A NOVEL APPROACH FOR DIFFERENTIAL DIAGNOSIS OF ACUTE FLACCID PARALYSIS (AFP) VIA DATA SCIENCE

Author

Uzair Rasheed

00000319186

MS-19 (CSE)

Supervisor

Dr. Usman Qamar

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL AND MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

SEPTEMBER, 2021

A NOVEL APPROACH FOR DIFFERENTIAL DIAGNOSIS OF ACUTE FLACCID PARALYSIS (AFP) VIA DATA SCIENCE

Author

Uzair Rasheed

00000319186

A thesis submitted in partial fulfillment of the requirements for the degree of MS Computer Software Engineering

Thesis Supervisor:

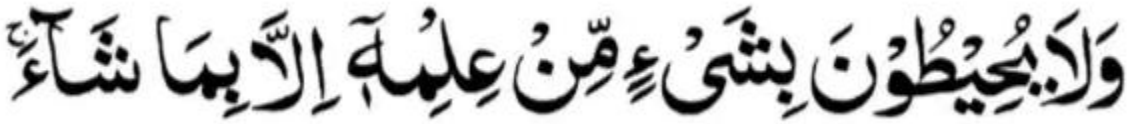
Dr. Usman Qamar

Thesis Supervisor's Signature: _____

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD

SEPTEMBER 2021

In the name of Allah most beneficent most merciful



And they can't encompass anything from His knowledge, but to extend He wills [2:255]

Declaration

I certify that this research work, titled "*A Novel Approach for Differential Diagnosis of Acute Flaccid Paralysis (AFP) via Data Science"* is my own work. This work is not presented elsewhere for assessment. The material used from other sources are properly acknowledged/ referred.

Signature of Student

Uzair Rasheed

2021-NUST-Ms-Soft-19

Language Correctness Certificate

This thesis has been read by an English expert and is free of typing, syntax, semantic, grammatical and spelling mistakes. The thesis is also according to the format given by the university.

Signature of Student

Uzair Rasheed

Registration Number

00000319186

Signature of Supervisor

Dr. Usman Qamar

Copyright Statement

Acknowledgements

I am thankful to Allah Almighty for giving me countless blessings and guidance in this research work. I want to thank my supervisor Dr. Usman Qamar for encouraging and supporting me throughout my research. He is a source of great knowledge and his expertise helped a lot in achieving my goals. I also want to thank Dr. Wasi Haider Butt and Dr. Usman Akram for their unlimited guidance and support. I want to thank my mother, siblings, and friends for encouraging me throughout my degree program and motivating me to overcome obstacles and difficulties. I would like to express my gratitude to the Department of Computer and Software Engineering and the management of College of Electrical and Mechanical Engineering for assisting me throughout my journey.

Dedicated

to a courageous and strong woman,

my mother

Abstract

Acute Flaccid Paralysis (AFP) is caused due to diseases like Poliomyelitis, Acute Transverse Myelitis (ATM), Guillain-Barre-Syndrome (GBS), Traumatic Neuritis (TrN), and others of similar nature. These diseases can damage various muscles or spinal cord, or peripheral nerves. Pakistan is among the few countries which are struggling to eradicate Poliomyelitis. The conventional diagnosis takes a lot of effort and time. A solution to this problem is provided in this research using the methods from Data Science. ***Objective:*** *Diagnosis of AFP is done via two approaches, used by the framework proposed in this research: 1) Prediction of diagnosis for non-polio AFP cases, using patient's dataset; 2) Prediction of clinical diagnosis for AFP, using the developed rule base data based on a standard chart, provided by the World Health Organization (WHO) for differential diagnosis of AFP.* ***Method:*** *1) The laboratory diagnosis method predicts the diagnosis of non-polio AFP cases using patient data. The data was preprocessed and converted into a dataset; which was used for model training and prediction of diagnosis for non-polio AFP. 2) The clinical diagnosis method is based on the symptoms given in the standard table of WHO. This table differentiates between diseases like Poliomyelitis, GBS, TrN, TM, and Acute Childhood Hemiplegia. A rule base was devised according to this standard diagnosis table, with the five diseases acting as labels of the dataset. The rules helped in diagnosis for a particular AFP case. Decision Tree and Random Forest are the two algorithms used for model training and prediction in this research. The trained model predicts the disease for a particular AFP case.* ***Results:*** *For the first method, the prediction accuracy was 87.76% with the decision tree algorithm and 92.46% with random forest. For the second dataset, using the rule base dataset, the accuracy with the decision tree algorithm was 98.54% and 100% with random forest.*

Keywords: *Acute Flaccid Paralysis, AFP, Differential Diagnosis of AFP, Polio, WHO, Data Science*

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION 15

1.1 MOTIVATION 17
1.2 PROBLEM STATEMENT 17
1.3 AIMS AND OBJECTIVES 18
1.4 OUTLINE 19
1.5 SUMMARY 19

CHAPTER 2 LITERATURE REVIEW 20

2.1 INTRODUCTION 20
2.2 DISEASES ASSOCIATED WITH AFP 20
2.2.1 POLIOMYELITIS 20
2.2.2 GUILLAIN-BARRÉ SYNDROME (GBS) 23
2.2.3 ACUTE TRANSVERSE MYELITIS (ATM) 24
2.2.4 TRAUMATIC NEURITIS (TRN) 26
2.2.5 HYPOKALEMIA 26
2.2.6 ACUTE CHILDHOOD HEMIPLEGIA 28
2.3 VIRUSES ASSOCIATED WITH AFP 29
2.3.1 POLIOVIRUS 29
2.3.2 VACCINE ASSOCIATED PARALYTIC POLIOMYELITIS (VAPP) 30
2.3.3 VDPV & CVDPV 31
2.3.4 ENTEROVIRUSES 32
2.3.5 NEUROTROPIC VIRUSES 33
2.3.6 ADENOVIRUSES 34
2.4 APF SURVEILLANCE 34
2.4.1 THE TWO KEY INDICATORS OF AFP SURVEILLANCE 35
2.4.2 AFP STATISTICS BY WHO (GLOBAL) 36
2.4.2.1 GLOBAL AFP CASES 38
2.4.2.2 GLOBAL POLIOVIRUS CASES 38
2.4.2.3 GLOBAL NON-POLIO AFP RATE 39
2.4.2.4 GLOBAL COMPATIBLE CASES 39
2.4.2.5 GLOBAL STOOL SAMPLE ADEQUACY 40
2.4.2.6 GLOBAL CASES OF CVDPV 40
2.4.2.7 AFP DATA (PAKISTAN) 41
2.4.2.8 AFP CASES IN PAKISTAN 42
2.4.2.9 POLIOVIRUS CASES IN PAKISTAN 42
2.4.2.10 NON-POLIO AFP RATE IN PAKISTAN 43
2.4.2.11 COMPATIBLE CASES IN PAKISTAN 43
2.4.2.12 STOOL SAMPLE ADEQUACY IN PAKISTAN 44
2.4.2.13 CVDPV CASES IN PAKISTAN 44
2.4.2.14 COMPARISON OF WILD POLIOVIRUS CASES (PAKISTAN VS. GLOBAL) 45
2.4.2.15 GRAPHICAL COMPARISON OF WILD POLIOVIRUS CASES (PAKISTAN VS. GLOBAL) 46
2.4.2.16 GRAPHICAL REPRESENTATION OF WILD POLIOVIRUS CASES PROPORTION 46
2.4.2.17 GRAPHICAL COMPARISON OF WILD POLIOVIRUS CASES (PAKISTAN VS. GLOBAL) 47
2.4.2.18 WILD POLIOVIRUS CASES DEVELOPMENT (PAKISTAN VS. GLOBAL) 47
2.5 WHO'S STANDARD DIAGNOSIS PROCEDURE FOR AFP 48
2.6 AFP CASE HANDLING 49
2.6.1 ORDER OF AFP DIFFERENTIAL DIAGNOSIS 51

2.6.2 TIME-FRAME FOR AFP DIAGNOSIS 53
2.7 SUMMARY 54

CHAPTER 3 METHODOLOGY 55

3.1 INTRODUCTION 55
3.2 AFP LINE LISTING DATA 55
3.3 DATASET 1: AFP LINE LISTING 59
3.4 WHO DIFFERENTIAL DIAGNOSIS TABLE 62
3.5 DATASET 2: AFP RULE BASE 66
3.6 INTRODUCTION TO MACHINE LEARNING 71
3.6.1 ALGORITHMS IN MACHINE LEARNING 73
3.6.1.1 DECISION TREE 73
3.6.1.2 RANDOM FOREST 74
3.7 METHODOLOGY 74
3.7.1 PROPOSED FRAMEWORK 74
3.8 IMPLEMENTATION 76
3.8.1 MODULE I: LABORATORY DIAGNOSIS 76
3.8.2 MODULE 2: CLINICAL DIAGNOSIS 78
3.9 SUMMARY 80

CHAPTER 4 RESULTS 81

4.1 INTRODUCTION 81
4.2 EVALUATION METRICS 81
4.2.1 ACCURACY 82
4.2.2 PRECISION 82
4.2.3 RECALL 82
4.2.4 F-MEASURE 83
4.3 RESULTS MODULE 1: LABORATORY DIAGNOSIS 83
4.4 RESULTS MODULE 2: CLINICAL DIAGNOSIS 86
4.5 SUMMARY 88

CHAPTER 5 CONCLUSIONS 90

5.1 INTRODUCTION 90
5.2 APPLICATIONS OF THIS RESEARCH 90
5.3 CONCLUSION 90

REFERENCES 93

LIST OF EQUATIONS

EQUATION 1 NON-POLIO AFP RATE 35
EQUATION 2 ACCURACY 82
EQUATION 3 PRECISION 82
EQUATION 4 RECALL 82
EQUATION 5 F-MEASURE 83

LIST OF FIGURES

FIGURE 1 GLOBAL AFP CASES [28] 38
FIGURE 2 GLOBAL WILD POLIOVIRUS CASES [28] 38
FIGURE 3 GLOBAL NON-POLIO AFP RATE [28] 39
FIGURE 4 GLOBAL COMPATIBLE CASES [28] 39
FIGURE 5 GLOBAL STOOL SAMPLE ADEQUACY [28] 40
FIGURE 6 GLOBAL CASES OF CVDPV [28] 40
FIGURE 7 AFP CASES IN PAKISTAN [28] 42
FIGURE 8 WILD POLIOVIRUS CASES IN PAKISTAN [28] 42
FIGURE 9 NON-POLIO AFP RATE IN PAKISTAN [28] 43
FIGURE 10 COMPATIBLE CASES IN PAKISTAN [28] 43
FIGURE 11 STOOL SAMPLE ADEQUACY IN PAKISTAN [28] 44
FIGURE 12 CPDPV CASES IN PAKISTAN [28] 44
FIGURE 13 COMPARISON OF WILD POLIOVIRUS CASES [28] 46
FIGURE 14 POLIOVIRUS CASES PAKISTAN [28] 46
FIGURE 15 POLIOVIRUS CASES (PAKISTAN VS. REST OF WORLD) [28] 47
FIGURE 16 WILD POLIOVIRUS CASES (2010-2020) [28] 47
FIGURE 17 CLASSIFICATION OF AFP CASES [2][15][17] 48
FIGURE 18 PROCEDURE FOR HANDLING AFP CASE [21] 49
FIGURE 19 FLOW DIAGRAM OF AFP CASE INVESTIGATION [2] 53
FIGURE 20 AFP LINE LISTING TABLE [17] 56
FIGURE 21 MACHINE LEARNING FRAMEWORK 72
FIGURE 22 PROPOSED FRAMEWORK 75
FIGURE 23 MODULE 1: LABORATORY DIAGNOSIS 76
FIGURE 24 MODULE 2: CLINICAL DIAGNOSIS 78

LIST OF TABLES

TABLE 1 AFP DATA, GLOBALLY [28] 36
TABLE 2 AFP DATA, PAKISTAN [28] 41
TABLE 3 WILD POLIOVIRUS CASES (PAKISTAN VS. GLOBAL) 45
TABLE 4 TABULAR CLASSIFICATION OF AFP CASES [2] 50
TABLE 5 AFP DIAGNOSIS CATEGORIES [2] 52
TABLE 6 PATIENT DATASET 59
TABLE 7 WHO'S DIFFERENTIAL DIAGNOSIS TABLE FOR AFP [16][17][21] 63
TABLE 8 RULE BASE DATASET 66
TABLE 9 CONFUSION MATRIX, DECISION TREE, MODULE 1 84
TABLE 10 CONFUSION MATRIX, RANDOM FOREST, MODULE 1 84
TABLE 11 MODULE 1: LABORATORY DIAGNOSIS RESULTS (A) 85
TABLE 12 MODULE 1: LABORATORY DIAGNOSIS RESULTS (B) 85
TABLE 13 CONFUSION MATRIX, DECISION TREE, MODULE 2 86
TABLE 14 CONFUSION MATRIX, RANDOM FOREST, MODULE 2 87
TABLE 15 MODULE 2: CLINICAL DIAGNOSIS RESULTS (A) 87
TABLE 16 MODULE 2: CLINICAL DIAGNOSIS RESULTS (B) 88

CHAPTER 1
INTRODUCTION

Being a combination of three words, Acute (rapid progression of paralysis) Flaccid (floppy; loss of muscle tone) Paralysis (weakness; loss; diminution); notably known by AFP, is a condition of disability in the human body. In other words, when sudden and progressive paralysis occurs in the human body and the muscle tone is slowly lost, specific organs of the human body are fully or partially disabled, known as AFP [1][2][3].

One of the reasons for AFP is virus attack; like the Wild Poliovirus, the Enterovirus, and echovirus [4][5][6][7][8][9][10][11][12][31]. These viruses usually attack the Central Nervous System (CNS) [3], the brain's grey matter [13], the spinal cord, and other parts of the body related to the nervous system of the human body [1]. Different organs get damaged by the attacks of these viruses; which cause multiple diseases. These diseases differ from each other on the basis of their symptoms. Every disease is responsible for a specific disorder in the body; such as, when the virus attacks the Anterior Horn Cell, it causes Enteroviral Myelitis [10], Poliomyelitis [13] and GBS [6][14][49]. Similarly, if the virus damages the spinal cord, it causes diseases like ATM, Ascending Myelitis, Traumatic Spinal Injury [3][9][13]. Furthermore, when the muscles are under attack by a virus, it causes Periodic Paralysis, Hypokalemia, and many other diseases; which are some major diseases in consideration of AFP [10][33][36].

Differential Diagnosis of AFP means to differentiate among all these diseases that cause AFP [36]. In other words, the goal of differential diagnosis of AFP is to correctly identify

the diseases which cause paralysis in a human body [21]. There is a standard procedure of WHO for the differential diagnosis of AFP [27]. In this procedure, two stool specimens of an AFP patient are collected for laboratory tests. The laboratory tests are performed to confirm the existence of Wild Poliovirus; the results can be interpreted in several ways. In case the Wild Poliovirus is found in the stool sample, the diagnosis is completed and the patient is confirmed as a Poliomyelitis victim. If there is no existence of Wild Poliovirus, the experts may discard the case, otherwise, the case can declared as Polio compatible [2][15][17].

In the field of medicine, data science has provided countless solutions in terms of disease diagnosis. A lot of work has been done using classifier algorithms; it includes correctly predicting the diagnosis of diseases by the trained models [18][19][70][71]. This research proposed a similar approach. The differential diagnosis of AFP is, currently, a manual process and it is being automated in this research. This process, in reality, is very complicated, but the data regarding the AFP cases or patients, generated by the manual approach, is used in this research by applying the concepts of data science. A framework is proposed, which uses the concepts of supervised learning to diagnose the diseases associated with AFP, correctly.

Supervised learning techniques have been used in disease diagnoses recently. Some of the well-known supervised learning techniques are Artificial Neural Network (ANN), Logistic Regression, Naïve Bayes, K-Nearest Neighbor (KNN), Decision Trees, and Random Forest [60]. In this research, the datasets used for model training had categorical data; hence, Decision Trees and Random Forest algorithms were used; as they are best suited for training

and prediction of such kind of data. Two datasets were used for two different types of diagnosis, using the selected algorithms.

1.1 MOTIVATION

According to WHO, Poliomyelitis mainly affects children below the age of 5. One out of two hundred infections has irreversible paralysis. Among these paralyzed people, 5% to 10% die due to disabled breathing muscles. There has been a 99% decrease in the cases of Wild Poliovirus since 1988; most of the countries have been successful in eradicating Poliomyelitis; but Pakistan, Afghanistan, and Nigeria are the countries still struggling to eradicate Poliomyelitis [20][22][24][37][38][39][40][41]. More than 18 million people walking today would have been disabled if the extensive Polio vaccination had not been done over the last three decades [42]. As long as a single child remains infected, children around the globe are at risk of contracting Poliomyelitis. If WHO fails to eradicate Poliomyelitis from these three countries, it could result in around 200,000 new cases annually within a decade [38]. Since the threat of Poliomyelitis persists, especially in Pakistan [39][40], new methods should be initiated, which can help to eradicate Poliomyelitis. Differential diagnosis of AFP is the primary concern of WHO in the eradication of Poliomyelitis. Therefore, this research aims to propose a novel automated solution for the differential diagnosis of AFP, using the concepts of data science.

1.2 PROBLEM STATEMENT

The manual process for the diagnosis of AFP is complex, steady, and lengthy. It takes a lot of time and effort to diagnose a patient with AFP. Also, the precise diagnosis of

Poliomyelitis, in the manual procedure, is possible only when there are adequate stool samples and when the Wild Poliovirus is detected; otherwise, in most cases, it takes months to diagnosis a particular patient of AFP via the manual process.

1.3 AIMS AND OBJECTIVES

This research aims to propose a framework to automate the now-manual process for the differential diagnosis of AFP. The proposed framework incorporates two modules:

1. The first module is termed as "Laboratory Diagnosis"; as the dataset involves laboratory examination results
2. The second module is termed as "Clinical Diagnosis"; as the dataset is derived from the differential diagnosis table provided by WHO. This table includes clinical questions regarding the diseases associated with AFP.

The following are the primary objectives of the research associated with each module:

1. Laboratory Diagnosis
    a. To use the AFP patient data as a dataset for model training
    b. To predict the differential diagnosis of AFP by classifiers, trained using the AFP patient dataset
2. Clinical Diagnosis
    a. To convert the standard table for differential diagnosis of AFP into a dataset
    b. To predict the differential diagnosis of AFP by classifiers, trained using the AFP rule base dataset

1.4 OUTLINE

Following is the order of the chapters in this research and their context breakdown:

2. Literature Review: diseases and viruses associated with AFP; surveillance of AFP; statistics regarding AFP; diagnosis procedure of AFP; machine learning techniques.
3. Methodology: proposed framework; details of two datasets involved in the proposed framework and its implementation
4. Results: evaluation criteria; results
5. Conclusions

1.5 SUMMARY

A novel automated framework for the differential diagnosis of AFP is introduced in this research. This proposed framework is based on the concepts of artificial intelligence. The manual process of the differential diagnosis has limitations; it is stagnant and costs a lot of time and effort. Machine learning algorithms are utilized in the proposed framework; which provide an efficient and rapid solution for the diagnosis of AFP. The algorithms selected for training and validation of the models are Decision Tree and Random Forest. The proposed framework has two parts: laboratory diagnosis and clinical diagnosis; each sub-portion work on a different dataset.

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

This chapter presents the literature available on various aspects of AFP. It explains diseases associated with AFP; followed by history, symptoms, etiology, and causes of each in detail. It also entails the differences among diseases, the history of breakouts caused by each disease, and the vaccines associated with AFP. Lastly, WHO's standard procedure for the differential diagnosis of AFP is also discussed here.

2.2 DISEASES ASSOCIATED WITH AFP

2.2.1 POLIOMYELITIS

The two Greek words “Polio” and “myelon” makes up the word Poliomyelitis, which means “grey marrow”. The Poliovirus is 25 to 30 nanometers in diameter. The external cover is made up of 60 protomers; each is formed by virion proteins VP1, VP2, VP3 and VP4. All the virions are composed of 8 protein strands. Three stereotypes of Polio type 1, 2, and 3 have been recognized so far. When Poliovirus is carried to a human mouth, it reaches the oropharynx and starts multiplying from tonsils to the small intestine. The incubation time of this virus varies from 2 to 35 days. The Wild Poliovirus spreads when it sheds in stool after 3 to 5 days of entering a human digestive system [47].

