E-HEALTH SYSTEM FOR EARLY DIAGNOSIS OF INHERITED PRENATAL DISORDERS

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

IN

COMPUTER SCIENCE & ENGINEERING

By

Pushpa

Registration Number: 11312729

Supervised By
Dr. Anil Sharma
Professor
School of Computer Applications
Lovely Professional University
Phagwara, Punjab, India

Co-Supervised By
Dr. Lalit Garg
Professor
Department of Computer Information System
University of Malta
Msida, Malta

LOVELY PROFESSIONAL UNIVERSITY

PUNJAB

2024

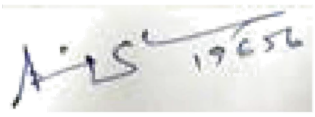
DECLARATION

I, hereby declare that the presented work in the thesis entitled "E-Health System for Early Diagnosis of Inherited Prenatal Disorders" in fulfilment of degree of **Doctor of Philosophy (Ph.D)** is outcome of research work carried out by me under the supervision of Dr. Anil Sharma working as Professor in the School of Computer Applications of Lovely Professional University, Punjab, India and Dr. Lalit Garg working as Professor in the Department of Computer Information Systems of University of Malta, Msida, Malta. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)
Pushpa
Registration No.:11312729
School of Computer Science and Engineering
Lovely Professional University
Phagwara, Punjab, India

CERTIFICATE

This is to certified that the work reported in the Ph.D thesis entitled "**E-Health System for Early Diagnosis of Inherited Prenatal Disorders**" submitted in fulfilment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D)** in the Computer Science and Engineering, is a research work carried out by Pushpa, Registration No: 11312729, is bonafide record of her original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.



(Signature of Supervisor)

Dr. Anil Sharma
Professor
School of Computer Application
Lovely Professional University
Phagwara, Punjab, India

(Signature of Co-Supervisor)

Dr. Lalit Garg
Professor
Department of Computer Information System
University of Malta
Msida, Malta

Abstract

Information and technology-based e-Health systems play a vital role in healthcare. From remote consultancy systems to teaching and learning systems, all these have been widely deployed to serve different sectors of the medical domain along with associated remote patients. As a result, e-Health systems are increasingly being used as both a passive and active strategy to eliminate the challenges facing the healthcare industry (such as a lack of resources, staff, and awareness) and reported significant improvement in communal health of different countries. As a result, different kinds of e-Health systems have been developed and utilized to accomplish various tasks in various areas of the medical domain, such as remote consultancy, electronic health record management, hospital management, health awareness, resource tracking, teaching, and learning; mother and child care is one of them, which is emerging as an upcoming area in e-Health, where a lot of research projects including check-up reminder system, prenatal nursing, Antenatal Care (ANC), Expanded Programme on Immunization (EPI), knowledge updates regarding mother and newborn care and insulin calculation are under the spotlight; but research has rarely been found in predicting possible chances of affecting offspring from genetic or inherited haematological or bleeding disorder such as Haemophilia and Thalassemia at the prenatal stage. Haemophilia is a severe condition where a person with Haemophilia bleeds for a very long time, even from small wounds or injuries, as their blood does not coagulate. In contrast, thalassemia is a disorder in which red blood cells cannot synthesise haemoglobin (oxygen-carrying molecules in the blood). Due to this, the body makes fewer healthy red blood cells and less or insufficient haemoglobin (an iron-rich protein in red blood cells) compared to normal levels. Hence, the oxygen does not reach all body parts properly because of the insufficiency of haemoglobin in the blood, which causes starvation of oxygen in various organs, increasing the inability to function correctly. Thalassemia often leads to anaemia and causes symptoms like fatigue or tiredness, an enlarged spleen, slow growth in children, bone pain, a tendency to break bones, shortness of breath, pale skin, lack of appetite, dark urine, and jaundice. Therefore, it is a

crucial area of concern in the global medical system because Haemophilia and thalassemia have been steadily increasing over time, quickly overtaking other diseases in terms of mortality rates, and placing a financial burden on patients' families as well as the governments of India and other countries around the world. This indicates that an early diagnosis and prediction of the probability of offspring inheriting Haemophilia or thalassemia from parents at the prenatal stage is an essential requirement so that, in time, informed decisions or actions can be taken under the supervision of an expert doctor (s). Analysis of various medical records, such as pathological records, family history records, and genetic analysis records, may cause clinicians to take longer to make decisions or result in incorrect evaluations due to a lack of patient data or information. To overcome these difficulties and ultimately aid medical practitioners in diagnosing, a variety of e-Health systems have been developed using computational models constructed using machine learning techniques. In light of this, the thesis suggests an e-Health system that utilises a computer model created using machine learning techniques to provide an early diagnosis of Haemophilia and thalassemia at the embryonic stage.

The purpose of this work is to provide remote patients with e-services to assess the possibility of having a child with Haemophilia or thalassemia at the prenatal stage, to assist medical professionals in the evaluation process for inherited bleeding disorders like Haemophilia and thalassemia, to effectively analyse complex and ambiguous patient health examination data, to overcome the staffing and resource shortage, and to reduce the cost, time, and effort needed. This is accomplished with the development of a proposed system that predicts the probability that offspring will inherit Haemophilia or thalassemia based on medical history, family history, physiological symptoms history, and obstetric history of the couple planning a family and also provides a list of expert doctors for further consultation. To design the suggested system, hybrid and rule-based prediction models have been developed employing a variety of machine learning methods, including Bayesian Networks, Decision Trees, and Gaussian Naive Bayes. Later, a mobile application-based e-Health system was created using a rule-based prediction model to assist patients and doctors in anticipating the possibility of having a child with Haemophilia and thalassemia at an early stage. The effectiveness of

the models was then evaluated using a confusion matrix and statistical techniques such as the Receiver Operating Characteristic (ROC) Curve, which establishes the validity of the suggested system in making an early diagnosis of inherited prenatal disorders like Haemophilia and thalassemia. As a result, the current study can be used to predict the chance of acquiring other inherited blood illnesses, such as sickle cell anaemia, and it can serve as the basis for premarital analytic models or programmes that evaluate a couple's medical compatibility.

Acknowledgements

A single person has never produced an invention. This thesis was developed through teamwork, just like any truly exceptional piece of work. On the successful completion of the research work and this thesis, I really would like to thank my Lord for giving me the privilege of writing this dissertation with immense devotion and a profound sense of gratitude. The work presented in this thesis would not have been possible without my close relationships with numerous people. The help and cooperation of numerous people allowed the current job to be completed. With the utmost humility, I would want to use this opportunity to offer my profound gratitude to everyone who made this research thesis possible.

First and foremost, I am thankful to my supervisor Dr. Anil Sharma, Professor, School of Computer Science, Lovely Professional University, Phagwara, Punjab, for his dedicated support, timely advice, inspiration, encouragement, and continuous support throughout my Ph.D. I gratefully acknowledge my co-supervisor Dr. Lalit Garg, Associate Professor, Faculty of Information & Communication Technology, University of Malta, Msida, Malta for his valuable suggestions. I wish to thank my Ph.D. doctoral committee members for providing valuable comments and suggestions during this research work. I wish to express my sincere thanks to the management and staff of the Division of Research and Development, Lovely Professional University for extending the support and facilities to carry out my research work under their supervision I had started my research work.

I would like to express my respect, sincere gratitude, and debt of gratitude to Dr. Shruti Kakkar (MD), Associate Professor, Department of Paediatrics, Paediatric Hematology/Thalassemia Unit, Dayanand Medical College & Hospital, Ludhiana, Punjab, India, for her direction, encouragement, and support in the collection of the data set and the selection of parameters required for the current research work.

I would like to thank and gratefully acknowledge the efforts of Dr. Anubha Bharthuar (MD), Medical Oncologist, Hematology, Patel Cancer & Super Speciality Hospital, Jalandhar, Punjab, and Dr. Bhavneet Kaur, MD (OBS and Gyne), Tagore Hospital & Heart Care Centre, Jalandhar, Punjab, for their kind support, encouragement, and suggestions during Ph.D. research work.

I wish to express my sincere thanks to Mr. Deepak Singhal, Secretary, Haemophilia Society Dehradoon Chapter, Dehradoon, Uttarakhand, India; for extending the

support and facilities to carry out my research work. Without the encouragement and support of my mother Mrs. Kalawati Sharma, father Mr. Chander Ballabh Sharma, and my husband Mr. Harish Kumar, who took on all of my responsibilities throughout my work, I would not have been able to finish the thesis.

I struggle to find the right words to thank my brother Mr. Raj Kumar Sharma, who helped me in completing my data-gathering work, with all of my heart. I also appreciate Mr. Mahesh Kumar Sharma, my younger brother, for being there for me. Additionally, I want to express my gratitude to our two young children, Krishna and Jaitik, to whom I haven't given enough time because of my thesis writing task. I appreciate my friends for continuously encouraging and pushing me to finish this project, especially Ms. Navneet Kaur, Dr. Naina Handa, and Mr. Amit Sharma.

Lastly and most importantly, I want to express my gratitude to all the Haemophilia and thalassemia patients who assisted me in completing the data-collecting assignment by providing answers to the posed questions so that I could complete my work. I also want to express my gratitude to everyone who helped make this project a success by participating directly or indirectly in its completion.

Pushpa

Contents

List of Figures

List of Tables

Abbreviations

ANC	Antenatal Care
APTT	Activated Partial Thromboplastin Time
AUC	Area Under the Curve
CBC	Complete Blood Count
CVS	Chorionic Villus Sampling
DAG	Directed Acyclic Graph
e-Health	Electronic Health
EPI	Expanded Programme on Immunization
GNB	Gaussian Naive Bayes
HA	Haemophilia A
HB	Haemophilia B
HC	Haemophilia C
ICT	Information and Communication Technology
KNN	K-Nearest Neighbors Algorithm
MIDD	Maternally Inherited Diabetes and Deafness
MPSGC	Mandatory Pre-Marital Screening and Genetic Counselling
NGO	Non-Governmental Organization
ROC	Receiver Operating Characteristic
PND	Prenatal Diagnosis
PT	Prothrombin Time
RBC	Red Blood Count

TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Dedicated to my beloved family...

Chapter 1

Introduction

Electronic Health (e-Health) is a fantastical idea that has arisen as a result of the use of Information and Communication Technologies (ICT) in healthcare, including networks, the internet, computers, and wireless[1] [2] [3] [4]. The successful implementation of e-Health has transformed the way that healthcare is delivered by enabling the storage of health records, retrieval of medical records and information, and transmission of medical services in digital form [5]. Additionally, consistent e-Health use in the healthcare industry has been linked to notable improvements in the general health of several nations [5]. Besides this, e-Health systems are increasingly being used as a passive and active solution to lessen numerous healthcare industry barriers (such as a lack of employees, funding, and awareness) [4]. Due to all of these factors, e-Health solutions are replacing traditional methods of delivering health services in areas including medical consulting (such as patient treatment, appointment scheduling, and patient record administration), medical learning, and health management [6] [7].

While e-Health is becoming more and more popular as a relatively new method of delivering medical services through the use of communications technology, it has been in utilise in some form or type for over thirty years. The National Aeronautics and Space Administration (NASA) introduced the idea of e-Health under the term

telemedicine in the early 1960s[8] [9] [10] [11] [12] [13] [14] [15]. After that, in 1989, the National Aeronautics and Space Administration (NASA) carried out the first global telemedicine programme; "Space Bridge to Armenia/Ufa," for carrying out telemedicine consultations [8] [9] [10] [11] [12] [13] [14] [15]. The term "e-Health" is now used to refer to the global digitization of the health sector. Additionally, the increased desire to create new e-Health systems to influence all facets of healthcare and society is being driven by e-Health's improved performance, efficiency, and cost-effectiveness [16].

Different kinds of e-Health systems have been identified in the literature for providing services regarding remote consultancy [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30]electronic health record management [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42], hospital management [43] [44] [45] [46], health awareness [47][48] [49] [50] [51] [52] [53], resource tracking[54], teaching and learning [55]. Similar to how mother and child care [5] is developing as an upcoming field in e-Health, many types of research are currently in the spotlight (such as check-up reminders [56], prenatal nursing [56], Antenatal Care (ANC) [56], expanded programme on immunization (EPI), knowledge updates regarding mother and newborn care [54][57][58][59], and insulin calculation [60]); but research is rare has found in predicting possible chances of effecting expecting offspring from different problems such as genetic or inherited bleeding disorder at early or prenatal stage [56][60].

This identified gap becomes the motivation to conduct current research work and to propose an e-Health system for the early diagnosis of inherited disorders; which will predict the probability of inheritance of inherited bleeding disorders such as Haemophilia and Thalassemia from parents to offspring at the prenatal stage or even at pre-conception stage based on medical, physiological and family or ancestral history of couple planning to extend their family and served as the scope of study [60]. So that they may make the required decisions (such as whether to continue the pregnancy or not) with the guidance and advice of qualified medical

professionals (such as a haematologist, gynaecologist, paediatric haematologist, or medical oncologist haematologist). In this way, the proposed system will help to provide services like self-awareness, diagnosis, and prediction of inherited disorders services to the remote public (especially to target couples on the way to extend their family), electronically.

The leftover part of this chapter is structured as follows: Section 1.1 discusses the brief background of the medical domain. Section 1.2 briefly discusses the basic concepts of Machine Learning (ML). Section 1.3 provides fundamental knowledge regarding the Bayesian Network. Section 1.4 discusses the concept of K-Nearest Neighbours (KNN). Section 1.5 introduces the concept of Decision Tree and Gaussian Naive Bayes. Section 1.6 discusses the Hybrid Approach. Section 1.7 highlighted to research challenge. Section 1.8 provides details of research motivation. The objectives of the current research work have been presented in section 1.9. Section 1.10 depicted the research methodology, briefly. The contribution of the thesis has been discussed in section 1.11. The scope and limitations of the study have been discussed in section 1.12. The outline of the thesis is presented in Section 1.13.

1.1 Medical Background

1.1.1 Inherited Disorder

The term "inherited disorder" refers to a disease that is carried from parents to their children through mutations in genes or chromosomes [53] [61] [59]. All these disorders cause birth defects (like abnormalities in physical and mental structure or function) and other kinds of complexities in prenatal [53] [61] [59]. These defects may be inherited or hereditary in origin (like Haemophilia, Thalassemia, Sickle Cell Anaemia, Cystic fibrosis, Diabetes, and Maternally Inherited Diabetes

and Deafness (MIDD)) or maybe the outcome of infections (such as rubella); disclosure to deadly chemicals and radiation (affect the development of the foetus), mother's diet, drinking and drug intake habits (such as foetal alcohol syndrome) as well as maternal illnesses [53]. The disorders that occur due to inheritance are divided into several types such as single gene disorders, chromosomal or mitochondrial disorders, multifactorial gene disorders, and hereditary disorders [59]. A change in a single gene contained in the Deoxyribonucleic Acid (DNA) results in a single gene condition or disorder. Contrarily, chromosomal disorders also called mitochondrial genetic disorders occur when a chromosome or a portion of a gene's chromosomes are removed or replaced in the DNA structure. Similar to this, complicated illnesses or disorders develop when several genes (presented in DNA) are altered [59]. Although lots of work has been reported for the prevention of these inherited disorders as well as for improving mother and child health; the efforts in the direction of empowering parents for the early detection, prevention, and management of inherited disorders (especially in the context of Haemophilia and Thalassemia) are limited and required more efforts in the context [53]. Hence, the development and deployment of an electronic healthcare system are required for educating and empowering parents (especially mothers) for the early detection of inherited disorders such as Haemophilia and Thalassemia causing birth defects (like abnormalities in physical and mental structure or function) in prenatal. So that, parents can make early informed decisions regarding their pregnancy with the consultation of expert medical professionals.

1.1.2 Haemophilia

Haemophilia is an X-linked recessive genetic trait or disorder that occurs due to an altered gene on the X chromosome and can be inherited by the child from parents [59]. As a result, males with an XY genotype are affected by Haemophilia and are referred to as patients, but females with a XX genotype are carriers of

the condition since they have one Haemophilia gene that is functional and one that is not. But, sometimes females may also be affected if alteration on both X chromosomes has happened which causes bleeding symptoms and is called the symptomatic carrier or affected Patient of Haemophilia. Hence, it can be said that an alteration in the copy of a gene on a single X chromosome causes the disorder in males as they have only one X chromosome, whereas females require mutations on both X chromosomes to be affected by the disorder. In contrast, if only one X chromosome has mutated in females then they will be Carriers of Haemophilia otherwise Patients of Haemophilia [62][63]. It is additionally broken down into three subtypes [64]. One of these is Haemophilia A (HA), Haemophilia B (HB), and Haemophilia C (HC), which is caused by a deficiency or absence of clotting factors VIII, IX, and XI (together with Platelet-Rich Plasma, or PRP) in the blood, respectively [65] [63] [64]. In terms of the prevalence of Haemophilia, it has been found that Haemophilia A (HA) affects one in 5,000 male live births, compared to Haemophilia B (HB), which affects one in 30,000 male live births [65] [63]. Based on the level of clotting activity found in a subject's blood, the severity of different types of Haemophilia is occasionally further categorised into Mild, Moderate, and Severe categories. According to the classification system used for adults, Mild Haemophilia is defined as having a factor activity level between 6% and 49%, moderate Haemophilia as having a factor activity level between 1% and 5%, and severe Haemophilia as having a factor activity level of less than 1% [65] [42]. Severe Haemophiliacs frequently experience aberrant bleeding (such as prolonged bleeding) and frequent spontaneous bleeding in response to even minor wounds. Internal bleeding in the kidneys, brain, and gastrointestinal system is not unusual, although spontaneous bleeding happens most frequently in the joints. Joints can become irreversibly damaged by chronic joint bleeding, which can also irritate the joint and harm the cartilage. Muscle hematomas and internal bleeding from minor injuries can develop days after the initial trauma and may manifest as symptoms. Although Haemophilia is an inherited illness, a small

number of new cases are also seen in this environment, with about one-third of infants receiving a diagnosis of Haemophilia without any known family history. Despite all this, Haemophilia is not considered a common disorder as the report by the Centers for Disease Control and Prevention of the United States (published in August 2022) depicted that only 33,000 Americans are haemophilic [66]. Additionally, the report further depicted that approximately 10 in 100,000 people have Haemophilia A, 3 in 100,000 have Haemophilia B and 1 in 100,00 has Haemophilia C [66] [67][68]. There are several methods (including screening and clotting factor testing) that can be used to diagnose Haemophilia. Blood tests that determine whether or not the blood is clotting properly include screening tests. Complete Blood Count (CBC), Prothrombin Time (PT), Activated Partial Thromboplastin Time (APTT), and Fibrinogen Test are among the screening tests on the list. The diagnosis of Haemophilia depends heavily on the results of the tests outlined [66] [68] [67]. A Complete Blood Count (CBC) test measures haemoglobin levels; red blood cell size and quantity, type of white blood cell, and platelet count in the blood [66] [68] [67]. In cases of Haemophilia, this test reveals low haemoglobin and red blood cell counts, particularly when the Haemophiliac bleeds frequently or for an extended period [66] [68] [67]. The Activated Partial Thromboplastin Time (APTT) Test measures the ability of factors VIII, IX, and XI to clot blood as well as how long it takes for blood to clot [82-85]. The probability of Haemophilia can be anticipated if a person's factors have a very low clotting ability and it takes longer for blood to clot [66] [68] [67]. Similarly, the Prothrombin Time (PT) Test also represents to clotting ability of factors I, II, V, VII, and X. In contrast, the Fibrinogen Test has been done when the test results of PT or APTT have been found abnormal. Besides this, Clotting Factor Tests are another test that tells the type of Haemophilia and the severity [66] [68] [67]. Haemophilia diagnosis is also influenced by family history (whether the condition runs in the family or not); A diagnosis can be made early in the pregnancy, during labour and delivery, or after delivery if there is a family history of the illness [63] [66]. There are several

ways to determine whether a child or an expected child has a chance of acquiring Haemophilia from their parents. Genetic counselling and testing are one of the methods which have been utilized to know the likelihood of inheritance of disorder by offspring in early pregnancy or even before conception [78] [66] [68] [67] [69][63] In genetic testing, a tissue or blood sample may be analyzed for the indications of genetic alteration that results in Haemophilia. In contrast, genetic counselling is a way that can be followed at the preconception or prenatal stage to know the occurrence of Haemophilia in offspring [70] [71]. In this process, the detailed medical and family history of the client has been taken by the genetic counsellor to discuss and predict the hereditary condition of Haemophilia [68] [70]. Besides this, genetic counselling also provides information about testing options, resources, and management of the disorder. Risk assessment and psychological support are some other services that are provided by genetic counsellors [70]. For this process, pedigree (which is one kind of diagram) has been utilized for the investigation of a genetic condition like Haemophilia and the inheritance of the condition throughout the family. It also shows different family members and their relation with each other, their medical condition, and their risk of having or developing disorders in a later stage [70][71]. In addition to this, several other ways (such as Chorionic Villus Sampling (CVS) and Amniocentesis) are there that can be opted during pregnancy for the diagnosis of inherited disorders [70]. CVS is a method that uses a tiny sample of the placenta extracted from the womb and tested to know the alteration gene or inheritance of Haemophilia genes from parents to offspring. This prenatal diagnosis is usually conducted between the first trimesters of pregnancy [70]. In contrast, Amniocentesis uses amniotic fluid for testing of Haemophilia between 15 to 20 weeks of pregnancy. All these prenatal diagnosis have been done under the supervision of a doctor as there is a chance of complications such as miscarriage or early labour [82]. Furthermore, if the inheritance of Haemophilia is suspected in newborns due to family history then further blood testing such as umbilical cord blood testing has been done immediately after the birth of a child

to know the level of the clotting factors [70]. In cases when the family history is unknown, the diagnosis of Haemophilia will differ. In this instance, the disease is typically noticed as a youngster begins to crawl because they may have bruising on their joints or joint bleeding. These early symptoms can therefore aid in the diagnosis of a condition. However, in severe cases, this is possible as mild cases did not exhibit many symptoms in the beginning and can be recognised in the later stages by some of the worst symptoms, such as prolonged bleeding during an accident or surgery [70]. The absence of a long-term cure for this bleeding disease is its main drawback. The only solution is to stop bleeding episodes and replace any missing blood clotting factors [63]. Therefore, underdiagnosis and improper treatment of the illness result in lifelong significant arthropathy, which places a financial, social, psychological, and physical burden on society as a whole as well as on the parents. Early detection and effective care can reduce the risk of this permanent impairment [72] [73]. This calls for prenatal diagnosis, but it must first be followed by appropriate genetic counselling, risk assessment of the potential carrier, and support during the diagnostic process [65].

