random forest.

# TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1

# Introduction

## 1.1 Introduction

Bangladesh is a poor area, but over recent years, its health has progressively improved. It has the lowest infant and child mortality, the lowest vaccine rates and the most general family planning services, although it is slightly poorer than Pakistan, India and Nepal. In the course of these last few decades children in Bangladesh continue to experience many diseases, despite considerable advances achieved by private medicine companies and NGOs. Malaria, diarrhea and vomiting, chicken pox, pneumonia, respiratory infections, preterm birth complications and neonatal encephalopathy are the most frequent diseases among infants. In 13 of Bangladesh's 64 districts, endemic malaria transmission has been reported. In the late 1990s, up to 70000 research facilities affirmed and 900000 clinical cases were enlisted, with over 500 deaths each year. The quantity of unreported cases may have been just about as high as 250 000 every year. In the 13 malaria-endemic districts, a cross-sectional survey conducted in 2007 revealed a crude prevalence of 4%. In the 21st century in developing countries diarrhea is one of the main motives of morbidity and mortality in children. Diarrhea is most common in babies and young children. Between 2014 and 2019 in Bangladesh children who have diarrhea have collected 387 fecal specimens. In order to detect bacterial infections, traditional methods of bacteriological, biochemical and molecular sequence analyses were used. The study of molecular sequence recognition by polymerase chain reaction and reverse transcriptase polymerase chain reaction of DNA virus and RNA virus, respectively, was tested. Molecular test of the series confirmed the identification by polymerase chain reaction and the reverse transcriptase-polymerase chain reaction of DNA virus and diarrheal virus. [1] The Bangladeshi infant mortality rate in Southern Asia has decreased by 63 percent in the last 20 years and has dropped by the child mortality rates of the world — for children under the age of 5. According to the 2019 Saving the Children's Global Childhood Study, Since 2000 Bangladesh has lowered infant mortality amongst infants under 5 years of age to 63% and 77% since 1990, while the country is on course to reach the global target of 25 or lower deaths per 1,000 births within years leading to the 2030 deadline for Sustainable Development Goals (SDGs). According to the WHO Global Health Observatory, 532,000 children died in Bangladesh in 1990, but that number has now dropped to about 100,000 a year, according to BSS. [2] Varicella

(Chicken Pox) fatality rates range from 1 in 100,000 for 1 to 14-year-old, to 6 in 100,000 for 15-19 year-olds and to 21 in 100,000 for adults. Immune competent children and adults account for the majority of deaths.

## 1.2 Motivation

Between 2014 and 2019 in Bangladesh children who have diarrhea have collected 387 fecal specimens. Chicken Pox fatality rates range from 1 in 100,000 for 1 to 14-year-old, to 6 in 100,000 for 15-19 year-olds and to 21 in 100,000 for adults. The bulk of deaths are caused by immune-related children and adults. There is no information infrastructure capable of forecasting such diseases. This is why our work was inspired.

## 1.3 Problem Definition

In today's ICT world, the word "machine learning" is extremely relevant. Machine learning will aid in the development of our medical field. It is important to identify the problems and related criteria in this area in order to have an adequate solution. It's also important to know about government policies and regulations, as well as software industry standards and curriculum approaches for applying machine learning in the medical field.

## 1.4 Research Questions

The key questions that this study focuses on are given below:

- From where did we get the data?
- What is the research objective?
- Based on which issue is the study being developed?

## 1.5 Research Methodology

In the section of methodology of our research paper, we have collected Data, then we analyzed them, classified the data, selected algorithms, then implemented them, after these steps we evaluated them. At the end of this chapter performance of the proposed model will be described.

## 1.6 Research Objectives

Our aim is to categorize child disease based on a survey of several Bangladeshi villages in order to determine the key causes of newborn baby disease and minimize infant mortality.

Some technical issues are pointed below:

- Develop an efficient model to identify the main reasons of child disease.
- To teach software developers how to use the model to work with machine learning.
- Encourage the government to take appropriate measures to prevent young marriages.
- Models can be integrated into mobile applications and websites.

# CHAPTER 2

# Background

## 2.1 Introduction

There is no work or science in our country that can accurately predict disease and have a solution. As a result, the context is the current state of infant mortality and the use of Machine Learning in Bangladesh's medical field.

## 2.2 Related Works

Machine learning is extensively applied to resolving problems that are subject to forecasting. For taking actions toward the death of children, a lot of work has been done applying ML. ML has made this strategy extremely convenient.

S. Grampurohit et. al. [3] in this paper, they use a variety of classification methods to predict disease. For the research, they used 4920 records from patients who were diagnosed with 41 diseases. There were 41 diseases in the dependent variable. They chose 95 out of 132 independent variables (symptoms) that are closely linked to diseases and optimized them. They employed Machine learning algorithms as Decision Tree classifiers, Random Forest classifiers, and Naive Bayes classifiers.

O. Altay et. al. [4] they discussed Autism Spectrum Disorder Discussed (ASD). ASD is essential for young people to be diagnosed. This research on children 4 to 11 years old has used the classification system for ASD diagnosis. The LDA and K-Nearest Neighbor (KNN) algorithms are used in classification. The KNN is the most commonly used algorithm. The algorithms were tested for 30 percent of the data package. 30% of the datasets were selected for test data, and 70% were chosen for training for algorithm testing. The work has produced an accuracy of 90.8% for the LDA algorithm, while the precision of the KNN algorithm is 88.5%. The LDA algorithm has a sensitivity of 0.95 and a specificity of 0.08744. The KNN algorithm is used for these values of 0.97.