Poliovirus is a prototypical enterovirus which spreads with fecal excretions and pharyngeal secretions, mostly by hand to hand and hand to mouth. Only humans can carry the

Poliovirus in the digestive route of the body. Out of the total infected people, 79% were paralyzed due to Wild Poliovirus type 1, 8% through type 2, while the remaining 13% of the total cases were victims of Wild Poliovirus type 3 [44].

At the end of the 19th century, Poliomyelitis appeared as an epidemic in some of the European and North American countries and then, this disease went global. There was no cure in the first half of the 20th century, but after the formation of the Poliovirus vaccine in 1955, a stepwise, gradual decline started in the cases of Poliomyelitis. The United Stated eliminated the Wild Poliovirus in 1972. The elimination protocol was followed by many countries and the cases dropped from 600,000 annually to only a thousand in the year 2000. In the year 1999, the native type 2 Wild Poliovirus was also eradicated but the circulation of Wild Poliovirus type 1 and 3 in still continuous in 4 countries of Africa and Asia [44].

Poliovirus is a prototypical enterovirus, which spreads with the excretion in feces and in pharyngeal secretions, mostly by hand to hand and hand to mouth. Only humans can carry the Poliovirus in the digestive route of the body. 79% of the people paralyzed due to Wild Poliovirus type 1, 8% through type 2 while the remaining 13% of the total cases were victims of Wild Poliovirus type 3 [44].

Poliomyelitis mostly affects children. After the onset of paralysis, which lasts a few days, the recovery phase can take months; including treatment of respiratory problems and orthopedic problems. In the initial days of Poliomyelitis, the target of the virus is motor neuron of anterior horn cell of brain stem and spinal cord. It leads to muscular paralysis, where some cases might be more severe than others. After the initial onset, the severity level

of the paralysis increases within 48 hours. Sometimes the patients feel discomfort in respiration and sometimes it affects the bulbar. The infection phase is very fast as compared to the recovery phase [46].

Currently polymerase chain reaction (PCR) is used for the diagnosis of Wild Poliovirus. Throat swabs, blood samples, stool samples and cerebrospinal fluid (CSF) are used for diagnosis through PCR. Primarily, the stool sample of the patient is used extensively as a source for PCR. The virus is excreted by the patient, 1 to 2 months after the day of infection; mostly in the first two weeks; and from the third week onwards, the excretion ratio declines. This is the reason why it is necessary to collect two stool samples with 14 days of infection. The stool samples are collected twice in a gap of 24 hours. Collecting the stool samples early increases the chances for the isolation of Wild Poliovirus. Isolating the virus with throat swab is only possible at initial times of the infection because it is present in the oropharynx in the early times. The CSF is used for isolation of virus only in aseptic meningitis cases, while isolation of Wild Poliovirus through blood sample is conducted from 3 to 5 days. All these methods are not equally significant for diagnosis of Poliomyelitis [47].

Herd immunity is developed up to 95% in the people who receive the Oral Polio Vaccine (OPV); which was licensed in 1962. It was developed after a lot of hard work by Salk and Sabin, who conducted trials on pets, school children, and their relatives [46]. OPV carry live-attenuated Poliovirus strains (type 1, 2 and 3). This vaccine has been used for the eradication of Poliomyelitis as a basic tool globally. The OPV helps the recipient of the OPV dose to develop immunity from 4 to 6 days. The OPV can sometimes affect the CNS

and can cause paralysis similar to Poliomyelitis caused by the Wild Poliovirus; also known as vaccine associated paralytic Poliomyelitis (VAPP) [48].

Initially, VAPP affected people in Belarus and United States from 13% to 55% cases. Based on the assumption of 1 VAPP case per 2-4 million births, 250-500 cases were estimated by WHO in 2002. After the withdrawal of Sabin type 2 vaccine virus OP2, and the introduction of inactivated Poliovirus vaccines (IPV), in November 2012, all the countries were recommended to use a minimum of 1 IPV dose in the basic immunization program along with the OPV. Furthermore, studies show that VAPP is eliminated in countries where infants received several doses of IPV before receiving doses of OPV. But the countries that followed the Global Polio Immunization schedule of receiving an IPV dose along with 3 OPV dose before the age of 14 weeks, did not shown any significant signs of mitigating the risk of VAPP [48].

2.2.2 GUILLAIN-BARRÉ SYNDROME (GBS)

Apart from Poliomyelitis, GBS is the most frequent one among the other causes of AFP. It is so severe that around 20% of GBS patients remain disabled, while around 5% of the patients die. It affects the cerebrospinal fluid of humans. Its variant, Miller Fisher syndrome, can cause disorders in the peripheral nerve, which results in weakness in limbs [6][33][49].

The most frequent cause of AFP is acute inflammatory plyneuropathy GBS [32][33]. It is a very common and severe type of acute paralytic neuropathy. Annually, about 100,000 people are affected by this disease around the world. GBS have many variants, some being more severe than others. All these variants have different features in terms of clinical

analysis and pathology. 20% to 30% of the cases of GBS are considered most severe as they cause respiratory failures. Exchange of plasma or intravenous immunoglobulin is considered as the best possible treatment for a patient with other sorts of care in parallel as well [50].

GBS have four main subcategories, the Acute Inflammatory Demyelinating Polyradiculoneuropathy (AIDP), the Acute Motor Axonal Neuropathy (AMAN), Acute Motor Sensory Axonal Neuropathy (AMSAN), and Acute Sensory Neuropathy (ASN). The antibodies of AIDP are still unknown, while among the other subtypes of GBS, AMAN has GM1, GM1b, and GD1a antibodies; AMSAN has GM1, GM1b, GD1a and GalNac-GD1a; ASN has GD1b antibodies. Among the variants of GBS, Oropharyngeal has GT1a, while the variant Fisher's syndrome has antibodies GT1a and GQ1b [51].

2.2.3 ACUTE TRANSVERSE MYELITIS (ATM)

The acute inflammation causes acute spinal cord injury which results in neurological disorder in a human body. It may be acute or a slow disorder. In the case of ATM, it takes only few hours to develop symptoms which can get worse in days to come. The situation of patients usually gets worse in a few weeks. About 33% of the patients face severe disabilities, 33% have moderate residual disabilities, while 33% recover from ATM [52].

The differential diagnosis of ATM is extensive. A very smart approach is required for the diagnosis of ATM, doctors must be conscious about the possible etiologies, and their process should be very efficient and orderly. The approach for the diagnosis of ATM has to be cost-effective and according to the history, clinical examination, immunological findings, and magnetic resonance imaging (MRI) of the patient; MRI is the radiological approach for

getting images of the brain and the spinal cord. These things are necessary as the diagnosis differentiates etiologies like demyelination, neoplastic, paraneoplastic, infection, and vascular [53].

The basic difference between myelitis and myelopathy is that myelitis in the disease due to spinal cord inflammation, while myelopathy is an extensive word for bruise or lesion which affects the spinal cord. The next term is “transverse”, which makes the myelitis a bit different which categorizes the disease in acute and sub-acute disorder in the spinal cord that causes weakness in muscles and involuntary impairment beneath the lesion [36][54].

ATM can be further categorized in types [54]:

1. Acute Complete Transverse Myelitis (ACTM): ACTM can weaken the upper and lower sides of the body and can cause sensory disorders; defined by sensory levels. It can also cause the involuntary reduction beneath the lesion level. While analyzing the MRI in most of the cases, there is a single bruise between 1 or 2 segments of the backbone and on axial sections. It can affect the spinal cord partially in the center or with full thickness.
2. Acute Partial Transverse Myelitis (APTM): APTM causes asymmetric neurologic impairment, which is limited to spinal cord. It affects the anatomic tract between 2 segments of the backbone, which disturbs the spinal cord on axial sections.
3. Longitudinally-Extensive Transverse Myelitis (LETM): This affects 3 or more segments of the backbone. It usually targets the center of the spinal cord and sometimes two-third area of the spinal cord.

4. Secondary Transverse Myelitis (STM): this is where the inflammatory autoimmune disorder happens; such as lupus and Sjögrenr syndrome. This category is a subcategory of ACTM.
5. Idiopathic Transverse Myelitis (ITM): it does not have a definite etiology but with clinical findings and investigations, a patient can be declared as the case of ITM.

2.2.4 TRAUMATIC NEURITIS (TRN)

Traumatic Neuritis is a type of AFP where one of the two limbs of the patient is paralyzed. When a patient is reported as a case of AFP and only one limb is affected by paralysis, then the patient is suspected as a case of TrN; if there is a history of an injection with the patient 24 hours before the onset of the paralysis, [55].

Symptoms of TrN include hypothermia (very low body temperature) and severe pain in the affected limb of the patient. In the diagnosis of AFP, it is difficult to differentiate among TrN and Poliomyelitis. However, the lack of CSF pleocytosis (an increase in cell count) and sensory deficits hints at the diagnosis of TrN. Some of the cases of Poliomyelitis are considered as TrN, which is a misdiagnosis and it is probable. Moreover, the residual sensory deficits hint at the diagnosis of injection neuritis, which is another subject of AFP [55].

2.2.5 HYPOKALEMIA

The next disease associated with AFP is Hypokalemia. Hypokalemia is among the diseases diagnosed in the differential diagnosis process of AFP [10][30][36]. In a human body, when the levels of the potassium serum drop beyond normal, it causes Hypokalemia. Among the

diseases where the body electrolyte is disturbed, Hypokalemia is a very common, especially if the patients are hospitalized. In some cases, it turns to be a serious issue and emergency medical care is needed. A low consumption of potassium or high expulsion of potassium from the body can be the common causes of Hypokalemia [56].

The normal range of Potassium is 3.5 to 5mEq/L, while a level of potassium serum less than 2.5mEq/L can be dangerous to a human life and it is the severe form of Hypokalemia. 14% of the patients which are not hospitalized, but engaging in laboratory tests, have a moderate level of Hypokalemia; while 20% of the hospitalized patients have moderate Hypokalemia. Only 4-5% of the hospitalized patients have remarkably low levels of potassium serum, a severe form of Hypokalemia, which means it is not very common. The severity of Hypokalemia in those patients reveal the possibility of some systematic diseases in them. Hypokalemia is equally distributed among males and females without any significant distinction [56].

With the use of diuretic drugs, potassium is excreted in excess in the urine and can cause disorders in kidney. Along with excretion through urine, potassium is lost through gastrointestinal way; vomiting or diarrhea, when prolonged, can be the reasons which cause infections in intestine. Thus, the shift of potassium between the cells becomes the cause of severe Hypokalemia. The other important factor in the severe Hypokalemia patients is the deficiency of concomitant magnesium, frequently observed in more than half of the clinically significant Hypokalemia [56].

Hypokalemia which can be linked to various outcomes [56]:

- In the nervous system, it can cause Ascending Paralysis, leg cramps and Paresis (weakness in muscles).
- In the renal system, it causes Hypokalemic kidney disease, Rhabdomyolosis, and Metabolic Acidosis.
- In the gastrointestinal system it can cause intestinal paralysis.
- In the cardiovascular system it can cause heart failure, arrhythmias and changes in U and T waves of ECG.
- It can cause respiratory failure in terms of Respiratory system.

2.2.6 ACUTE CHILDHOOD HEMIPLEGIA

When an arm, leg or trunk is wholly paralyzed on one side of human body, it is diagnosed as hemiplegia; whereas the weakness in a muscle is hemiparesis, which is a partial paralysis. In hemiplegia, the response of the CNS is affected, it does not work properly. Hemiplegia and hemiparesis are more common in adults than in children. Cerebrovascular stroke usually causes hemiparesis in adults, while the hemiparesis is not limited to CNS infection in children. There are other disorders in children, diagnosed by the "acute hemiplegia of childhood"; it causes trauma, neoplastic intracranial space-occupying lesions (ICSOL) and anomalies of the human brain, which increases over time [57].

Acute childhood hemiplegia is when a child (from 3 months to 3 years of age), who was previously healthy and having no genetic inclination towards disorders, suddenly becomes sick. The etiology of this disease is not known but can be diagnosed by its concurrent infectious disorders. Initial signs of this disease are vomiting or convulsions, fever along

with a little hindrance and paranomia in speech temporarily. In rare cases, eye muscles are paralyzed too. Epileptic spams occurs early or later after the onset of the Acute Childhood Hemiplegia [58].

The causes Of Acute Childhood Hemiplegia can be vascular or post trauma effects, and different variants of encephalitides (diseases related to spinal cord). The other causes include arteritis, small vessel disease, cardiac embolus, and angiomata. Children with prolonged epilepsy, face severe motor impairment (loss of movement of legs and arms), repeated fits, hyperkinetic behavior (uncontrollable movement of muscles), and other defects related to cognition [58].

2.3 VIRUSES ASSOCIATED WITH AFP

2.3.1 POLIOVIRUS

The cause of Poliomyelitis is the three stereotypes of the Wild Poliovirus. It is a neurotropic RNA virus; belong to Picornaviridae and Enterovirus genus. The three stereotypes of the Wild Poliovirus are termed as type 1, 2 and 3. Type 1 virus has caused the most paralytic infection among all the stereotypes. Type 1 has been observed more frequent in causes of the paralysis disease epidemics. Compared to type 1, the other two types are not much neurovirulent. Type 2 has been eradicated already in the year 1999 in America. In 1999, type 3 virus was the reason of a major epidemic of paralytic disease in Angola [1].

Wild Poliovirus is transmitted from one person through feces or enters through mouth. Food is the source which normally carries this virus to mouth; usually via milk and water. The incubation time of this virus is 7 to 14 days, can range from 3 to 35 days. A person, who is

exposed to this virus and is not immune, will be affected by Poliomyelitis disease. 72% of the people are in-apparently infected, while 24% of the people face "Abortive Poliomyelitis", which is a minor illness. 4% of the patients are infected by non-paralytic Poliomyelitis. Infected people who develop disease of paralysis are only 1 per 100 to 1 per 1000 [1].

The clinical symptoms in the initial stages of the onset include pain in limbs, stiffness in neck, fatigue, high fever, vomiting, headache, and constipation. Poliomyelitis is found in young children under the age of 15 years. To distinguish the paralytic Poliomyelitis from other types, some characteristics have to be observed; which include fever at onset and quick development in the paralysis from 24 to 48 hours. The distribution of paralysis to both limbs is asymmetric (not same). The other distinguishing characteristics include the preservation of the sensory nerve function along with myalgia and residual paralysis after 60 days. Generally, the death rate of paralytic Poliomyelitis is 5-10%, varies from region to region. It was very high at some point from 20-30%. The reason of death for most of the Poliomyelitis patients is the bulbar region complications, such as failure in respiration. The most frequent victims of Wild Poliovirus are children under age 5 [1].

2.3.2 VACCINE ASSOCIATED PARALYTIC POLIOMYELITIS (VAPP)

The OPV has been extensively used for immunization against the Poliomyelitis, recommended by WHO. This vaccine was helpful in the eradication of Wild Poliovirus around the globe, along with breaking up the transmission of the virus. Sometimes the live attenuated strain of the virus in OPV reverts to neurovirulent strain; which is a reason for a

lot of paralytic disease cases similar to the paralytic Poliomyelitis caused by the Wild Poliovirus; known as the VAPP disease. The risk ratio of VAPP is a case per 2.5 million oral doses of OPV in the United States. This ratio differs in other regions such as Romania, where the risk is 14 times higher than the United States; a VAPP case per 183,000 OPV doses [1][2][35][47][61].

People who are at risk of VAPP can be divided into three different groups [1][35]:

- The first group includes the recipients of the OPV (infants).
- The second group includes the unvaccinated people who are in contact with the OPV recipients.
- The third group includes the immune compromised people.

2.3.3 VDPV & CVDPV

Vaccine Derived Poliovirus **(**VDPV) is a strain of Poliovirus OPV; it is different from OPV genetically. In certain circumstances, the OPV changes to a different strain which may cause VDPV in humans. When a child is given the dose of that particular OPV, he may end up as a victim of VDPV. VDPV sustains and circulates in the environment, known as Circulating Vaccine Derived Poliovirus (CVDPV). It can transmit from person to person and can circulate and persist up to six months [2][10][13][61][62].

Cases of Wild Poliovirus type 1 (WPV1) exists in Pakistan and Afghanistan only. Cases of two types of CVDPV exist in 25 countries currently; CVDPV1 exist in 23 of the 25 countries; Madagascar and Yemen are the two countries where cases of CVDPV2 exists [24].

2.3.4 ENTEROVIRUSES

Paralytic diseases similar to Poliomyelitis are caused by non-poliomyelitis enteroviruses. These diseases include hand-foot-and-mouth disease, acute hemorrhacgic conjunctivitis, and aseptic meningitis. The non-polio viruses, which causes paralytic diseases like Poliomyelitis, include echovirus, enterovirus 70 and 71, and coxsackieviruses. In Thailand, Taiwan, Panama, and India, epidemics of radiculomyelitis (a paralytic disease) and hemorrhagic conjunctivitis were caused by enterovirus 70. The symptoms of disease caused by non-poliomyelitis enterovirus 70 were muscle wasting and weakness, which are permanent and severe in most of the cases [1][31][45].

Enterovirus 71 [5][6][7][8][11] is the most involved virus in the epidemic CNS disease and most of the AFP cases. This virus was discovered in California (1969-1973). Enterovirus 71 grabbed the attention of the world when it was found as the cause of severe epidemics related to CNS disease in 1973 Japan and 1975 Bulgaria. In Bulgaria, 21% of the total patients infected by enterovirus 71 were paralyzed, while 29% were dead. The most frequent victims of enterovirus were children under age 5. Apart from these epidemics, other countries also faced outbreaks; such as Hungary 1978, two states of the US, and Pennsylvania and Philadelphia in 1987. From 1988-1990, enterovirus 71 was the cause of acute neurolofic disease in Brazil. Malaysia and Taiwan were also infected by enterovirus in 1997 and 1998 respectively [1].

The symptoms of the illness are reported 7 to 14 days prior to onset of AFP. These symptoms include vomiting, lethargy, fever, irritability, diarrhea, anorexia, and nuchal

rigidity. Residual paralysis, with wasting and weakness of muscles, was observed in patients at a 60-day follow up. In the AFP investigation process, Poliovirus disease is found more frequent than diseases like Poliomyelitis upon isolation. Clinically, the non-polio enterovirus cannot be differentiated from Poliovirus; laboratory tests are required for a clear diagnosis [1].

2.3.5 NEUROTROPIC VIRUSES

AFP associated to GBS and ATM can be caused by Herpesviridae. It is an umbrella of neurotropic DNA viruses, such as Epstein Barr, cytomegalovirus, and varicella zoster. One of the common forms of the non-epidemic viral encephalitis around the globe is the Herpes simplex virus. Annually, it infects 1 to 3 people per million, irrespective of gender and age differences [1].

Rabies is another type of the neurotropic viruses. After two months of incubation time, Rabies becomes evident. The nervous system becomes abnormal, some cases turns into a paralytic disease. Sphincter involvement and sensory disturbances can be observed in AFP cases of paralytic rabies when bitten. Furious rabies does not cause as much deaths as the patients with longer sickness of bulbar and respiratory paralysis [1].

Neurotropic virus, such as togovirus, Herpes viridae, paramyxovirus, parasites, and arbovirus may cause viral meningoencephalitis. Meningoencephalitis becomes an obvious symptoms of CNS such as coma, disorientation and upper motor neuron lesions. Another case of meningoencephalitis is the invasion of brain cells when rabies or herpeviridae viruses are involved [1].

The next virus causing AFP is the Japanese encephalitis. The infection of this virus is similar in behavior with Poliomyelitis in clinical symptoms and is caused by the anterior horn cell damage. The symptoms, which are similar to Poliomyelitis, include the weakness of muscles after a 60-days follow up [1].

2.3.6 ADENOVIRUSES

Adenoviruses cause infection of mild severity usually, but sometimes it can cause severe neurological disorders. Adenovirus is investigated in AFP diagnosis. 1.05% of the AFP cases shed adenovirus in feces, mostly infants. HAdV-C is a variant of adenoviruses, which is detected in 85% of the adenoviruses infected AFP cases. 40% of the AFP cases with adenoviruses are diagnosed with TrN. The variants of adenoviruses are: HAdV-A, HAdV-B, HAdV-C, HAdV-D, HAdV-E and HAdV-F [59].