1.1.3 Thalassemia

Thalassemia is one kind of inherited haemoglobin or bleeding disorder [74]. It occurs when hemoglobin (oxygen-carrying molecules in the blood) cannot be synthesized by red blood cells [74]. Due to this, the body makes fewer healthy red blood cells and less or insufficient hemoglobin (it is an iron-rich protein in red blood cells) as compared to normal levels. Hence, the oxygen does not reach all the parts of the body properly because of insufficiency of haemoglobin in the blood and causes starvation of oxygen in various organs which increases the inability to function properly [75]. Thalassemia often leads to anemia and causes symptoms like fatigue or tiredness, an enlarged spleen, slow growth in children, bone pain, a tendency to break bones, shortness of breath, pale skin, and lack of appetite,

dark urine, and jaundice [75] [74]. People with Thalassemia can have mild or severe anemia [76]. Thalassemia is a severe yet preventable inherited haematological disorder [74]. Thalassemia also runs in the family as it is an inherited disorder [75]. Inherited disorder means at least one parent is a carrier of this disorder. Patients with Thalassemia Minor, which has relatively moderate symptoms such as modest anaemia or no symptoms at all, occur if just one parent is a carrier [75]. In contrast, in cases where both parents have the characteristic (almost 25% of cases), the patient develops a substantial type of condition known as Thalassemia Major [75]. Despite being an inherited condition, it can also be caused by genetic deletion or mutation [75]. As a result, Thalassemia affects between 45 and 70 million people in South Asian countries, or 1% and 5% of the world's population [74]. Similarly, 90% of the births of infants with severe haemoglobin abnormalities occur in developing countries, which account for roughly 500,000 annual cases [74], and it is likely that during the coming few years, the prevalence of Thalassemia will rise in most developing countries. Due to the rise in occurrence, Thalassemia is now considered to be a global public health concern [74]. A Complete Blood Count (CBC) and a haemoglobin electrophoresis test are the initial steps in the diagnosis of Thalassemia. These tests are mostly used to measure the fractions and other types of haemoglobin, including haemoglobin A, A2, F, H, and E [76]. In addition to this, in-depth analysis of hemoglobin has been conducted through hemoglobin electrophoresis test and sometimes high-performance liquid chromatography can also be used [93]. But, sometimes these screening tests show the overlapping of mutations and result in an incorrect diagnosis or a false negative [76]. Genetic tests for Thalassemia or gene mutations are required in this illness. For this, parents and siblings are tested [76], which reveals the presence of the Thalassemia trait, which is brought on by an autosomal dominant inheritance pattern, in which a child only needs one copy of the altered gene from either or both parents to develop the genetic condition or trait. These screening tests are only advised if the patient or a family member exhibits any type of Thalassemia-related sign or symptom,

including anaemia, fatigue, tiredness, an enlarged spleen, slow growth in children, bone pain, a tendency for broken bones, shortness of breath, pale skin, lack of appetite, dark urine, and jaundice [74] [75]. Thus, each one of these screening tests are essential in the diagnosis of Thalassemia or also in ruling out the causes of some common symptoms such as anemia; which occurs due to some kind of nutritional deficiency, illness, or infection, and medications [76]. To evaluate if low haemoglobin is a temporary condition or an indication of Thalassemia, a doctor should review all of this information with a patient during a consultation and ask about their medical history, specifically about low haemoglobin [77]. As a result, all of these tests have been applied in this way when patients display particular signs and symptoms [78]. Based on the type and severity of Thalassemia, patients have received the necessary treatment after receiving a diagnosis [94-96]. For instance, routine blood transfusions have been used as a type of treatment for those with severe or substantial Thalassemia [79]. Additionally, they have received iron chelation therapy and folic acid medication [79]. Chelation therapy is primarily used to prevent the body's organs from accumulating too much iron as a result of repeated blood transfusions, which can lead to heart or liver disease, infections, and osteoporosis. Deferoxamine, deferiprone, and deferasirox are three different types of iron chelators that are used for chelation therapy [79]. In addition to this, Bone Marrow Transplantation or Haematopoietic Stem Cell Transplantation is another curative option for treatment of Thalassemia [79]. It involves the management of healthy hematopoietic stem cells in patients with dysfunctional or depleted bone marrow [96-97]. However, the issue with this choice is that it is pricey and comes with a high risk of death and morbidity [74] [76] [80] [78] [81]. Furthermore, most patients were unable to locate a donor who was a good match [74] Additionally, the majority of developing countries lack the medical expertise and resources necessary to perform hematopoietic stem cell transplantation [74]. As a result, the cost of normal Thalassemia treatment and therapy is high for an ordinary household with low income, laying a heavy financial and psychological strain [74]. As a

result, some countries have introduced different Thalassemia prevention strategies, such as mandatory Pre-Marital Screening and Genetic Counselling (MPSGC) and Prenatal Diagnosis (PND), which gives women the option of ending a pregnancy if they are affected[74]. In nations like Cyprus, Italy, Greece, Turkey, and Iran, PND with therapeutic abortion is the method that is used as one of the preventive methods, and has attained an impressive success rate in stopping the births of offspring with Thalassemia [74]. However, due to restrictions imposed by culture and religion, this option is typically not offered in several other nations [74]. Pre-marital screening, which analyses a couple's medical history and family history through genetic counselling before marriage and determines the likelihood that the couple will give birth to a child with Thalassemia and is also used in case of Haemophilia diagnosis, is another preventive measure. Through this procedure, the couple is also allowed to learn more about their options for conception and to make informed decisions regarding their marriage [74]. Genetic counselling is a helpful tool for individuals and families who have been diagnosed with or are at risk of developing an inherited or hereditary disorder. It deals with reproductive issues at various stages of pregnancy and also offers vital services to patients who are struggling with issues related to Thalassemia and its various forms, such as disease and trait [76]. As a result, genetic counselling must be done by a licenced genetic counsellor who gives genetic counselling based on medical or pedigree (For example, a three-generation family history) of genetic illnesses and also offers preliminary education on the natural history, signs, and symptoms of disorders that should be addressed immediately [93]. In addition to this, genetic counsellors have been providing continuing information about services that are accessible in the context of emotional and social support, available resources (such as research projects, support groups, and advocacy organisations), and patient-to-patient or parent-to-parent linkages [76]. Even after getting genetic counselling, high-risk couples still choose to get married because they do not have enough information about the consequences of having thalassemic kids [74]. The prevalence of thalassemia

is increasing, mostly because so few individuals are aware of it [74]. The rising prevalence of Thalassemia is largely attributed to a lack of genetic counselling resources and professionals [81] [77]. As a result, it can be inferred that hereditary bleeding diseases like Haemophilia and Thalassemia are important fields of research in the worldwide system of healthcare given that their frequency is rising over time and necessitates greater focus on their prevention with early diagnosis . Although there are numerous pathological and blood testing options available for the diagnosis of Haemophilia and thalassemia, these options are not helpful in the early diagnosis (such as pre-conception or pre-marital stage); as these are chosen to either diagnose a pregnant woman or a person who was exhibiting symptoms and do not offer a straightforward method for the general public to independently identify potential hereditary bleeding problems. While genetic counselling is a technique that may be used to assess a person's risk for Haemophilia and Thalassemia based on their whole medical and family history [69] [70]. However, some limitations in the field of genetic counselling were discovered. First of all, it works best when the client is well aware of their medical history and family history of bleeding disorders (reaching at least three generations). This means that a person who doesn't have all of the required information won't be able to take advantage of the predictions made by genetic counselling [71]. Second, genetic counselling is not widely known and is now not readily available to the general public due to the small proportion of genetic counsellors in the population [71]. The same may be seen in MaryAnn Abacan's research, which states that by 2018, at least 7000 genetic counsellors were employed by no less than 28 countries [71]. Last but not least, there are times when one partner in a pair chooses not to reveal their medical and obstetric history (such as past miscarriages, abortions, or adoptions) in front of the other, which results in inaccurate prediction. The resulting gap led to the development of an e-Health system with the specific goal of predicting the likelihood of inherited disorders (namely Haemophilia and thalassemia) at the prenatal stage. This system will take into account both cases where the client

already has a medical history of the disorders and cases where the client does not. Additionally, by making e-services available to the remote public in this context, it also gets beyond the limitation of genetic counsellors.

1.2 Machine Learning

Machine learning is a branch of artificial intelligence that works on creating models and methods that let computers learn from data on their own. It applies statistical methods to analyse data, draw conclusions, and forecast outcomes [82][83] [84] [85]. Thus, the focus of machine learning is on the models and techniques that let computers learn from data or historical experiences to maximise performance criteria [83]. Scalability and efficiency are enhanced by the combination of new computer technology with machine learning techniques. Natural language processing, recommender systems, image and audio recognition, and many more fields employ machine learning [85]. Modern machine learning models can now be used to anticipate anything from disease outbreaks to stock price fluctuations. As a result, machine learning innovations have the power to transform several industries, including the health industry, which makes substantial contributions to public health with the implementation of e-Health [85].

1.2.1 Role of Machine Learning in e-Heath

The discipline of e-Health is being revolutionised by machine learning, which has opened up several options and breakthroughs for bettering healthcare results [86]. One significant use of machine learning in e-Health is diagnostics [104]. These models can learn to recognise patterns and predict diagnoses with high accuracy by being trained on vast datasets of medical records, symptoms, and test results [86]. This helps doctors diagnose patients more quickly and accurately, and it

also enables early disease diagnosis, which could potentially save lives [85]. Personalised medicine is another area of e-Health where machine learning excels [87]. Machine learning algorithms can find trends and prescribe specific treatments by studying vast volumes of patient data, including genetic data, lifestyle factors, and treatment outcomes. This can aid medical professionals in selecting the best drugs, doses, and treatment plans for each patient based on their unique requirements, enhancing outcomes and reducing side effects . To track and forecast patient health, machine learning is essential [86]. Machine learning algorithms may recognise minor changes in a patient's vital signs and forecast the risk of adverse events like heart attacks or strokes by using data from wearable devices, electronic health records, and real-time sensor data [104]. This enables medical professionals to take preventive action and offer early interventions, enhancing patient safety and lowering hospital readmission rates. The management of resources and healthcare operations are both improved by machine learning. Machine learning algorithms can effectively schedule appointments, forecast patient demand, and distribute resources by examining past patient data, appointment histories, and resource usage trends [85]. Reduced wait times, greater staffing levels, and better resource allocation ultimately result in higher patient satisfaction and lower costs [85]. By improving diagnoses, enabling the personalised treatment, easing patient monitoring, and optimising healthcare operations, machine learning is revolutionising e-Health [85]. e-Health is positioned to make great leaps in providing effective, individualised, and high-quality healthcare services to people all over the world thanks to ongoing improvements in data collection and machine learning algorithms. As a result, machine learning greatly advances the field of e-Health by making it easier to conduct complex data analysis, forecast, and make decisions for a variety of healthcare applications, as shown below [88].

- **Drug Discovery and Development**:Machine learning algorithms can analyze vast amounts of molecular and genetic data to identify potential drug candidates, predict their efficacy, and optimize drug discovery processes.

This can significantly accelerate the development of new drugs and therapies, potentially leading to breakthrough treatments for various diseases [88].

- **Disease Prognosis and Risk Assessment**: Machine learning models can analyze patient data and clinical parameters to predict the progression of diseases, assess the risk of complications, and estimate patient outcomes. This information helps healthcare providers make informed decisions about treatment plans and interventions, improving patient care and management [89].

- **Medical Imaging Analysis**: Machine learning algorithms excel in analyzing medical images such as X-rays, MRI scans, and histopathological slides. By training on large datasets of annotated images, these models can detect abnormalities, assist in diagnosing diseases, and provide quantitative measurements for better treatment planning [89].

- **Natural Language Processing** (NLP): NLP techniques combined with machine learning enable the extraction and analysis of valuable information from unstructured medical text, such as electronic health records, clinical notes, and research articles. This helps in improving information retrieval, clinical decision support systems, and research insights [89].

- **Health Behaviour Analysis**: Machine learning algorithms can analyze patient behaviour patterns and data from wearable devices to assess lifestyle factors, monitor adherence to treatment plans, and provide personalized recommendations for healthier habits. This promotes patient engagement, self-management, and preventive care [87].

- **Fraud Detection and Cyber Security**: Machine learning models can detect anomalies and patterns in healthcare claims data, identifying potential fraudulent activities and enhancing cyber security measures. This helps

in preventing insurance fraud, protecting patient data, and ensuring the integrity of e-Health systems [87].

- **Telemedicine and Remote Monitoring**: Machine learning algorithms facilitate remote patient monitoring by analyzing real-time data streams from wearable devices and sensors. They can detect abnormal patterns, alert healthcare providers, and enable timely interventions, making telemedicine more effective and scalable [87].

- **Classification and Prediction**: Machine learning algorithms can be used for classification tasks, where they learn to categorize patients into different classes based on their characteristics or symptoms. For example, machine learning models can predict the likelihood of a patient developing a particular disease or assess the risk of readmission to the hospital. These predictions can help healthcare providers make proactive interventions and personalize treatment plans [87].

- **Personalized Medicine**: Machine learning enables the development of predictive models that consider individual patient characteristics, genetic data, and treatment history to personalize medical interventions. By analyzing large-scale patient data and identifying patterns, machine learning can help in recommending the most effective treatments for specific patient populations. This approach supports precision medicine, where treatments are tailored to individual patients based on their unique characteristics [87].

- **Health Monitoring and Wearable Devices**: Wearable devices equipped with sensors, such as smartwatches and fitness trackers; generate vast amounts of data about an individual's health and activities. Machine learning algorithms can analyze this data to monitor vital signs, detect anomalies, track sleep patterns, and identify trends in physical activity. This information can aid in the early detection of health problems, remote patient monitoring, and preventive care [87].

- **Disease Outbreak Prediction**: Machine learning models can be trained to analyze large-scale epidemiological data, social media feeds, and other relevant sources to predict disease outbreaks and identify potential hotspots. By analyzing patterns and correlations, these models can help public health authorities and healthcare providers allocate resources and implement targeted interventions to control the spread of infectious diseases [87].

All of the aforementioned e-Health applications have used a sequential approach, which includes data collection, data pre-processing, feature extraction and selection, model training, model evaluation, prediction and decision-making, and continuous learning and improvement, to perform complex data analysis, forecasting, and decision-making. This approach is explained below:

- **Data Collection**: Machine learning in e-Health begins with the collection of relevant data [88]. Electronic health records (EHRs), medical imaging, wearable technology, patient-generated data, and patient-provided data are some of the sources of this information [88].

- **Data Pre-Processing**: Before applying machine learning algorithms, the collected data needs to be pre-processed [89]. This involves cleaning the data, handling missing values, normalizing or scaling features, and addressing any other data quality issues [88].

- **Feature Extraction and Selection**: The obtained data frequently includes several traits or variables [88]. To find the most pertinent and instructive features for the particular healthcare job at hand, feature extraction and selection procedures are applied. This process aids in reducing dimensionality and boosts the effectiveness and precision of machine learning models.

- **Model Training**: Once the data is prepared, machine learning models are trained using algorithms such as supervised learning unsupervised learning, or reinforcement learning. While unsupervised learning algorithms discover

patterns in unlabeled data, supervised learning algorithms discover patterns from labeled data. Reinforcement learning algorithms learn through interactions with an environment to optimize actions [88].

- **Model Evaluation**: After training, the efficiency of the machine learning models is evaluated using appropriate evaluation metrics. This step ensures that the models generalize well to unseen data and perform effectively for the intended purpose [88].

- **Prediction and Decision-Making**: Once the models are trained and evaluated, they can be deployed to make predictions and support decision-making in various e-Health applications. For example, machine learning can assist in diagnosing diseases, predicting patient outcomes, suggesting treatment plans, identifying patterns in medical images, and personalizing healthcare interventions [88].

- **Continuous Learning and Improvement**: Machine learning models can be continuously updated and refined based on new data and feedback from healthcare professionals. This iterative process helps improve the accuracy and effectiveness of the models over time [88].

1.3 Bayesian Network

The Bayesian Network technique is an efficient strategy for detecting illnesses in machine learning [90]. A Directed Acyclic Graph (DAG) is used in this probabilistic graphical model to show how parameters or variables are related to one another [109]. The variables in the context of disease prediction indicate various aspects of the disease or its symptoms [90]. The Bayesian Network is built first by the algorithm, which entails specifying the variables and their connections based on data or domain expertise [90]. Every variable in the network stands for a distinct

characteristic or symptom that may be related to the illness [109]. The graph's directed edges, which show conditional dependencies, describe the relationships between variables [109]. Once the Bayesian Network is constructed, the algorithm can be used for disease prediction [90]. Given a set of observed symptoms or evidence, the algorithm calculates the probability distribution over the possible states of the disease variable [90]. This is done by utilizing Bayes' theorem and the conditional probabilities defined in the network [90]. The algorithm begins by setting the observed evidence in the network, which corresponds to the symptoms present in the patient [90]. It then propagates this evidence through the network, updating the probabilities of all other variables based on their conditional dependencies. This process is known as probabilistic inference [90]. The output of the algorithm is the probability distribution over the possible states of the disease variable [90]. This distribution represents the likelihood of each possible disease given the observed symptoms [90]. The algorithm can provide the most likely disease as the prediction, based on the highest probability in the distribution [109]. One advantage of the Bayesian Network algorithm is its ability to handle uncertainty and incomplete information [91]. It can make predictions even when some symptoms are missing or uncertain [90]. The algorithm uses the probabilistic framework to incorporate all available information and provide a robust prediction [109].

1.4 K-Nearest Neighbors

The k-nearest neighbours algorithm, sometimes referred to as KNN or k-NN, is a non-parametric supervised learning classifier that groups individual data points based on closeness to classify or predict them. Many applications, most notably in classification, have made use of the k-NN algorithm. Finance, healthcare, and pattern detection are a few of these use cases.

1.4.1 Role of K-Nearest Neighbors Algorithm in e-Health

In the area of e-Health, the K-Nearest Neighbours (KNN) algorithm is essential for illness prediction [91]. Electronic health records, telemedicine, wearable technology, and health monitoring systems are all examples of e-Health, which utilizes information and communication technologies [91]. Here's how the KNN algorithm contributes to disease prediction in the e-Health domain:

- **Patient Similarity**: e-Health systems gather a wealth of data about patients, including their medical history, symptoms, vital signs, and genetic information. This information is used by the KNN algorithm to compare the attributes of different patients to determine how similar they are. By comparing a patient's characteristics with those of previously diagnosed patients, the algorithm can identify similar cases and predict the likelihood of specific diseases [91].

- **Personalized Medicine**: e-Health platforms often aim to provide personalized healthcare solutions. The KNN algorithm can be used to recommend appropriate treatment plans and interventions based on the similarity of patients' features and the outcomes of past treatments. By considering the experiences of similar patients, the algorithm helps healthcare providers tailor their approach to individual patients, leading to more effective and personalized disease management [91].

- **Early Diagnosis**: Early detection of diseases is crucial for successful treatment and improved patient outcomes. e-Health systems equipped with the KNN algorithm can analyze patients' symptoms and medical records to identify patterns associated with specific diseases. By recognizing the symptoms and risk factors shared by previously diagnosed patients, the algorithm can alert healthcare professionals to potential diseases at an early stage, enabling timely intervention and treatment [91].

- **Decision Support**: The KNN algorithm in e-Health systems acts as a decision support tool for healthcare providers. By presenting predictions and recommendations based on patient similarity, the algorithm assists in clinical decision-making. For instance, when faced with a complex case or ambiguous symptoms, healthcare professionals can rely on the algorithm's insights to aid in disease diagnosis and treatment planning [91].

- **Continuous Learning**: e-Health systems generate vast amounts of data over time, providing an opportunity for continuous learning. The KNN algorithm can adapt to new data by dynamically updating its model without requiring a complete retraining process. This enables the algorithm to incorporate the latest information and improve its predictive accuracy over time [91].

1.5 Decision Tree and Gaussian Naive Bayes

Supervised learning techniques such as decision trees are used for regression modelling and classification. Given that regression is a subset of predictive modelling, these trees are employed for both data classification and future prediction. Decision trees, like flowcharts, start with a root node that has a particular data question and branch out to include potential answers. The branches then lead to decision (internal) nodes, which generate more responses and ask more questions. This keeps going until the data ends at a terminal node, also called a "leaf" node. Decision trees are a helpful tool for decision-making in machine learning since they provide the problem and its possible solutions. Developers can evaluate the possible outcomes of a decision by using algorithms that can predict results for data in the future when they get access to more data.

A Gaussian distribution and a probabilistic approach provide the foundation of the machine learning classification technique known as Gaussian Naive Bayes (GNB).