K. Kosasih et. al. [5] in their paper, they have discussed the Wavelet-Augmented Cough Analysis for Rapid Diagnosis of Childhood Pneumonia. Pneumonia kills more than a million children per year all over the world. They discuss problems and suggest a method for screening pneumonia based on a statistical study of cough sounds in this paper. They gathered 815 cough sounds from 91 patients suffering from respiratory diseases including influenza, asthma, and bronchitis. Based on wavelet features alone, the methods proposed in this paper had a sensitivity and specificity of 94 percent and 63 percent, respectively, in distinguishing pneumonia patients from non-pneumonia patients. When the wavelets are combined with features from our previous work, the sensitivity and specificity increase to 94 percent and 88 percent, respectively.

M. M. Faizin et. al. [6] this paper describes the use of artificial intelligence in a medical decision support system for diagnosing dehydration in children. A total of 92 medical records were collected for the dataset, which was then divided into two subsets: training set (57 records) and test set (57 records) (35 records). General appearance, skin, respirations, turgor, and mucous membranes are used as input variables for medical symptoms of dehydration, whereas the output variable is the degree of dehydration, which is divided into three categories: extreme dehydration, some dehydration, and no dehydration. The intelligent system's classification success was compared to the doctor's diagnosis for validation. The classification output of the intelligent system was mapped using the confusion matrix and evaluated using accuracy and the value of error rate. The results show that using artificial intelligence in a medical decision support system has a 91 percent accuracy rate and an error rate of 0.085714286.

M. U. Khan et. al. [7] Prematurity is the world's leading cause of death and morbidity in children. In order to investigate uterine contractions, the most promising biophysical signature is the non-invasive surface uterine electromyogram (sEMG), called the Electrohysterogram (EHG). This can prove to be a marker for premature birth detection, which enables us to diagnose premature birth before work begins. They proposed a method for signal treatment to prevent preterm birth with raw, short-term EHG signals (1 min). The raw EHG record is first preprocessed and segmented using Empirical Mode Decomposition by selecting only the first intrinsic mode function. Only four features extracted from segmented EHG records are provided in the Support Vector Machine Classification: Shannon Energy, Log Energy, Medium Frequency and Lyapunov Exponents. The device achieves 95.5 percent accuracy with the publicly available Term-Preterm EHG database.

M. E. H et. al. [8] the study aims to see how easily and accurately the useful artificial intelligence (AI) of the chest X-ray images can detect COVID-19. This paper aims to develop an accurate method for detection of COVID-19 pneumonia using pre-trained deep-learning algorithms from digital chest X-ray images while optimizing detection precise. By integrating numerous public databases and collecting images from recently published papers, researchers have created a public database. The database contains 423 COVID-19, 1485 pneumonia and 1579 regular chest X-ray images. Networks were able to identify the various pneumonic types: normal and pneumonia COVID-19; and (ii) normal, bacterial and pneumonia COVID-19 with or without picture augmentation. Both schemes had 99.7% classification accuracy, precision, sensitivity, and specificity of 99.7%, 99.7%, and 99.55% and 97.9%, 97.95%, 97.9%, and 98.8%, respectively.

L. Malihi et. al. [9] this research shows a method in giemsa-stained blood samples for identifying malaria parasites. The red blood cell mask will be retrieved in the first step to improve detection accuracy. This is the case because the majority of malaria parasites live in red blood cells. Following this, stained blood elements are removed, such as red blood cells, parasites and white blood cells. A red blood cell mask is attached to the extracted stained material to distinguish the external parasites. Finished by extracting and using features such as color histogram, granulometry, pitch and flat texture as input of classification. In this research, the nearest medium (NM), K nearest neighbors (KNN), 1-NN and Fisher supporting vector machines (SVM) have been used. The KNN classification of the closest neighbor in this study was the most accurate and scored 91%.

M. A. M. Ali et. al. [10] this research has shown that monitoring vital signs and early intervention can reduce mortality between children and newborns by up to 50 percent. Effective real-time care is required to ensure the right treatment for a sick baby. Further attention is therefore needed if all signs or symptoms of infant diseases are to be identified. This proposes a system that combines several functions into one unit, analyzes sensor data and device interaction data to inform about the conditions decided in standard clinical trials. With less overt intervention for non-serious conditions and early intervention for serious conditions, the proposed system could be able to reduce the burden on hospitals and parents. The system would offer technology to support current medical procedures as well as a real-time surveillance system, special care and assistance system for newborns.

From discussion of above work we can find some correlation and variations with our work. In our work we tried to solve the child death unnecessarily by predicting future price.

## 2.3 Bangladesh Perspective

The maternal mortality rate is the number of women who die during or within 42 days of pregnancy per 100,000 live births. For estimating data using the regression model, data have been used to estimate the ratios of maternal deaths from non-Aids deaths of women aged 15 to 49, fertility, birthright and GDP.

1. In 2017, the maternal mortality rate in Bangladesh was 173.00, down 6.99% from 2016.
2. The maternal mortality rate in Bangladesh in 2016 was 186.00, down 7% from 2015.
3. The maternal mortality rate in Bangladesh in 2015 was 200.00, down 6.54 percent from 2014.
4. The maternal mortality rate in Bangladesh in 2014 was 214.00, down 5.73 percent from 2013.