2.4 APF SURVEILLANCE

Disease surveillance is a constant and systematic gathering of information regarding diseases and their influencing factors on long-term basis, along with time to time analysis of data to take intervention measures and assess their effectives [34]. In other words, surveillance is the process of monitoring some activity or behavior of a particular planned system, to ensure performance through some indicators, and WHO also defined some performance measuring indicators for surveillance [2][3][21][27][29][30][37].

Surveillance of AFP means monitoring the cases of AFP worldwide. The rate of non-polio AFP tells us about the sensitivity of surveillance [21][26], which means detection of non-polio AFP cases per hundred thousand in children aged less than 15 years. The standard rate

for satisfactory surveillance performance is detecting 1 case per hundred thousand generally [21][27][29][30]. In the surveillance process, if the rate of cases per hundred thousand is less than 1, it means that the surveillance performance is not good enough. This is a very important indicator; it helps in measuring the sensitivity of the implemented surveillance system.

The second important indicator is the adequacy of stool samples. For the stool sample examination, earlier the minimum stool sample adequacy indicator was 60%, but recently, as the system evolved, the rate has been increased by 20 %. This means 80% of the total AFP cases should have adequate stool samples i.e. collection of stool sample should be done in the defined time after onset of paralysis, which is 14 days, otherwise the stool sample is inadequate [2][15][17][21]. Table 1 depicts global data for the last 10 years about number of AFP cases, rate of non-polio AFP per hundred thousand as well as number of Wild Poliovirus confirmed cases, number of compatible cases, number of CVDPV cases [13][24], and stool adequacy [28].

2.4.1 THE TWO KEY INDICATORS OF AFP SURVEILLANCE

1. The non-Polio AFP detection rate in Equation 1: [2]

Equation 1 Non-polio AFP Rate

$$\textbf{Non–polio AFP rate} = \frac{\textbf{number of reported non–polio AFP rates} < \textbf{15 years of age}}{\textbf{total number of children} < \textbf{15 yrs of age}}$$

2. At least 80% of AFP cases should have 2 adequate stool samples collected within 14 days of onset of paralysis. [2]

2.4.2 AFP STATISTICS BY WHO (GLOBAL)

The data in Table 1 is about AFP cases in the last decade (2010-2020). The second column is about Wild Poliovirus, these are concrete statistics about number of Wild Poliovirus confirmed cases per year and the data in the fourth column is the number of compatible AFP cases during that period. The important thing is the non-polio AFP rate [43] in the third column. AFP rate shows the surveillance sensitivity, as discussed earlier. Looking at the column, according to this data, at least more than 5 AFP cases, per hundred thousand, have been detected each year; this indicates a good surveillance. If any case in the AFP is a Wild Poliovirus, emergency in that area will triggered and investigations undertake in that particular area to isolate the detected virus in order to stop its spread along with OPV vaccination.

Table 1 AFP Data, Globally [28]

Year	No. of AFP Cases	Wild Poliovirus Cases	Non-polio AFP Rate	Compatible Cases	Adequate Stool Collection	Cases of CVDPV
2010	98788	1352	5.56	579	85%	60
2011	104260	650	6.01	480	86%	66
2012	106457	223	6.0	315	88%	71
2013	101665	416	5.36	403	90%	65
2014	103974	359	5.45	285	89%	56
2015	100222	74	5.16	277	89%	32
2016	109821	37	5.77	211	90%	5

2017	104060	22	5.46	259	89%	96
2018	97315	33	5.11	170	86%	104
2019	106175	176	5.55	229	84%	368
2020	75840	140	3.84	216	87%	1093

The importance of non-polio AFP surveillance can be understood by its sensitivity [26]. If these AFP cases are not closely monitored, Polio can re-emerge on the globe, or at least in that region. Diseases like Poliomyelitis, ATM, TrN and GBS etc. are the subject of AFP surveillance across the globe [2][15][16][37]. AFP surveillance is the backbone of Polio eradication as it monitors the viruses associated with AFP. In the surveillance process, these viral diseases are monitored and timely reported as the new cases for differential diagnosis. Various data entry forms generates the data of a particular case which is gathered and summarized in the line listing form [17][21][23].

The aim of AFP surveillance is monitoring and isolation of the circulating Poliovirus, which was recognized for the first time in 1975 in United States. The program for eradicating Polio was not set until WHO got the vaccine in 1988, and since then, with this surveillance and vaccination system of WHO, Polio has a decrease of 99%. All countries are now free of this virus, except Pakistan, Afghanistan and Nigeria [20][22][37][39][40][61][62]. Figure 1 depicts the data regarding the annually reported AFP cases worldwide (2010-2020) [28].

2.4.2.1 GLOBAL AFP CASES

Figure 1 illustrates the total number of AFP cases reported globally from 2010 to 2020.

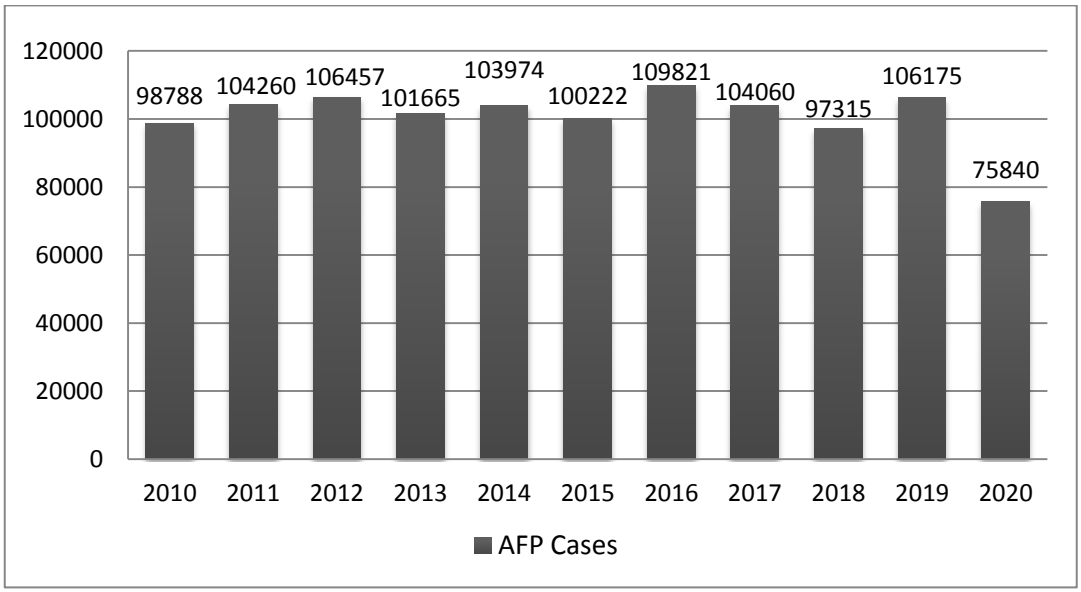


Figure 1 Global AFP Cases [28]

2.4.2.2 GLOBAL POLIOVIRUS CASES

Figure 2 shows the number of Wild Poliovirus cases reported globally from 2010 to 2020.

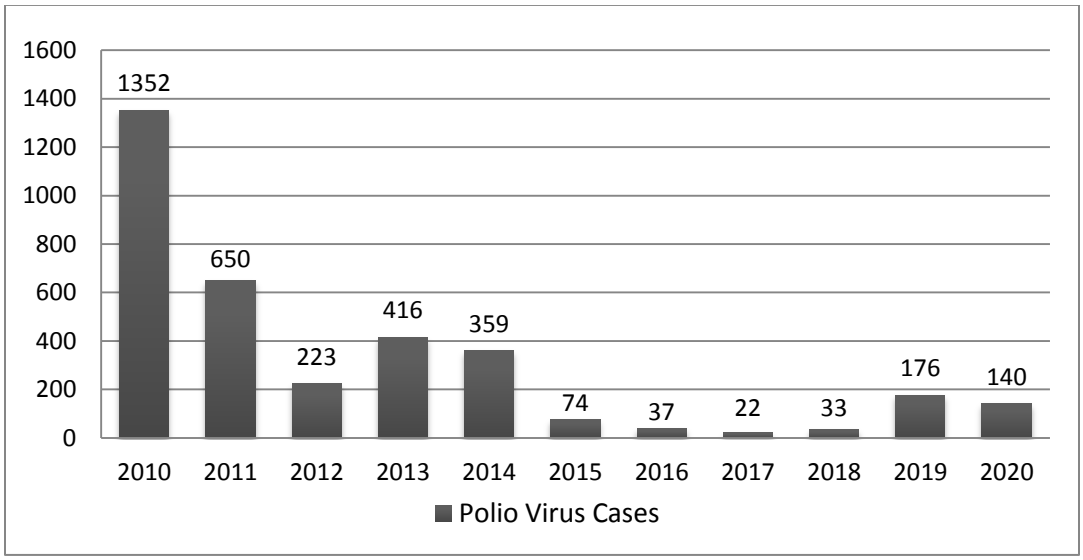


Figure 2 Global Wild Poliovirus Cases [28]

2.4.2.3 GLOBAL NON-POLIO AFP RATE

Figure 3 demonstrates the rate of non-polio AFP globally from 2010 to 2020.

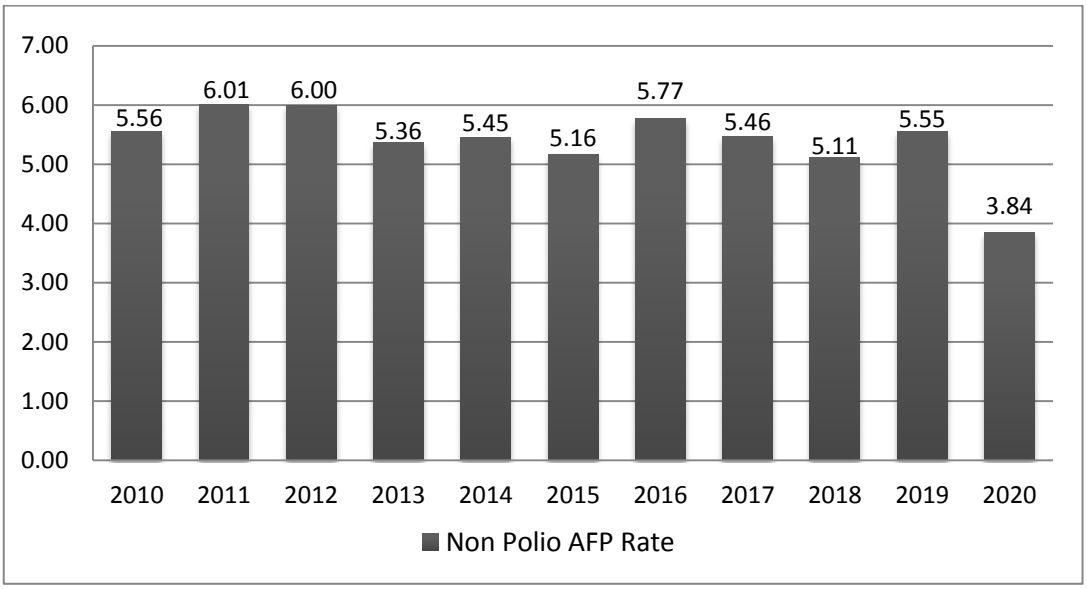


Figure 3 Global Non-polio AFP Rate [28]

2.4.2.4 GLOBAL COMPATIBLE CASES

Figure 4 shows the number of compatible cases reported globally from 2010 to 2020.

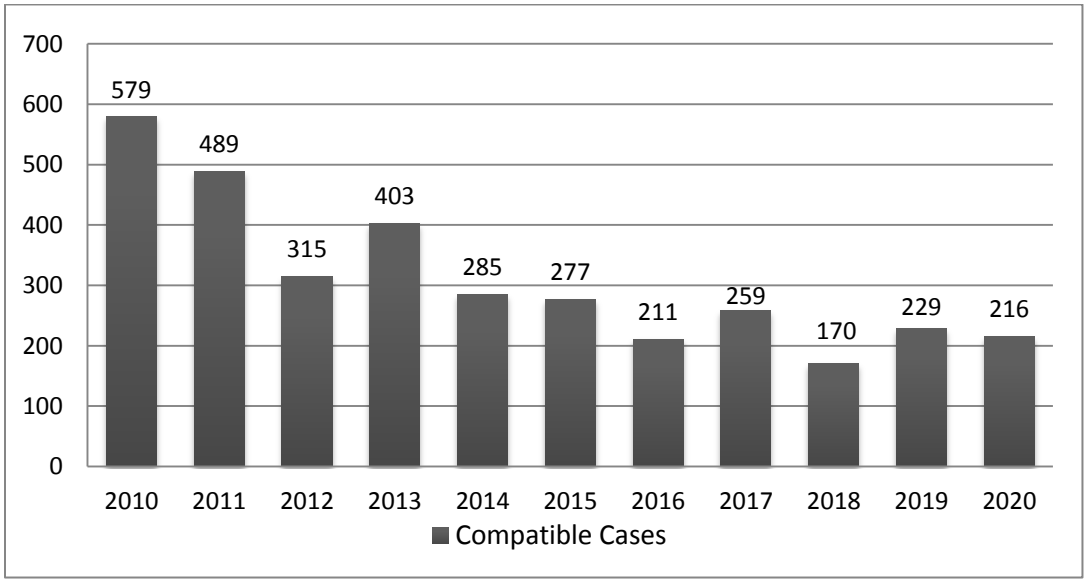


Figure 4 Global Compatible Cases [28]

2.4.2.5 GLOBAL STOOL SAMPLE ADEQUACY

Figure 5 reflects the global stool sample adequacy percentage from 2010 to 2020.

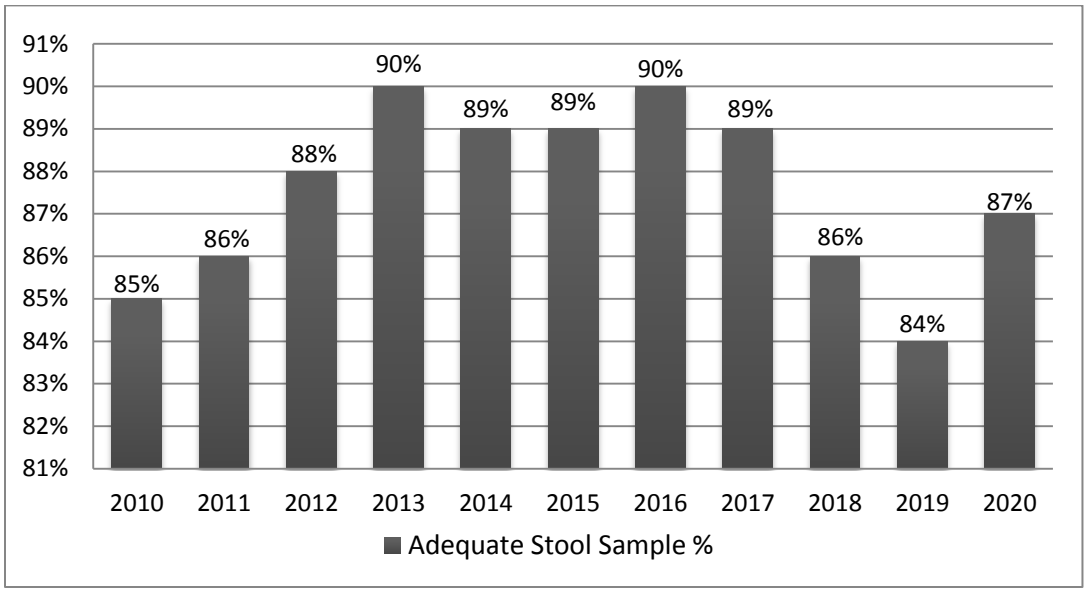


Figure 5 Global Stool Sample Adequacy [28]

2.4.2.6 GLOBAL CASES OF CVDPV

Figure 6 reflects the number of CVDPV cases reported globally from 2010 to 2020.

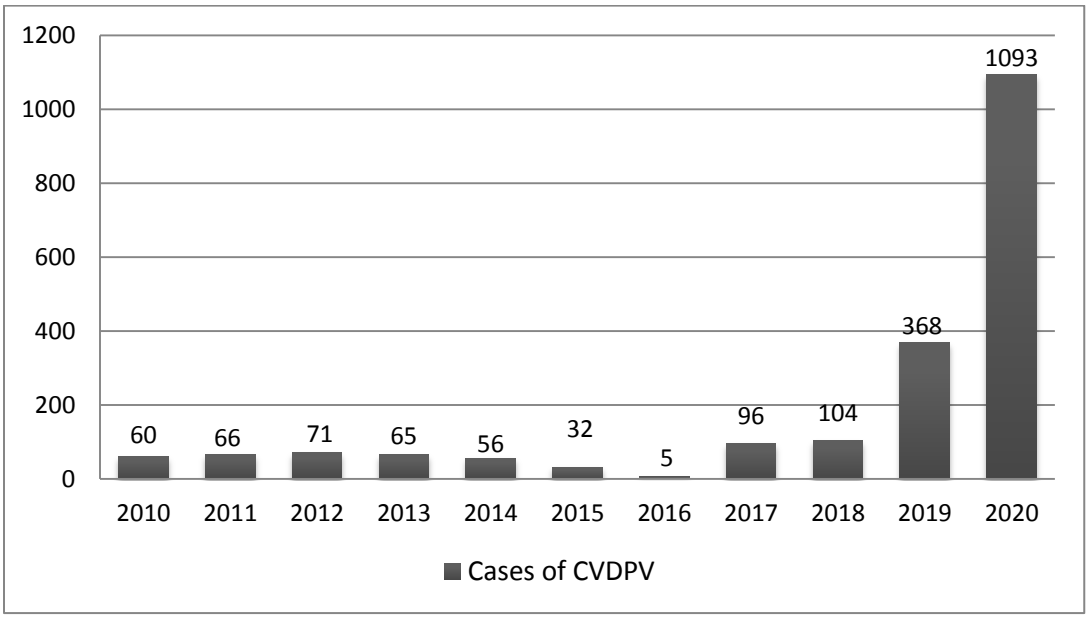


Figure 6 Global Cases of CVDPV [28]

2.4.2.7 AFP DATA (PAKISTAN)

Table 2 illustrates the AFP data in Pakistan from year 2010-2020. This comprehensive table represents statistics about number of AFP cases, Wild Poliovirus cases, Non-polio AFP rate, compatible cases, stool sample adequacy, and CVDPV cases.

Table 2 AFP Data, Pakistan [28]

Year	No. of AFP Cases	Wild Poliovirus Cases	Non-polio AFP Rate	Compatible Cases	Adequate Stool Collection	Cases of CVDPV
2010	5393	144	8.87	24	88%	0
2011	5762	198	9.35	31	88%	0
2012	5036	58	8.36	6	89%	16
2013	4790	93	7.72	24	90%	48
2014	5369	306	8.25	45	88%	22
2015	5807	54	9.2	11	87%	2
2016	7843	20	12.77	5	87%	1
2017	10315	08	16.84	2	86%	0
2018	12257	12	20.01	1	87%	0
2019	15192	147	24.54	6	87%	22
2020	11961	84	18.77	29	86%	135

2.4.2.8 AFP CASES IN PAKISTAN

Figure 7 reflects the total AFP cases reported in Pakistan from 2010 to 2020.