Any parameter, sometimes called a feature or predictor, can independently predict the output variable, according to Gaussian Naive Bayes. The probability of a dependent variable falling into each group can be predicted by it. The total of all the forecasts for each parameter is the final prediction that increases the likelihood that the dependent variable will be placed in each group. The final classification is made according to the group with a greater probability.

1.6 Hybrid Approach

The hybrid approach is often required to predict diseases in e-Health due to the complex and multifaceted nature of healthcare data. To assist the delivery of healthcare, e-Health refers to the use of electronic tools including electronic health records (EHRs), medical imaging, wearable technology, and genetic data [92]. Here are some reasons why a hybrid approach is beneficial in disease prediction:

- **Data Variety**: Healthcare data is diverse and comes from various sources, including structured data (such as demographics and lab results) and unstructured data (such as clinical notes and medical images). To obtain a more complete picture of the patient's health and boost prediction accuracy, a hybrid approach incorporates several data kinds and sources, utilising both structured and unstructured data [92].

- **Feature Extraction**: Extracting meaningful features from raw data is crucial for accurate disease prediction. Different machine learning techniques are suited for different data types. For example, structured data can be analyzed using traditional statistical methods, while unstructured data may require NLP techniques. A hybrid approach combines these techniques to extract relevant features from various data types effectively [92].

- **Prediction Models**: Different machine learning algorithms have different strengths and limitations. A hybrid approach combines multiple models, such as decision trees, support vector machines, neural networks, or ensemble methods, to exploit the unique characteristics of each model and improve the overall predictive performance. Hybrid models can leverage the strengths of each component model and mitigate their weaknesses [92].

- **Interpretability**: In healthcare, interpretability and explainability are essential for gaining trust and acceptance from healthcare professionals. Hybrid models can provide a balance between accuracy and interpretability. For example, using a combination of machine learning models and rule-based systems allows for both accurate predictions and transparent decision-making [92].

- **Scalability and Generalization**: Healthcare datasets are often limited in size and can suffer from class imbalance or data sparsity. Hybrid approaches can leverage transfer learning, where knowledge gained from one task or dataset is transferred to another related task or dataset, to improve the prediction performance on small or imbalanced datasets. This enables the models to generalize better and perform well on unseen data [92].

1.7 Research Challenges

- During a literature review, various e-Health systems were found to offer various healthcare services, including remote consulting [17] [18] [19] [20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30]electronic health record management [31] [32] [33] [34] [35] [36] [37] [38] [39] [40] [41] [42], hospital management [43] [44] [45] [46], health awareness [47][48] [49] [50] [51] [52] [53], resource tracking[54],

teaching and learning [55]. Similar to how mother and child care [5] is developing as an upcoming field in e-Health, many types of research are currently in the spotlight (such as check-up reminders [56], prenatal nursing [56], Antenatal Care (ANC) [56], expanded programme on immunization (EPI), knowledge updates regarding mother and newborn care [54][57][58][59], and insulin calculation [60]); but research is rare has found in predicting possible chances of effecting expecting offspring . It is extremely important to get an early diagnosis of hereditary bleeding disorders since they can arise not only genetically but also as a result of related gene deletion or mutation. Therefore, it is advantageous if the process of diagnosis and prediction is automated. It decreases the mental strain, time, and human effort required to recognise and categorise illnesses (namely Haemophilia and Thalassemia), as well as the likelihood that kids of the parents would inherit those disorders during the prenatal or preconception stage. The current study focuses on this, which proved to be a difficult one.

- There was no secondary dataset on the internet for Haemophilia and Thalassemia. Therefore, original datasets were gathered from a variety of places, including hospitals, medical schools, and non-governmental organisations (NGOs). So, the biggest obstacle to the current research is the lack of a secondary dataset. Hence, the difficulty motivates the researchers to gather more data sets. So, more sophisticated e-Health systems must be created to stop or control the rising prevalence of Haemophilia and Thalassemia.

- For the diagnosis of Haemophilia and Thalassemia, there are several pathological or blood test alternatives (such as Complete Blood Count, Haemoglobin Electrophoresis Test, Chorionic Villus Sampling, and Fibrinogen Test, etc.), genetic testing, and counselling are available [66]. Genetic counselling, among other things, is better suited for the early diagnosis of inherited bleeding disorders because it forecasts the likelihood of having a child with Thalassemia or Haemophilia based on medical and family history of the disorder [69].

Despite this, the current study only found a small number of limitations to genetic counselling. The first situation where it works effectively is when the client has detailed information about his or her medical history and family history (spanning at least three generations) about bleeding disorders. It implies that a person who does not have all of the requested information will not be able to benefit from the genetic counselling's prediction services. Second, due to the limited number of genetic counsellors in the existing population, genetic counselling is not well recognised [70] and has not been made readily available to the general public up to this point. The same is seen in MaryAnn Abacan's research, which claims that by 2018, no fewer than 28 nations had at least 7000 genetic counsellors on staff [70]. Last but not least, there are times when one partner in a pair chooses not to reveal their medical and obstetric history (such as past miscarriages, abortions, or adoptions) in front of the other, which results in inaccurate prediction. The challenge has thus been accepted to conduct the present research task.

1.8 Research Motivation

Globally, the term "e-Health" has been used to refer to the digitalization of the healthcare industry. Additionally, the increased desire to create new e-Health systems to influence all facets of healthcare and society is being driven by e-Health's improved performance, efficiency, and cost-effectiveness. Although, lots of work has been reported for the prevention of these disorders namely Haemophilia and Thalassemia as well as for improving mother and child health; the efforts in the direction of empowering parents for the early detection, prevention, and management of inherited disorders (especially in the context of Haemophilia and Thalassemia) are limited and required more efforts in this context. Haemophilia and Thalassemia also known as haematological disorders are generally caused by the inheritance and mutation of genes [55]. The presence of inherited prenatal disorders

such as Haemophilia and Thalassemia is progressively growing over the duration. The World Health Organisation (WHO) states that both of these conditions are widespread and that greater attention is needed to prevent them [93]. Each year, more than 300,000 infants are born with significant haemoglobin anomalies [93]. Although up to 50,000 cases of Haemophilia A may exist, only 11,586 cases have been recorded [93]. Statistics also show that between 3,000 and 5,000 children are born each year with this issue [93]. This makes it an important issue of concern in the field of medicine worldwide. Therefore, to address this at its most basic level, novel methods or tools are needed. There are a variety of diagnosis alternatives accessible in medical science, including screening tests, blood tests, prenatal diagnoses, premarital counselling, and genetic counselling. Despite this, the prevalence of disorders is still rising because of a lack of knowledge about the disorders and their effects, as well as a lack of access to crucial services like premarital and genetic counselling [70] [71], as only 7000 genetic counsellors were reported in no less than 28 countries by 2018 [71]. Furthermore, the general population is not well-informed about genetic counselling [71] [70]. These reasons led to the current study project's motivation.

1.9 Objectives

The broad outline of the proposed objectives is as follows:

1. To collect a dataset of identified ancestral patterns, and physiological and pathological symptoms.
2. To integrate Bayesian network and machine learning techniques for the prediction of prenatal disorders.
3. To evaluate and validate results using intelligent statistical techniques.
4. To develop an application for the diagnosis of prenatal disorders.

1.10 Research Methodology

Research methodology is a way to solve the research problem systematically [94]. It is a step-by-step procedure used by researchers to describe the research work [94]. The actual execution of the research methodology utilises the mixed methodology research approach (which refers to the integration of qualitative and quantitative approaches) to attain the predefined objective of the current research work [94]. Additionally, different kinds of research methods have been utilised for the execution of current research work. These methods are divided into two main categories [94]. One of the methods is concerned with data collection and the other one is concerned with data processing and analysis [94]. The primary data set connected to the suggested diseases, namely Haemophilia, and Thalassemia, are collected as part of the present study activity using methods such as field surveys, interviews, discussion, and literature surveys since secondary data related to disorders are not yet accessible. After that, data pre-processing and analysis have since been carried out [113]. Data has been pre-processed by deleting duplicate records and null values. After that, classification techniques such as bayesian networks and machine learning techniques were used to construct individual or integrated models for the identification of hereditary illnesses at the prenatal stage [95] [96]. Here, the term "classification method" refers to the procedure of categorising an object or item based on its resemblance to precedent instances of other things or items [95] [96]. Then, a rule-based prediction model using a Bayesian network and Python was created to predict the likelihood of having a child with Haemophilia and thassaemia.

This model makes use of the rule set developed during the current research work and will serve as the basis for the creation of the proposed e-Health system (in the form of a mobile application) for the early diagnosis of inherited or hereditary disorders. So, using the Android platform, the design and development of the suggested application have begun. Last but not least, testing has been

carried out using the Receiver Operating Characteristic (ROC) Curve and other clever statistical approaches to assess and validate results. The suggested research methodology's top-down sequential approach is depicted in Figure 1.1 beginning with the use of the Android platform.
The proposed work, as shown in Figure 1.1, follows the mentioned methodology as below.

1. The very first step started with the literature survey focusing on e-Health systems

2. In the second step, the problem has been identified and objectives have been defined.

3. This step focused on the data collection task as the data set has been utilized for the accomplishment of the proposed study and the very first objective of the current study. To achieve the first objective, the primary data set has been collected from medical colleges and hospitals and Non-Governmental Organisations (NGOs) through field surveys.

4. To attain the second objective, different algorithms such as Bayesian Network, K-Nearest Neighbour (KNN), Decision Tree, and Gaussian Naive Bayes were implemented individually and interestedly for the development of a model for the prediction of inheritance of Haemophilia and Thalassemia; which was not achieved with this model and as a result a rule-based prediction model has been developed with the utilization of bayesian network and rule set developed during current study and proved successful in the prediction of the probability of having a child with Haemophilia and Thalassemia.

5. To achieve the third objective, intelligent statistical techniques such as the Receiver Operating Characteristic (ROC) Curve have been utilized.

6. Finally, to achieve the fourth objective, a mobile application has been developed as a utility of current research work. For this, the Android studio platform has been utilized along with the Kotlin programming language.

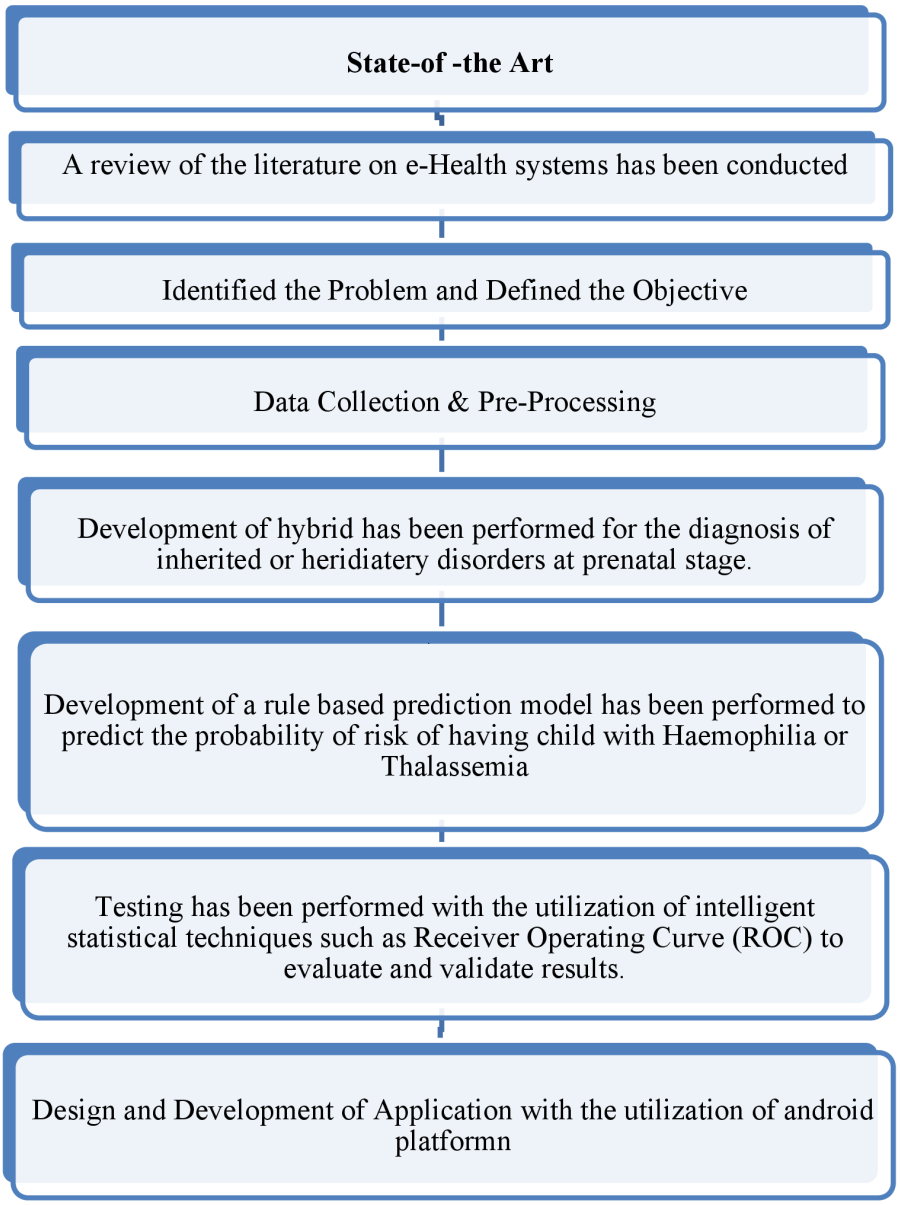


FIGURE 1.1: Research Methodologies

1.11 Contribution of Thesis

The main contributions of this thesis are:

1. The review of the literature carried out during the current research work will assist researchers in understanding the current state-of-the-art existing systems and for the development of a proposed e-Health system, especially for inherited or hereditary disorders such as Haemophilia and Thalassemia. To achieve this, many electronic or telemedicine system types related to various domains and disorders have been covered and examined to determine the benefits, drawbacks, and limits of existing systems related to e-Health.

2. Even though the frequency of Thalassemia and Haemophilia is continually rising. Although this is the case, both of these conditions are uncommon since patients with these two bleeding disorders are seldom [72]. The Centre for Disease Control and Prevention (CDC) study from the United States provides strong evidence for this, showing that 33,000 Americans out of the country's whole population have Haemophilia. Among them, 10 per 100,000 persons have Haemophilia A and 3 have Haemophilia B [72]. Hence, it can be said that inherited bleeding disorder such as Haemophilia and Thalassemia is rare bleeding disorders that cause a less number of patients. The data needed for the current scientific endeavor especially that Thalassemia and Haemophilia, was not discovered as a result. Online data repository sites (like the University of California Irvine machine learning repository) have searched for the required data but have not been successful in finding it. Because the data is not publicly available and most institutions are unable to offer it owing to legal issues and patient shortages, the dataset-gathering procedure has taken a lengthy time (about 3 years); which utilizes field survey along with interview and discussion methods. As a result, the dataset gathered during the current research study is a significant contribution to

the thesis since it will serve as a foundational resource for further study in this area.

3. For the diagnosis of bleeding diseases including Haemophilia and thalassemia, decision trees and the Gaussian Naive Bayes (GNB) technique have been combined to create a hybrid diagnosis model that may capitalise on the advantages of both approaches while attenuating their shortcomings. It is crucial to correctly diagnose both Thalassemia and Haemophilia, two bleeding disorders. Combining these two techniques allows us to divide the data into smaller subgroups according to the most useful attributes using the decision tree, and then use GNB for each subgroup to get the class probabilities. By dividing the data into smaller, more homogenous subgroups, this method can manage correlated characteristics while lowering the danger of overfitting. Additionally, because GNB is less impacted by unimportant variables, a greater variety of features may be included in the decision tree without degrading performance. The Decision Tree can be simple and understandable but may not be able to capture complicated connections, whereas the Bayesian network can describe complex dependencies between features but may suffer from computational complexity. We may make use of the advantages of both approaches and outperform them individually by combining the Decision Tree and GNB. As a consequence, the model developed for the current inquiry is the thesis' main contribution.

4. A rule set-based inherited prenatal disorder prediction model has been developed with the utilisation of the bayesian network and existing inheritance patterns; which were further extended by including physiological symptoms and obstetric history along with medical history to deal with the cases where medical and family history is not known clearly. This is a component that was absent from the previous inheritance pattern, As a result, this model was used to further build the recommended e-Health System for the early

identification of genetic bleeding disorders such as Haemophilia and Thalassemia.

5. A mobile-based application was developed that calculates the risk that a child would inherit a bleeding issue from one or both parents using a rule-based inherited prenatal condition predictor model. Genetic counsellors or medical experts will benefit from this mobile application-based e-Health system for result prediction based on necessary parameters such as the patient's medical and family history. It also saves the time and energy of genetic counsellors by automating rules required for prediction while providing counselling. It is also helpful when patients are unclear of their medical history or family history of proposed disorders. As a result, the suggested models make a significant contribution to the delivery of health care by assisting the gynaecologists, haematologists, and paediatricians as well as the general public in determining the level of risk of having a child with Haemophilia or Thalassemia.

6. The proposed effort makes a substantial contribution to educating the public and offering online premarital and genetic counselling services to the remote public.

7. The proposed study will provide great help in preventing the birth of inherited disorders by providing preventive measures at the prenatal or premarital stage to the general public of underserved areas; so that they can take informed decisions regarding the continuation of their pregnancy with the further consultation of medical experts.

1.12 Scope and Limitation of the Study

The development of a mobile-based e-Health system for predicting the likelihood that genetic illnesses, namely Thalassemia, and Haemophilia, would be inherited

by offspring during the prenatal stage from their parents comes under the scope of the current study and is limited to looking at the variables that affect the prediction results, such as medical history, family history, obstetric history, and history of symptoms related to Haemophilia and thalassemia in couples planning a child, as well as the couples' desire to align medically before getting married. This allows informed decisions to be made under the guidance of a medical professional after receiving the prediction results through the proposed system.

1.13 Outline of the thesis

This thesis is structured into seven chapters which are presented briefly below:

- **Chapter 1** introduces e-Health systems, the medical history of inherited disorders, such as Haemophilia and Thalassemia, as well as the basic concepts of machine learning. It also discusses the role, working, and contribution of machine learning in e-Health, the fundamentals of the Bayesian network algorithm, the function of the KNN algorithm, and the hybrid approach. Additionally, this chapter performed and offered a brief explanation of the research challenges, motivation, objectives, methodology, contribution of thesis, scope and limitation of the study and thesis outline.

- **Chapter 2** discusses the necessity for literature reviews and delivers a study on e-Health systems that includes a system that offers e-services to various sectors of the healthcare industry. This chapter includes a problem statement, a research gap that was discovered during a review of the literature, and findings at the end of the chapter. The following lists the papers' publications that are related to this chapter.
  Sharma, P., Shivaram, T. R., & Sharma, A. (2016), "A methodical review of e-Health systems developed for Indian Healthcare Sector," Indian Journal of Science and Technology, 9(44), 1-6.

Sharma, P.& Sharma, A. (2022), "e-Health Systems for Mother and Child Care Domain: A Systematic Review," Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1, 439-447.

- **Chapter 3** discusses the proposed research methodology in detail; which comprises of research approach, research strategy, time horizon, sampling strategy, data collection methods, data analysis methods, research process, and limitation of proposed research methodology. The publications of the articles that are connected to this chapter are listed below.
  Sharma, P., & Sharma, A. (2019), "Collection and Analysis of Patients Dataset to Develop an e-Health System for the Prediction of Inherited Prenatal Disorders," Journal of Computational and Theoretical Nanoscience, 16(12), 5118-5121.

- **Chapter 4** discusses the brief introduction of proposed disorders and e-Health system. After this, a description of the dataset and feature selection criteria has also been discussed in this chapter. Besides this, the development and implementation of two individual models (implemented through classification algorithms such as the Bayesian network algorithm and KNN algorithm for the diagnosis of Haemophilia and Thalassemia) and a hybrid model (with the utilization of decision trees and the Gaussian Naive Bayes) has been discussed in this chapter. Furthermore, an evaluation of the performance of classification models has been depicted in the chapter. For this, the Confusion matrix has been used to assess the effectiveness of models developed during the current research work. Besides this, the chapter also depicted the comparison of the accuracy of individual models and hybrid models. Statistical analysis with the utilization of the ROC (Receiver Operating Characteristic) curve has also been presented in this chapter. The efficiency of models has been assessed through the cross-validation method and discussed under the heading of hypothesis testing of this chapter. After

this, the execution and outcome of the newly developed hybrid diagnosis model have been presented.

- **Chapter 5** discusses the rules based prediction model in detail along with the implementation and outcome of all the rules developed during the current research. The publications of the articles that are connected to this chapter are listed below.
  Sharma, P., Sharma, A., & Kakkar,S. (2024), "A multi-layered framework for analysis and prediction of Haemophilia using machine learning techniques based e-health system.", Tuijin Jishu/Journal of Propulsion Technology, 45 (3), 2295-2307.

- **Chapter 6** also discusses the design and implementation details of the proposed mobile application developed during current research work along with screen shots of prediction results generated after taking input from users. The chapter also discuss the evaluation of innovative mobile phone based e-Health system and the process of including this system in real life treatment protocol.

- **Chapter 7** discusses the conclusion and findings of the current thesis. There are also some new directions for investigation which are depicted in this chapter briefly.

Chapter 2

Review Of Literature

This chapter is organised as follows: In Section 2.1, the need for a literature analysis is discussed, and in Section 2.2, an overview of e-Health systems is given. These systems include those that use information and communication technology to provide remote consultations, electronic health record management, hospital administration, health awareness, resource tracking, teaching and learning, and mother and child care. Section 2.3 addresses a problem statement as well as the research gap that was found throughout the literature review. Ultimately, the findings are presented in Section 2.4.