# CHAPTER 3

# Research Methodology

## 3.1 Introduction

The methodology covers an absolute of five steps which conclude our research that is displayed in Fig. 3.1. The steps are the following:

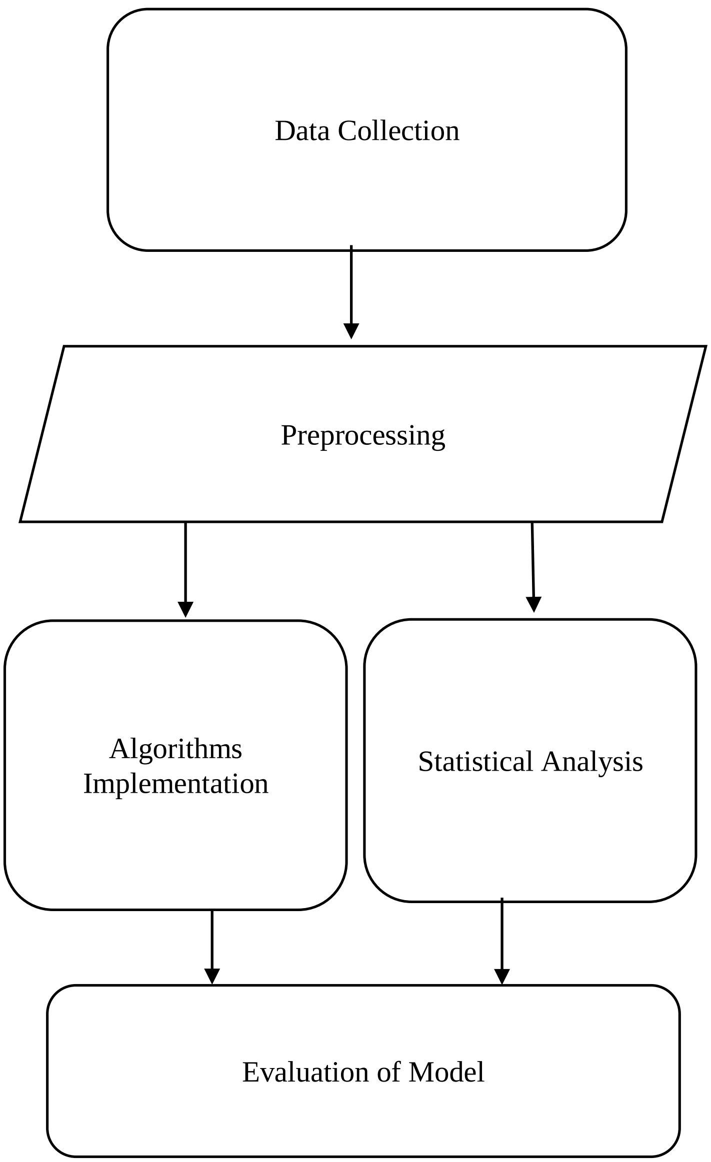



Figure 3.1.1 Methodology diagram

## 3.2 Data Collection

We have surveyed and collected almost 1200 data. This survey was performed only for Women. We selected Name, Age, Location are common for all and the parameters that are responsible for occurring diseases of newborn babies. Our work is not only predicting diseases but also a statistical analysis of these parameters that cause these types of diseases. So each of our selected attributes are very important for our work.

So we carefully collected these parameters by questionnaire that is given bellow:

1. What is the age of the child?
2. Whether the baby was infected with any disease after birth?
3. If yes, what is the name of the disease?
4. At what age did the children get the disease?
5. At what age was the mother of the child married?
6. At what age did the mother give birth to the child?
7. Whether the mother was suffering from any disease?

## 3.3 Preprocessing

1. Data cleaning: Name and Location are included in our initial data collection, but they are not needed for training our model. As a result, we combine all of the names and positions from the main dataset.
2. Transformation: To obtain high-quality information, data transformation is an integral component of any dataset. [1] Since machine learning algorithms do not understand strings, all strings must be converted to numbers. We used 1 for yes and 0 for no in this case.

## 3.4 Algorithm Implementation

TABLE 3.5.1. PARAMETER USAGE

| Algorithms | Details |
|---|---|
| Naïve Bayes Classifier | Random_state = 1, classifier = GaussianNB |
| Neural Network | Classifier= MLPClassifier, activation = relu, solver= adam, max_itter = 1000, alpha = 1 |
| SVM | Kernel=linear |
| Random Forest | n_estimators=100 |

Implementing algorithms, we attained Random Forest produced the highest 96.17% accuracy by using 30% data usage rate. The other three algorithms also accomplished very well. As Random Forest provided the best efficiency, we determined to apply this algorithm to forecast the child diseases.

## 3.5 Statistical Analysis

In this portion we actually tried to find out the causes of child disease of new born baby. Is it really depending on mother health or daily routine? For this purpose, we focused on our independent attribute.

Age: We surveyed between the ages of 14 to 30 and we collected the age when they have given birth to a child.

### 3.5.1 Disease vs. Delivery Age

This graph represents the age of mother’s delivery (X-axis) and number of affected children (Y-Axis). Here, we can see, which mother’s delivered their children at the age of 16, the numbers of affected children are more than 160+. The number of affected children is reduced with age but there are also some exceptions.