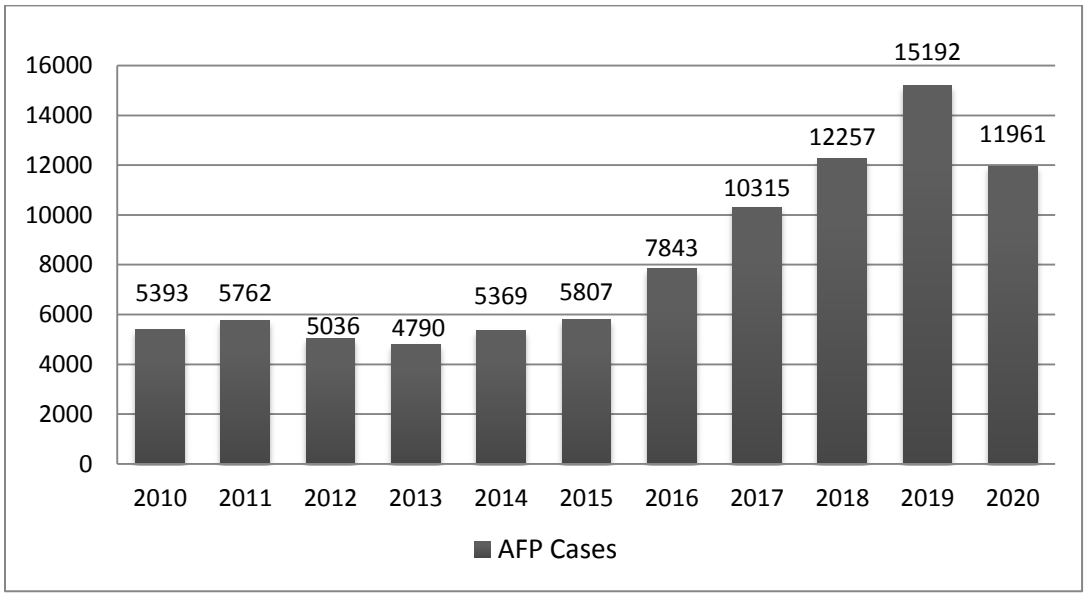


Figure 7 AFP Cases in Pakistan [28]

2.4.2.9 POLIOVIRUS CASES IN PAKISTAN

Figure 8 shows the number of Wild Poliovirus cases reported in Pakistan from 2010 to 2020.

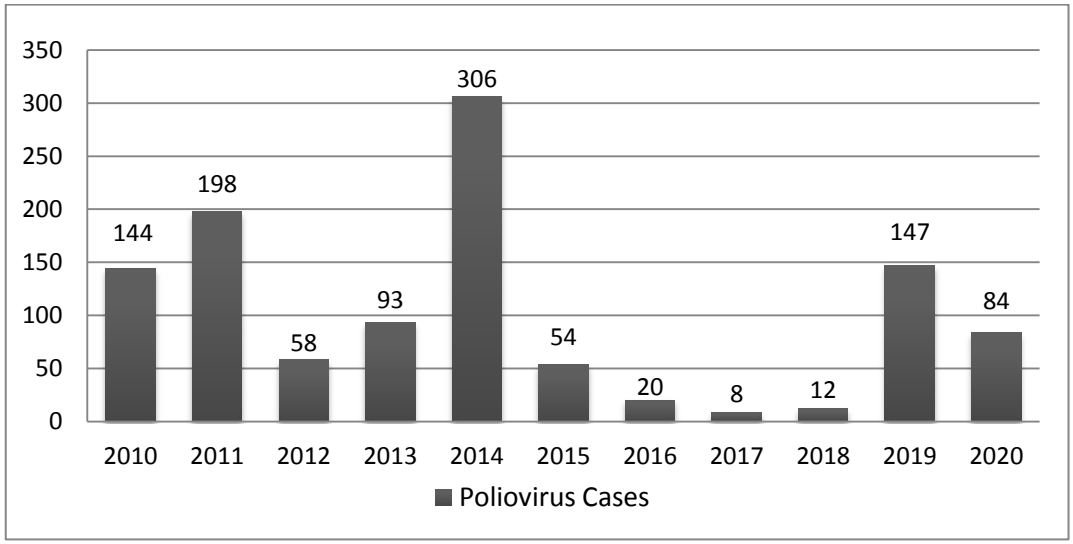


Figure 8 Wild Poliovirus Cases in Pakistan [28]

2.4.2.10 NON-POLIO AFP RATE IN PAKISTAN

The rate of non-polio AFP is in Pakistan from 2010 to 2020 can be observed in Figure 9.

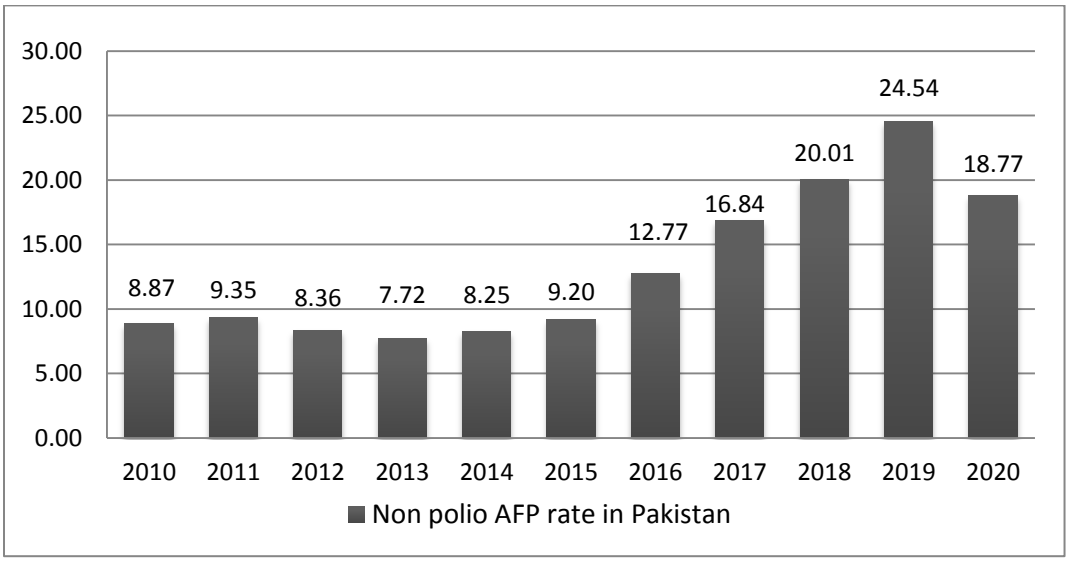


Figure 9 Non-polio AFP Rate in Pakistan [28]

2.4.2.11 COMPATIBLE CASES IN PAKISTAN

Figure 10 shows the number of compatible cases reported in Pakistan from 2010 to 2020.

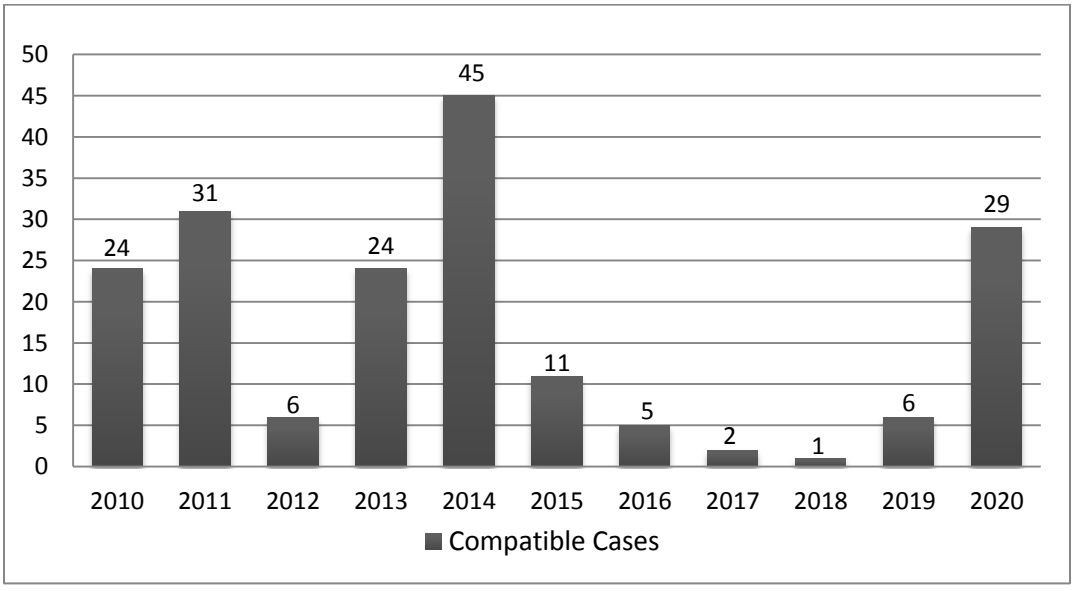


Figure 10 Compatible Cases in Pakistan [28]

2.4.2.12 STOOL SAMPLE ADEQUACY IN PAKISTAN

Figure 11 reflects stool sample adequacy percentage in Pakistan from 2010 to 2020.

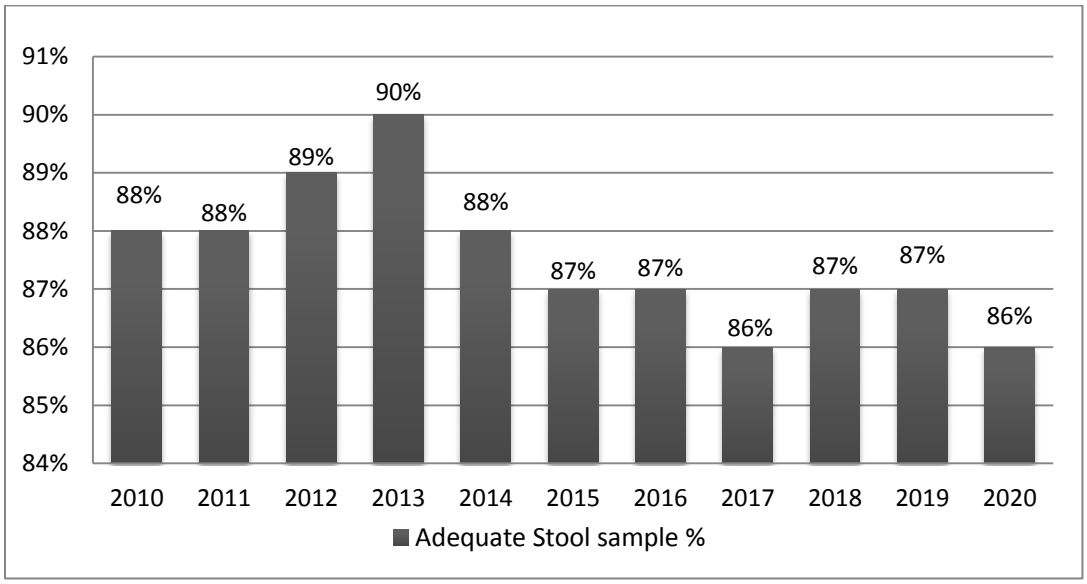


Figure 11 Stool Sample Adequacy in Pakistan [28]

2.4.2.13 CVDPV CASES IN PAKISTAN

Figure 12 reflects the number of CVDPV cases reported in Pakistan from 2010 to 2020.

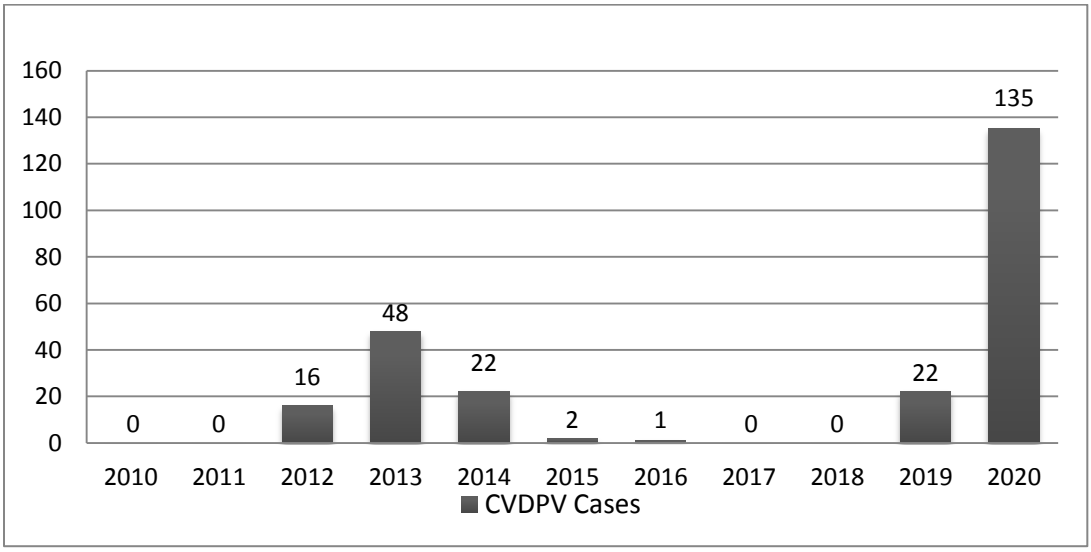


Figure 12 cPDPV Cases in Pakistan [28]

2.4.2.14 COMPARISON OF WILD POLIOVIRUS CASES (PAKISTAN VS. GLOBAL)

Table 3 exhibits the comparison between Pakistan and the rest of the world in Wild Poliovirus cases from 2010-2010. The 4th column reflects the percentage of Wild Poliovirus cases in Pakistan. The case percentage has increased over the years, although the number of reported cases has decreased.

Table 3 Wild Poliovirus Cases (Pakistan Vs. Global)

Year	Wild Poliovirus Cases Pakistan	Wild Poliovirus Cases Global	Pakistan's Cases Percentage
2010	144	1352	10.65
2011	198	650	30.46
2012	58	223	26
2013	93	416	22.35
2014	306	359	85.24
2015	54	74	72.97
2016	20	37	54
2017	08	22	36.36
2018	12	33	36.36
2019	147	176	83.52
2020	84	140	60

2.4.2.15 Graphical Comparison of Wild Poliovirus Cases (Pakistan Vs. Global)

Figure 13 illustrates the comparison between Pakistan and rest of the world in terms of reported Wild Poliovirus cases each year from 2010 to 2020.

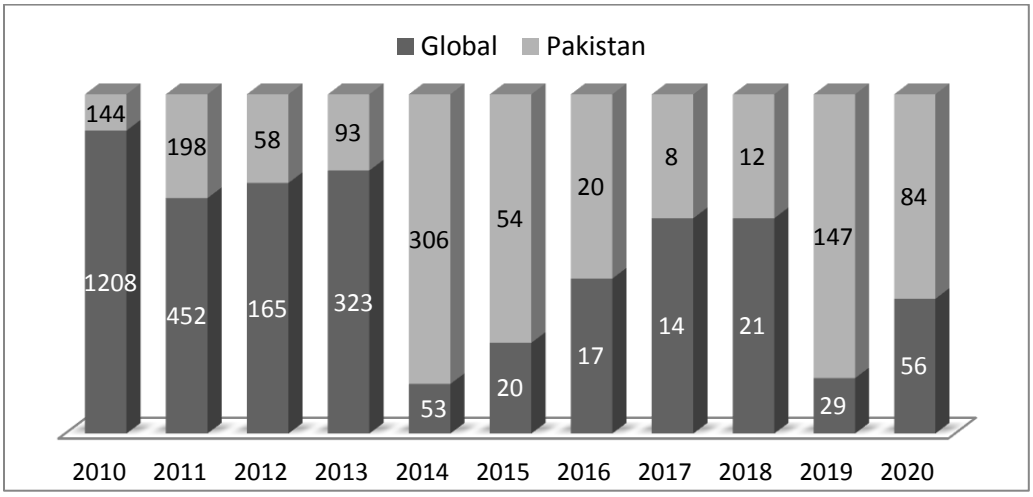


Figure 13 Comparison of Wild Poliovirus Cases [28]

2.4.2.16 Graphical Representation of Wild Poliovirus Cases Proportion

Figure 14 shows that in recent years, Pakistan has a huge percentage of the reported Wild Poliovirus cases worldwide.

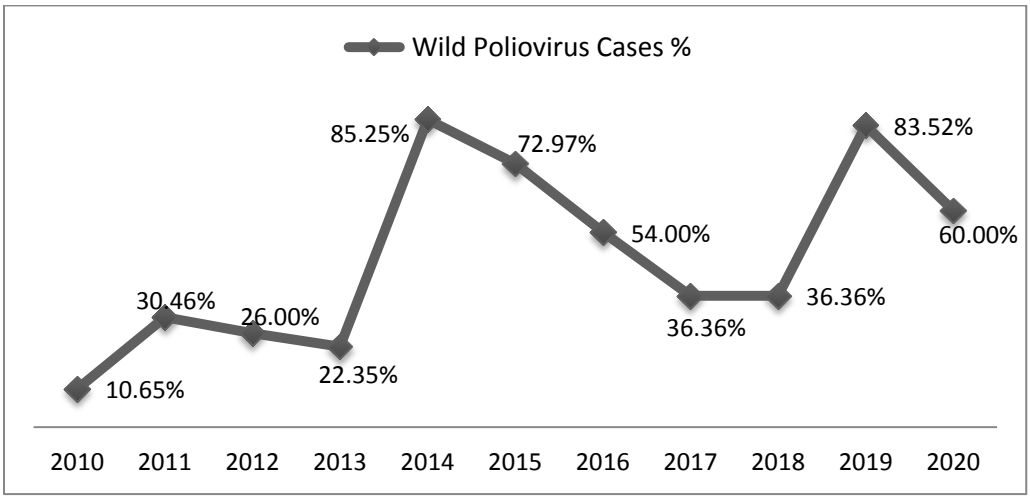


Figure 14 Poliovirus Cases Pakistan [28]

2.4.2.17 Graphical Comparison of Wild Poliovirus Cases (Pakistan Vs. Global)

Figure 15 represents a graphical comparison between Pakistan and rest of the world in terms of the rise and fall in the percentage of Wild Poliovirus cases from 2010 and 2020.

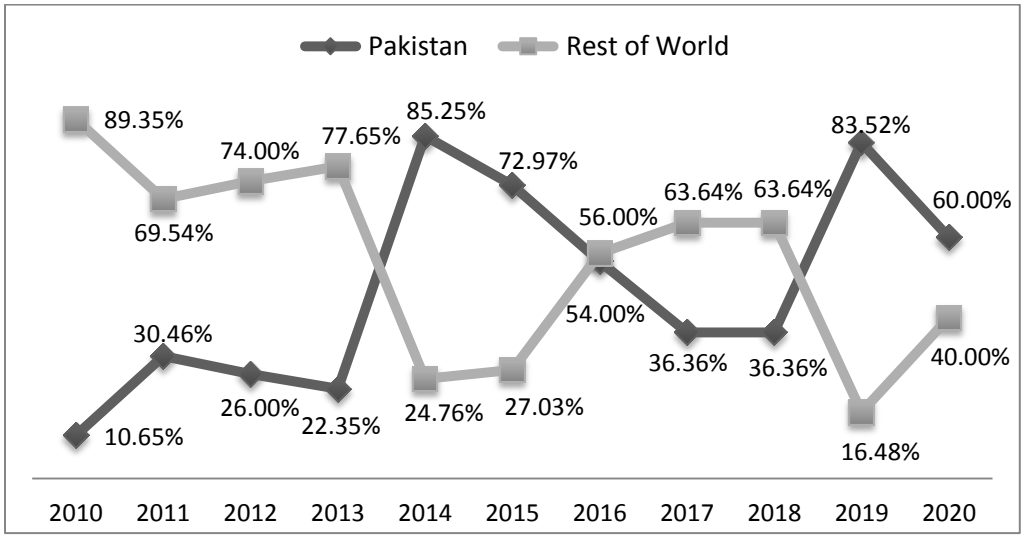


Figure 15 Poliovirus Cases (Pakistan vs. Rest of World) [28]

2.4.2.18 Wild Poliovirus Cases Development (Pakistan Vs. Global)

3482 cases of Wild Poliovirus have been reported worldwide (2010 to 2020), 32% of them were reported in Pakistan, as shown in Figure 16.