2.1 Introduction

e-Health, according to the World Health Organisation, is the use of information and communication technology (ICT) in the healthcare industry for patient care, research, student education, illness tracking, and public health monitoring [97] [56]. In other words, it may be claimed that e-Health systems manage various healthcare-related tasks despite inadequate resources (such as personnel, funding, and infrastructure) [56]. Additionally, it offers necessary medical treatments via

electronic media to the patients [56]. But in reality, there are many different interpretations of the term "e-Health" that may be found in earlier literature. In this context, Mary A. Curran and Kent E. Curran stated that the e-Health has created new opportunities for the patient [97]. Similar to this, Everett J. and Kerr D. characterised e-Health in 2010 as a digital instrument that sends the data necessary for patient treatment, health education, and healthcare sector administration [30]. In 2010 Jon Avalon, Professor Victor Lane, Dr. Peter Hayward, and Jim Snaith described an e-Health system as a tool for helping healthcare professionals by organising and communicating expert information throughout the organisation [51]. In addition, Sharma Kalpa in 2012 characterised e-Health as a useful instrument for enhancing lives. This is how many researchers have described e-Health in a variety of ways based on their observations.

Numerous environmental problems in the healthcare sector (such as a lack of personnel, resources, and infrastructure), as well as how they affect social health, have been uncovered by the thorough assessment of the literature on e-Health. To address all these problems in the healthcare industry, researchers from across the world are working nonstop. As a result of their continuous efforts e-Health has appeared as a digital tool for reducing diverse healthcare problems and also for improving the level of healthcare. Even though the review of the literature confirms that e-Health systems are a recognized area of study, it focuses on the various e-Health systems that have been developed in any format (such as mobile applications, websites, portals, models, frameworks, methods, frameworks, and tools) to serve the various healthcare sectors and to offer e-Health services to underserved populations in remote areas. to identify potential service areas within the healthcare domain for which e-Health systems have not yet been established or identified. As a result, the chapter discusses the numerous systems found throughout the literature review as well as potential issue areas for which creative systems could be created.

2.2 e-Health Systems

Information and communication technologies have conducted several innovations and developments in various fields . From this perspective, e-Health has been a front-line development in the healthcare sector. It refers to electronic systems developed for the management and organisation of medical resources. The deployment of e-Health systems has made healthcare delivery fast, approachable, and affordable for remote patients. This has increased the significance of e-Health systems in terms of decision-making processes about individuals' health, digitization of medical records, and management of healthcare resources across the globe [14]. In addition to this, consistent utilization of e-Health systems is reducing medical errors; increasing administrative efficiencies; decreasing paperwork, and expanding the accessibility of healthcare services at a reasonable cost [14]. Thus, the information provided led to the decision to focus on e-Health systems as a major field of current study, and a thorough literature analysis was carried out to pinpoint a particular area to establish a particular research need for the construction of a problem statement. In this case, a systematic research approach was employed [14]. It started with the exploration of reliable academic databases (IEEE, Google Scholar Research Gate, and Willey) using a variety of e-Health-related keywords, such as "e-Health," "Telemedicine," "e-Health Systems," "e-Health Services," and "e-Health Challenges" [14]. Afterward, 153 papers were selected using the title and abstract basis exclusion technique; these were then analysed. A total of 59 systems have been seen to provide a variety of services (such as remote consulting, hospital management, teaching and learning, resource tracking, and health monitoring) across multiple healthcare domains. These observations have been recorded for additional analysis. After the assessment of a few systems, the final phase involves documenting the findings to determine a potential research field and formulating a problem statement. This process is addressed in the sections that follow in this chapter. According to the preliminary findings of the literature

research, the term "e-Health" was first proposed by the National Aeronautics and Space Administration (NASA) in the early 1960s under the name telemedicine. Subsequently, a multitude of telemedicine or e-Health systems have been developed and utilised for the digitization of the healthcare sector due to its advantages (such as improved efficacy, cost-effectiveness, and performance), which have been extensively documented in scholarly literature [16] and some of them also presented in Table 2.1.

TABLE 2.1: Telemedicine or Remote Consultancy Systems

S. No	**Systems**	**Years**	**Services**	**Ref.**
1	Tele-surgery	2005	Gives remote medical personnel access to expert information so they can quickly complete the surgical task.	[15]
2	Tele-dermatology	2005	Gives distant dermatologists a way to receive clinical and laboratory data for use in further diagnosing and treating patients.	[98] [13]
3	Tele-cardiology	2006	Delivers accurate and quick reports to a distant cardiologist so that they can treat patients from underserved areas.	[13]
4	Tele-Psychiatric	2008,2014	Offer tele-psychiatric services to those with neurological ailments or mental illnesses.	[6], [17]
5	Tele-homecare	2009, 2012	The study introduces the concept of geriatric healthcare to old patients for monitoring their health at home.	[12], [13]
6	Tele-Health	2010, 2012	Provides effective healthcare services to the children of aboriginal communities.	[6], [18], [25]

7	Tele-radiology	2013	Electronically transmits radio graphic images and text from one location to another for healthcare consultancy.	[99]
8	Tele-pathology	2013	Allows access to highly specialised pathology knowledge through communication channels with nearby medical centers.	[99]
9	e-Sanjeevani	2013	Provides the facility of medical consultancy to patients from a distant site, using a telecommunication interface.	[99]
10	Tele-Ophthalmology	2013	Provides online ophthalmology services through ISDN and Broadband connectivity.	[99]
11	Teleconsultancy	2013	Provides remote facility of medical consultancy through using electronic media.	[99]
12	Transitional care	2014	Provides post-hospital/transitional care for old patients.	[6]
13	eMedical Help	2015	Provides services for diagnosis of stress.	[52]

All of the systems included in Table 2.1 electronically provide expert knowledge or health services to patients who are in remote locations as well as to medical professionals in an emergency. In this way, the public's time and complexity were reduced by all these systems. Furthermore, it has been found that e-Health systems and telemedicine that provide remote consultations are more advantageous in areas where there is a recognised shortage of medical resources (e.g., financing, infrastructure, manpower). The notion behind an e-Health system offers more services than merely remote or teleconsultation, which led to the necessity for a

more thorough assessment of the literature. For example, it offers post-hospital care for elderly patients [100] and stress diagnosis [87]. It also allows for the online scheduling of doctor consultation appointments. This is made possible by the development of web-based e-Health systems like the Online Registration System (ORS)[101], Health You Card [102], and Practo [103], which all seek to reduce patients' waiting times. Because they save time, apps like Practo [137] are becoming more and more popular with the general public and raising the need for these systems. This fact therefore served as motivation for identifying the unidentified region for which a cutting-edge e-Health system can be constructed; as a result, more analysis was carried out to accomplish the goal. As the review of literature progresses, it has become clear that e-Health systems, through a variety of system types, provide a great deal more services to the general public in underprivileged areas, including hospital administration, digital health record keeping, health education, resource monitoring and online training [13] [15] [17]. Some of them have already been discussed in this section; they are grouped or clustered based on the services they provide and are tabulated to make it easier to understand the hierarchy for selecting a specific area for planned improvement.

The Electronic Health Record (EHR) and Hospital Management systems , which include all the systems that provide services like patient record maintenance and effective management of all the activities involved in running the hospital, are the first cluster to be identified during the literature review and covered in this section [104]. Thus, it can be said that Hospital Management Systems (HMS) and Electronic Health Records (EHR) systems [104] have a significant impact on hospital operations and resource management [105]. This has been noted in a number of the systems displayed in Table 2.2; which provides the facility of recording and saving of medical data of patients for further analysis and treatment.

TABLE 2.2: Electronic Health Record and Hospital Management Systems

S. No	**Systems**	**Years**	**Services**	**Ref.**
1	Electronic Health Record System (EHR).	2005	Helpful in collecting and saving patients' health records for future use.	[106] [35]
2	Health NET	2008	The system made it possible for hospital staff to store and route patient histories, lab results, and prescriptions within a hospital.	[105]
3	Health Vault	2015	The system gives patients the ability to collect, save, use, and share health information.	[107]
4	Health Memo	2015	Users can upload and maintain their health records electronically.	[108]
5	Health PIE	2015	Allows users to make reminders and upload personal health records.	[109]

Furthermore, several systems like Health Net-permit hospital departments to exchange medical information, including test results, patient prescriptions, and medical histories [105]. Table 2.2 also shows how systems like Health PIE, Health Vault, and Health Memo allow for the electronic uploading of individual health records; as a result, significant development has been observed in the context of such kinds of systems. Comparably, another class or cluster of systems that were seen during the literature review was resource tracking systems. These systems are beneficial in locating or monitoring vital medical resources (like hospitals, ambulances, medications, blood banks, etc.) that are needed for medical emergencies or treatments. These systems not only track the resources but also tell about the

availability of resources and much other information which are mentioned briefly in Table 2.3 along with the name of concerned systems.

TABLE 2.3: List of Resource Tracking Systems

S. No	**Systems**	**Years**	**Services**	**Ref.**
1	e-Grama	2008	Provide details on how far the target hospital is from the source community.	[21]
2	Kolkata Medical Emergency System (KMES)	2014	Details the availability of critical care, specialty-specific beds in surrounding large private hospitals, and ambulances with equipment details.	[110]
3	Malaria Early Epidemic Detection System	2014	Offers services for mapping malaria cases and cases by geographic area.	[111]
4	1mg	2015	The system provides information about the availability of medicine with their cost as well as provides the facility of online purchase of medicines.	[112]
5	WebMD	2015	Offers the ability to track symptoms, details on medications and treatments, information on first aid, and a directory of nearby medical facilities.	[109] [113]
6	5 Minute Consult	2015	Provides the fastest information about the resources required by healthcare professionals to obtain the most likely diagnosis, treatment, and management for different diseases and conditions.	[114]

7	NHP-Health Directory Services.	2015	Provides information about the availability of blood banks in India.	[115]

The examination of Table 2.3 thus demonstrated that Resource Tracking e-Health Systems have improved healthcare outcomes [66, 130] and also provide timely and significant information about needed health resources without any delay [131]. This has increased public acceptance of e-Health systems and culminated in the discovery of a new type of cluster, known as Health Awaring e-Health Systems. This cluster consists of all the systems that inform the general people about various diseases, their symptoms, available treatments, and disease patterns intending to enable individuals to take charge of their health. The brief descriptions of these systems have been depicted in Table 2.4.

TABLE 2.4: List of Health Awaring e-Health Systems

S. No	**Systems**	**Years**	**Services**	**Ref.**
1	Genetic Home Reference (GHR)	2010	Provides information resources on genetics.	[50]
2	Cardio Pulmonary Resuscitation (CPR)	2012	It is a mobile application that provides services for preventing /recovering from cardiac arrest.	[111]
3	Mswasthya	2014	Provides different Services like diabetes monitoring, vaccination alerts, and hospital search.	[99]
4	Health Central.Com	2015	Empowers patients and care givers with the information to permit better health outcomes.	[116]
5	Medscape	2015	Provides the latest news and information related to the healthcare field to users.	[113]

6	Isabel Symptom Checker	2015	Provide a list of diseases and their symptoms.	[117]
7	Veegilo	2015	Provides the facility to see trends and patterns of diseases at a glance.	[118]
8	e-medicine	2016	Provides information related to medicine associated with disease.	[119]

The analysis of every system shown in Table 2.6 showed that the systems significantly contribute to raising the general public's level of knowledge [118] by giving users detailed information about various diseases so they can make educated decisions. This is accomplished because the public has been using the systems regularly over the last few years, particularly between 2010 and 2016. In addition, e-Health systems offer an online platform for health workers to receive critical knowledge and information in conjunction with the development of new systems.

These systems collectively referred to as Teaching and Learning e-Health Systems, offer electronic teaching and learning features. As the name suggests, this cluster consists of all the systems that enable distant workers, medical students, or personnel to get online instruction, either in full or in part, so they may complete their tasks more quickly. Many systems were found in this context throughout the literature study, and Table 2.5 provides a brief discussion of them.

TABLE 2.5: List of e-Health Systems Related to Teaching and Learning Domain

S. No	**Systems**	**Years**	**Services**	**Ref.**
1	Geochat	2011	Provides the facility of chatting to community health workers in rural areas.	[120]

2	Tele- mentoring	2005	Define the concept of tele-mentoring for providing real-time and live interactive teaching techniques/ procedures to a student.	[98]
3	Tele-proctoring	2006	Provides the facility for evaluation of surgical trainees from distance to proctor.	[98]
4	Free Open-Access Medical Education (FOAM)	2014	Provides the latest information to care providers regarding medicine.	[48]
5	Mobile Alliance for Maternal Action (MAMA)	2015	Essential health information on maternal behaviour is provided.	[121] [56]

Despite the constraints of medical experts in underserved areas, the systems listed under the teaching and learning cluster do a remarkable job of teaching techniques and procedures to remote medical students or interns through communication technology. In addition, the cluster highlighted programmes like Geochat and the Mobile Alliance for Maternal Action (MAMA), which offers mother care and pregnancy-related services to lower the maternity rate and the health problems that accompany pregnancy and childbirth. Even though these systems are linked to the process of teaching and learning, their focus is particularly on issues about pregnancy, childbirth, and newborn care under the Mother and Child Care domain. This is a significant area of concern because a review of the literature revealed a higher-than-expected rate of maternal or foetal death during delivery, necessitating further investigation and development. This is what led to the decision to conduct research in this field. To identify the research deficit in this area, more review work has been done. Numerous distinct types of e-Health systems that support both prenatal and postpartum patients have been uncovered during the literature research and also presented in Table 2.6. These systems have been

grouped under the umbrella of mother and child care, which is further subdivided into two subclasses: e-Health systems for prenatal or mother care and e-Health systems for newborn care.

TABLE 2.6: e-Health Systems Related to Mother and Child Care Domain

S. No	**Systems**	**Years**	**Services**	**Ref.**	**Focused**
1	Neonatology	2013	Gives a healthcare practitioner time-sensitive information about neonatal care.	[57]	Newborn Care
2	Standard Treatment Protocol (STP)	2014	Instructions for managing an ill infant at a small hospital with few resources are given to doctors, nurses, and midwives.	[122]	Newborn Care
3	Essential Newborn Care	2014	Improve the knowledge and abilities of medical professionals in areas with limited resources about vital infant care.	[122]	Newborn Care
4	AIIMS-WHO CC ENBC	2014	Provides essential education in newborn nursing for small hospitals with limited resources.	[123]	Newborn Care
5	Essential Newborn Nursing	2014	Improve the knowledge and abilities of medical professionals in areas with limited resources about vital infant care.	[55]	Newborn Care
6	The Kangaroo Mother Care Initiative (KMC)	2014	Disseminate knowledge skills among healthcare providers and parents of low birth weight babies.	[124]	Newborn Care
7	New Born Care	2015	Provides information regarding newborn care to nurses.	[57]	Newborn Care

8	Sick New born	2015	Provides standard treatment protocols to nurses and doctors for managing sick neonatal.	[56]	Newborn Care
9	Essential Care for Every baby (ECEB)	2015	Improve the knowledge and abilities of healthcare professionals on vital infant care in areas with limited resources.	[56]	Newborn Care
10	A mother - and-child care module	2010	Provides prenatal care and an extended immunization campaign for the underserved population in the border region.	[125]	Mother/ Prenantal Care
11	OB Insulin	2014	Depending on user input such as maternal weight; determine the initial dose for a pregnant lady needing insulin.	[60]	Mother/ Prenantal Care
12	Mobile Alliance for Maternal Action (MAMA)	2015	Essential health information on maternal behaviour is provided.	[56] [121]	Mother/ Prenantal Care
13	Verboise	2011	It offers recommendations for infant care, checkup reminders, and monthly updates on what to anticipate during pregnancy.	[126]	Mother/ Prenantal Care

Hence, the field of mother and child care refers to the establishments that provide essential healthcare services to expecting mothers, new mothers who have given birth, and their young offspring. In this way, this domain is divided into sub-part prenatal care and post-natal care each of which has e-Health systems linked to it. Prenatal care systems, for instance, deal with the treatment given to expectant mothers and their unborn children during pregnancy. Conversely, post-natal care systems deal with the postpartum care given to a woman and her child.

Prenatal and postnatal care are therefore crucial areas of concern because these stages account for the bulk of maternal and newborn deaths [56]. Furthermore, due to the lack of prenatal care in rural or underdeveloped areas, about 250,000 babies are born each year with genetic abnormalities such as sickle cell anaemia, Haemophilia, and Thalassemia [56]. Hence, the fact motivated the selection of the area of prenatal care which comes under the mother care for the current study so an innovative system intended to be developed that will electronically provide prenatal services to the underserved public (especially to pregnant women) in the context of inherited disorders. However, the examination of Table 2.6 revealed important advancements in e-Health Systems about neonatal care. However, there aren't many systems available for prenatal care, which is insufficient because these systems give pregnant women the necessary care and attention so that they can give birth to their kids safely and successfully. Despite having been selected as the study's topic, the exact location of the problem formulation has not yet been determined. Hence, an in-depth analysis of the e-Health systems concerning mother and child care has been conducted based on specific variables covered by the systems. This has allowed for the identification of the precise area for problem formulation, especially in the context of prenatal care, and the purposeful development of an innovative system in this context.

2.3 Research Gap and Problem Statement

A factor-based review of the e-Health system (acquired from Table 2.6) connected to the mother and child care domain has been undertaken and is briefly shown in Table 2.7; which discusses the various e-Health system types pertinent to maternal or prenatal care, as well as child or newborn care, and which may offer a range of services necessary for the care of a mother, a pregnant woman, and her kid (expected or newborn) [56]. In addition, Table 2.7 shows which system covers how many different factors (factors here refer to the services that e-Health

systems provide) such as information updates on mother/newborn care, check-up reminder services (which fall under the category of prenatal care services offered to expectant mothers during the prenatal stage), newborn nursing resources (which contain all necessary information and treatment for the care of a newborn), training tool for neonatal (which includes all necessary services provided for the care of a newborn), expanded programme on immunization (EPI) (prenatal care offered during pregnancy) Insulin calculator (which determines the insulin dosages given to expectant mothers receiving prenatal care); inherited disorders (which comprises systems that determine the likelihood of a disorder being inherited by an offspring at the prenatal stage); neonatal care guidelines and low birth rates (which include systems that provide the services needed for neonatal care) [66]; Among of these factors, which factor does the majority of systems use less or not at all? Thus, factors that are underutilised or not used at all have been identified as research gaps and are the focus of the current study. The parameters considered are Mother/New-born (M/N) care, Checkup Reminders(CRs),Newborn Nursing Resources(NNRs), Training Tool for neonatal(TTn), Antenatal Care/ Expanded Programme on Immunization(Ac/EPI), Insulin Calculator(IC), Inherited Disorders(IDs),Guidelines for Neonatal Care (GNC), and Low Birth Weight Babies(LWB).

Table 2.7: Factor-Based Review of e-Health Systems Related to Mother and Child Care Domain

Factors/ Solution	**M/N care**	**CRs**	**NNRs**	**TTn**	**AC/ EPI**	**IC**	**IDs**	**GNC**	**LWB**
Modul for Mother and Child care(2010) [56]	✓	✓	✓	✗	✗	✓	✗	✗	✗
Verboise (2011) [118]	✓	✓	✓	✗	✗	✗	✗	✗	✗

Newborn Care (2013) [57]	✗	✗	✗	✓	✓	✗	✗	✗	✗
Neonatology (2013) [57]	✗	✗	✗	✓	✓	✗	✗	✗	✗
Essential Newborn Nursing (2014) [55]	✗	✗	✗	✓	✓	✗	✗	✗	✗
OB Insulin Application (2014) [60]	✗	✗	✗	✗	✗	✗	✓	✗	✗
Sick Newborn (2014) [56]	✗	✗	✗	✓	✓	✗	✗	✗	✗
Essential Care for Every Baby (2014) [56]	✗	✓	✗	✗	✗	✗	✗	✗	✗
Kangaroo Mother Care (KMC) (2014) [124]	✗	✓	✗	✗	✗	✗	✗	✗	✓
Standard Treatment Protocol (STP) (2014) [122]	✗	✗	✗	✗	✗	✗	✗	✓	✗

Mobile Alliance for Maternal Action (2015) [121]	✓	✓	✗	✗	✗	✗	✗	✗	✗
Essential Newborn Care (ENBC) (2015) [55]	✗	✗	✗	✓	✓	✗	✗	✗	✗

To achieve this, an evaluation of the systems listed in Table 2.7 was conducted and it was discovered that all the factors—apart from the inherited disorder—had been covered by one or more e-Health systems related to the mother and child care domain [56]. Nevertheless, there isn't currently a system in place that provides prenatal diagnosis services for inherited disorders like Haemophilia and Thalassemia, which are becoming more and more common. Each year, almost 300,000 neonates are affected by severe haemoglobin abnormalities and require more focus on prevention [56] [97]. In this way, the outcomes of the examination of Table 2.7 clearly show the flaws in the field of early pregnancy identification and avoidance of potentially inherited disorders (specifically, Thalassemia and Haemophilia) and serve as the basis for the research problem that is suggested and stated as in Figure 2.1.

Focusing mother and child care domain to find different types of prenatal disorders and their possible causes. And, out of them to further focus on inherited disorders which cause prenatal disorders.