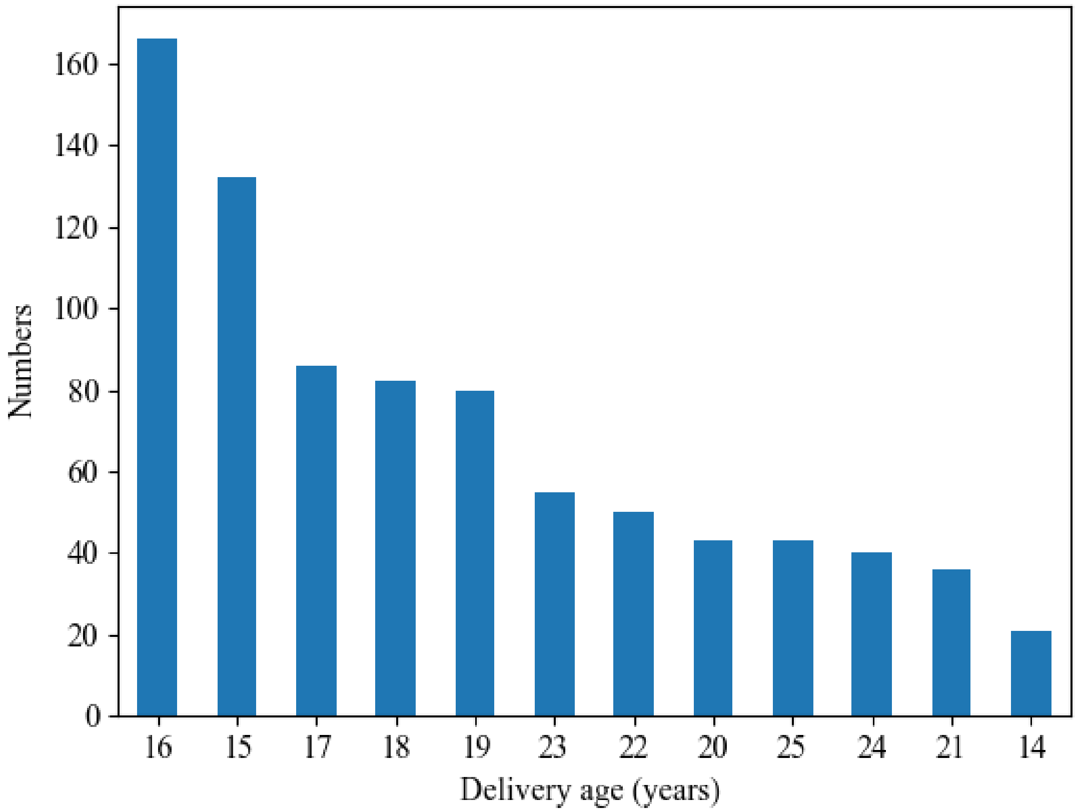



Fig. 3.5.1. Disease vs. Delivery Age (years)

### 3.5.2 Disease Percentage

In this research, there are 4 categories of diseases. They are- Chickenpox, Malaria, Diarrhea and no Disease (who are not affected). In our total 1200 data, the percentage of Chickenpox affected are 37.4%, 18.5% for Malaria. 12.5% for Diarrhea and 31.6% were with no disease. This is the representation of our dataset, which shows which disease affected children with what percentage.

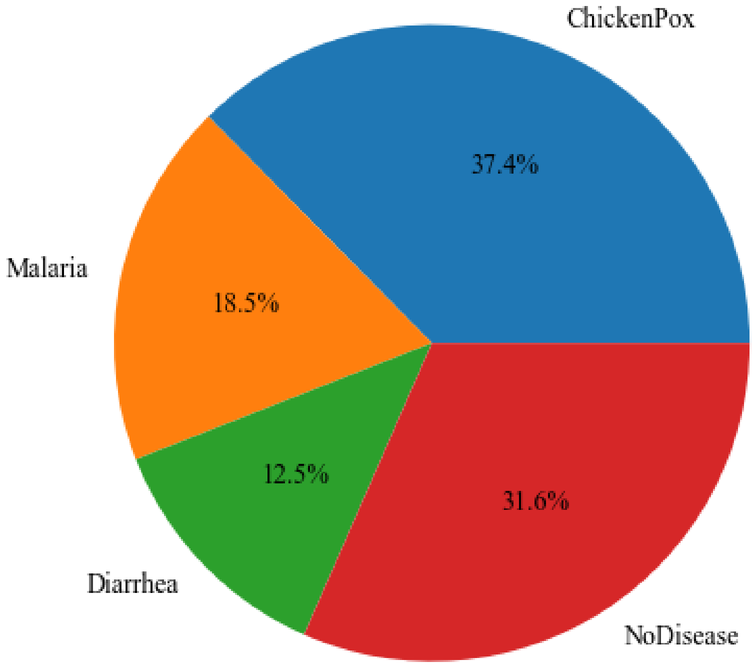



Fig. 3.5.2. Disease Percentage

### 3.5.3 Disease vs. Age

Here, the clue color represents the age of 0-3 years of children, orange represents the age of 4-6 years of children and green represents the age of 7-9 years of children. In the graph, we can see the number of chickenpox is more than 200 for the age of 0-3 years of children, then it slightly decreased for 4-6 years of children and it is reduced more for 7-9 years of children.
On the other hand, the scenario of Malaria Diseases is almost same like chickenpox. Here, we can see, the percentage is also larger for 0-3 years of children, then 4-6 and 7-9 are decreasing. For Diarrhea, The numbers are lower than other 2 diseases. But here, also the number for 0-3 years of children are bigger than 4-6 and 7-9 years of children.