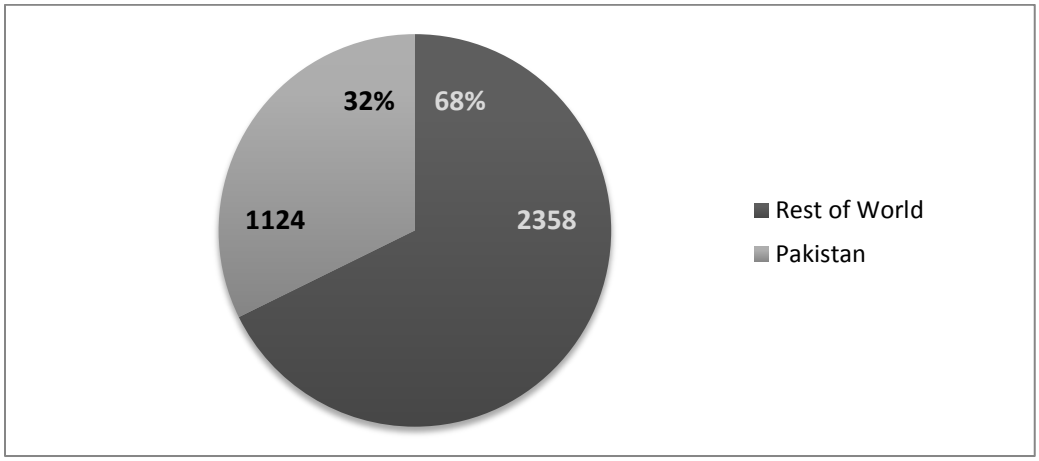


Figure 16 Wild Poliovirus Cases (2010-2020) [28]

2.5 WHO'S STANDARD DIAGNOSIS PROCEDURE FOR AFP

When a child is reported as a case of AFP, one of the initials steps is the collection of two stool samples in a period of 24 hours. If the collected stool samples are gathered in the first 14 days from onset of paralysis, then the samples are adequate, otherwise the stool samples are inadequate. Figure 17 shows the procedure of diagnosis where the first priority is the laboratory examination of the stool samples. If Wild Poliovirus is found in the specimen, the patient is declared as a victim of Poliomyelitis otherwise further procedure is followed.

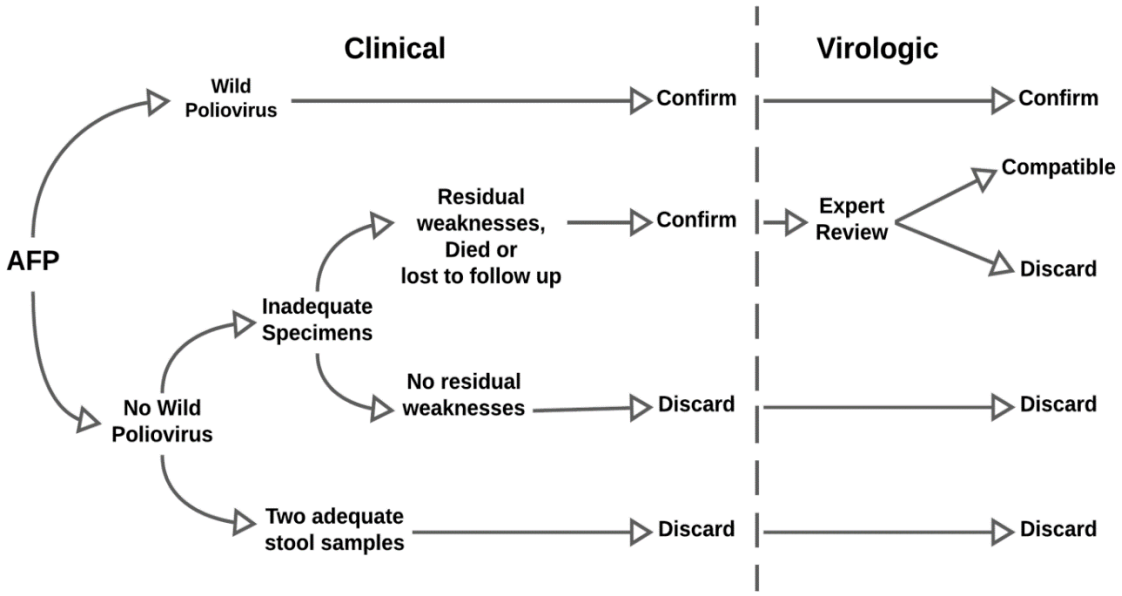


Figure 17 Classification of AFP Cases [2][15][17]

If the specimen shows no signs of Wild Poliovirus, the adequacy of the stool samples is questioned. If the two samples are adequate, then the case is discarded but if the stool samples are not adequate then other questions are asked while following up with the patient for months. The condition of the patient is monitored in the coming months; if there is no residual weakness in the patient then the case is discarded. While following up with the

patient, if the patient dies in the coming months or residual weakness is found in the patient, then medical experts decide for the case to be compatible or not [2][15][17][21] [25].

2.6 AFP Case Handling

When a child is reported as a suspect of Polio, the process shown in Figure 18 is exercised by WHO. The first question to answer is, whether it is a trauma or tumor, if any of the two yields a "yes", the case is discarded straight away. Otherwise, the second question arises of whether the stool sample is adequate (taken inside 14 days after onset of paralysis); if the answer is yes, the sample will be tested in the laboratory, looking for the presence of Wild virus. If the test result for the virus is negative, the case will be discarded; otherwise it's a confirmed Polio case and further investigations will take place.

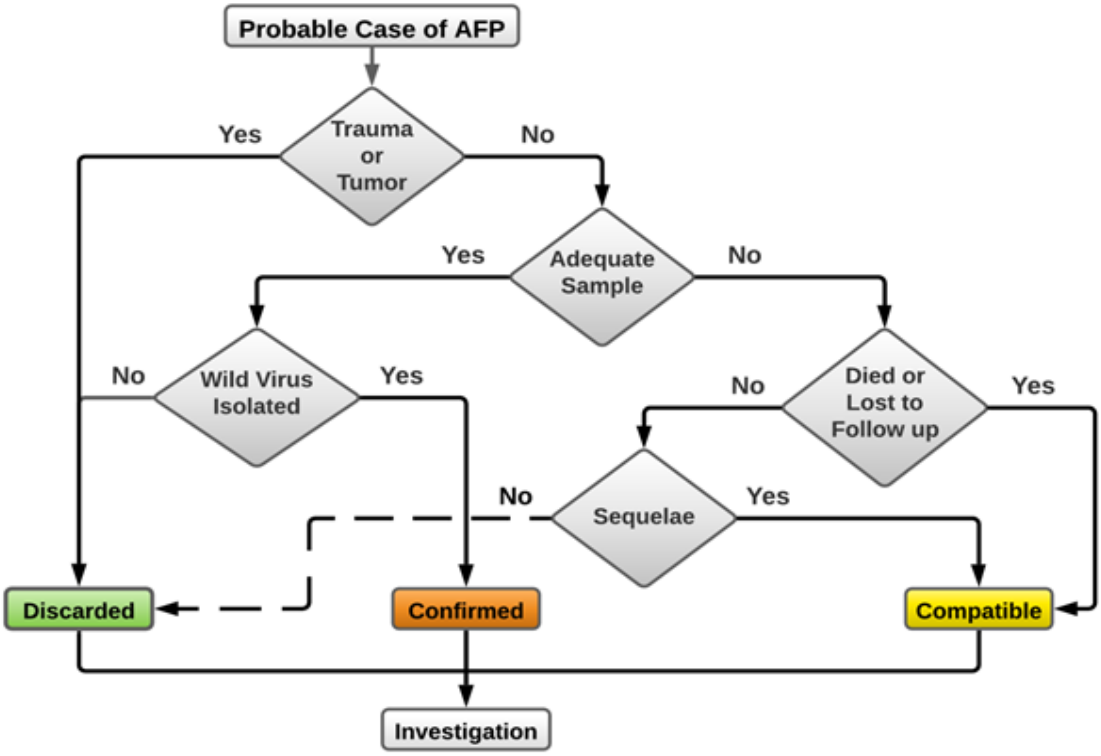


Figure 18 Procedure for Handling AFP Case [21]

On the contrary, if the stool sample is inadequate then a long process gets initiated. The case is followed and checked for further situations; if the status of the case is 'lost' or 'died' to follow up day (60 days after onset of paralysis), it will be considered as a Polio compatible case [21][29]. If these two conditions are not met, it means the case is decidable at this stage. Then the symptoms are checked until the end of follow up period. If they prevail, the case will be declared as compatible, otherwise it will be discarded [2][17][21][25]. Table 4 further elaborates the classification of AFP cases.

Table 4 Tabular Classification of AFP Cases [2]

Status	Classification	Reason
Final	Confirmed (Wild Polio Virus)	Wild Polio Virus detected in stool sample.
	Confirmed (VDPV)	VDPV detected in stool sample which has residual paralysis in 60 day follow-up, or clinically confirmed.
	Compatible	• AFP case lost to follow-up at 60 days. • Death of the AFP case reported due to the illness within 60 days • Residual paralysis for which no other medical reason is evident
	Discarded	• Two adequate negative stool specimens with 14 days of onset of paralysis • No Wild Polio Virus detected in stool sample and no residual paralysis in 60 day follow-up • Confirmed alternative diagnosis • Non-polio enterovirus isolated

		• No virological investigation, and a clinical picture • incompatible with polio
Pending	Inadequate Information	Due to incomplete information, the Polio Eradication Committee is unable to make a decision. The committee gives 30 more days to investigation team. The committee will take the final decision in the next meeting.
	60 day follow-up not yet done	Final decision will be taken in the next meeting by Polio Eradication Committee.

2.6.1 ORDER OF AFP DIFFERENTIAL DIAGNOSIS

When examining records, the surveillance team officers should look for diagnosis and symptoms. Some of the common symptoms and diagnosis are highlighted in Table 5. While investigating a patient and reviewing the logbooks, if any diagnosis or symptoms found like the ones mentioned in Table 5, then further details should be checked in the medical books. During this process if the records are not necessarily matching the case or indicate the possible of AFP, the surveillance officer should re-examine the patient once more. The surveillance officer (head of district or province) should immediately visit the patient's home for a physical examination and collect a second stool sample (with in sample adequacy time limit), if a case has been discharged too early or not investigated thoroughly. The following is the 3 step method for the differential diagnosis [2]:

- Step 1: Examination of AFP suspect. If the symptoms are related to AFP, perform step 2.

- Step 2: Look for the diagnosis that should always be investigated as AFP, if the case is diagnosed then the process is over in 2 steps, otherwise step 3 will be perused.
- Step 3: Look for the diagnosis that sometimes present as AFP.

Table 5 AFP Diagnosis Categories [2]

Step 1	Symptoms that should alert further investigation	• Paralysis, paresis (weakness), flaccid (floppy) paralysis (in combination with any other words) • Weakness (of limb, of unclear origin, etc.) • Frequent falls, gait disturbance, cannot walk, etc. • Muscle hypotonia (hypotonia means loss of muscle tone due to some other cause)
Step 2	Diagnoses that should always be investigated as AFP	• Poliomyelitis, rule out polio, suspect polio • Guillain-Barre' Syndrome • Transverse myelitis • Traumatic neuritis
Step 3	Diagnoses that sometimes present as AFP	• Hypokalemic paralysis • TB of the spines (Pott's disease) • Meningitis / encephalitis

2.6.2 TIME-FRAME FOR AFP DIAGNOSIS

The time for the manual procedure for the diagnosis of AFP is flexible. It can take from 3 weeks to 3 months, depending upon the outcomes of laboratory examinations and expert's reviews [2].

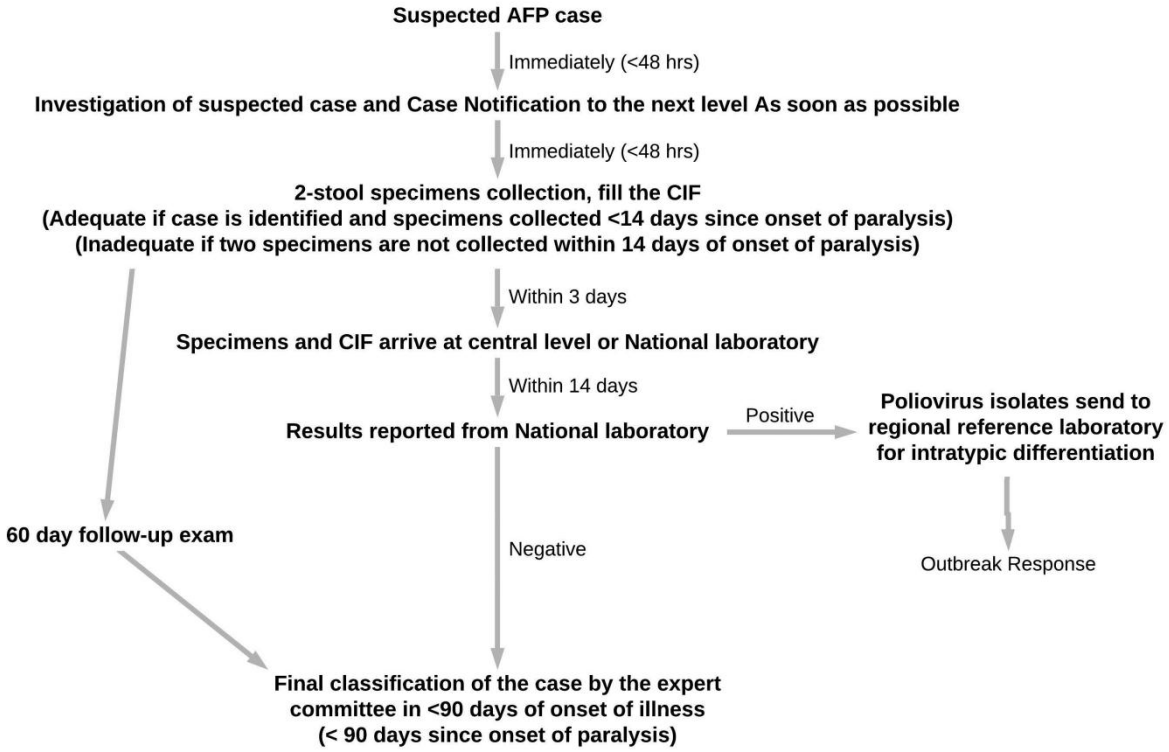


Figure 19 Flow Diagram of AFP Case Investigation [2]

The following is the step by step elaboration of time-frame for AFP Diagnosis procedure as demonstrated in Figure 19:

1. When a case is reported as AFP suspect, investigation takes place as soon as possible, ideally within 2 days.
2. The case is registered and two stool specimens collected 24 hours apart.

3. The stool samples are adequate if collected within 14 days after onset of paralysis, otherwise not adequate.
4. If the stool samples are inadequate, then a 60 day follow-up examination takes place, which lead to a decision taken by experts committee within 90 days.
5. If the stool samples are adequate, then laboratory examinations will be conducted and result will be reported in the next 2 weeks.
6. If the result shows a positive Poliovirus then the diagnosis is completed almost in 3 weeks.
7. If the result is negative, then it can take up to 3 months after further follow-up of 60 to 90 days, where experts committee is involved in the diagnosis process.

2.7 SUMMARY

In this chapter, the diseases, viruses, and vaccines associated with AFP were reviewed. After getting through the surveillance and diagnosis procedures of AFP, gaps have been identified. The process for differential diagnosis of AFP is manual and take months in the diagnosis process for most cases. The second most important point is the inadequacy of stool samples. If the stool sample is inadequate, the Poliovirus cannot be detected in the stool sample examination, and the patient will be followed up for 60 days at least. After the follow-up months, the diagnosis will be finalized within 90 days by the experts committee. Firstly, the current diagnosis of AFP is manual; secondly, it takes a lot of time and effort in most cases.

CHAPTER 3
METHODOLOGY

3.1 INTRODUCTION

In this chapter, the transition from the current manual procedure for the differential diagnosis of AFP to a new method is represented. First of all, the procedure of data gathering and handling is elaborated; followed by the discussion of the line listing data and AFP patient dataset, which is used for laboratory diagnosis in this research. Afterwards, the standard diagnosis table for the differential diagnosis of AFP is discussed in detail. The rule base dataset, derived from the standard diagnosis table is explained before the discussion of the proposed frame work and its implementation.

3.2 AFP LINE LISTING DATA

In the literature review chapter, the approach was discussed for deciding if the case is Polio confirmed or compatible or neither. There are the other diseases as well, as discussed earlier in the literature review chapter. To deal with a newly reported case of AFP, WHO has a standard procedure in which the data of every newly reported AFP patient is generated through various forms [17][21]. These forms are used for different purposes, such as the stool collection form, which contains the details of stool samples and whether the sample is adequate or not, and some other information necessary for the diagnosis of that particular patient. Other forms include case investigation form, laboratory request form, the form for final classification of case by experts etc. [21][23]. From time to time, after the onset of

paralysis, additional information is gathered and generated through all these forms and then all the data about a particular AFP case is entered in the final form called the Line Listing form.

A Line listing form is a complete table that contains the complete data of an individual AFP case in a single row and in the end some conclusion is given in the last column and that is the diagnosis of each AFP case [17]. The diagnoses in the final column can be diseases like as ATM, GBS, TrN or any other disease that comes under AFP surveillance. The structure of Line Listing form is described in Figure 20 along with all of its attributes.

EPIDNO	Patient name	District	OPV R (1)	OPV S (2)	Date (3)							Fever at onset (11)	Rapid paralysis (12)	Assym paral (13)	Res of FU (14)	Laboratory Results				Final Class (17)
					DOB (4)	Onset (5)	Not (6)	Inv (7)	S1 (8)	S2 (9)	FU (10)					P (15)	P2 (15)	P3 (15)	E (16)	

Figure 20 AFP Line Listing Table [17]

The following are the 17 features of the line listing table:

- **OPV-R**

  Number of OPV doses given through Routine Immunization, reported by history.

- **OPV-S**

  OPV doses given through Supplementary Immunization

- **Date**

  Data of below information

- **DOB**

  Date of Birth

- **Onset**

  Date of onset of paralysis

- **Not**

  Date of case report

- **Inv**

  Date of case investigation

- **S1**

  Date of 1st stool sample collection

- **S2**

  Date of 2nd stool sample collection

- **FU**

  Date of 60 days examination.

- **Fever at Onset**

  Presence of fever at onset of paralysis

  (a) 1= yes

  (b) 2= No

  (c) 9= Unknown

- **Rapid Paralysis**

Rapid progression of paralysis within 4 days

(a) 1= yes

(b) 2= No

(c) 9= Unknown

- **Asymmetric Paralysis**

  Is paralysis symmetric?

  (a) 1= yes

  (b) 2= No

  (c) 9= unknown

- **Res of FU**

  Result of 60 days examination

  (a) 1= presence of residual paralysis

  (b) 2= no residual paralysis

  (c) 3= lost to follow-up

  (d) 4= died before follow-up

- **P, P2, P3 Laboratory Results**

  Results of Poliovirus isolation

  (a) 1= Wild

  (b) 2= Sabin-like

  (c) 3= pending intratypic differentiation

  (d) 4= negative

  (e) 5= not processed

- **E Laboratory Results**

  Results of Enterovirus isolation

  (a) 1= positive

  (b) 2= negative

- **Final Class**

  Results of final classification

  (a) 1= confirmed Polio

  (b) 2= Polio compatible

  (c) 3= discarded

3.3 DATASET 1: AFP LINE LISTING

The data in the line listing form has many features but not all the features of this data are required, therefore a preprocessing step on the line listing data was necessary. After preprocessing the rough data of Line Listing form, the unnecessary features were removed. As a result of the preprocessing step, a final dataset was created for training and validation of the model for diagnosis prediction. Table 6 elaborates the dataset structure in terms of labels, features and attributes.

Table 6 Patient Dataset

Feature No.	**Feature Name**	**Feature Values**
1	Sex	Male, Female
2	Age (in weeks)	1- 200

3	Type Site	1,2,3,4,5
4	No of Doses	0-15
5	Adequate	Adequate, Inadequate
6	OPV 1st	EV, NVI, P1, P2, P3, P1+P3
7	OPV 2nd	EV, NVI, P1, P2, P3, P1+P3
8	FUP	Died, Lost, Yes, No
9	URGNSC	Urgent, N U
10	Fever	1,2,9
11	ASYM	1,2,9
12	Progression	1,2,9
13	Diagnosis (Label)	Transverse Myelitis, GBS, Hypokalemia, Traumatic Neuritis

The features discussed in the Table 6 are selected after the preprocessing on the rough Line Listing data. In the preprocessed dataset:

- The first feature is a simple one: differentiating among genders, with values Male and Female.
- The second feature is Age in weeks; value varies from 1 week to 200.
- The third feature is the Type Site; it indicates the sensitivity of a geographic in terms of AFP cases; its values are 1, 2, 3, 4 and 5.