FIGURE 2.1: The Basis for Research Problem

Hence, the suggested study will be useful in addressing a variety of medical conditions, such as prenatal disorders caused by the inheritance of genetic abnormalities. Additionally, it will provide parents with more control by simplifying the identification of inherited prenatal disorders in foetuses, such as Thalassemia and Haemophilia [56]. This will enable informed decisions regarding pregnancy to be made or not if there is a chance of inherited disorders being found, under the guidance of a qualified medical professional. Thus, the data prompted the development of an intended system that will use the medical, family, physiological, and obstetric history of a couple planning for a baby; to identify and predict the probability of inheritance of inherited disorders (like Haemophilia and Thalassemia) from parents to offspring at the prenatal stage. These hereditary bleeding disorders (Haemophilia and Thalassemia) were chosen for the planned study because their frequency has been steadily rising. Both of these, according to the World Health Organisation (WHO), occur often and require more focus on prevention [56] [97]. Each year, more than 300,000 infants are born with serious haemoglobin abnormalities [56] [97].

2.4 Conclusion

This chapter has contributed to the medical field by providing a study on e-Health systems used by the healthcare industry and general public; so that various issues, such as staff shortages, resource shortages, and the rapid and effective diagnosis, management, and treatment of inherited prenatal disorders like Haemophilia and Thalassemia, can be resolved. Besides this, the chapter also provides fundamental knowledge about inherited disorders, Haemophilia, and Thalassemia; which would be beneficial for researchers in creating more effective e-Health systems for the diagnosis, management, and treatment of hereditary bleeding disorders. To create related e-Health Systems, one must be well-versed in the use of information and communication technology as well as a medical background in inherited

bleeding disorders. Thus, the chapter offers other scholars a foundation upon which to build their future research endeavors in this domain. The initiation of this chapter started with a search of different articles. For this, different keyword indices such as "e-Health", "Telemedicine", "Haemophila", "Thalassemia", "AI or Machine Learning Based tool for Haemophilia" and "AI based Framework for Thalassemia" has been utilized. After that, findings acquired from these articles have been organized in tabular form for further analysis. Consequently, patterns and trends have emerged from the literature review, illustrating the development of various e-Health systems for various healthcare domains. The study also revealed the advantages of current e-Health solutions that are intended to help the general population as well as the medical community. Additionally, it is noted that e-Health systems created with the use of information and communication technology resourcefully manage imprecise, incomplete, and dynamic patient data. These platforms also benefited the general public, patients, and physicians by acting as wise resources for making choices. Hopefully, this study has succeeded in meeting the goal of a literature review in the following ways: It gives specifics on the research papers that were published about e-Health, telemedicine, mother and child care, Haemophilia and Thalassemia; it details the articles published with their benefits and services; it also identifies research gaps based on key findings, services offered by existing e-Health systems; and what kind of parameters has been taken for development. It also plays a fundamental role in formulating the research problem, preparing the research design, collecting, and analysing the data. It is believed that the work put forward in this study would help other researchers in the creation of future e-Health systems that will properly and precisely detect and treat various diseases. The publications of the articles that are connected to this chapter are listed below.

Sharma, P., Shivaram, T. R., & Sharma, A. (2016). A methodical review of e-Health systems developed for Indian Healthcare Sector. Indian Journal of Science and Technology, 9(44), 1-6.

Sharma, P., & Sharma, A. (2022). e-Health Systems for Mother and Child Care Domain: A Systematic Review. Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1, 439-447.

Chapter 3

Research Methodology

The goal of the current study is to create an electronic healthcare system that can predict prenatally if a child will have haemophilia or Thalassemia. For this, four distinct objectives have been established to achieve the primary aim of the proposed research activity. The following is a general description of the suggested objectives:

1. To collect a dataset of identified ancestral patterns, and physiological and pathological symptoms.
2. To integrate bayesian network and machine learning technique for the prediction of prenatal disorders.
3. To evaluate and validate results using intelligent statistical techniques.
4. To develop an application for the diagnosis of prenatal disorders.

Thus, to achieve the suggested objectives of the current study, a systematic research approach has been applied. In this way, this chapter discusses the research methodology of the proposed work in detail, including the research approach, the research type, the research strategy, the time horizon, the sampling strategy, the data collection methods, the data analysis methods, and the proposed work. This

chapter is arranged as follows: The research process, as well as the research approach, strategy, time horizon, sample plan, data collection methods, and data analysis techniques, are covered in Section 3.1, which also presents the research methodology. The strategy intended for the use of proposed work is covered in Section 3.2. In Section 3.3, the final findings are made.

3.1 Introduction

According to J. Creswell (2002), research methodology is a plan or strategy to conduct research work in a scientific way and link methods to outcomes [94]. In other words, it can be said that research methodology is a way to systematically solve the research problem [94]. Hence, the current research work also utilises the research methodology, which includes the research approach, research type, research strategy, time horizon, sampling strategy, data collection methods, and data analysis methods. The detailed description of each step associated with research methodology has been specified in the following sections:

3.1.1 Research Approach

The research approach refers to the collection of procedures that decides the overall process of research, which includes various methods and techniques associated with data collection, analysis, interpretation, and utilisation. Research approaches are commonly divided into two main categories: qualitative and quantitative [94]. But, sometimes combination of two can be used as a mixed approach. Hence, the selection of a research approach entirely depends on the research problem, audience, and expected goal of the purposeful research work. Consequently, the suggested study employed a mixed methodological approach to achieve the established goals of the ongoing research [56]. Mixed methodology refers to the use of a qualitative as well as quantitative approach for the conduction and completion of

proposed research work [94] [83]. The use of a qualitative as well as quantitative approach helped in understanding the real situation and the demands and expectations of users of e-Health [94]. A brief introduction of qualitative and quantitative approaches and justification for using them in the proposed work is defined below:

3.1.1.1 Qualitative Approach

A Qualitative approach provides insight into the problem as well as helpful instrument (questionnaire) design for the succeeding stages [94]. It is concerned with the subjective assessment of attitudes, opinions, and behaviour [94]. The techniques of interview, literature review, and discussion are utilised by this approach [94]. Hence, the proposed study utilized the literature review, informal discussion, and interview methods of this approach.

Why is this approach useful for the proposed problem?

Various factors support the use of a qualitative approach at the initial stage, which are listed below:

- To gain rich insight into inherited disorders and preventive measures.
- To take input from patients and doctors related to haemophilia, Thalassemia, and inheritance patterns.
- To know the various types of parameters and which are more suitable for the early diagnosis of inherited disorders.

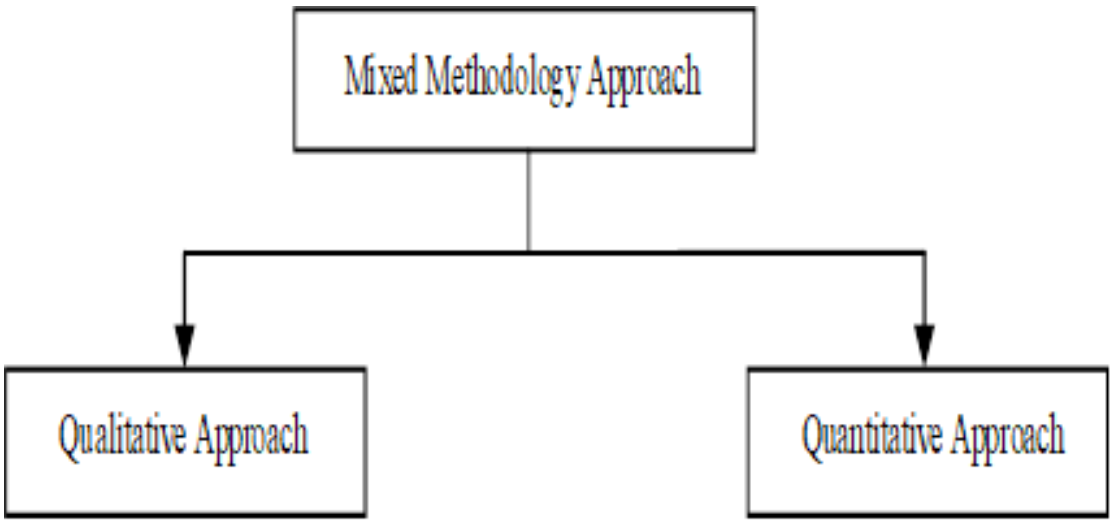


FIGURE 3.1: Mixed Research Approach

3.1.1.2 Quantitative Approach

This approach involves the generation of data in quantitative form which can be subjected to rigorous quantitative analysis formally and rigidly [94]. A quantitative approach is further sub divided into inferential, experiment, and simulation approaches. This approach has used survey and case study methods for the conduction of research work [94]. The characteristics of the quantitative approach provide great motivation for selecting the survey and case study method as an important tool for proposed research work [94]. Hence, a separate survey of the target audience will be administered to collect requirements and responses from the general public as well as professionals. Based on the survey outcome suitable suggestions can be made to design a proposed system and application.

Why is this approach useful for the proposed work?

- To study a sample of the population for quantifying the demands of target users (Doctors, patients, and medical students) for designing user-specific systems and an application for caring for mothers and children in the context of inherited prenatal disorders.

3.1.2 Research Strategy

The research approach establishes the methodology for the study [127]. Various research strategies are available, including action, grounded, ethnographic, case, and experimental research. Any individual's decision is dependent upon the study's objectives [151]. The current study combines the case study and the ethnographic research technique. A case study is a comprehensive analysis of a single subject or person [127]. By applying case studies of different patients about the context of sickness symptoms and supporting the development of a system with a defined objective, case studies will be very helpful in gathering the necessary data. Similarly, the ethnographic research approach is the process of watching and recording participants' experiences and views in their natural environments to enhance research outcomes [127]. Therefore, the choice to employ case studies and ethnographic research strategies to finish the current study project was motivated by their inherent qualities.

3.1.3 Time Horizon

The term "time horizon" describes the amount of time needed to gather data for a specific project. In this case, there are two different types of options: cross-sectional, or short-term, and longitudinal, or long-term. In cross-sectional data collection, a brief interval is designated for data collection. On the other hand, longitudinal time horizon refers to data that is continuously collected over an extended period after a certain interval. Since haemophilia and Thalassemia are rare conditions, there aren't many persons with them; therefore, the current study collected data longitudinally. As a result, to obtain enough data, researchers must wait and then resume the procedure [127].

3.1.4 Sampling Strategy

Sampling strategy refers to methods used for the selection of participants and collection of data for proposed research work which is further divided into two categories such as probability and non-probability sampling. In probability sampling, participants are selected randomly from a population. In contrast, non-sampling supports the method of selecting participants based on ease of access or non-randomly [127]. To obtain the necessary data for the current research project, a non-probability sampling strategy was chosen. The sample participants were chosen based on their knowledge, connections, and experience related to specific diseases, such as Thalassemia and haemophilia [128]. The sample participants in this study, which includes patients and affiliated physicians, are chosen based on their level of active participation, past medical history, and sufficient and pertinent experience about the symptoms, diagnosis, treatment, and management of thalassemia and haemophilia. Furthermore, experts in medicine from renowned hospitals in Jalandhar and Ludhiana, as well as the head of a nongovernmental organisation dedicated to haemophilia located in Dehradoon, were chosen as study participants based on their research background and comprehension of raw data regarding regions [128]. The details of these experts are provided in Table 3.1:

Table 3.1: List of Specialists Involved in the Present Research Study

S. No	**Name**	**Address**	**Domain**	**Location**
1	Dr. Shruti Kakkar MD Fellowship in Hematology/Oncology Associate Professor	Department of Paediatrics, Paediatric Haematology/Oncology Unit Dayanand Medical College and Hospital Ludhiana, Punjab India.	Thalassemia	Ludhiana Punjab India
2	Dr. Anubha Bharthuar, Medical Oncologist, MD ABIM Fellow (US) Head Medical Oncology Haematology	Patel Cancer and Super Speciality Hospital, Jalandhar, Punjab, India.	Haemophilia	Jalandhar Punjab India

3	Dr. Bhavneet Kaur MBBS, MD (Obs and Gyne)	Tagore Hospital and Heart Care Centre, Jalandhar, Punjab, India.	Gynecologists	Jalandhar Punjab India
4	Mr. Deepak Singal Chief Functionary Officer and Secretary,	Haemophilia Society Dehradoon Chapter, Dehradoon, Uttarakhand, India.	Haemophilia	Dehradoon, Uttarakhand, India.
5	Ms. Somya Coordinator	Haemophilia Day Care Christian Medical College and Hospital, Ludhiana, Punjab, India.	Haemophilia	Ludhiana Punjab India

All of the aforementioned experts helped comprehend the basic information about intentional bleeding disorders such as haemophilia and thalassemia, as well as other facts and numbers related to the current study. They also provided patient data to start more interviews and discussions. To obtain a data set about physiological, family, symptom, and obstetric history. For this, open-ended questions were used in both in-person and telephone interviews with patients and their families. The participants' prior consent was obtained before the interview began. As a result, those who were prepared were interviewed to collect data; 501 patients' records were made for specific purposes. In this way, patients and their families actively participated in data sampling in addition to doctors and subject experts.

3.1.5 Data Collection Methods

Data collection is an important task and is associated with the collection of required data for the successful completion of research work [94]. A variety of data collection methods, including discussions, case studies, questionnaires, interviews, and observation, can be used to gather the data. To do the assignment with minimal complexity, the right kind of data and procedures must be used [94]. For this reason, the in-depth interview and discussion method was used to collect data for the current study. As the secondary data related to the suggested study was not available publically. Thus, to collect the primary data, the in-depth interview and

discussion method was employed [94].

Interviews with concerned, knowledgeable doctors from hospitals and medical schools, as well as patients and their families, were done to obtain information related to the data set. To gather data, an outline of the proposed work was given to target participants, such as patients, along with their families and medical professionals, and their comments were noted to create a data set. Consequently, comments from patients (with their previous consent) regarding physiological, medical, family, and obstetric factors have been documented as part of the planned work's data set. Alongside this, open discussions were held to get their feedback regarding the intended system to design a more focused system for the current study. This makes it possible to conclude that the in-depth talks and interviews were very important in assembling the data set for the intended study.

The rationale behind choosing these two is that because they are both inherently informal and personal, a direct and intimate relationship has been formed between the interviewer and the interviewee [128]. However, to lower participant non-response rates, the in-depth interview and discussion strategy requires the requisite abilities to conduct a good interview [128]. That has been addressed by using a semi-structured questionnaire to start the interview and discussion, which serves as the researcher's interview guide [128]. To achieve the predetermined goal of data collection, a set of basic questions was prepared with input from domain-specific experts in medicine. Additional questions could be asked based on the specifics of the discussion and interview, though these were the suggested starters [128]; some of them are depicted in Table 3.2.

Table 3.2: List of Questions for Patients

Q. No	**Questions**
1	What is your name?
2	What is your Age?
3	When did you find out you have thalassemia or haemophilia?
4	What are the initial symptoms?
5	Currently, what kind of symptoms do you have?
6	Do you have a family history of the disorder? If, yes then please mention the relation and what kind of symptoms they have.

7	Are you married?
8	Does your partner also have haemophilia or thalassemia?
9	Do you have any kids? What is the child's thalassemia or haemophilia status (if either)?
10	What kind of treatment is going on?

TABLE 3.3: List of Questions for Doctors

Q. No	**Questions**
1	What are the major symptoms of haemophilia or Thalassemia?
2	What are the early symptoms that are used for the diagnosis of haemophilia or Thalassemia?
3	What kind of pattern or parameters (such as pathological, physiological or medical Family history and obstetric history) should be utilized for the accomplishment of proposed work?

All of the aforementioned queries are posed to the patient or members of their family. Though there have been extra questions added based on each person's unique situation, they were constantly monitored. Before recording the response, the patient requested permission by briefly outlining the goal of the work. Following that, every response was entered into a tabular format, with columns containing the following information: name, age, disorder status (Thalassemia or Haemophilia), age, gender, parents' and partners' details, family history related to the condition, obstetric history (about a child with Haemophilia and Thalassemia), and symptoms related to the condition. Now, under the direction of knowledgeable physicians, the complete column head will be chosen, with a focus on symptoms. Expert medical advice was sought before deciding to prioritise physiological characteristics above pathological ones. For this, a telephone conversation or in-person interview will be conducted based on a few questions, which are presented in Table 3.3. All of the aforementioned inquiries will help gather data, create a table for data collection, choose parameters for specific tasks, and create an e-Health system with specific goals.

3.1.6 Data Analysis Methods

Data assessment is a technique used to examine dataset gathered through in-person or telephone interviews and conversations [128]. After the completion of the data collection task, pre-processing will be carried out. Data Pre-processing is a very significant step followed during the development of any machine learning model. In this process, data will be normalized by reducing unwanted data and classified for further analysis and comparison [128]. In which, the data get transformed and brought into a state that can be efficiently understood and interpreted by the machine learning algorithm. So that further results can be produced; which can be evaluated using statistical techniques? Therefore, the next action taken after gathering the data is pre-processing and analysis. Appropriate methods, such as the correlation coefficient approach, were applied for this. Subsequently, the data set was divided into training and testing sets and features were selected to facilitate the ensuing analysis. Subsequently, advanced statistical and computational techniques such as the ROC curve have been utilised to evaluate the results of concentrated labour.

3.1.7 Research Process

Research technique is the methodical planning of steps and their sequential execution to achieve the predetermined goal and goals of the study [94]. As a result, the section outlines the different procedural stages that the current research project uses. The following lists the many procedures that were taken to carry out the suggested work:

1. The research process started with a literature review of e-Health Systems.
2. In the next step, a research gap was identified and a problem statement was formulated. Besides this, objectives have been defined.

3. In this step, data collection and pre-processing tasks have been conducted.

4. A hybrid model, which combines Gaussian naive Bayes and Decision Tree, has been developed for the categorization of inherited prenatal bleeding disorders, specifically Thalassemia and Haemophilia.

5. A new set of rules has been developed for assessing the chances of inheritance of Haemophilia or Thalassemia from parents to their offspring; this is particularly useful in situations where one spouse is unsure of their medical history, physiological makeup, or family history of illnesses. The foundation for this has been the inheritance pattern's current rules.

6. Based on the existing pattern of inheritance, a novel set of rules has been developed to determine the likelihood that a child will inherit either Thalassemia or Haemophilia from their parents. This new rule set is particularly useful in situations where one partner is unsure of their own medical history or physiological or family history of disorders.

7. This step is associated with the testing phase that was carried out using sophisticated statistical methods like the ROC curve.

8. An e-Health system that runs on a smartphone has been created to forecast the likelihood that an expecting child would inherit Thalassemia and Haemophilia from their parents.

9. In the end, a conclusion has been drawn, and then thesis writing has been accomplished.

3.2 Proposed Work

A systematic research process, as illustrated in section 3.1.7 above, has been employed to construct an e-Health system for the prediction of hereditary prenatal

illnesses, such as Haemophilia and Thalassemia [56]. Specifically, the current proposed e-Health system development includes feature selection, data collection, hybrid model development, creation of a new rule set to determine the likelihood of inheriting inherited disorders like Thalassemia and Haemophilia, creation of a rule-based prediction model that uses the newly created rule set in conjunction with pre-existing inheritance patterns, model performance testing, and creation of a mobile-based e-Health system.

3.2.1 Data Collection

Primary data collection was the first step in the research procedure because there was no secondary data set accessible about inherited bleeding disorders, such as Thalassemia and Haemophilia. For this reason, numerous laboratories were first contacted for data collection. As a result, a pathological data set was gathered and subsequently analysed; however, it was later determined that the pathological data set was irrelevant to the system's intended use [53] because the system will be used for the general population, and such data is difficult to comprehend. Furthermore, this data set was not determined to be pertinent for the early identification of a hereditary prenatal disease. As a result, additional datasets have been acquired, encompassing physiological history, obstetric history, family history, and history of symptoms linked to thalassemia and Haemophilia [53].

To gather inputs for the required data set, a variety of medical facilities, hospitals, and nongovernmental organisations (NGOs) were visited. Planned questions from a semi-structured questionnaire—which is illustrated in section 3.1.7 of this chapter—were used to conduct in-depth interviews and facilitate discussions. For this, Patients, patient families, concerned professionals, and knowledgeable doctors from hospitals and medical institutes have all been questioned. In addition, open discussions with patients and experts have taken place. Hence, to generate a data set, the target participants—patients, family members, and medical professionals—were given an outline of the proposed study.

These people's responses were then noted to create a data set. Consequently, 501 patients' responses have been recorded (with their prior authorization) as part of the proposed work, regarding physiological, medicinal, family, and obstetric aspects. However, due to a lack of patients, most medical institutions did not have enough data, and those that did have the data were not eager to assist in this way, making the process of acquiring data sets very difficult. Thus, after much work, information on 501 patients with thalassemia and Haemophilia has been collected from the Haemophilia Society Dehradun Chapter in India, Dayanand Medical College, and CMC (Christian Medical College & Hospital), all located in Ludhiana, Punjab, India.

These details consist of the patient's physiological history, obstetric history, family history, and medical history. As a result, the thesis's primary contribution is the compilation of a dataset of 501 individuals with thalassemia and Haemophilia, as secondary material about these conditions is not easily found on internet portals. With the compilation of a data set comprising 501 patients with Haemophilia and Thalassemia, the initial stage of the purposeful research work process, as shown in Figure 3.2, has thus been accomplished.

3.2.2 Data Pre-processing

Following the conclusion of the data-gathering task, pre-processing—the second most crucial stage of the intended research work, as shown in figure 3.2 has been carried out. Data pre-processing is a crucial step in creating any machine learning model. It involves transforming the data into a format that the machine learning algorithm can properly read and comprehend.

A dataset is a collection of data objects, including entities, vectors, samples, observations, points, and records. Features may also be referred to as fields, variables, dimensions, traits, and characteristics. A feature collection is called a data object. Any quantifiable trait or attribute of an object under observation can be referred to as a feature. Dimension, colour, and form are a few examples of table features.