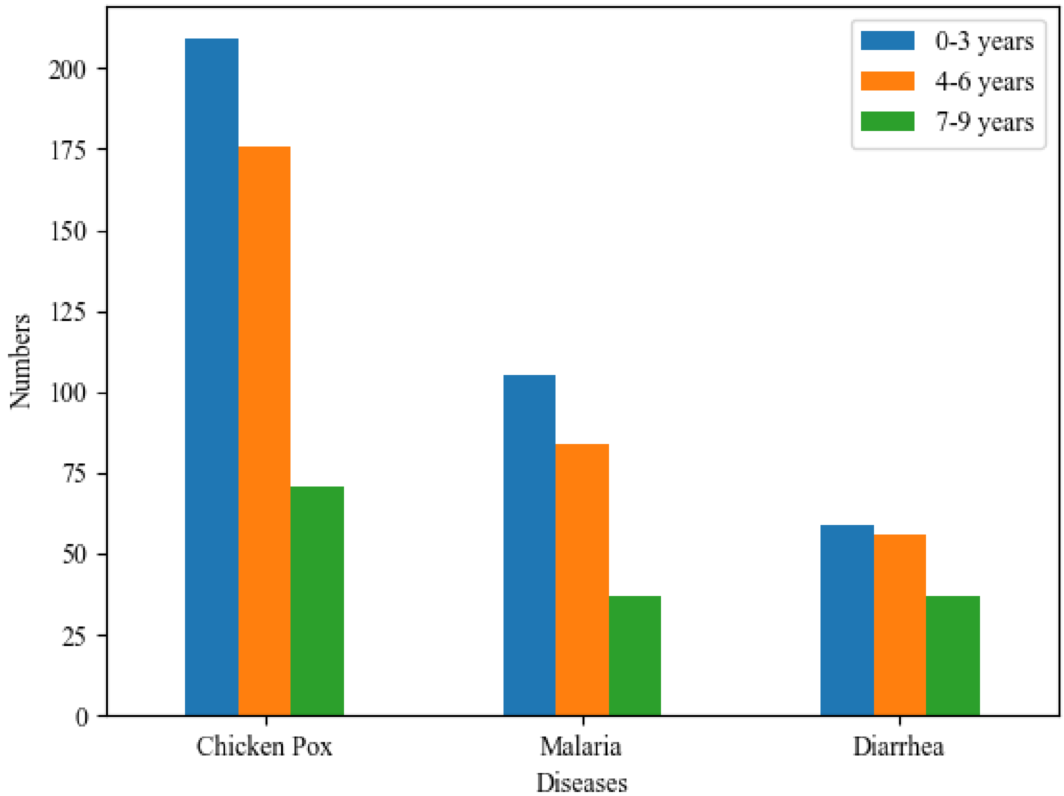



Fig. 3.5.3. Disease vs. Age (years)

### 3.5.4 Disease Yes vs. Affected Age

This graph shows at which years of age they get affected. Here, 0 represents 0-3 years old of children, 1 represents 4-6 years old of children and 2 represents 7-9 years old of children. When they are 0-3 years of age, the probability of getting affected is 44.7%, at the age of 4-6 years of old the percentage is 37.9% and at the age of 7-9 years of old the probability is 17.4%.

So it can be said that the percentage of getting affected decreases with age.

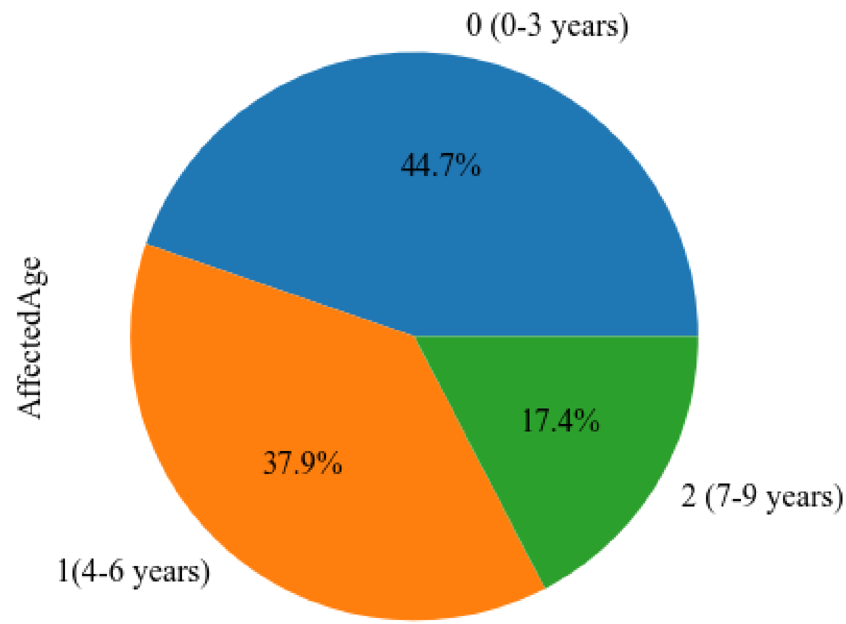



Fig. 3.5.4. Disease Yes vs. Affected Age (years)

### 3.5.5 Child disease vs. Disease During mother’s pregnancy

The graph represents the percentage of children getting affected when their mothers’ are sick during pregnancy.

Here, 1 represents the percentage of getting affected by the children is 70.4% when their mothers’ were sick at the time of pregnancy. And 0 represents that only 29.6% of children are healthy and without any disease though their mothers’ were affected with different diseases.