- The fourth feature describes the number of Oral Polio Vaccine doses given to a particular AFP patient; it contains values from 0 to 15 doses.
- The fifth column has two values: Adequate and Inadequate; which means a stool sample collected was adequate or not.
- Columns six and seven are about the type of OPV vaccine given to a patient, but the slight difference is that OPV 1^{st} is the vaccine given in routine immunization while OPV 2^{nd} is given through supplementary immunization. The values of feature six and seven varies because of the variation in vaccination, which are EV, NVI, P1, P2, P3 and P1+P3 both.
- In feature eight, FUP means follow up of an AFP case after sixty days; the values are Died, Lost, Yes, No. If a patient is Lost, has Died, or if the value in this column is yes, these three values mean that the case is compatible, but if the value is no, the case get discarded for polio.
- Column nine is about the case report urgency; it elaborates if the case was reported Urgent or Not Urgent.
- Fever in column ten tells about having fever at onset; values are 1, 2 and 9 per WHO standard, which means 1 for yes, 2 for no and 9 for unknown.
- The eleventh column is telling whether a case is Asymmetric; if yes, the value is 1; if no, the value is 2; while 9 means unknown.
- The twelfth column describes the rapid progression in the first four days of paralysis; value 1 for yes, value 2 for no and 9 for unknown.

- The thirteenth column is the label column, consists of four labels, Transverse Myelitis, GBS, Hypokalemia and Traumatic Neuritis. These diseases are discussed already in this paper. These four diseases are considered as non-Polio AFP cases. The whole dataset contains data about these four diseases include in AFP cases.

3.4 WHO DIFFERENTIAL DIAGNOSIS TABLE

Apart from the line listing table (form), there is differential diagnosis table, used for the clinical differential diagnosis of AFP. The table has information about diseases like Poliomyelitis, GBS, ATM and TrN. The table provides the basis for differential diagnosis of AFP. Table 7 is the standard diagnosis table provided by WHO [16][17]. This table is a global standard for differential diagnosis of AFP. The second half of the proposed framework is based on this table. There are 14 different attributes in one column on the left, each of them is actually a question asked by medical experts during the clinical diagnosis of a particular AFP patient.

In the first column all the diseases are mentioned namely Poliomyelitis, GBS), TrN, ATM and Acute Childhood Hemiplegia. The table contains different possibilities and against every variation the diagnosis changes from one disease to another. This table is considered for this research and a rule base dataset devised for the automatic prediction of a particular AFP case.

Muscle Tone	Flaccid Paralysis	Fever at Onset	Installation of Paralysis	*Diagnosis*
Reduced or absent in affected limb	Acute, usually asymmetrical, principally proximal	High, always present at onset of flaccid paralysis, gone by the following day	24 to 48 hours Onset to Full Paralysis (acute)	***Polio Myelitis***
Global Hypotonia	Generally acute, symmetrical and distal	Not Common	Hours to ten days (acute-sub acute)	***Guillain-Barre-Syndrome (GBS)***
Reduced or absent in affected limb	Asymmetrical, Acute and affecting only one leg	Commonly Present before, during and after Flaccid Paralysis	Hours to four days (acute)	***Traumatic Neuritis***
Reduced in affected limbs	Asymmetrical-Hemiphegia more common on right	High and accompanied by convulsion	Hours to 1-2 days and sometimes > 3 days (acute)	***Transverse Myelitis***
Hypotonia in lower limbs	Acute, lower limbs, symmetrical	Rarely Present	Hours to four days + Pain (acute)	***Acute Childhood Hemiplegia***

Table 7 WHO's Differential Diagnosis Table for AFP [16][17][21]

Autonomic Signs & Symptoms	Respiratory Insufficiency	Cranial Nerve Involvement	Sensation	Deep-tendon Reflexes
Rare	Only when bulbar involvement is present	Only when bulbar involvement is present	Severe Myalgia, backache, no sensory changes	Decreased to absent
Frequent blood pressure alterations, sweating, blushing and body temperature fluctuations	In severe cases, enhanced by bacterial pneumonia	Often present affecting nerves VII, IX, X, XI, XII	Cramps, Tingling, hypo-anesthesia of palms and soles	Globally absent
Hypothermia in affected limb	Absent	Absent	Pains in gluteus, hypothermia	Decreased to absent
Absent	Absent	Facial nerve (VII)	Decreased in affected limbs	Decreased in Hemiplegic limbs early & hyper-reflexia late
Present	Sometimes	Absent	Anesthesia of lower limbs with sensory level	Absent in lower limbs early & hyper-reflexia late

Sequelae at Three Months and up to a	EMF at Three Weeks	Nerve Conduction Velocity: Third Week	Bladder Dysfunctio	Cerebro-Spinal Fluid
Severe, asymmetrical, atrophy, skeletal deformities developing later	Abnormal	Abnormal: anterior horn cell disease (normal during the first 2 weeks)	Absent	Inflammatory
Symmetrical atrophy of distal muscles	Normal	Abnormal: slowed conduction, decrease motor amplitudes	Transient	Albumin-Cytologic dissociation
Moderate atrophy, only in affected lower limb	Normal	Abnormal: axonal	None	Normal
Moderate atrophy	Normal	Normal	None	Depends on Aetiology: Infectious or
Flaccid diplegia, atrophy after years	Normal	Normal or abnormal, no diagnostic value	Present	Normal or Mild in cells

3.5 DATASET 2: AFP RULE BASE

The WHO standard table for the differential diagnosis of AFP has been transformed into a rule base dataset. The final column of the dataset is the diagnosis column which contains the labels of the dataset. The labels are the five diseases which are considered as a possible AFP case. The labels (diseases) are Poliomyelitis, GBS, Traumatic Neuritis, Acute Childhood Hemiplegia and Transverse Myelitis.

A total of 27 features have been extracted from the standard table, followed by the development of a rule base containing 5498 different combinations. Table 8 elaborates the dataset structure in terms of labels, features and values.

Table 8 Rule Base Dataset

No.	Feature Name	Feature Values
1	SEVERE Sequelae at 3 months & upto 1 Year	No, Yes
2	Sequelae Deformity Type	Atrophy, Atrophy years 1, Atrophy years 2, Atrophy years 3, Atrophy years 4, Distal Muscles, Lower Limb, Skeletal
3	Sequelae at 3 months & upto 1 Year	Asymmetric, Flaccid Diphlegia, Moderate Atrophy, Symmetrical
4	EMG at 3 weeks	Abnormal, Normal
5	Nerve Conduction Velocity Reason	Axonal Damage, Decreased Motor Amplitude, Interior Horn Cell, None, Normal, Slowed

		Conduction
6	Nerve Conduction Velocity	Abnormal, Normal
7	Bladder Dysfunction	Absent, No, Present, Transient
8	Cerebro-Spinal Fluid	Albumin-Cytologic dissociation, Infectious, Inflammatory, Normal, Vascular
9	Auto Signs & Symptoms	Absent, Blushing, Body Temperature Fluctuations, Frequent BP Alterations, Hypothermia, No, Present, Sweating
10	Respiratory Insufficiency	Absent, Bacterial Pneumonia, Bulbar Nerve, Yes
11	Cranial Nerve Involvement	Absent, Bulbar Nerve, Facial Nerve VII, Nerve VII, Nerve IX, Nerve X, Nerve XI, Nerve XII
12	Sensation Effect	Back Ache, Cramps, Decreased in Limbs, Hypothermia, Lower Limbs, No Sensory Change, Pains, Severe Myalgia, Tangling
13	Sensation Anesthesia	Lower Limbs, No, Palms, Soles
14	Deep-Tendon Reflexes EARLY	Absent, Decreased
15	Deep-Tendon Reflexes LATE	Absent, Decreased
16	Muscle Tone	Absent, Global Hypotonia, Lower Limb

		Hypotonia, Reduced
17	Flaccid Paralysis Effect	All, Distal, Lower Limbs, One Limb, Right Limbs
18	Flaccid Paralysis Symmetry	Asymmetric, Symmetrical
19	Flaccid Paralysis Proximal	No, Yes
20	Flaccid Paralysis Acute	No, Yes
21	Fever Convulsion	No, Yes
22	Fever After Onset	No, Yes
23	Fever Before Onset	No, Yes
24	Fever at Onset	High, No, Normal
25	Installation Of Paralysis Type	No, Yes
26	Installation Of Paralysis Pain	No, Yes
27	Installation Of Paralysis Time	3 days, 4 days, 5 days, 6 days, 7 days, 8 days, 9 days, 10 days, 6 hrs, 12 hrs, 18 hrs, 24 hrs, 30 hrs, 36 hrs, 42 hrs
28	Diagnosis (Label)	Polio Myelitis, Transverse Myelitis, GBS, Acute Childhood Hemephlegia, Traumatic Neuritis

Every Feature of the dataset has binomial or polynomial values, as shown in Table 8.

- The first attribute is *SEVERE Sequelae at 3 months & upto 1 Year,* have binary values *No and Yes*.
- The second attribute of the dataset is *Sequelae Deformity Type* have 8 polynomial values from *Atrophy* to *Skeletal.*
- The third attribute S*equelae at 3 months & upto 1 Year,* this attribute have 4 polynomial values.
- The fourth attribute is *EMG at 3 weeks;* have binary values *Abnormal and Normal.*
- The fifth attribute of the dataset is *Nerve Conduction Velocity Reason* with 6 polynomial values.
- The sixth attribute is *Nerve Conduction Velocity*, have 6 polynomial values.
- The seventh attribute *Bladder Dysfunction* and have 4 polynomial values.
- The eigth attribute is *Cerebro-Spinal Fluid* with 5 polynomial values.
- The ninth attribute *Auto Signs & Symptoms* with a huge variety of 8 attribute values.
- The tenth attribute of the rule base is *Respiratory Insufficiency* have 4 polynomial values.
- The eleventh attribute is *Cranial Nerve Involvement* have a variety of 8 polynomial values.
- the twelfth attribute *Sensation Effect* has 9 polynomial values.
- The thirteenth attribute is *Sensation Anesthesia* have 4 polynomial values.

- The next two attributes *Deep-Tendon Reflexes EARLY* and *Deep-Tendon Reflexes LATE* have similar binomial values *Absent* and *Decreased.*
- The sixteenth attribute is *Muscle Tone* with 4 polynomial values.
- The seventeenth attribute of the dataset *Flaccid Paralysis Effect* with 5 polynomial values.
- The eighteenth attribute is *Flaccid Paralysis Symmetry* with binomial values *Asymmetric* and *Symmetrical*.
- The next five attribute have binomial values *No* and *Yes* including *Flaccid Paralysis Proximal*, *Flaccid Paralysis Acute*, *Fever Convulsion*, *Fever after Onset* and *Fever before Onset*.
- *Fever at Onset* is the twenty-third attribute of the dataset with 3 polynomial values
- The next two attributes, *Installation of Paralysis TYPE* and *Installation of Paralysis PAIN* both have binomial values *No* and *Yes*.
- The last attribute is *Installation of Paralysis Time*, have a range on values with different timings starting from 6 hrs and ending up with the attribute value of *10 days*.
- Finally, the last header with the name of Diagnosis is the label column with five labels *Polio Myelitis, Transverse Myelitis, GBS, Acute Childhood Hemephlegia, and Traumatic Neuritis*.

Based on all these 27 attribute with their values, different combinations were created and a dataset of 5498 different training examples was developed. The next step was the training of

model based on this dataset for prediction of differential diagnosis. Decision Tree and Random Forest were the two classifiers selected for training and validation on the data.

3.6 INTRODUCTION TO MACHINE LEARNING

"Learning from data" is the short definition of Machine learning. It consists of numerous algorithms, which are used for learning from the pattern of data for decision making; that is why availability of data is a requirement for machine learning. Data can be defined as raw facts and figures, while machine learning utilizes these facts and figures to learn and make decisions. Data used in machine learning, usually available in rows and columns, is called a dataset. The rows of a dataset are the training examples, while columns are called features or attributes; which contain the properties of every example. Attributes of a dataset can be discrete or continuous. The difference between discrete and continuous is that discrete has a finite set of values while continuous attributes have infinite values [63].

The datasets can exist in two general categories, labeled dataset and unlabeled dataset. Data with the unique attribute is labeled, while the one without a unique attribute is unlabeled. When utilizing labeled data, the type of learning is known as supervised learning [63]. In case, the values of the unique attributes are numeric, e.g. predicting house prices, it is called a regression problem. If the values are categorical, such as categorizing emails as "spam" or "not spam", it is called classification problem.

On the contrary, if machine learning deals with unlabeled data, it is called unsupervised learning [64]. Unsupervised learning is utilized for extracting the hidden functional patterns in a dataset. Classification and regression are the problems of supervised learning while

unsupervised learning has two methods, clustering and association learning. Association learning is used to fin the relationship among the training dataset and its features while clustering is used for grouping the items with related properties in same groups.

The labeled datasets are used by the algorithms of machine learning to build a model which can learn the relationship exists among input and output features. The training data is given to the model for making prediction accurately with the help of an optimization algorithm. Apart from the training data, the two other types of data used in supervised learning are test and validation data. The test data is used for checking the performance of the model, while the validation data is used for improvement. Figure 21 illustrates machine learning framework.

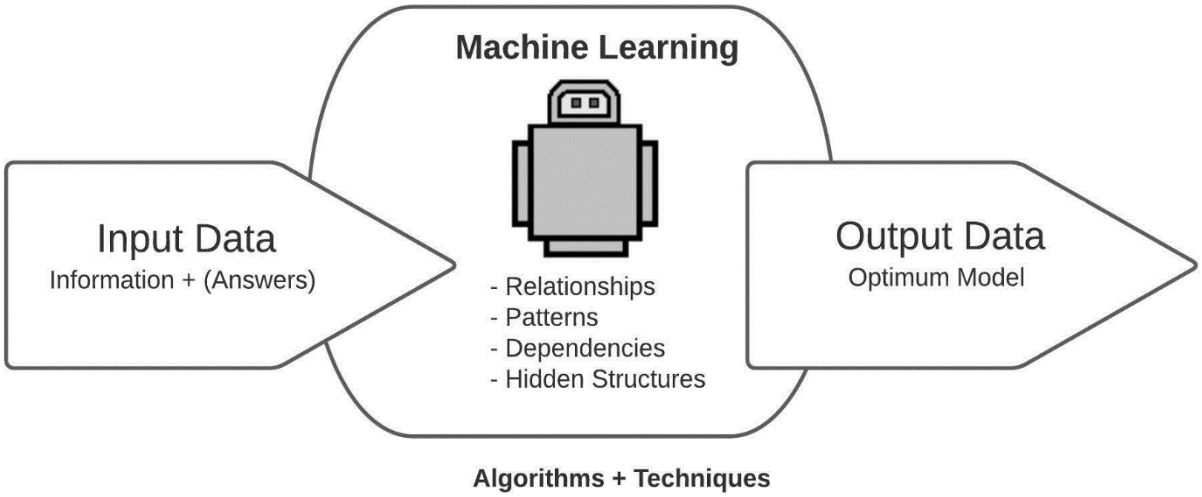


Figure 21 Machine learning framework

3.6.1 ALGORITHMS IN MACHINE LEARNING

Machine learning has two types of algorithms, linear and non-linear. In this research non-linear algorithms will be used. Linear regression is a well-known example of linear algorithms, along with logistic regression. Linear models cannot perform on non-linear separable data, which can be handled using non-linear algorithms. The well-known non-linear algorithms include K-Nearest Neighbor (KNN) classifier, Naïve Bayes Classifier, Neural Networks, Decision Tree classifier and Random Forest. In this research Decision Tree and Random Forest are used on the categorical data of the two datasets.

3.6.1.1 DECISION TREE

Decision tree is type of the predictive modeling methods applied in machine learning and data mining. Decision tree model can be defined as classification trees where the target variable takes a finite set of values. The branches of the trees are the links of attributes which lead the way to class labels, while the leaves of the tree are labels. The node of decision tree represents the attributes of the dataset. The topmost node in the tree is the parent node or root node. Attribute testing is performed on every internal node and the branches carry the result of testing while the leaf is the class label that possesses the result. [65]. Decision trees are normally used for categorical data but it can be utilized for numeric data as well. The performance of decision trees can be evaluated using confusion matrix. Accuracy, Precision, Class recall and F-measure can be calculated using the confusion matrix. Decision Tree is extensively used in providing solutions to health care problems [18][19][66][67].

3.6.1.2 RANDOM FOREST

Random Forest is an ensemble of Decision Tree classifier. The name itself is depiction of a forest of trees. Random Forest takes two parameters for training along with the training data; the first is the number of trees, while the second is the number of features selected randomly for the evaluation of each tree node. Random Forest is a powerful supervised classification ensemble with qualities like higher accuracy and robustness to noise. It can identify non-linear patterns in the data. As an ensemble, it uses the bagging method without pruning to create a forest of classifiers, where voting takes place for a certain class [68][69].

In order to compute the recall, precision and F-measure, Random Forest allows the adjustment of voting threshold. The accuracy is calculated using the Out of Bad Error (OOB). The OOB calculates the average misclassification ratio of unseen data to measure the accuracy. The performance of Random Forest can be evaluated similar to Decision Tree, using the 4 metrics of evaluation, accuracy, precision, recall and F-measure. Every tree in Random Forest has a single vote [68][69]. Every input is assigned to the most likely class label. Similar to this research of providing a novel solution to a specific health problem, Random Forest is extensively used in providing solutions to health care problems [70][71].

3.7 METHODOLOGY

3.7.1 PROPOSED FRAMEWORK

The proposed framework is composed of two modules. The first module is termed as Laboratory Diagnosis, while the second is termed as Clinical Diagnosis in this research. Figure 22 illustrates a generic understanding of the proposed framework. This diagram

reflects the diseases associated with AFP such as Poliomyelitis, GBS, ATM, and TrN etc. Besides these diseases, the diagram reflects the viruses associated with these diseases such as Poliovirus, Echovirus, and Coxsackie virus and enterovirues etc. Moreover Figure 22 elaborates the implementation of the two modules for the differential diagnosis of AFP; the laboratory diagnosis and the clinical diagnosis.

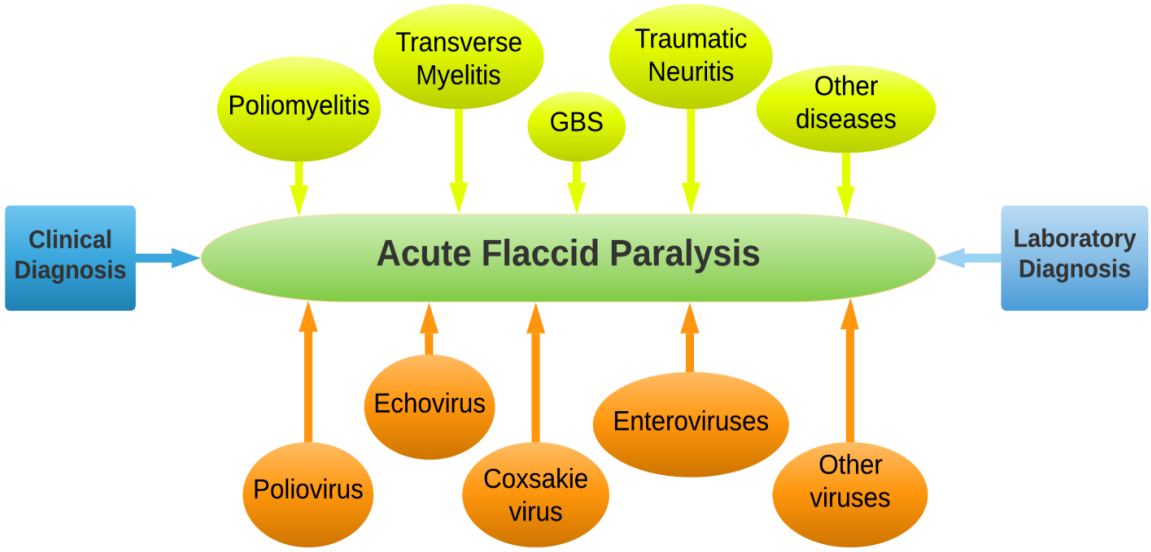


Figure 22 Proposed Framework

The reason for two different modules in the proposed framework is the involvement of two different datasets used for predicting the differential diagnosis of AFP. The first dataset termed as patient data, which consists of the AFP patient data from the line listing forms of WHO. Some features of this set include laboratory examination results, which is why this module is termed laboratory diagnosis in the proposed framework. The second module uses the rule base dataset for model training, which is derived from the standard differential diagnosis table of WHO. This module is termed clinical diagnosis in the proposed framework.