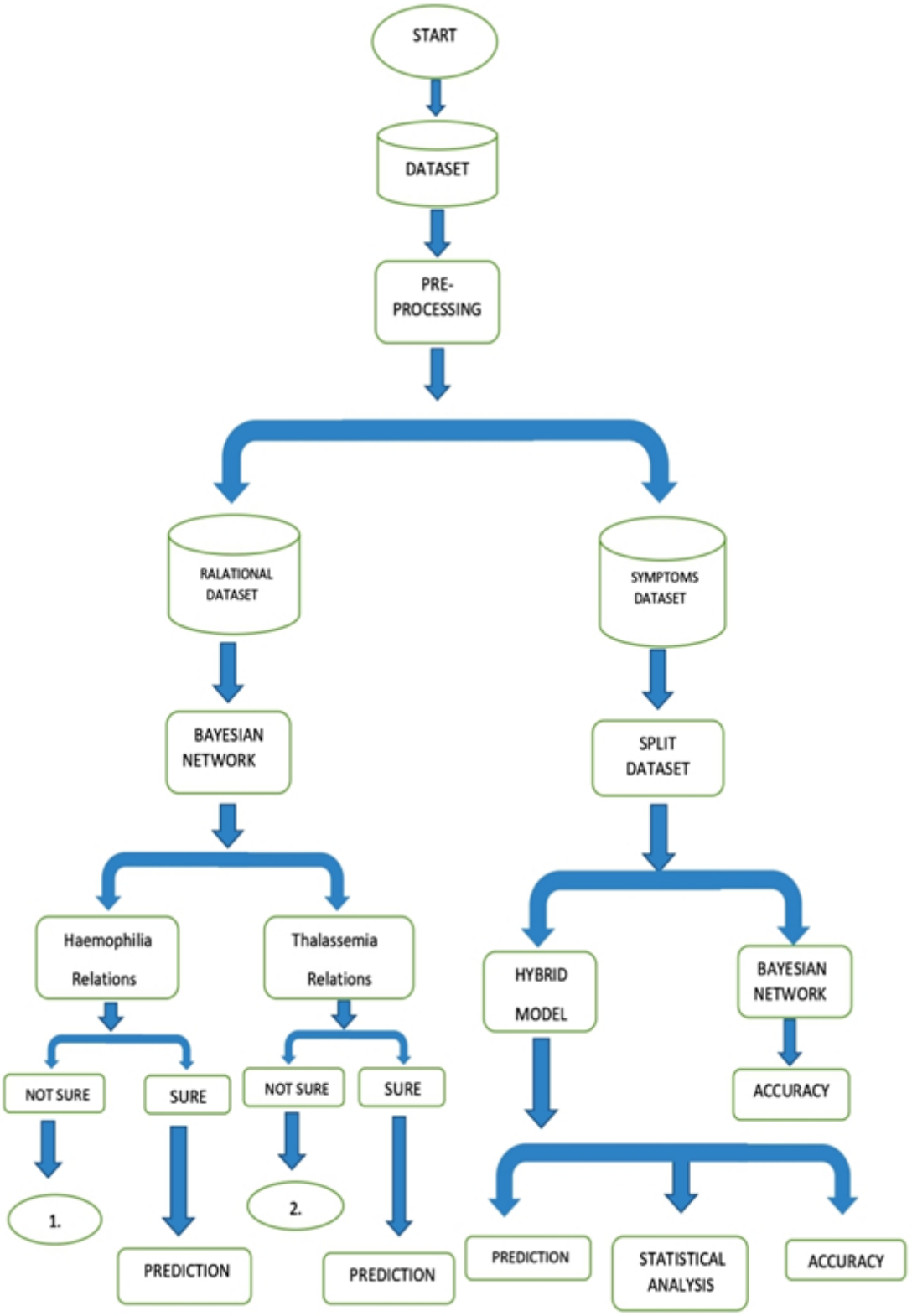


FIGURE 3.2: Research Process for Proposed Work

These characteristics can be further divided into Nominal, Ordinal, Interval, and Ratio categories within the larger Categorical and Numerical categories.
The fixed collection of values serves as the source of values for the category features. Boolean set: False, True, for example, is a categorical variable since the values must come from the set. The features with integer values are known as numerical features. These are mostly defined by numbers and have numerical characteristics. Therefore, the study's intended application will make use of pre-processing's categorical feature.
The current scenario's data pre-processing includes normalising, scaling, eliminating missing values, and balancing the dataset. These steps enable us to adjust the dataset's numeric column values to conform to a common scale. Numerical values, like 0 and 1, have been generated from the gathered responses. Following this, the feature selection work will be completed using the suitable methodology. So that the most significant features can be selected for the development of the intended system.

3.2.3 Splitting of Dataset

The data set has been divided into training and testing sets at this point. The term "training set" refers to the data set that was used to train the intended model. This data set was further separated into two categories: the symptom-based dataset related to thalassemia and Haemophilia and the relational dataset for both, as is evident in Figure 3.2. On the other hand, the term "testing data set" describes the set of data used to test the planned model and ensure that the results are accurate.

3.2.4 Development and Prediction of Models

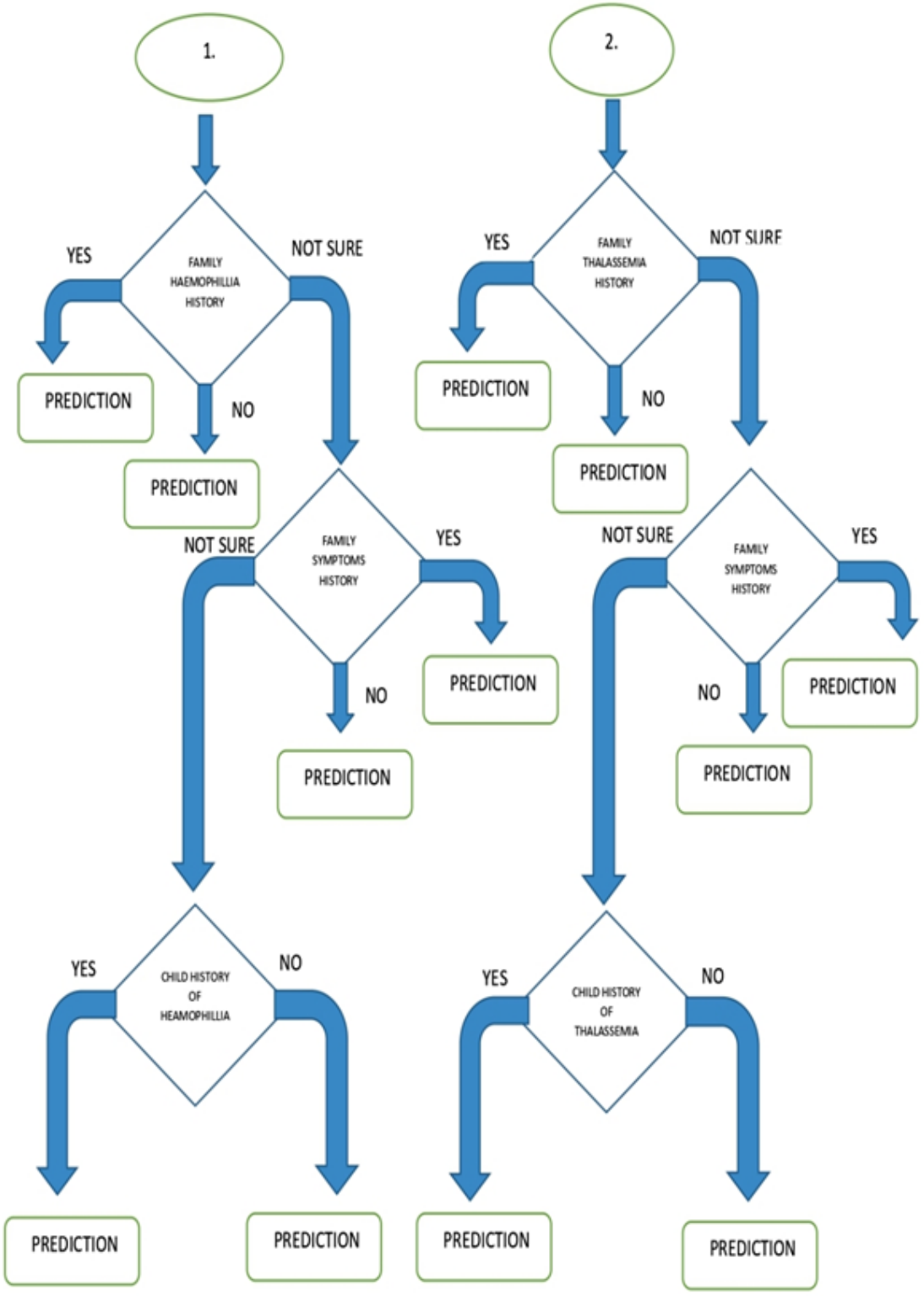


FIGURE 3.3: Research Process for Proposed Work

TABLE 3.4: Inheritance Pattern of Haemophilia

Case No	**Father's /Husband's Status of Haemophilia**	**Mother's / Wife's Status of Haemophilia**	**Probability of Inheritance**	**Ref.**
1	Yes	No	# All sons will be normal. # All daughters will be carriers of Haemophilia.	[66] [66] [69] [74]
2	No	Yes(carrier)	# 25% risk of having a daughter with Haemophilia-related genes. # 25% chance of having son with Haemophilia. # 50% likelihood that an offspring (boy or female) will not carry the Haemophilia gene	[66] [66] [69] [74]
3	Yes	Yes(carrier)	# 25% chance of having a son with Haemophilia. # 25% chance of having a normal son. # 25% chance of having a daughter who is a carrier of Haemophilia. # 25% chance of having a daughter who has Haemophilia.	[66] [66] [69] [74]
4	No	No	# All sons will be normal. # None of the children (daughter/son) have Haemophilia or carry the gene	[66] [66] [69] [74]

During this phase, a classification model was developed using the set of symptoms obtained from a dataset containing patients with thalassemia and Haemophilia. Various machine learning algorithms, including Bayesian networks, KNNs (K-Nearest Neighbour), and decision trees, were also implemented individually for the development of a diagnosis or classification model, to select the most appropriate combination of algorithms for the development of a hybrid model. As a result, a hybrid model has been developed with the integration of gaussian naive bayes and a decision tree for the diagnosis of Haemophilia and Thalassemia based on the parameters classified according to their symptoms. Despite effectively classifying the patients into Haemophilia and Thalassemia based on their symptoms, the model was unable to predict the likelihood that their offspring will acquire

inherited illnesses like Haemophilia and Thalassemia from their parents. To address this issue, a prediction model has been created using a Bayesian network, current inheritance patterns that have been obtained from literature, and a newly developed rule set in the next level, which is also shown in Figure 3.3. This model predicts the likelihood that a child will be born with Haemophilia or Thalassemia based on the medical, family, physiological, and obstetric history of the parents. The rule set that was used to develop this model has been very important because it predicts the likelihood of inheriting Haemophilia and Thalassemia by analysing user-provided values in the context of predefined parameters like medical history, family history, symptom history, and obstetric history in a hierarchical manner. This allows the intended system to make predictions in situations where one or both partners are unsure of the history of Haemophilia and Thalassemia, which is not possible in the case of the existing inheritance pattern because these patterns are limited to situations in which the couple is aware of the medical history of inherited disorders like Haemophilia and Thalassemia. Tables 3.4 to 3.9 provide a brief explanation of these rules that were produced during the current research investigation as well as information on the previous inheritance patterns.

Table 3.4 makes it very clear that the cases of Haemophilia's inheritance pattern that have been documented in the literature are few and only apply to individuals or couples who are already aware of their Haemophilia status (whether they suffer from the disorder or not) [83-85] [90][99] [66] [66] [69] [74] [67] [73] [77] and does not cover the cases where the individuals or couples are not sure about the medical history of Haemophilia. Therefore, to solve this issue, a new set of rules was created during the current research work and is shown in table 3.5. This represents the main contribution of the current study. It makes use of the medical history, family history, physiological history, or history of symptoms in the family, and the obstetric history of the couple. These patterns give guidelines for estimating the likelihood that children may inherit Thalassemia and Haemophilia from their parents.

TABLE 3.5: New Rules and Predictions Results of Haemophilia (when a wife is unsure of her medical history)

S. No	**Husband**	**Wife**	**Do you have a family history of haemophilia ?**	**Your / Family Member/ Relatives having the symptoms of haemophilia**	**Obstetric History (Any Child with Haemophilia)**	**Prediction Results**
1	Yes	Not Sure	Yes	—	—	The husband has Haemophilia and the wife is not sure about the disorder. But, she has a family history. So, this has raised the chances of inheritance from parents to offspring. Hence, further testing and diagnosis of a wife are required for the confirmation as she is not sure about the disorder, and if she is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.

2	Yes	Not Sure	No	—	—	Husband has the Haemophilia and no family history/obstetric history of Haemophilia has been observed in the case of the wife. Hence, it is assumed that she will not have Haemophilia. But, for the confirmation further testing and diagnosis of her is required, and if she is found negative for Haemophilia then the chances of inheritance will be: # All sons of a couple will be normal. # All daughters will carry Haemophilia.
3	Yes	Not Sure	Not Sure	Yes	—	The husband has Haemophilia and the wife is not sure about the disorder. But, she has a family history. So, this has raised the chances of inheritance from parents to offspring. Hence, further testing and diagnosis of a wife are required for the confirmation as she is not sure about the disorder, and if she is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.

4	Yes	Not Sure	Not Sure	Not Sure	Yes	The husband has Haemophilia and the wife is not sure about the disorder. However, she has an obstetric history in the context of disorder. So, this has raised the chances of inheritance from parents to offspring. Hence, further testing and diagnosis of a wife are required for the confirmation as she is not sure about the disorder, and if she is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.
5	Yes	Not Sure	Not Sure	Not Sure	No	The husband has Haemophilia and the wife is not sure about the disorder. Besides this, No family and obstetric history related to Haemophilia has been found in the case of the wife. Hence, further testing and diagnosis of a wife are required for the confirmation as she is not sure about the disorder and if she is found negative for Haemophilia then the chances of inheritance will be: # All sons of a couple will be normal # All daughters will carry Haemophilia.

6	No	Not Sure	Yes	—	—	No medical history of Haemophilia has been observed in the context of the husband and wife not sure about the disorder. However, she has a family history of disorder. Hence, further testing and diagnosis of a wife are required for the confirmation as she is not sure about the disorder, and if she is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child.
7	No	Not Sure	No	—	—	The husband did not have Haemophilia. No family history of Haemophilia has been observed in the case of the wife. Hence, it is assumed that she will not have Haemophilia. But, for the confirmation further testing and diagnosis of her is required, and if she is found negative for Haemophilia then the chances of inheritance will be: # None of the children (Daughter/Son) will have Haemophilia or carry the gene.

8	No	Not Sure	Not Sure	Yes	—	The husband did not have a medical history of Haemophilia and the wife did not sure about the disorder, but she confirmed the existence of symptoms of the disorder in the family. Hence, further diagnosis and testing are required, and if she is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25%(1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child.
9	No	Not Sure	Not Sure	Not Sure	Yes	The husband did not have a medical history of Haemophilia. His wife is not sure about the disorder. But she has an obstetric history. Hence, further diagnosis and testing of her are required, and if she is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child.
10	No	Not Sure	Not Sure	No	—	The husband has no Haemophilia. No family history of Haemophilia has been observed in the case of the wife. Hence, it is assumed that she will not have Haemophilia. But, for the confirmation further testing and diagnosis of her is required, and if she is found negative for Haemophilia then the chances are that: # None of the children (Daughter/Son) will have Haemophilia or carry the gene.

11	No	Not Sure	Not Sure	Not Sure	No	The husband is normal as he has no Haemophilia. No family history of Haemophilia has been observed in the case of the wife. Hence, it is assumed that she will not have Haemophilia. But, for the confirmation further testing and diagnosis of her is required, and if she is found negative for Haemophilia then the chances are that: # None of the children (Daughter/Son) will have Haemophilia or carry the gene.
12	No	Not Sure	Not Sure	Not Sure	Yes	The husband is normal as he has no Haemophilia. But the wife has an obstetric history of disorder. Hence, it is assumed that she will have Haemophilia. But, for the confirmation further testing and diagnosis of her is required, and if she is found positive of Haemophilia then the chances is that: # 25%(1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child.

The recently created rule set was further split into two groups: the rule set for Haemophilia and the rule set for Thalassemia. They have been further separated into two subcategories. The first category of rules refers to the scenario where the wife is unsure of her medical history and the husband is aware of his. In this instance, the wife is asked about her family history; if she is aware of it, the result will be predicted; if not, the wife is asked about her history of symptoms or physiological symptoms; if she is unsure, then obstetric history has been asked and prediction will be made based on inputs provided by her. A similar procedure is shown in Table 3.5, which lists all 12 rules for determining the likelihood of a

child being born with Haemophilia in situations where the woman is unsure about her and her family's history of the condition. Moreover, unlike the previous rules, the recently formed set of rules makes use of more than two participants, such as spouses and family members.

Similarly, the second group pertains to the rules that determine the likelihood of inheritance when the wife is aware of her medical history and the husband is unsure of his. In this instance, the husband's family history is enquired about; if he is aware of it, the result will be predicted; if not, the physiological symptoms have been asked; if he is not sure about it then obstetric history will be questioned, and ultimately the result about inheritance will be predicted. The same can be seen in following table 3.6. Hence, the study proved helpful in circumstances when the spouse is unsure of his medical history, table 3.6 shows the total of 23 rules that will be put into practice for the creation of the intended prediction model. This model will predict the likelihood that the foetus would acquire Haemophilia from the parents during the prenatal stage.Table 3.6 represents the rules and predictions in case of Haemophilia. The considered parameters are Wife Status (WS), Husband Status (HS), "Do you have family history of Haemophilia?"(HFH), "Your / Family Member/ Relatives having the symptoms of haemophiliaand" (HSYM),"Obsteric History(Any child with Haemophilia)"(HOBS) etc.

TABLE 3.6: New Rules and Predictions Results of Haemophilia (when husband is unsure of her medical history)

S. No	**WS**	**HS**	**HFH?**	**HSYM**	**HOBS**	**Prediction Results**
1	No	Not Sure	Yes	—	—	No medical history of Haemophilia has been observed in the context of the wife and the husband did not sure about the disorder. But, he has a family history of disorder. Hence, further testing and diagnosis are required for the confirmation as he is not sure about the disorder, and if he is found positive for Haemophilia then the chances of inheritance will be: # All sons of a couple will be normal. # All daughters will carry the Haemophilia genes.

2	No	Not Sure	No	—	—	The wife did not have Haemophilia and the husband did not sure about the medical history of the disorder. However, he did not have a family history of the disorder. Hence, it is assumed that he will not have Haemophilia. But, for the confirmation further testing and diagnosis is required, and if he is found negative for Haemophilia then the chances of inheritance will be: # None of the children (Daughter/Son) will have Haemophilia or carry the gene.
3	Yes (carrier)	Not Sure	Yes	—	—	The wife is a carrier of Haemophilia and the husband is not sure about the disorder. But, he has a family history. So, this has raised the chances of inheritance from parents to offspring. Hence, further testing and diagnosis of the husband are required for the confirmation as he is not sure about the disorder, and if he is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.
4	(carrier)	Not Sure	No	—	—	The wife is the carrier of Haemophilia and her husband is not sure about the Haemophilia, he has no family history of the disorder. Hence, it is assumed that he will not have Haemophilia. But, for the confirmation further testing and diagnosis of him is required, and if he is found negative for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child

5	Yes (carrier)	Not Sure	Not Sure	Yes	—	The wife is the carrier of Haemophilia and the husband is not sure about the disorder. But, he has a family history of physiological symptoms. So, this has raised the chances of inheritance from parents to offspring. Hence, further testing and diagnosis are required for the confirmation as he is not sure about the disorder, and if he is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.
6	Yes	Not Sure	Not Sure	Not One	No	The wife is the carrier of Haemophilia and the husband is not sure about the disorder. Besides this, he did not have a family and obstetric history of disorder. So, this has raised the chances of inheritance from parents to offspring. Hence, further testing and diagnosis are required for the confirmation as he is not sure about the disorder, and if he is found negative for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child
7	Yes (carrier)	Not Sure	Not Sure	Not Sure	No	The wife is the carrier of Haemophilia and the husband is not sure about the medical and family history of the disorder. But, his wife did not have any obstetric history of disorder. Hence, further testing and diagnosis are required for the confirmation as he is not sure about the disorder, and if he is found negative for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child

8	No	Not Sure	Not Sure	Yes	—	The wife did not have a medical history of Haemophilia and the husband did not sure about the disorder, but he confirmed the existence of symptoms of the disorder in the family. Hence, further diagnosis and testing are required, and if he is found positive for Haemophilia then the chances of inheritance will be: # All sons of the couple will be normal. # All daughters will carry the Haemophilia genes.
9	No	Not Sure	Not Sure	Not Sure	Yes	The wife did not have a medical history of Haemophilia. The husband is not sure about the disorder. But, his wife has an obstetric history. Hence, further diagnosis and testing are required, and if he is found positive for Haemophilia then the chances of inheritance will be: # All sons of the couple will be normal. # All daughters will carry the Haemophilia genes.
10	No	Not Sure	Not Sure	No	—	Wife did not have Haemophilia. No family history of Haemophilia has been observed in the case of the husband. Hence, it is assumed that he will not have Haemophilia. But, for the confirmation further testing and diagnosis is required, and if he is found negative for Haemophilia then the chances are that: # None of the children (Daughter/Son) will have Haemophilia or carry the gene.
11	No	Not Sure	Not Sure	Not Sure	No	The wife is normal as she did not have Haemophilia. No family history of Haemophilia has been observed in the case of the husband. Hence, it is assumed that he will not have Haemophilia. But, for the confirmation further testing and diagnosis is required, and if he is found negative for Haemophilia then the chances are that: # None of the children (Daughter/Son) will have Haemophilia or carry the gene.