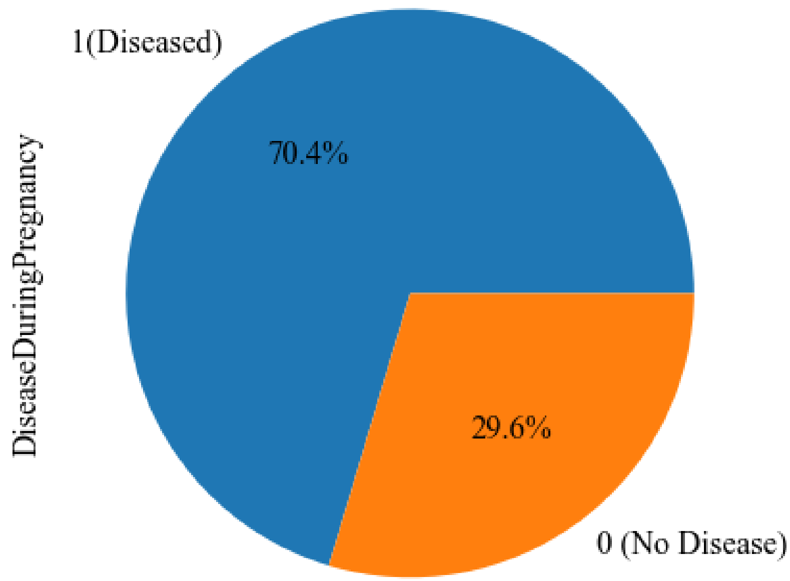



Fig. 3.5.5. Child disease vs. Disease During mother's pregnancy

## 3.6 Evaluation

For our research, at first, we assembled our dataset from an expedient source. Then we estimated our data by employing four ML algorithms to see which algorithm provides the best efficiency.

# CHAPTER 4

## Result comparison and analysis

To predict child disease, we have applied several Machine Learning algorithms. We have used Naive Bayes classifier, Support Vector Machine (SVM), Neural Network algorithm, and Random Forest for prediction. On the basis of accuracy and F1 score, we compared these algorithms. From the accuracy table, for 30% data usage rate, we got 85.25% for SVM, 96.17% for Random Forest, for 40% data usage rate, we got 92.01% for Neural Network, and for 70% data usage rate, we got 74.94% accuracy in Naïve Bayes.

Table 4.1.1 ACCURACY TABLE

| Data usage rate | Algorithms | | | |
|---|---|---|---|---|
| | Naive Bayes | SVM | Neural Network | Random Forest |
| 30% | 70.22% | 85.25% | 89.89% | 96.17% |
| 40% | 71.11% | 84.43% | 92.01% | 95.29% |
| 50% | 72.13% | 81.15% | 73.11% | 95.25% |
| 60% | 71.86% | 82.24% | 82.92% | 95.08% |
| 70% | 74.94% | 81.50% | 82.55% | 93.91% |

Table 4.1.2 describes the F1-score matrix table, for 30% data usage rate, we got 84.26% for SVM, 95.67% for Random Forest, by using 40% data usage rate, we got 82.20% for Neural Network, and for 70% data usage rate, and we got 72.76% accuracy in Naïve Bayes. Considering results from all the algorithms, we chose the Random Forest for its outstanding performance.

TABLE 4.1.2. F1-SCORE MATRIX TABLE

| Data usage rate | Algorithms | | | |
|---|---|---|---|---|
| | Naive Bayes | SVM | Neural Network | Random Forest |
| 30% | 68.31% | 84.26% | 81.74% | 95.67% |
| 40% | 69.35% | 83.50% | 82.20% | 94.63% |
| 50% | 70.55% | 81.31% | 79.84% | 95.38% |
| 60% | 70.26% | 81.26% | 81.00% | 95.21% |
| 70% | 72.76% | 80.28% | 80.28% | 93.71% |