3.8 IMPLEMENTATION

3.8.1 MODULE I: LABORATORY DIAGNOSIS

The data of AFP patients from the line listing forms was transformed into a dataset via data preprocessing. Earlier in this chapter the patient dataset was discussed in detail, that dataset was used for training and validation the models using two algorithms, Decision Tree and Random Forest.

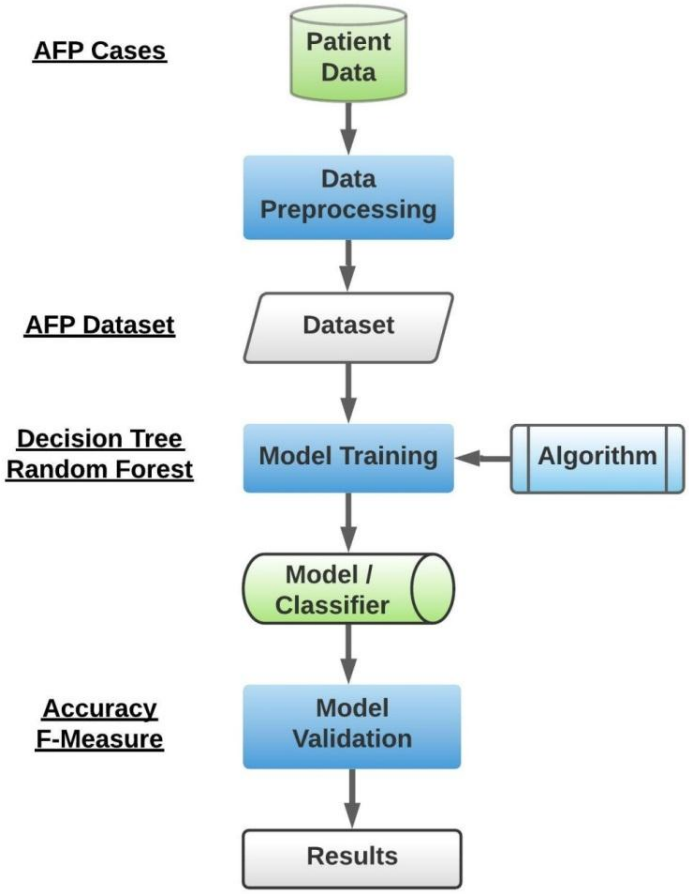


Figure 23 Module 1: Laboratory Diagnosis

Figure 23 illustrates the implementation process of laboratory diagnosis, which is one of the two modules of the proposed framework. The performance of the models was evaluated using the evaluation metrics Accuracy, Precision, Recall and F-measure [67].

The following are the implementation steps of module 1:

1. **Data Preprocessing:**
    (a) The line listing data was in raw form. The data was preprocessed; unnecessary rows were removed.
    (b) Data rows with relevant labels (GBS, Hypokalemia, Traumatic Neuritis, and Transverse Myelitis) were selected.
    (c) Rows with missing values were removed and the data was ready for the next step of feature selection.
    (d) Feature selection was performed on the basis of removing irrelevant features such as EPID Number, Name, and District etc. These features do not impact the model training and prediction process.
    (e) The dataset is finally ready for the next step, which is model training.
2. **Model Training:**
    Model training was performed with two different classifiers, Decision Tree and Random Forest, using the preprocessed dataset.
3. **Model Validation:**
    10-folds cross validation was performed and results were generated in a confusion matrix. The evaluation criteria includes: Accuracy, Precision, Recall and F-measure.

3.8.2 MODULE 2: CLINICAL DIAGNOSIS

Earlier in this chapter the rule base set was discussed in detail. This dataset is derived from the standard differential diagnosis table of AFP. After the initial step of transforming the table in a rule base dataset, it was used for training and validation the models using two algorithms, Decision Tree and Random Forest.

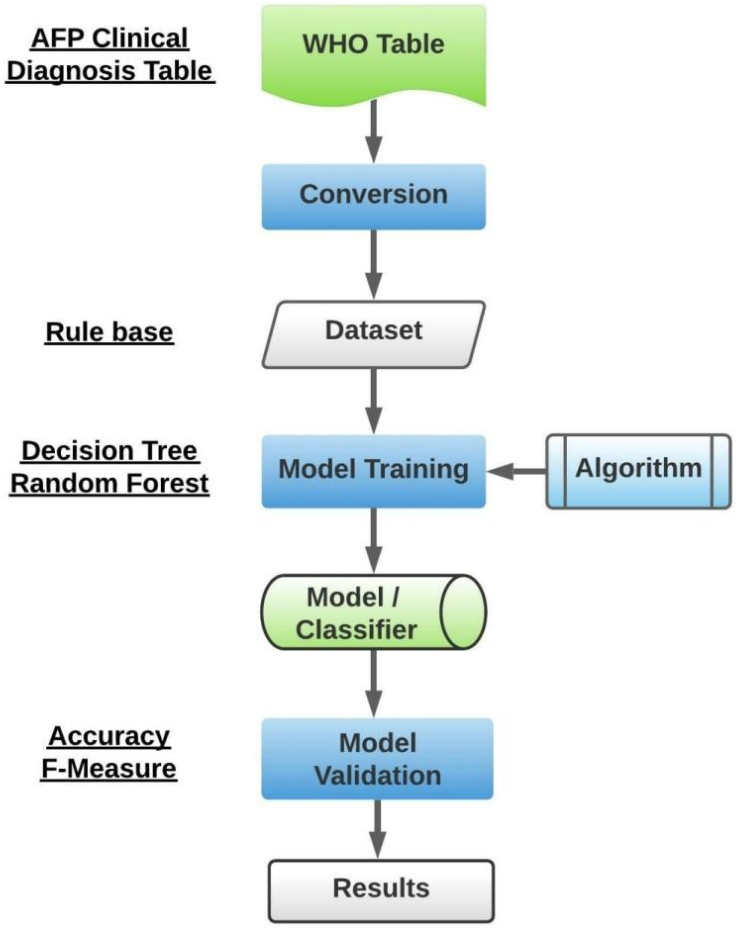


Figure 24 Module 2: Clinical Diagnosis

Figure 24 illustrates the implementation process of clinical diagnosis, which the second module of the proposed framework. The performance of the two classifier evaluated using the evaluation metrics Accuracy, Precision, Recall and F-measure [67].

The following are the implementation steps of module 1:

1. **Rule base creation:**
    (a) The 14 rows of the AFP differential diagnosis table were converted into rows.
    (b) Those 14 columns headings are treated as the features of the rule base dataset but first some steps will be performed.
    (c) Some of the features holding multiple conditions further divided and new features derived from the original 14.
    (d) The rule base data has now a total of 27 features.
    (e) The rows of differential diagnosis table for AFP were converted into columns and the column headings of the original table are now the labels of the rule base, which are: Polio, GBS, Traumatic Neuritis, Hemiplegia, and Transverse Myelitis.
    (f) Now the rule base is consisting of 28 columns, 27 of them are features and the final column represents the labels.
    (g) Finally, based on the differential diagnosis table, a total of 5498 rules were generated.
    (h) With the completion of step (g), the differential diagnosis table for AFP is converted into a rule base dataset; ready for model training.
2. **Model Training:**
    Model training was performed with two different classifiers, Decision Tree and Random Forest, using the rule base dataset.

3. **Validation:**

   10-folds cross validation was performed and results were generated in a confusion matrix. The evaluation criteria includes: Accuracy, Precision, Recall and F-measure.

3.9 SUMMARY

In this chapter, two datasets were discussed in details. These datasets were used for the two modules of the proposed framework. The laboratory diagnosis module used the dataset termed as the patient data, while the clinical diagnosis module used the rule base dataset derived from the standard differential diagnosis table of AFP. The two datasets were used for model training to predict the diagnosis of diseases associated with AFP. 10-folds cross validation was performed and evaluation metrics like accuracy, precision, recall and F-measured the performance of the classifiers. The results of the two modules are discussed in the next chapter.

CHAPTER 4
RESULTS

4.1 INTRODUCTION

The results of the two implemented methods are discussed in this chapter. Decision Tree and Random Forest were the algorithms selected for training and validation of the prediction model. The following is the evaluation criteria, based on which the results were calculated.

4.2 EVALUATION METRICS

Two different datasets were used for two different kinds of purposes. The first model was trained and evaluated on the patient data while the second model was trained and evaluated on the rule based developed from the standard table of WHO. In order to evaluate the models, there were four different performance measurement metrics known as accuracy, precision, recall and F-measure [67]. These four methods were used in this research to evaluate the performance of the two models, trained by two different datasets using the two classification algorithms, Decision Tree and Random Forest.

Accuracy, Precision, Recall, and F-measure are calculated using confusion matrix. A confusion matrix holds four types of values, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These terminologies hold inherent meaning [67]:

- TP: When the model predicts a label (positive), which is actually positive, called TP.
- TN: When the model predicts a label (negative), which is actually negative, called TP.
- FP: When the model predicts a label (positive), which is actually negative, called FP.

- FN: When the model predicts a label (negative), which is actually positive, called TP.

4.2.1 ACCURACY

Accuracy should be high for a prediction model; it illustrates the performance of the model as how many labels are truly predicted out of the total predicted labels.

Equation 2 Accuracy

$$\boldsymbol{Accuracy} = (\boldsymbol{TN} + \boldsymbol{TP})/(\boldsymbol{TN} + \boldsymbol{FP} + \boldsymbol{TP} + \boldsymbol{FN})$$

4.2.2 PRECISION

A high precision value means the classifier is performing well. The highest precision value could be 1, which means the prediction is 100% precise. The precision value can be 0 to 1. Precision shows, how many of the total positive predicted values are predicted truly positive.

Equation 3 Precision

$$\boldsymbol{Precision} = \boldsymbol{TP}/(\boldsymbol{TP} + \boldsymbol{FP})$$

4.2.3 RECALL

Similar to precision, a high recall value means the classifier is performing well. The highest recall value could be 1, which means the denominator and numerator are equal and the result is 100%, when the FN is zero.

Equation 4 Recall

$$\boldsymbol{Recall} = \boldsymbol{TP}/(\boldsymbol{TP} + \boldsymbol{FN})$$

4.2.4 F-MEASURE

Also known as F1-Score, the F-measure takes the other two metrics, Precision and Recall, and combine into a formula. The formula reveals that the value of F-measure could be equal to 1 only when the value of both, Precision and Recall is 1. Basically F-measure is the harmonic means of the other two metrics and it's a better measure than the others. A high F-measure reflects the good performance of a classifier.

Equation 5 F-measure

$$F\text{–}measure = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right)$$

4.3 RESULTS MODULE 1: LABORATORY DIAGNOSIS

Decision Tree and Random Forest were used for model training and validation in method 1 using the patient dataset of AFP. The results of Decision Tree and Random Forest classifiers were generated in the form of confusion matrix. Accuracy, precision, recall and F-measure were calculated. Random Forest proved to be the better among the two classifiers in terms of performance, given the evaluation criteria, including the four metrics. Table 9 and 10 shows the confusion matrices for Decision Tree and Random Forest respectively.

The heading row of the confusion tables represents 'True' labels of the dataset. The labels include GBS, Hypekalemia, Traumatic Neuritis, and Transverse Myelitis. The first column of the confusion tables represents the predicted values of the label mentioned earlier. The Last column is holds the values of class precision while the last row holds the values of class recall.

Table 9 Confusion Matrix, Decision Tree, Module 1

	True GBS	True Hypokalemia	True Traumatic Neuritis	True Transverse Myelitis	Class Precision
Predicted GBS	286	36	23	12	80.11%
Predicted Hypokalemia	7	55	11	0	75.34%
Predicted Traumatic Neuritis	30	23	617	5	91.41%
Predicted Transverse Myelitis	0	0	0	5	100%
Class Recall	88.54%	48.25%	94.78%	22.73%	

Table 10 Confusion Matrix, Random Forest, Module 1

	True GBS	True Hypokalemia	True Traumatic Neuritis	True Transverse Myelitis	Class Precision
Predicted GBS	304	26	7	8	88.12%
Predicted Hypokalemia	1	68	2	0	95.77%
Predicted Traumatic Neuritis	18	20	642	2	94.13%
Predicted Transverse Myelitis	0	0	0	12	100%
Class Recall	94.12%	59.65%	98.62%	54.55%	

Table 11 Module 1: Laboratory Diagnosis Results (a)

Label	Decision Tree		Random Forest	
	Precision	Recall	Precision	Recall
GBS	80.11	88.54	88.12	94.12
Hypokalemia	75.34	48.25	95.77	59.65
Traumatic Neuritis	94.78	91.41	94.13	98.62
Transvers Myelitis	100	22.73	100	54.55

Table 11 illustrates the combined results of the two confusion matrices generated as a result of the 2 classifiers, Decision Tree and Random Forest. The first column is consists of the dataset labels. The second column illustrates the precision and Recall results for Decision Tree, while the last column represents the precision and Recall results for Random Forest.

Table 12 Module 1: Laboratory Diagnosis Results (b)

Algorithm	Accuracy	F-measure
Decision Tree	86.76%	73.28
Random Forest	92.43%	84.7

Table 12 clearly shows that Random Forest has better average Class Precision and Recall percentages. The accuracy of Decision Tree is 86.76% while the best suited algorithm for this AFP dataset, Random forest, predicted the diseases of differential diagnosis with an accuracy of 92.43%. As the data in the dataset is imbalanced therefore F-measure is important for this dataset and Random Forest has a better F-measure value of 84.7.

4.4 RESULTS MODULE 2: CLINICAL DIAGNOSIS

Model training was performed with the second dataset using algorithms Decision Tree and Random Forest. The confusion matrix in Table 13 shows the performance results of Decision Tree classifier, while Table 14 illustrates the performance of Random Forest. Accuracy, precision, recall and F-measure were calculated. Random Forest proved to be the better among the two classifiers, in terms of performance given the evaluation criteria, including the four metrics.

Table 13 Confusion Matrix, Decision Tree, Module 2

	True Polio	True GBS	True Traumatic Neuritis	True Hemiplegia	True Transverse Myelitis	Class Precision
Predicted Polio	120	0	0	0	0	100%
Predicted GBS	0	5120	0	0	32	99.38%
Predicted Traumatic Neuritis	0	0	160	0	0	100%
Predicted Hemiplegia	0	0	0	18	0	100%
Predicted Transverse Myelitis	0	0	0	0	48	100%
Class Recall	100%	100%	100%	100%	60%	

The heading row of the confusion tables represents ‘True’ labels of the dataset. The labels include Polio, GBS, Traumatic Neuritis, Hemiplegia and Transverse Myelitis. The first column of the confusion tables represents the predicted values of the label mentioned

earlier. The Last column is holds the values of class precision while the last row holds the values of class recall.

Table 14 Confusion Matrix, Random Forest, Module 2

	True Polio	True GBS	True Traumatic Neuritis	True Hemiplegia	True Transverse Myelitis	Class Precision
Predicted Polio	120	0	0	0	0	100%
Predicted GBS	0	5120	0	0	32	100%
Predicted Traumatic Neuritis	0	0	160	0	0	100%
Predicted Hemiplegia	0	0	0	18	0	100%
Predicted Transverse Myelitis	0	0	0	0	80	100%
Class Recall	100%	100%	100%	100%	100%	

Table 15 Module 2: Clinical Diagnosis Results (a)

Label	Decision Tree		Random Forest	
	Precision	Recall	Precision	Recall
Polio	100	100	100	100
GBS	100	99.38	100	100
Traumatic Neuritis	100	100	100	100
Hemiplegia	100	100	100	100
Transvers Myelitis	60	100	100	100

Table 15 represents the combine results of the two confusion matrices generated as a result of the 2 classifiers; Decision Tree and Random Forest. The first column is consists of the dataset labels. The second column illustrates the precision and Recall results for Decision Tree while the last column represents the precision and Recall results for Random Forest.

Table 16 Module 2: Clinical Diagnosis Results (b)

Algorithm	Accuracy	F-measure
Decision Tree	99.4%	95.2
Random Forest	100%	100

Table 16 illustrates the result of the performance for both classification models. Decision Tree has an accuracy of 99.4%, while Random Forest has 100% accuracy. As the data in the dataset is imbalanced therefore F-measure is important for this dataset. The F-measure value is also mentioned in table 17. Decision Tree has 95.2 which a high value and has a good performance, where 4 out of 5 labels has 100% Precision or Recall value. On the other hand, Random Forest has 100% Precision and Recall, that is why its F-measure value is 100 and it is even more efficient then Decision Tree.

4.5 SUMMARY

In this chapter, the results of two classifiers were discussed. Two datasets were used in the process. Separate models were trained using those datasets. In the Laboratory diagnosis method, the model was trained using the AFP patient while in the Clinical diagnosis method the model was trained using the dataset developed from the standard table of WHO. The

performance of each model was evaluated. In the laboratory diagnosis method Decision Tree algorithm has 86.76% accuracy and F-measure value of 73.28, while Random Forest algorithm has 92.43% accuracy and F-measure value of 84.7. In the Clinical diagnosis method Decision Tree algorithm has 99.4% accuracy and F-measure value of 95.2 while Random Forest algorithm has 100% accuracy and 100 F-measure value.

CHAPTER 5
CONCLUSIONS

5.1 INTRODUCTION

In this research, AFP, its associated diseases, viruses, and vaccines were reviewed. The methods of WHO regarding the differential diagnosis of AFP were also reviewed. An automated novel framework of two distinct methodologies was proposed and implemented as an alternative to the current manual procedure for the differential diagnosis of AFP. The results were analyzed; now this research is concluded in this chapter along with the discussion of future work.

5.2 APPLICATIONS OF THIS RESEARCH

Pakistan, along with a few other countries, is struggling to eradicate Poliomyelitis. Diagnosis of AFP is a long and expensive process. This research can help in:

- Automated Clinical Differential Diagnosis of AFP Diseases including Poliomyelitis
- Automated Differential Diagnosis of non-polio AFP based on patient data

5.3 CONCLUSION

Polio eradication is one of the most important achievements of WHO as it has been eradicated from most parts of the world. AFP surveillance played an integral part is the eradication of Poliomyelitis from the globe, but still countries like Pakistan, Afghanistan and Nigeria are not declared as polio-free states. Every year, thousands of AFP cases are

being diagnosed, which is a threat to other countries as it can spread. AFP surveillance ensures the isolation of viruses which causes AFP. In this regard, the differential diagnosis process is the core focus of WHO. Diagnosis among different diseases like Poliomyelitis, GBS, ATM, TrN along with other enterovirus diseases has a slow and lengthy process as it is performed manually and gradually. The nature of manual differential diagnosis process is data driven. Data through multiple forms generated for every reported AFP case which makes it possible for data science to efficiently diagnose the diseases of a particular AFP patient.

In this research, a framework was proposed against the manual diagnosis process currently being used by WHO. The purpose of the framework was to automate the diagnosis process in the most efficient manner to save time. Both clinical and laboratory diagnosis were looked upon in detail within this research and a single automated framework was created to accurately imitate the manual diagnosis process for both.

The rough data of AFP cases in the line listing table was preprocessed, and then based on that dataset, model training was performed with two different classifiers, the Decision Tree and Random Forest. The performance of the each model was evaluated using the metrics Accuracy, Precision, Recall and F-measure. Decision Tree algorithm has 86.76% accuracy and F-measure value of 73.28 while Random Forest algorithm has 92.43% accuracy and F-measure value of 84.7. Random Forest proved to be the better classifier.

As discussed in methodology section, the proposed framework in this research had two distinct methods. The clinical diagnosis method used the standard differential diagnosis

table of WHO, which was converted into a rule base dataset. Using the rule base dataset model training was performed using Decision Tree and Random Forest algorithms. In the Clinical diagnosis method, Decision Tree algorithm had 99.4% accuracy and F-measure value of 95.2 while Random Forest algorithm had 100% accuracy and 100 F-measure value. Random Forest proved to be the better classifier once again.

REFERENCES

[1] Marx, A & Glass, J & Sutter, R. (2000), **Differential Diagnosis of Acute Flaccid Paralysis and its Role in Poliomyelitis Surveillance.** Epidemiologic reviews. 22.298-316. 10.1093/oxfordjournals.epirev. a018041.