12	No	Not Sure	Not Sure	Not Sure	Yes	The wife is normal as she did not have Haemophilia and the husband is not sure about his and his family history of the disorder. But, his wife has an obstetric history of disorder. Hence, it is assumed that he has Haemophilia. But, for the confirmation further testing and diagnosis is required, and if he is found positive for Haemophilia then the chances are that: # All sons of the couple will be normal. # All daughters will carry the Haemophilia genes.
13	Yes (carrier)	Not Sure	Not Sure	Not Sure	No	Wife is the carrier of Haemophilia. The husband was not sure about the disorder and he has no family history, either. Hence, further diagnosis and testing of him are required and if he is found negative for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child
14	Yes (carrier)	Not Sure	Yes	—	—	Wife is the carrier of Haemophilia. The husband was not sure about the disorder. But, he has a family history. Hence, further diagnosis and testing of him are required, and if he is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.

15	Yes (carrier)	Not Sure	Not Sure	Yes	—	The wife is the carrier of Haemophilia But, no medical/family history of Haemophilia has been observed in the case of the husband. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia and if he is found negative for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child.
16	Yes (carrier)	Not Sure	Not Sure	Yes	—	Wife is the carrier of Haemophilia. But, a family history of physiological symptoms of Haemophilia has been observed in the case of the husband. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia, and if he is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.

17	Yes (carrier)	Not Sure	Not Sure	Not Sure	Yes	The wife is the carrier of Haemophilia and the husband is not sure about the disorder. However, the obstetrics history of his wife was found positive in the context of Haemophilia. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia, and if he is found positive for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a daughter, who has Haemophilia. # 25% (1/4) chances of having a son with Haemophilia. # 25% (1/4) chances of having a normal boy.
18	Yes (carrier)	Not Sure	Not Sure	Not Sure	No	The wife has a medical history of Haemophilia and the husband has not family history of Haemophilia. Hence, further testing and diagnosis of a husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia and if he is found negative for Haemophilia then the chances of inheritance will be: # 25% (1/4) chances of having a carrier daughter. # 25% (1/4) chances of having a son with Haemophilia. # 50% chance of having a normal child.
19	No	Not Sure	Not Sure	Yes	—	Wife did not have Haemophilia. But, the husband has a family history of the disorder. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia and if he is found positive for Haemophilia then the chances of inheritance will be: # All sons of the couple will be normal # All daughters will carry Haemophilia.

20	No	Not Sure	No	—	—	The wife did not have a medical history of Haemophilia and the husband is not sure about the disorder. Besides this, he did not have a family history of disorder. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia and if he is found negative for Haemophilia then the chances of inheritance will be: # None of the children (Son/Daughter) will have Haemophilia or carry the gene.
21	No	Not Sure	Not Sure	No	Yes	The wife did not have Haemophilia and the husband did not sure about the disorder. However, the obstetric history of his wife found positive. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia, and if he is found positive for Haemophilia then the chances of inheritance will be: # All sons of the couple will be normal # All daughters will carry Haemophilia.
22	No	Not Sure	Not Sure	Not Sure	Yes	The wife did not have Haemophilia and the husband did not sure about the disorder. However, the obstetric history of his wife found positive. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia, and if he is found positive for Haemophilia then the chances of inheritance will be: # All sons of the couple will be normal # All daughters will carry Haemophilia.

23	No	Not Sure	Not Sure	No	No	Chances of inheritance of Haemophilia are rare as the wife did not have Haemophilia and the husband did not sure about his medical status and family history of Haemophilia. However, obstetric history of his wife found negative. Hence, further testing and diagnosis of the husband are required for the confirmation of the disorder as he is not sure about the status of Haemophilia and if he is found negative for Haemophilia then the chances of inheritance will be: # None of the children (Son/Daughter) will have Haemophilia or carry the gene.

Similar to this, the inheritance pattern of Thalassemia acquired from literature, which is shown in table 3.7, restricts its prediction to a small number of instances (i.e., only four). Only two participants—father and mother or husband and wife—were used in this study and had a known history of the disorder, especially a medical history; which is further extended for multiple participants such as couple, their family member, and their kids and for the cases where any one of the couples does not sure about his or her medical history with the development of a rule set associate to Thalassemia which is depicted in Table 3.8 and Table 3.9. When determining the likelihood of producing a child with Thalassemia, these Thalassemia inheritance patterns are quite helpful, especially in situations where the couple has a known history of the disease. These patterns have therefore been used to develop intended predictions. However, these patterns are not applicable in situations where one partner is unsure of the other's medical background. Hence, a new set of rules has been developed for this, which is briefly shown in tables 3.8 and 3.9. As with Haemophilia, newly developed rules associated with Thalassemia have also been further separated into two sub categories. The first category of rules refers to the scenario where the wife is unsure of her medical history and the husband is aware of his. In this instance, the wife is asked about her family history; if she is aware of it, the result will be predicted; if not, the wife is

asked about her history of symptoms or physiological symptoms; if she is unsure, then obstetric history has been asked and prediction will be made based on inputs provided by her. The process of assessing the prediction regarding the inheritance of Thalassemia can be understood by observing the rules and prediction results depicted in Table 3.8, which lists all 20 rules for determining the likelihood of a child being born with Thalassemia in situations where the woman is unsure about her and her family's history of the condition. Moreover, unlike the previous rules, the recently formed set of rules makes use of more than two participants, such as spouses and family members. Thus, to achieve the intended prediction model and mobile-based e-Health system, a total of 73 rules for evaluating the inheritance of haemophilia and Thalassemia have been developed. These rules will predict the likelihood of having a child with Thalassemia and will be implemented in conjunction with existing rules obtained from the literature. Therefore, it can be concluded that the recently established set of rules will be very helpful when one partner is unsure about their medical history of haemophilia and thalassemia and wants to know their odds of having a child with one of the targeted disorders at a prenatal stage.

TABLE 3.7: Inheritance Pattern of Thalassemia

Case No	**Husband's Status of Thalassemia**	**Wife's Status of Thalassemia**	**Probability of Inheritance**	**Ref.**
1	Carrier	Normal	# The likelihood of having a disordered child is 0%. # 50% likelihood of having a carrier kid. # 50% likelihood of having a normal kid.	[99]
2	Normal	Carrier	# 0% probability of having a disordered child. # 50% likelihood of having a carrier kid. # 50% likelihood of having a normal kid.	[99]

3	Carrier	Carrier	# 25% chance of having a child with a disorder. # 50% chances that child is carrier. # 25% chance of having normal child.	[99]
4	Patient	Patient	100% chance of having a child with a disorder.	[99]
5	Patient	Carrier	# 50% chance of having a child with disorders. # 50% chance that the child is a carrier.	[99]
6	Carrier	Patient	# 50% chance of having a child with disorders. # 50% chance that the child is a carrier.	[99]
7	Normal	Normal	# None of the children will have Thalassemia or carry the gene.	[99]
8	Patient	Normal	# 100% chance of having a child with Thalassemia trait or child will carry the genes of Thalassemia.	[99]
9	Normal	Patient	# 100% chance of having child with Thalassemia trait or child will carry the genes of Thalassemia.	[99]

The considered parameters are Husband Status, Wife Status, Do you have a Family history of Thalassemia in case of Wife(Wife's FHT), Your / Family Member/ Relatives having the symptoms of Thalassemia in the case of wife (Wife's TSH), and Any Child with Thalassemia has Obstetric History(OBST).

TABLE 3.8: New Rules and Predictions Results of Thalassemia (when the wife is unsure of her medical history)

S. No	**Husband Status**	**Wife Status**	**Wife's FHT**	**Wife's SHT**	**OBST**	**Prediction Results**

1	Carrier	Not Sure	No	—	—	The husband is the carrier of Thalassemia and no family history and obstetric history of Thalassemia have been observed in the case of a wife. Hence, further testing and diagnosis of her are required as she is not sure about the disorder and if she is found negative for Thalassemia then the chances of inheritance will be: # 0% chance of having a child with a disorder # 50% chance is that the child carries the gene.
2	Carrier	Not Sure	Yes	—	—	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. But, she has a family history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier

3	Carrier	Not Sure	Not Sure	Yes	—	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. But, she has a family history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier
4	Carrier	Not Sure	Not Sure	No One	Yes	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. No Family history of Thalassemia has been found in the case of the wife. But, she has an obstetric history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier

5	Carrier	Not Sure	Not Sure	Not Sure	Yes	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. No Family history of Thalassemia has been found in the case of the wife. But, she has an obstetric history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier.
6	Carrier	Not Sure	Not Sure	Not Sure	No	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. She did not have a family or obstetric history of Thalassemia. Hence, it is assumed that she did not have Thalassemia. Despite this, further testing and diagnosis of her is required. So that, the status of disorder in her can be confirmed if she finds a negative disorder then the chances of inheritance will be: # 0% chance of having a child with the disorder # 50% chance is that the child carries the gene.

7	Carrier	Not Sure	Not Sure	No One	No	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. She did not have a family or obstetric history of Thalassemia. Hence, it is assumed that she did not have Thalassemia. Despite this, further testing and diagnosis of her is required. So that, the status of disorder in her can be confirmed if she found a negative disorder then the chances of inheritance will be: # 0% chance of having a child with a disorder # 50% chance is that the child carries the gene.
8	Patient of Thalassemia	Not Sure	No	—	—	The husband is a patient of Thalassemia and the wife is not sure about the disorder. However, she did not have a family history of the disorder. Hence, it can be assumed that she is normal. But, for the confirmation of the disorder further testing and diagnosis are required and if she is found normal then the chances of inheriting the genes of Thalassemia will be: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.
9	Patient of Thalassemia	Not Sure	Yes	—	—	The husband is a patient of Thalassemia and the wife is not sure about the status of the disorder. But, she has a family history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 50% chance of having a child with the disorder. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 100% chance of having a child with a disorder.

10	Patient of Thalassemia	Not Sure	Not Sure	Yes	—	Husband is the patient of Thalassemia and wife did not sure about the disorder. But, she has a family history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 50% chance of having a child with the disorder. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 100% chance of having a child with the disorder
11	Normal	Not Sure	Not Sure	Not Sure	No	The husband is normal and the wife is not sure about her medical history, family history, and physiological symptoms of Thalassemia. Her obstetric history is also negative and it can be assumed that she will not have Thalassemia. Hence, further testing and diagnosis of a wife are required for the confirmation of the disorder as he is not sure about the status of Thalassemia and if she is found negative for Thalassemia, then the chances of inheritance will be: # None of the children will have Thalassemia or carry the gene.

12	Carrier	Not Sure	Not Sure	Not Sure	Yes	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. No Family history of Thalassemia has been found in the case of the wife. But, she has an obstetric history of Thalassemia. Hence, further testing and diagnosis of her are required for the confirmation of the disease, and if she is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with a disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier But, if she found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier
13	Carrier	Not Sure	Not Sure	Not Sure	No	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. She did not have a family or obstetric history of Thalassemia. Hence, it is assumed that she did not have Thalassemia. Despite this, further testing and diagnosis of her is required. So that, the status of disorder in her can be confirmed if she finds a negative disorder then the chances of inheritance will be: # 0% chance of having a child with the disorder # 50% chance is that the child carries the gene.

14	Carrier	Not Sure	Not Sure	No One	No	The husband is the carrier of Thalassemia and the wife is not sure about the disorder. She did not have a family or obstetric history of Thalassemia. Hence, it is assumed that she did not have Thalassemia. Despite this, further testing and diagnosis of her is required. So that, the status of disorder in her can be confirmed if she found a negative disorder then the chances of inheritance will be: # 0% chance of having a child with a disorder # 50% chance is that the child carries the gene.
15	Patient of Thalassemia	Not Sure	Not Sure	No One	No	The husband is a patient of Thalassemia and the wife is not sure about the disorder. She did not have a family or obstetric history of Thalassemia, also. Hence, it is assumed that she did not have Thalassemia. Despite this, further testing and diagnosis of her is required. So that, the status of the disorder in her can be confirmed and if she finds a negative disorder then the chances of inheriting the genes of Thalassemia by offspring will be: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.

16	Normal	Not Sure	No	—	—	The husband is normal and the wife is not sure about Thalassemia. Her family history of disorder is also negative and it can be assumed that she will not have Thalassemia, also. Hence, further testing and diagnosis of a wife are required for the confirmation of the disorder as he is not sure about the status of Thalassemia and if she is found negative for Thalassemia, then the chances of inheritance will be: # None of the children will have Thalassemia or carry the gene. of thchildren will have Thalassemia or carry the gene.
17	Normal	Not Sure	Not Sure	Yes	—	The husband is normal and the wife is not sure about her medical and family history of Thalassemia. But, the physiological symptoms of the disorder have been observed in her family and it can be assumed that she could have Thalassemia as she has a family of disorders. Hence, further testing and diagnosis are required for the confirmation of the disorder and if she found a carrier of Thalassemia, then the chances of inheritance will be: # 0% chance of having a child with the disorder # 50% chance is that the child carries the gene. But, if she found a patient with Thalassemia then the chances of inheriting the Thalassemia by offspring would be: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.

18	Normal	Not Sure	Not Sure	Not Sure	Yes	The husband is normal and the wife is not sure about her medical history, family history, and physiological symptoms of the disorder. But, she has a positive obstetric history of the disorder and it can be assumed that she could have Thalassemia. Hence, further testing and diagnosis are required for the confirmation of the disorder and if she found a carrier of Thalassemia, then the chances of inheritance will be: # 0% chance of having a child with a disorder # 50% chance is that the child carries the gene. But, if she found a patient with Thalassemia then the chances of inheriting the Thalassemia by offspring would be: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.
19	Normal	Not Sure	Yes	—	—	The husband is normal and the wife is not sure about Thalassemia. But, her family history of the disorder is positive and it can be assumed that she may have Thalassemia as she has a family history of the disorder. Hence, further testing and diagnosis of her are required for the confirmation of the disorder and if she is found carrier of Thalassemia, then the chances of inheritance will be: # 0% chance of having a child with a disorder # 50% chance is that the child carries the gene. But, if she found a patient with Thalassemia then the chances of inheriting the Thalassemia by offspring would be: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.

20	Normal	Not Sure	Not Sure	No	—	The husband is normal and the wife is not sure about her medical and family history of Thalassemia. Thus, it can be assumed that she will not have Thalassemia. Hence, further testing and diagnosis of a wife are required for the confirmation of the disorder as he is not sure about the status of Thalassemia and if she is found negative for Thalassemia, then the chances of inheritance will be: # None of the children will have Thalassemia or carry the gene.

Similarly, the second group pertains to the rules that determine the likelihood of inheritance when the wife is aware of her medical history and the husband is unsure of his. In this instance, the husband's family history is enquired about; if he is aware of it, the result will be predicted; if not, the physiological symptoms have been asked; if he is not sure about it then obstetric history will be questioned, and ultimately the result about inheritance will be predicted. The same can be seen in Table 3.9. Hence, the study proved helpful in circumstances when the spouse is unsure of his medical history by providing 18 distinct rules associated with this. The considered parameters are "Do you have a family history of Thala- ssemia in the case of husband" (HFHT),"Your / Family Member/ Relatives having the symptoms of Thalassemia (In the case of husband)"(HSYMT), and "Obstetric History (Any Child with Thala- ssemia)"(OHT). Thus, to achieve the intended prediction model and mobile-based e-Health system, a total of 73 rules for evaluating the inheritance of haemophilia and Thalassemia have been developed. These rules will predict the likelihood of having a child with Thalassemia and will be implemented in conjunction with existing rules obtained from the literature. Therefore, it can be concluded that the recently established set of rules will be very helpful when one partner is unsure about their medical history of haemophilia and thalassemia and wants to know their odds of having a child with one of the targeted disorders

at a prenatal stage.

Table 3.9: New Rules and Predictions Results of Thalassemia (when the husband is unsure of her medical history)

S. No	**Wife**	**Husband**	**HFHT**	**HSYMT**	**OHT**	**Prediction Results**
1	Carrier	Not Sure	No	—	—	The wife is the carrier of Thalassemia and the Husband is not sure about the disorder. But, the family history of Thalassemia is negative in the case of the husband. Thus, it can be assumed that he is not positive for Thalassemia. Hence, further testing and diagnosis of him are required as he is not sure about the disorder and if he is found negative for Thalassemia then the chances of inheritance will be: # 0% chance of having a child with a disorder. # 50% chance is that the child carries the gene.
2	Carrier	Not Sure	Not Sure	No One	—	The wife is the carrier of Thalassemia and Husband is not sure about his medical history and family history of the disorder. But, the physiological symptoms of the disorder have not been observed in the family. Thus, it can be assumed that he is not positive for Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder and if he is found negative for Thalassemia then the chances of inheritance will be: # 0% chance of having a child with a disorder. # 50% chance is that the child carries the gene.

3	Carrier	Not Sure	Not Sure	Not Sure	No	The wife is the carrier of Thalassemia and the Husband is not sure about his medical history, family history, and physiological symptoms of the disorder. But, the obstetric history of his wife in the context of Thalassemia is negative. Thus, it can be assumed that he is not positive of Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder and if he is found negative for Thalassemia then the chances of inheritance will be: # 0% chance of having a child with a disorder. # 50% chance is that the child carries the gene.
4	Carrier	Not Sure	Yes	—	—	The wife is the carrier of Thalassemia and the husband is not sure about his medical history of the disorder. But, he has a family history of Thalassemia. Hence, further testing and diagnosis of him are required for the confirmation of the disease, and if he is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier. But, if he found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier

5	Patient of Thalassemia	Not Sure	No	—	—	The wife is a patient of Thalassemia and the Husband is not sure about the disorder. A family history of Thalassemia is also negative in the case of the husband. Thus, it can be assumed that he has not Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder, and if he is found negative for Thalassemia then: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.
6	Carrier	Not Sure	Not Sure	Yes	—	The wife is the carrier of Thalassemia and the husband is not sure about his medical history and family history of the disorder. But, the physiological symptoms of the disorder have been observed in his family members. Hence, further testing and diagnosis are required for the confirmation of the disease, and if he is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier. But, if he found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier

7	Carrier	Not Sure	Not Sure	Not Sure	Yes	The wife is the carrier of Thalassemia and the husband did not sure about his medical history, family history, and physiological symptoms of the disorder. But, his wife has an obstetric history of disorder. Hence, further testing and diagnosis are required for the confirmation of the disease, and if he is found carrier of Thalassemia then the chances of inheritance will be: # 25% (1/4) chance of having a child with a disorder. # 25% (1/4) chance of having a normal child. # 50% chance that the child is a carrier. But, if he found a patient with Thalassemia then the chances of inheritance would be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier
8	Normal	Not Sure	Not Sure	Not Sure	Yes	The wife is normal and the husband is not sure about his medical history, family history, and physiological symptoms of the disorder. But, the obstetric history (they have already a child with a disorder) of his wife has been found positive in the context of Thalassemia. Hence, further testing and diagnosis are required for the confirmation, and if he found a carrier of Thalassemia then the chances of inheritance will be: # 0% chance of having a child with a disorder. # 50% chance is that the child carries the gene. But, if he found a patient with Thalassemia then the chances are that: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.

9	Patient of Thalassemia	Not Sure	Not Sure	No One	—	The wife is a patient of Thalassemia and the Husband is not sure about his medical and family history of the disorder. But, the physiological symptoms of the disorder have not been observed in the family. Thus, it can be assumed that he has not Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder and if he is found negative for Thalassemia then: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.
10	Patient of Thalassemia	Not Sure	Not Sure	Not Sure	No	The wife is a patient of Thalassemia and the husband is not sure about his medical history, family history, and physiological symptoms of the disorder in the family. But, the obstetric history of his wife is negative. Thus, it can be assumed that he did not have Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder and if he is found negative for Thalassemia then: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.
11	Patient of Thalassemia	Not Sure	Yes	—	—	The wife is a patient of Thalassemia and the husband is not sure about the disorder. But, he has a Family history of Thalassemia. Thus, it can be assumed that he has Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder, and if he finds a carrier of Thalassemia then the Chances will be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier But, if he found a patient with Thalassemia then the chances would be: # 100% chance of having a child with Thalassemia disorder.

12	Normal	Not Sure	No	—	—	The wife is normal and the husband is not sure about the disorder. Moreover, the husband has not family history of Thalassemia. Thus, it can be assumed that he did not have Thalassemia. Hence, further testing and diagnosis are required for the confirmation and if he is found negative for Thalassemia then the chances are that: # None of the children will have Thalassemia or carry the gene.
13	Patient of Thalassemia	Not Sure	Not Sure	Yes	—	The wife is a patient of Thalassemia and the husband is not sure about his medical history and family history of the disorder. However, the physiological symptoms of the disorder have been observed in his family. Thus, it can be assumed that he has Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder, and if he finds a carrier of Thalassemia then the Chances will be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier But, if he found a patient with Thalassemia then the chances would be: # 100% chance of having a child with a disorder.
14	Normal	Not Sure	Not Sure	no one	—	The wife is normal and the husband is not sure about his medical and family history of the disorder. But, the physiological symptoms of the disorder have not been observed in the family. Thus, it can be assumed that he did not have Thalassemia. Hence, further testing and diagnosis is required for the confirmation and if he found negative Thalassemia then the Chances are that: # None of the children will have Thalassemia or carry the gene.