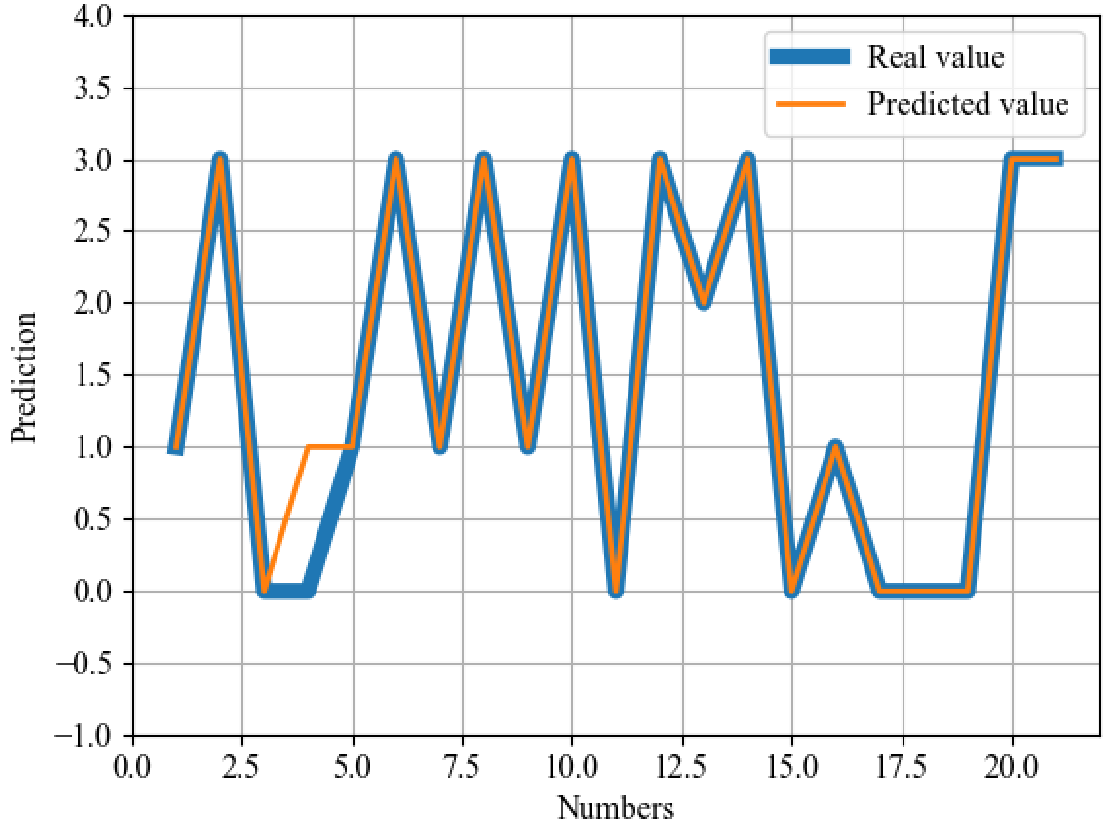



Fig. 4.1.1. Comparison between real and predicted class

Fig. 4.1.1 Illustrates, the real value is shown in blue color, while the predicted value is shown in orange color. In the overwhelming majority of cases, our expected value was similar to the actual value. Only for 1 case, our predicted values turned out to be wrong. In this case, it showed 1 rather than 0 and faded away from the original value line.

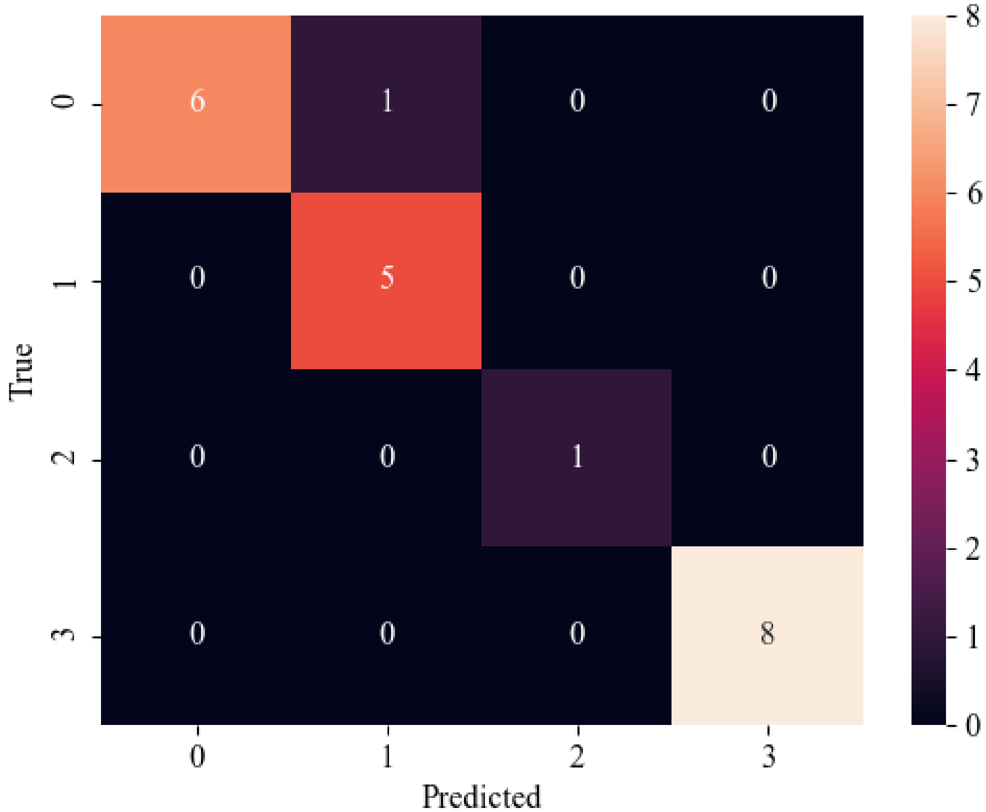



Fig. 4.1.2. Confusion Matrix

We can see in the confusion matrix that 0 number disease (chickenpox) has predicted 6 patients which is completely same with our dataset. Our algorithm has predicted total 6 numbers of 1(Malaria), among them 5 numbers are absolutely correct which is the same as our dataset and predicted 1 number as 1 which was 0 in our dataset. Which is only an error of our prediction, which is negligible. In number 2(Diarrhea), the machine has predicted only 1 malaria, which was also in our dataset. So the prediction was 100% accurate for malaria. Our machine has predicted a total 8 number of women who have no disease, which is also similar to our dataset. So the prediction is also accurate for no-disease patients.

# CHAPTER 5

## Conclusion and Future work

Bangladesh has one of the world's highest rates of infant mortality. Every newborn is the world's future. It is painful when your baby is unwell. It is our responsibility to ensure their safety. Babies are most vulnerable just after birth. They could be born with such congenital disorders or be susceptible to infections. If you're a new parent, you should be especially cautious about health issues that infants have as well as how to overcome them so that you can provide the correct healthcare for your kid. We would have achieved tremendous success if we could build a system that would reduce the incidence of baby diseases. To achieve the best outcome, we manipulated four common classification ML algorithms in our proposed model. After implementing these algorithms to the collection of data, we found that the algorithms' resolution was not the same. The Random Forest algorithm provided the best results in terms of accuracy. In comparison, the Random Forest had the highest accuracy, which we used for disease prediction in newborn babies and statistical analysis of maternal health. Child mortality should be reduced as much as possible. This work will help the government raise public awareness.