[2] PI Diseases Surveillance Guideline 3rd Edition (2015), EPI South Africa (EPISA). Last Accessed June 2021.

https://www.nicd.ac.za/assets/files/EPI%20Surveillance%20Manual_15Dec2015.pdf

[3] Jasem, Jagar & Marof, Kawa & Nawar, Adnan & Khalaf, Yosra & Al-Hamdani, Faisal & Ali, Sagvan & Kalil, Andre & Islam, Km. (2014), **An epidemiological analysis of acute flaccid paralysis and its surveillance system in Iraq, 1997-2011.** BMC infectious diseases. 2014.

[4] Kilmer, Michele & Shreve, Marilou & Jarrett, Anna. (2019), **Clarifying the Diagnosis of Acute Flaccid Myelitis. The Journal for Nurse Practitioners.** 15. 10.1016/j.nurpra.2019.01.007. Characterization of non-polio enterovirus isolates from acute flaccid paralysis children in Pakistan reflects a new genotype of EV-107.

[5] Shahmahmoodi, Shohreh & Mehrabi, Zahra & Eshraghian, Mohammad & Mokhtari-Azad, Talat & Tabatabaie, Hamideh & Yousefi, Maryam & Farrokhi, Kobra & Gouya, Mohammad & Esteghamati, Abolreza & Moosavi, Taha & Zahraie, Mohsen & Rad, Katayoon & Shokati, Zahra & Nategh, Rakhshandeh. (2008), **First detection of enterovirus 71 from an acute flaccid paralysis case with residual paralysis in Iran.** Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology. 42. 409-11. 10.1016/j.jcv.2008.02.013.

[6] Solomon, Tom & Willison, Hugh. (2003), **Infectious causes of acute flaccid paralysis.** 16. 375-81. 10.1097/01.qco.0000092807.64370.1e.Polio Myelitis Eradication Field Guide.

[7] Arita, Minetaro & Ling, Hua & Yan, Dongmei & Nishimura, Yorihiro & Yoshida, Hiromu & Wakita, Takaji & Shimizu, Hiroyuki. (2009), **Development of a reverse transcription-loop-mediated isothermal amplification (RT-LAMP) system for a highly sensitive detection of enterovirus in the stool samples of acute flaccid paralysis cases.** BMC infectious diseases. 9. 208. 10.1186/1471-2334-9-208.

[8] Saeed, Mohsan & Zaidi, Sohail & Naeem, Asif & Alam, Muhammad Masroor & Sharif, Salmaan & Shaukat, Shahzad & Angez, Mehar & Khan, Anis. (2007),

Epidemiology and clinical findings associated with enteroviral acute flaccid paralysis in Pakistan. BMC infectious diseases. 7. 6. 10.1186/1471-2334-7-6.

[9] Shaukat, Shahzad & Angez, Mehar & Alam, Muhammad Masroor & Jebbink, Maarten & Deijs, Martin & Canuti, Marta & Sharif, Salmaan & Vries, Michel & Khurshid, Adnan & Mahmood, Tariq & van der hoek, Lia & Zaidi, Sohail. (2014), **Identification and characterization of unrecognized viruses in stool samples of non-polio acute flaccid paralysis children by simplified VIDISCA.** Virology journal. 11. 146. 10.1186/1743-422X-11-146.

[10] Bitnun, Ari & Yeh, E.. (2018), **Acute Flaccid Paralysis and Enteroviral Infections.** Current Infectious Disease Reports. 20. 10.1007/s11908-018-0641-x.

[11] Chen, Feng & Li, Jian-Jun & Liu, Tao & Wen, Guo-Qiang & Xiang, Wei. (2013), **Clinical and neuroimaging features of enterovirus71 related acute flaccid paralysis in patients with hand-foot-mouth disease.** Asian Pacific journal of tropical medicine. 6. 68-72. 10.1016/S1995-7645(12)60203-X.

[12] **Causes of Acute Flaccid Paralysis (AFP) Worldwide. Health Protein Surveillance System.** Last Accessed June 2021.

https://www.hpsc.ie/a-z/vaccinepreventable/polio/acuteflaccidparalysisafp/guidance/File,14207,en.pdf

[13] Wang, Haibo & Luo, Hui-Ming & li, Machao & Fan, Chun-Xiang & Hao, Li-Xin & Ma, Chao & Su, Qi-Ru & Yang, Hong & Reilly, Kathleen & Wang, Hua-Qing & Wen, Ning. (2017), **Vaccine-derived Poliovirus surveillance in China during 2001-2013: The potential challenge for maintaining polio free status.** BMC Infectious Diseases. 17. 10.1186/s12879-017-2849-z.

[14] Halawa, Eman & Ahmed, Dalia & Nada, Mona. (2010), **Guillain-Barre syndrome as a prominent cause of childhood acute flaccid paralysis in post polio eradication era in Egypt. European journal of paediatric neurology** : EJPN : official journal of the European Paediatric Neurology Society. 15. 241-6. 10.1016/j.ejpn.2010.11.008. Differential diagnosis of flaccid paralysis in paediatric medicine.

[15] Soltani, Jafar & Esmailnasab, Nader & Roshani, Daem & Karimi, Mohamad & Amjadi, Mohamad-Jamil. (2014), **Acute Flaccid Paralysis and Its Differential Diagnosis in in Kurdistan Province, Western Iran; an 11-Year Surveillance.** Iranian journal of pediatrics. 24. 131-139. [20] 978-1-4612-6074-5_27 springer chapter.

[16] **Field Guide for supplementary activities and achieving polio eradication, 1996.** Last Accessed June 2021.

https://www.who.int/ihr/polio1996en.pdf?ua=1

[17] **WHO Expanded Program of Immunization AFP Field Manual, Republic of Iraq.** Last Accessed June 2021.

https://applications.emro.who.int/dsaf/libcat/EMROPD_2009_105.pdf

[18] Ghiasi MM, Zendehboudi S, Mohsenipour AA. **Decision tree-based diagnosis of coronary artery disease: CART model.** Comput Methods Programs Biomed. (2020), Aug;192:105400. doi: 10.1016/j.cmpb.2020.105400. Epub 2020 Feb 19. PMID: 32179311.

[19] Nahar, Nazmun & Ara, Ferdous. (2018), **Liver Disease Prediction by Using Different Decision Tree Techniques.** International Journal of Data Mining & Knowledge Management Process. 8. 01-09. 10.5121/ijdkp.2018.8201.

[20] Abdel-Fattaha,, Amgad & El-Gilany, Abdel-Hady & El-masry, Ragaa & Kanddeel, Amr. (2019), **Acute flaccid paralysis in North East Delta, Egypt: A retrospectiveanalysis of prospectively collected surveillance data.** Journal of Infection and Public Health. 10.1016/j.jiph.2019.03.016.

[21] **Polio Myelitis Eradication Field Guide. Third Edition.** Last Accessed June 2021.

https://www.paho.org/en/documents/Poliomyelitis-eradication-field-guide-third-edition-2006

[22] **Polio Eradication & Endgame Strategic Plan 2013-2018.** Last Accessed June 2021.

https://polioeradication.org/wp-content/uploads/2016/07/PEESP_EN_A4.pdf

[23] **SA Acute Flaccid Paralysis (AFP) Case Investigation Form (CIF).** Last Accessed June 2021.

https://www.nicd.ac.za/assets/files/AFP_CIF_and_Specimen_Collection_Guide.pdf

[24] **Polio Now. Polio Global Eradication Initiative.** Last Accessed September 2021.

https://polioeradication.org/polio-today/polio-now/

[25] Hobday LK, Thorley BR, Alexander J, et al. (2013), **Potential for the Australian and New Zealand paediatric intensive care registry to enhance acute flaccid paralysis**

surveillance in Australia: a data-linkage study. BMC Infect Dis. 2013;13:384. doi:10.1186/1471-2334-13-384.

[26] Watkins, Rochelle & Martin, P & Kelly, Heath & Madin, Ben & Watson, Charles. (2009), **An evaluation of the sensitivity of acute flaccid paralysis surveillance for Poliovirus infection in Australia.** BMC infectious diseases. 9. 162. 10.1186/1471-2334-9-162.

[27] **Best Practices In Active Surveillance For Polio Eradication.** Last Accessed June 2021. https://www.who.int/polio-transition/documents-resources/best-practices-active-surveillance.pdf?ua=1

[28] **World Health Organization Polis Data.** Last Accessed June 2021.

https://extranet.who.int/polis/public/CaseCount.aspx

[29] D'Errico, Marcello & Barbadoro, Pamela & Bacelli, Sonia & Esposto, Elisabetta & Moroni, Vania & Scaccia, Federica & Tantucci, Luana & Prospero, Emilia. (2008), **Surveillance of acute flaccid paralysis in the Marches region (Italy): 1997-2007.** BMC infectious diseases. 8. 135. 10.1186/1471-2334-8-135.

[30] Muscat, Mark & Fiore, Lucia & Busuttil, Ray & Gilles, Herbert. (2000), **Surveillance of Wild Polioviruses in patients with acute flaccid paralysis in Malta during 1998 and 1999.** European journal of epidemiology. 16. 1057-60. 10.1023/A:1010848925903.

[31] Shaukat S, Angez M, Alam MM, Sharif S, Khurshid A, Mahmood T, Zaidi SS. (2012), **Characterization of non-polio enterovirus isolates from acute flaccid paralysis children in Pakistan reflects a new genotype of EV-107.** Virus Res. 2012 Dec;170(1-2):164-8. doi: 10.1016/j.virusres.2012.09.010. Epub. PMID: 23041515.

[32] Molina-Giraldo P, Ulate-Campos A, Petanàs-Argemí J, Rebollo Polo M, González-Álvarez V. (2016), **Differential diagnosis of flaccid paralysis in paediatric medicine.** Neurologia. 31(7): 500-1. English, Spanish. doi: 10.1016/j.nrl.2014.12.012. Epub 2015 Feb 14. PMID: 25687680.

[33] Jasem, J., Marof, K., Nawar, A. *et al.* (2013), **Guillain-Barré syndrome as a cause of acute flaccid paralysis in Iraqi children: a result of 15 years of nation-wide study.** *BMC Neurol* **13,** 195. https://doi.org/10.1186/1471-2377-13-195

[34] Wen N. (2019), **Establishment and Development of the Disease Surveillance System.** In: Liang X. (eds) Immunization Program in China. Public Health in China, vol 3. Springer, Singapore. https://doi.org/10.1007/978-981-13-2438-3_3

[35] Fiore L, Novello F, Simeoni P, Amato C, Vellucci L, De Stefano D, Grandolfo ME, Luzzi I. (1999), **Surveillance of acute flaccid paralysis in Italy: 1996-1997.** AFP Study Group. Acute flaccid paralysis. Eur J Epidemiol. 15(8):757-63. doi: 10.1023/a:1007697421114. PMID: 10555620.

[36] Peluso, Christopher. (2020), **Importance of Selected Acute Flaccid Paralysis Diagnoses in an Emergency Department Setting for the Pediatric Population.** Current Emergency and Hospital Medicine Reports. 8. 10.1007/s40138-020-00217-2.

[37] Raji, I.A., Abubakar, A.U., Ahmad, A. *et al.* (2021), **Evaluation of acute flaccid paralysis surveillance indicators in Sokoto state, Nigeria, 2012–2019: a secondary data analysis.** *BMC Public Health* **21,** 1148. https://doi.org/10.1186/s12889-021-11238-1

[38] **News Room. World Health Organization.** Last Accessed June 2021.

https://www.who.int/news-room/fact-sheets/detail/Poliomyelitis

[39] Haqqi, A., Zahoor, S., Aftab, M. N., Tipu, I., Rehman, Y., Ahmed, H., & Afzal, M. S. (2021), **COVID-19 in Pakistan: Impact on global polio eradication initiative.** *Journal of medical virology*, *93*(1), 141–143. https://doi.org/10.1002/jmv.26240

[40] Shakeel, S.I., Brown, M., Sethi, S. et al. (2019), **Achieving the end game: employing "vaccine diplomacy" to eradicate polio in Pakistan.** BMC Public Health 19, 79. https://doi.org/10.1186/s12889-019-6393-1

[41] **Polio Endgame Strategy 2019-2023.** Last Accessed June 2021.

https://polioeradication.org/wp-content/uploads/2019/06/english-polio-endgame-strategy.pdf

[42] **Global Polio Eradication Initiative.** Last Accessed June 2021.

https://polioeradication.org/news-post/canada-announces-new-commitments-to-gpei-endgame-strategy/

[43] Mateen FJ, Black RE. (2013), **Expansion of acute flaccid paralysis surveillance: beyond poliomyelitis.** Trop Med Int Health. 18(11):1421-2. doi: 10.1111/tmi.12181. Epub 2013 Aug 29. PMID: 24033476.

[44] Nathanson N, Kew OM. (2010), **From emergence to eradication: the epidemiology of poliomyelitis deconstructed.** Am J Epidemiol. 172(11):1213-29. doi: 10.1093/aje/kwq320. Epub 2010 Oct 26. PMID: 20978089; PMCID: PMC2991634.

[45] Robert A Schwartz, Smeeta Sinha, Rajendra Kapila, Alexander Velazquez, Pratibha Dua. (2021), **Enteroviruses: Practice Essentials, Background, Pathophysiology.** Drugs and Diseases. Medscape. https://emedicine.medscape.com/article/217146-overview

[46] I. Laffont, M. Julia, V. Tiffreau, A. Yelnik, C. Herisson, J. Pelissier, (2010), **Aging and sequelae of poliomyelitis**, Annals of Physical and Rehabilitation Medicine, Volume 53, Issue 1, Pages 24-33, ISSN 1877-0657. https://doi.org/10.1016/j.rehab.2009.10.002.

[47] Mehndiratta MM, Mehndiratta P, Pande R. (2014), **Poliomyelitis: historical facts, epidemiology, and current challenges in eradication.** Neurohospitalist. 4(4):223-9. doi: 10.1177/1941874414533352. PMID: 25360208; PMCID: PMC4212416.

[48] Platt LR, Estívariz CF, Sutter RW. (2014), **Vaccine-associated paralytic poliomyelitis: a review of the epidemiology and estimation of the global burden.** J Infect Dis. 210 Suppl 1:S380-9. doi: 10.1093/infdis/jiu184. PMID: 25316859.

[49] Yuki N, Hartung HP. **Guillain-Barré syndrome. (**2012), N Engl J Med. 2012 Jun 14;366(24):2294-304. doi: 10.1056/NEJMra1114525. Erratum in: N Engl J Med. 367(17):1673. PMID: 22694000.

[50] Willison HJ, Jacobs BC, van Doorn PA. (2016), **Guillain-Barré syndrome.** Lancet. 2016 Aug 13;388(10045):717-27. doi: 10.1016/S0140-6736(16)00339-1. Epub. PMID: 26948435.

[51] Hughes RA, Cornblath DR. (2005), **Guillain-Barré syndrome.** Lancet 366(9497):1653-66. doi: 10.1016/S0140-6736(05)67665-9. PMID: 16271648.

[52] Bhat A, Naguwa S, Cheema G, Gershwin ME. (2010), **The epidemiology of transverse myelitis.** Autoimmun Rev. 9(5):A395-9. doi: 10.1016/j.autrev.2009.12.007. Epub 2009 Dec 24. PMID: 20035902.

[53] Jacob A, Weinshenker BG. (2008), **An approach to the diagnosis of acute transverse myelitis.** Semin Neurol. 28(1):105-20. doi: 10.1055/s-2007-1019132. PMID: 18256991.

[54] Beh SC, Greenberg BM, Frohman T, Frohman EM. (2013), **Transverse myelitis.** Neurol Clin. 31(1):79-138. doi: 10.1016/j.ncl.2012.09.008. PMID: 23186897; PMCID: PMC7132741.

[55] Singhi, S.C., Sankhyan, N., Shah, R. et al. (2012), **Approach to a Child with Acute Flaccid Paralysis.** Indian J Pediatr 79, 1351–1357. https://doi.org/10.1007/s12098-012-0831-8

[56] Kardalas, E., Paschou, S. A., Anagnostis, P., Muscogiuri, G., Siasos, G., & Vryonidou, A. (2018), **Hypokalemia: a clinical update.** Endocrine connections, 7(4), R135–R146. https://doi.org/10.1530/EC-18-0109

[57] Chinnabhandar, V., Singh, A., Mandal, A., & Parmar, B. J. (2018), **Acute Hemiplegia in Children: A Prospective Study of Etiology, Clinical Presentation, and Outcome from Western India.** Journal of neurosciences in rural practice, 9(4), 504–509. https://doi.org/10.4103/jnrp.jnrp_574_17

[58] Zubler F, Seeck M, Landis T, et al**. (**2003), **Contralateral medial temporal lobe damage in right but not left temporal lobe epilepsy: a 1H magnetic resonance spectroscopy study**. Journal of Neurology, Neurosurgery & Psychiatry 74:1240-1244.

[59] Ivanova OE, Yurashko OV, Eremeeva TP, Baikova OY, Morozova NS, Lukashev AN. (2011), **Adenovirus isolation rates in acute flaccid paralysis patients.** J Med Virol. 2012 Jan;84(1):75-80. doi: 10.1002/jmv.22265. Epub. PMID: 22052705.

[60] Rich Caruana and Alexandru Niculescu-Mizil. (2006), **An empirical comparison of supervised learning algorithms.** In Proceedings of the 23rd international conference on Machine learning (ICML '06). Association for Computing Machinery, New York, NY, USA, 161–168. DOI:https://doi.org/10.1145/1143844.1143865

[61] **Vaccine-associated paralytic polio (VAPP) and vaccine-derived poliovirus (VDPV). EPI, Polio Global Eradication Initiative.** Last Accessed June 2021.

https://www.who.int/immunization/diseases/poliomyelitis/endgame_objective2/oral_polio_vaccine/VAPPandcVDPVFactSheet-Feb2015.pdf

[62] **Annual Report 2020. Polio Global Eradication Initiative.** Last Accessed June 2021.

https://polioeradication.org/wp-content/uploads/2021/08/GPEI-2020-Annual-Report-ISBN-9789240030763.pdf

[63] Uddin, Shahadat, Arif Khan, Md Ekramul Hossain, and Mohammad Ali Moni. (2019), **Comparing different supervised machine learning algorithms for disease prediction.** BMC medical informatics and decision making 19, no. 1 (2019) 1-16

[64] Shailaja, K., B. Seetharamulu, and M. A. Jabbar. (2018), **Machine learning in healthcare: A review.** In 2018 Second international conference on electronics, communication and aerospace technology (ICECA), pp. 910-914. IEEE.

[65] Sharma, Himani & Kumar, Sunil. (2016), **A Survey on Decision Tree Algorithms of Classification in Data Mining.** International Journal of Science and Research (IJSR). 5.

[66] Patel, Harsh & Prajapati, Purvi. (2018), **Study and Analysis of Decision Tree Based Classification Algorithms.** International Journal of Computer Sciences and Engineering. 6. 74-78. 10.26438/ijcse/v6i10.7478.

[67] Hossin, Mohammad & M.N, Sulaiman. (2015), **A Review on Evaluation Metrics for Data Classification Evaluations.** International Journal of Data Mining & Knowledge Management Process. 5. 01-11. 10.5121/ijdkp.2015.5201.

[68] Chaudhary, Archana & Kolhe, Savita & Kamal, Raj. (2016), **An improved Random Forest Classifier for multi-class classification.** Information Processing in Agriculture. 3. 10.1016/j.inpa.2016.08.002.

[69] Petkovic D, Altman R, Wong M, Vigil A. (2018), **Improving the explainability of Random Forest classifier - user centered approach.** Pac Symp Biocomput. 23:204-215. PMID: 29218882; PMCID: PMC5728671.

[70] Guanglu Sun, Shaobo Li, Yanzhen Cao, and Fei Lang. (2017), **Cervical Cancer Diagnosis based on Random Forest.** [J]. Int J Performability Eng, 13(4): 446-457.

[71] Akhil, Jabbar & Deekshatulu, Bulusu & Chandra, Priti. (2016), **Intelligent heart disease prediction system using random forest and evolutionary approach.** Journal of network and innovative computing. 4. 175-184.