We concentrated on how to achieve the best possible outcome, although there are still a few more hurdles in our way. One of several primary drawbacks of our survey is that we did not survey the entire country. In Chittagong, Sylhet, Barisal, and Patuakhali, we surveyed nine villages. We will develop an intelligent system in the future that will predict child disease prediction and intelligently reduce the death rate of newborn babies. People would know what to do and what not to do after giving birth if they use this method. We will also seek to enhance our data collection by gathering data from around the country and developing an Android application that would be simpler when using.

## 5.1 Limitation

We tried our hardest to achieve the best possible result, but there are still a few roadblocks in our way. The insufficient amount of data is one of our work's major limitations. Our accumulated information was limited to a few locations. In the future, we'll try to improve our data collection by combining data from around the country and covering a longer period of time.

# APPENDIX

## Appendix A: Data collected from Survey.

We faced numerous challenges while conducting the analysis, the first of which was determining the methodological methodology for our study. It wasn't typical work, and there hadn't been much done in this field before. As a result, we were unable to receive much assistance from any source. We also began manually collecting data. We could do it after a long period of hard work.

# REFERENCE

[1] Mahmud Hossain Opu 'Report: Bangladesh cuts child death rate by 63% over 20 years' https://www.dhakatribune.com/health Mar 23, 2021

[2] Sharif N, Nobel NU, Sakib N, et al. Molecular and Epidemiologic Analysis of Diarrheal Pathogens in Children With Acute Gastroenteritis in Bangladesh During 2014-2019. The Pediatric Infectious Disease Journal. 2020 Jul;39(7):580-585. DOI: 10.1097/inf.0000000000002637.

[3] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130.

[4] O. Altay and M. Ulas, "Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, Turkey, 2018, pp. 1-4, doi: 10.1109/ISDFS.2018.8355354.

[5] K. Kosasih, U. R. Abeyratne, V. Swarnkar and R. Triasih, "Wavelet Augmented Cough Analysis for Rapid Childhood Pneumonia Diagnosis," in IEEE Transactions on Biomedical Engineering, vol. 62, no. 4, pp. 1185-1194, April 2015, doi: 10.1109/TBME.2014.2381214.

[6] M. M. Faizin, Subiyanto and U. M. Arief, "An intelligent support system for diagnosing dehydration in children," 2017 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Malang, Indonesia, 2017, pp. 108-112, doi: 10.1109/ICEEIE.2017.8328772.

[7] M. U. Khan, S. Aziz, S. Ibraheem, A. Butt and H. Shahid, "Characterization of Term and Preterm Deliveries using Electrohysterograms Signatures," 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2019, pp. 0899-0905, doi: 10.1109/IEMCON.2019.8936292.

[8] M. E. H. Chowdhury et al., "Can AI Help in Screening Viral and COVID-19 Pneumonia?," in IEEE Access, vol. 8, pp. 132665-132676, 2020, doi: 10.1109/ACCESS.2020.3010287.

[9] L. Malihi, K. Ansari-Asl and A. Behbahani, "Malaria parasite detection in giemsa-stained blood cell images," 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP), Zanjan, Iran, 2013, pp. 360-365, doi: 10.1109/IranianMVIP.2013.6780011.

[10] M. A. M. Ali, N. Md Tahir and A. I. Ali, "Monitoring Healthcare System for Infants: A Review," 2018 IEEE Conference on Systems, Process and Control (ICSPC), Melaka, Malaysia, 2018, pp. 44-47, doi: 10.1109/SPC.2018.8704143.

[11] M. Mehedi Hasan, M. Omar Faruk, B. Biswas Biki, M. Riajuliislam, K. Alam and S. Farjana Shetu, "Prediction of Pneumonia Disease of Newborn Baby Based on Statistical Analysis of Maternal Condition Using Machine Learning Approach," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 919-924, doi: 10.1109/Confluence51648.2021.9377169.

[12] Levine, Adam C., Richard M. Munyaneza, Justin Glavis-Bloom, Vanessa Redditt, Hannah C. Cockrell, Bantu Kalimba, Valentin Kabemba et al. "Prediction of severe disease in children with diarrhea in a resource-limited setting." PLoS One 8, no. 12 (2013): e82386.

[13] Olayinka, T. C., and S. C. Chiemeke. "Predicting paediatric malaria occurrence using classification algorithm in data mining." Journal of Advances in Mathematics and Computer Science (2019): 1-10.

[14] Crawley, Jane, Cindy Chu, George Mtove, and François Nosten. "Malaria in children." The Lancet 375, no. 9724 (2010): 1468-1481.

[15] Bevis, Douglas Charles Aitchison. "The antenatal prediction of haemolytic disease of the newborn." Lancet (1952): 395-8.

# PLAGIARISM REPORT

Prediction of Child Disease

ORIGINALITY REPORT

| 13% | 7% | 7% | 6% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Daffodil International University<br>Student Paper | 4% |
| 2 | Muhammad Umar Khan, Sumair Aziz, Sara Ibraheem, Aqsa Butt, Hira Shahid. "Characterization of Term and Preterm Deliveries using Electrohysterograms Signatures", 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2019<br>Publication | 2% |
| 3 | Leila Malihi, Karim Ansari-Asl, Abdolamir Behbahani. "Malaria parasite detection in giemsa-stained blood cell images", 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP), 2013<br>Publication | 1% |
| 4 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | 1% |
| 5 | ieeexplore.ieee.org<br>Internet Source | 1% |