15	Patient of Thalassemia	Not Sure	Not Sure	Not Sure	Yes	The wife is a patient of Thalassemia and the husband is not sure about his medical history, family history, and physiological symptoms of the disorder. But, he confirmed the obstetric history of his wife in the context of Thalassemia. Thus, it can be assumed that he has Thalassemia. Hence, further testing and diagnosis are required as he is not sure about the disorder, and if he finds a carrier of Thalassemia then the chances will be: # 50% chance of having a child with the disorder. # 50% chance is that the child is a carrier But, if he found a patient with Thalassemia then the chances would be: # 100% chance of having a child with a disorder.
16	Normal	Not Sure	Not Sure	Not Sure	No	The wife is normal and the husband is not sure about his medical history, family history, and physiological symptoms of the disorder. However, the obstetric history of his wife in the context of the disorder has been found negative. Thus, it can be assumed that he did not have Thalassemia. Hence, further testing and diagnosis are required for the confirmation and if he is found negative for Thalassemia then the chances are that: # None of the children will have Thalassemia or carry the gene.

17	Normal	Not Sure	Yes	—	—	The wife is normal and the husband is not sure about the disorder. However, he has a family history of Thalassemia. Thus, it can be assumed that he could have Thalassemia. Hence, further testing and diagnosis are required for the confirmation, and if he found a carrier of Thalassemia then the chances are that: # 0% chance of having a child with a disorder. # 50% chance is that the child carries the gene. But, if he found a patient with Thalassemia then the chances are that: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.
18	Normal	Not Sure	Not Sure	Yes	—	The wife is normal and the husband is not sure about his medical and Family history of disorder. But, the symptoms of Thalassemia have been observed in his family members Thus, it can be assumed that he could have Thalassemia. Hence, further testing and diagnosis are required for the confirmation, and if he found a carrier of Thalassemia then the chances are that: # 0% chance of having a child with a disorder. # 50% chance is that the child carries the gene. But, if he found a patient with Thalassemia then the chances are that: # 100% chance of having a child with Thalassemia trait or a child will carry the genes of Thalassemia.

3.2.5 Performance Evaluation and Testing

A model's performance can be measured via performance evaluation to see how well it operates given the inputs. It is used to assess if a model needs additional training or retraining, is prepared to proceed to the next testing phase, or can be more widely deployed. The effectiveness of machine learning models can be

evaluated using a variety of techniques, including confusion matrix, hypothesis testing, ROC (Receiver Operating Curve), accuracy, precision, recall, specificity, and F1 score. Confusion matrix, recall, precision, F1 score, hypothesis testing, and Receiver Operating Curve (ROC) are scheduled to be used in this part of the study to statistically analyse the intended model.

3.2.6 Development and Execution of Mobile-Based Application

This phase covers the development of a mobile application that will use a smartphone to provide electronic healthcare services to the underprivileged and remote public in areas where access to medical professionals (such as genetic counsellors and haematologists) who specialise in Haemophilia and Thalassemia is limited and awareness of these disorders is inadequate. For this, the Android Studio platform's Kotlin programming language and the rule-based prediction model created during the current study have been used. Every development process adhered to the design, which is represented in Figure 3.4 as a flow chart of the anticipated mobile application.

The intended system's operation is depicted in the flow chart shown in Figure 3.4. The program's login/signup screen, dashboard, and general information section on haematological or bleeding disorders (such thalassemia and haemophilia) are clearly visible in Figure 3.4. The dashboard of the application consists of connecting to the doctor, haemophilia, Thalassemia, report, and logout. All modules will perform various operations like providing general information about haemophilia and Thalassemia. In addition, the login and signup options will offer the ability to log in and register the application. To use this app, users must first register by providing personal and medical information, and then they must log in.In addition to this, the application also predicts the probability of inheriting bleeding disorders (such as haemophilia and Thalassemia) by offspring of users (expected couple planning for family) with the help of haemophilia and Thalassemia module.

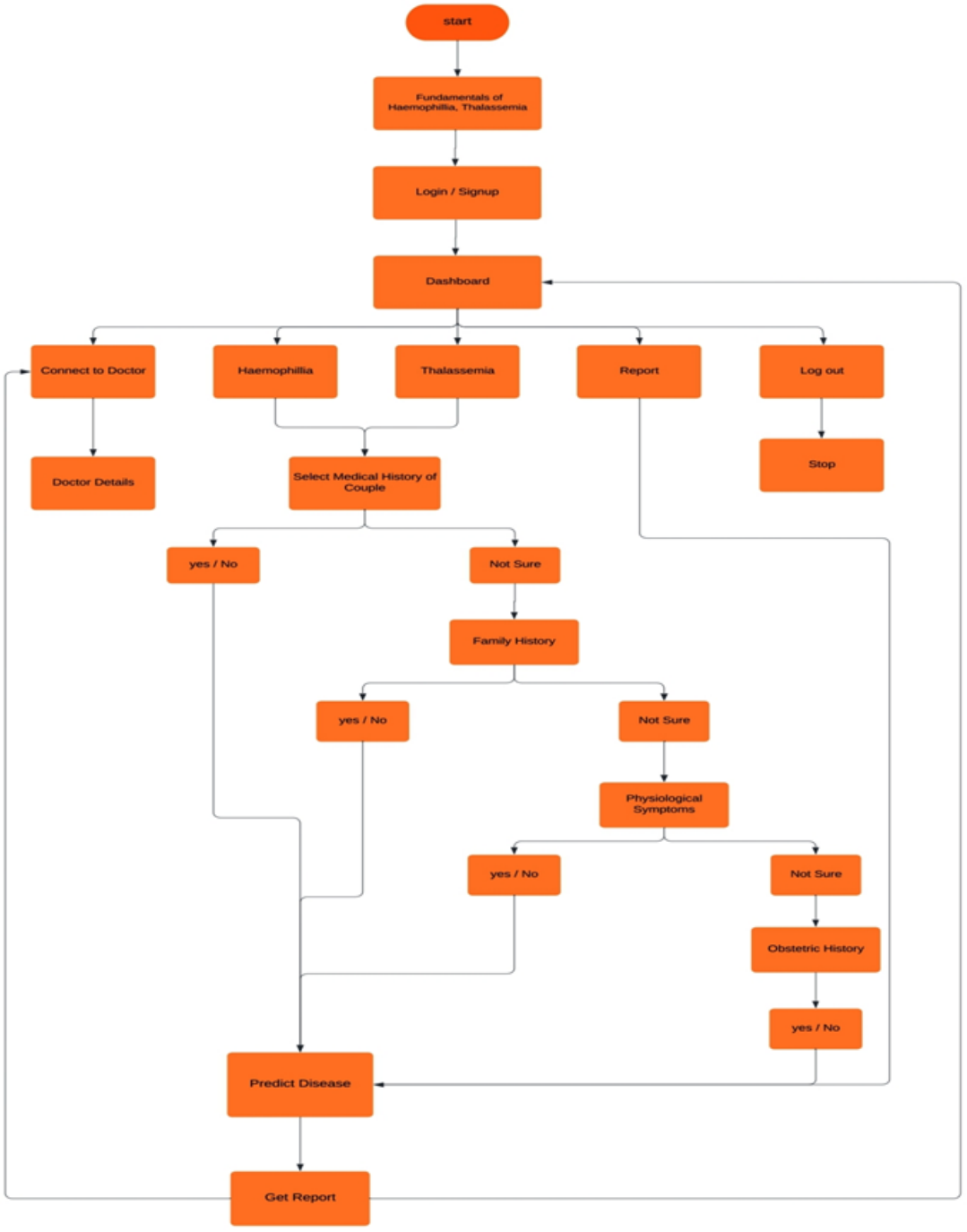


FIGURE 3.4: Flow Chart of Inherited Prenatal Disorder Predictor (IPDP)

Following creation, the programme will be tested by running it on an Android-based smartphone. This device will collect user input and produce results that will be compared to the rule set and prediction results, which were created and introduced in the chapter's prior section 3.2.4.

3.3 Conclusion

This chapter offers comprehensive details on the research methodology that was employed to carry out the intended study. This chapter also covers the significance of the research process, instruments, and methods for collecting data, time horizon, sample strategy, and research approach. Additionally, a methodical approach to study has been offered to achieve predetermined goals. In addition, several processes have been thoroughly covered with data collection, including data preprocessing, data splitting, model development and prediction, newly created rule sets, performance evaluation, model testing, and design and development of mobile applications. Therefore, the effort made in this chapter is meant to assist other researchers in gathering data and to offer a freshly formed set of guidelines for future research. The paper associated with this chapter and its publication information is shown below.

Sharma, P., & Sharma, A. (2019), "Collection and Analysis of Patients Dataset to Develop an E Health System for the Prediction of Inherited Prenatal Disorders," Journal of Computational and Theoretical Nanoscience, 16(12), 5118-5121.

Chapter 4

Symptoms based Classification Models

The structure of this chapter is as follows: The chapter is introduced in Section 4.1. The approach used to categorise hereditary bleeding diseases (such as Haemophilia and thalassemia), feature selection criteria and development of hybrid model is described in Section 4.2. Performance evaluation of models has been presented in section 4.3. Section 4.4 depicted the comparisons of accuracy of different machine learning algorithms utilized for the development of diagnosis models. Section 4.5 represents the statistical analysis of the diagnosis model. Section 4.6 presented the hypothesis testing of models developed during this current research work. Section 4.7 depicts the execution and outcome of the newly developed hybrid diagnosis model in this chapter. Section 4.8 concluded the findings.

4.1 Introduction

A condition that is passed down from parents to offspring due to changes in genes or chromosomes is referred to as an "inherited disorder," such as Haemophilia and Thalassemia [53] [59]. Haemophilia is an X-linked recessive genetic trait or disorder that results from a changed gene on the X chromosome and can be passed down to a child from parents [129] [62]. It is brought on by a lack or absence

of clotting foractors, which are substances in the blood that work with platelets to form clots. It is additionally broken down into three subtypes [130]. One of them is Haemophilia A (HA), Haemophilia B (HB), and Haemophilia C (HC), which is caused by a deficiency or absence of clotting factors (together with blood cells called platelets) VIII, IX, and XI, respectively, in the blood [63] [130] [64]. Haemophilia types are often further classified as mild, moderate, or severe according to the degree of clotting activity. Severe Haemophiliacs frequently experience spontaneous bleeding as well as aberrant bleeding (such as prolonged bleeding) in response to even minor wounds [65] [130]. Although internal bleeding in the kidneys, brain, and gastrointestinal tract is not unusual, spontaneous bleeding most often occurs in the joints. Recurrent joint bleeding may lead to cartilage loss, inflammation, and irreversible joint injury [63] [130] [64]. Minor injuries may cause internal bleeding or muscle hematomas, which may manifest days after the initial injury. There are several options (such as screening and clotting factor testing) available for the diagnosis of Haemophilia. The complete blood count (CBC), Activated Partial Thromboplastin Time (APTT) Test, Prothrombin Time (PT) Test, and Fibrinogen Test are among the screening tests on the list [68].

Thalassemia, on the other hand, is one type of genetic haemoglobin blood or bleeding condition [74]. It happens when red blood cells are unable to synthesise haemoglobin, which carries oxygen throughout the blood [74]. As a result, the body produces less or insufficient amounts of haemoglobin (a protein rich in iron that is present in red blood cells) compared to normal levels. Due to insufficient haemoglobin in the blood, oxygen cannot properly reach all of the body's parts and starves various organs, making them less able to function as they should [74]. Thalassemia frequently results in anaemia and manifests as fatigue or exhaustion, an enlarged spleen, sluggish growth in children, bone discomfort, a propensity for fractured bones, and other signs and symptoms [74]. Anaemia in people with Thalassemia can range from mild to severe [74]. The patient was diagnosed using a complete blood count (CBC) and haemoglobin electrophoresis testing. However,

occasionally, these screening tests show the overlap of mutations, resulting in an incorrect diagnosis or a false-negative result [76].

It is necessary to conduct a genetic analysis of the Thalassemia gene mutations in this condition. Testing of parents and siblings for this purpose [76] indicates the presence of the Thalassemia trait, which is caused by the autosomal dominant inheritance pattern, in which only one changed gene from either or just one parent is enough to pass on a genetic trait to the child. These screening tests are only recommended if patients display any type of Thalassemia-related signs and symptoms, such as anaemia, weariness, exhaustion, an enlarged spleen, or delayed growth in children. Even though all of these pathological or blood tests are frequently used to diagnose Haemophilia and thalassemia, they are complicated for the general public to use in aiding in the early diagnosis of an inherited bleeding problem (especially in the prenatal stage or even before that such age, premarital time). Numerous initiatives are put into place in various nations to prevent bleeding issues, such as pre-marital testing laws, genetic counselling, and prenatal assessments with the option of pregnancy termination if necessary [74]. With the help of genetic counselling, pre-marital testing assesses a couple's health and family history before they get married and predicts whether or not they will have a kid with Thalassemia [74]. This strategy also helps the couple to understand their numerous reproductive possibilities and make informed decisions regarding their marriage [74].

Similarly, Haemophilia also employed genetic counselling to provide advice and assistance to people and families who had been diagnosed with or were at risk of developing an inherited condition. Consequently, it can be said that genetic counselling is a technique that can be used at any point throughout pregnancy or even earlier, like before conception or marriage. A certified genetic counsellor is necessary for all of these. But sadly, there is a paucity of genetic counselling resources and counsellors, which raises the prevalence of bleeding disorders including Haemophilia and Thalassemia [81] [77]. Additionally, genetic counselling is

effective when the client already knows their health and their family's history of diseases (particularly those spanning three generations), which is not always the case [98]. Lastly, sometimes either of the couples doesn't want to disclose his or her medical history (such as previous miscarriages, abortions, adoption, or history of the disorder) in front of the other which leads to wrong predictions. Thus, these reasons drove the development of a focused e-Health system that will close the aforementioned gap by delivering necessary services electronically, including the diagnosis of the disorder at or even before that (such as at the preconception or premarital stage), as well as the ability to foretell the likelihood that the disorder will be inherited by offspring from them. Additionally, it will get around the lack of genetic counsellors [98] and the accessibility of online services like genetic counselling. Most importantly, because it is an electronic system, it will also function well in situations when patients or a couple are unsure of the history of a disorder in advance and will also offer a private setting for those who do not want to reveal their personal information in front of others.

Moreover, this kind of system saves medical professionals time by providing an automated method for diagnosis and prediction and reducing the likelihood of error due to the extensive testing done on the patient dataset using machine learning algorithms, which can be quickly retrieved at any time with the help of computer processing systems. As a result, a hybrid model that combines Decision Tree and Gaussian Naive Bayes has been developed to classify :haematological disorders into five categories Normal, Haemophilia Carrier, Haemophilia Diseases or Patient, Thalassaemia Major, and Thalassaemia Minor class. Before that, individual machine learning algorithms like Bayesian Network and KNN were used to construct individual classification models for the diagnosis of Haemophilia and thalassemia. However, these models were eventually abandoned in favour of the hybrid model due to their poor accuracy rates.

The experiment's findings demonstrated that the hybrid model had more successful prediction outcomes than individual models. Additionally, the model not only

produced exceptional precision and true positive rates but also boosted accuracy rates. This method employs a classification system to identify the illness using the symptoms as input; however, it is unable to predict the likelihood of inheriting a bleeding issue during the prenatal stage. To do this, a Bayesian network, a rule set, and an existing inheritance pattern have been combined to build a new rule-based prediction model called IPDP (Inherited Prenatal Disorder Predictor), which addresses the issue of people's lack of knowledge of their families' medical histories, which were not taken into account when determining their risk for certain diseases [67] [68] [73]. Then, a mobile-based application has been developed for the intended e-Health system that will estimate the possibility that bleeding disorders like Haemophilia and Thalassemia will be inherited from parents to their offspring at the prenatal stage. The details of rule based prediction model and a mobile based application developed during the current research work for the intended e-Health system has been presented in chapter 5 and 6 of this thesis. The remaining sections of this chapter provide a detailed explanation of dataset, feature selection criteria, different models developed during current research work (including individual and integrated models) with the utilization of Bayesian Network, K-Nearest Neighbours, Decision Trees and Gaussian Naive Bayes approach.

4.2 Development and Evaluation of Models

This section looked at the clinical and genetic characteristics of the two hereditary haematological disorders, Haemophilia and Thalassaemia, which affect millions of individuals worldwide. This section includes details about the dataset, the development of various models (individual and integrated models) throughout the current research project using the Gaussian Naive Bayes approach, K-Nearest Neighbours, Decision Trees, and Bayesian Network techniques and an analysis of the model's output. It also discusses the investigation's findings and their implications for diagnosis. Each of these is described in following sections.

4.2.1 Description of Dataset

The current study collected the required data set of Haemophilia and Thalassemia from DMC(Dayanand Medical College) and CMC (Christian Medical College & Hospital), Ludhiana, Punjab, India, and Haemophilia Society Dehradun Chapter respectively to perform further experimentation. This data set consists of 501 patient records with 34 columns that are likely to include details about proposed disorders (Haemophilia and Thalassemia) including patient characteristics, medical history, family history, obstetric history, symptoms history, and 5 target classes including normal, major Thalassemia, minor Thalassemia, Haemophilia carrier and Haemophilia disease or Haemophilia.

Each patient record in the data collection contains details about a person's medical history, including any family histories of disorders. This dataset's analysis can provide important new information about the incidence, severity, and outcomes of Haemophilia and Thalassemia in the patient community. Additionally, the construction of a hybrid diagnostic model that uses classification to identify Haemophilia and Thalassemia is greatly aided by this data set. For this, 33 different symptoms (such as spontaneous nose bleeding, joint bleeds/pain/ muscle bleeding, etc) have been utilized as a column head.

Figure 4.2 depicts the list of a total of 33 different symptoms associated with Haemophilia and Thalassemia which include spontaneous nose bleeding, joint bleeds/pain/swelling, muscle bleeding, bleeding after tooth removal, gastrointestinal bleeding, bleeding from knee and ankle, knee deformations, eye bleeding, swelling any one of the knees, bleeding gums, blood does not clot, tongue bite, bruises, slow growth in children, wide brittle bones, enlarged spleen, fatigue, weakness, yellow skin, heart problem, jaundice, frequent infection, low haemoglobin (HB), increased heart beat, shortness of breath, poor appetite, back/legs/joint pains, irritability, vomit after a feed or eating, bloated stomach or swelling in the stomach, lazy /lethargic, spleen extraction or removal of spleen and bone deformities. Additionally, the family history and obstetric history of patients have been

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 501 entries, 0 to 500
Data columns (total 34 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Spontaneous Nose Bleeding             501 non-null    int64
 1   Joint Bleeds/Pain/Swelling            501 non-null    int64
 2   Muscle Bleeding                       501 non-null    int64
 3   Bleeding After Tooth Removal          501 non-null    int64
 4   Gestrointestinal Bleeding             501 non-null    int64
 5   Bleeding from Knee and Ankle          501 non-null    int64
 6   Knee Deformation                      501 non-null    int64
 7   Eye Bleeding                          501 non-null    int64
 8   Swelling in any one of knee           501 non-null    int64
 9   bleeding Gums                         501 non-null    int64
 10  Blood Does not clot                   501 non-null    int64
 11  Toungebite                            501 non-null    int64
 12  Bruises                               501 non-null    int64
 13  Slow Growth in Children               501 non-null    int64
 14  Wide_Brittle Bones                    501 non-null    int64
 15  Enlarged Spleen                       501 non-null    int64
 16  Fatique                               501 non-null    int64
 17  Weakness                              501 non-null    int64
 18  Yellow Skin_E1                        501 non-null    int64
 19  Heart Problem                         501 non-null    int64
 20  Jaundice                              501 non-null    int64
 21  Frequent Infection                    501 non-null    int64
 22  Low HB                                501 non-null    int64
 23  Increased Heart Beat                  501 non-null    int64
 24  Shortness of Breath                   501 non-null    int64
 25  Poor Appetite                         501 non-null    int64
 26  Back LegsJoint Pain                   501 non-null    int64
 27  Irritability                          501 non-null    int64
 28  Vomit After Feed or Eating            501 non-null    int64
 29  Bloated Stomach or Swelling in stomach 501 non-null   int64
 30  Lazzi_ Lethargic                      501 non-null    int64
 31  Spleen Extraction or Removal 0f Spleen 501 non-null   int64
 32  Bone Deformities                      501 non-null    int64
 33  Target                                501 non-null    int64
dtypes: int64(34)
memory usage: 133.2 KB
```

FIGURE 4.1: Description of Dataset

compiled to construct a prediction model for determining the likelihood of having children with Haemophilia and thalassemia. Initial tabular records of all the information about the intended work have been made.

Pre-processing is followed by a random division of the dataset into training and testing sets, each of which contains 80% and 20% of the data, respectively, for further analysis and development. This division validates the proposed diagnostic model and reduces the biases associated with records. In this way, it can be said that the dataset has the potential to shed light on the epidemiology and path physiology of Haemophilia and Thalassemia, as well as to guide clinical judgment and public health initiatives about these disorders.

4.2.2 Feature Selection Criteria

The next stage in realising the suggested system is feature selection, which refers to selecting a small subset of the most essential characteristics by eliminating unnecessary, redundant, or noisy features from the initial feature set [153]. Thus, feature selection can result in improved learning outcomes, such as increased learning accuracy, reduced processing expenses, and improved model interpretability [153]. There are several different methods available in machine learning for feature selection such as filter methods, information gain, chi-square test, fisher's scores, correlation coefficient, variance threshold, mean absolute difference, and wrapper methods. The correlation function was used to analyse a dataset of 34 columns and 501 patient records, identifying 13 features associated with Haemophilia (such as joint bleeds/pain/swelling, bleeding after tooth removal, swelling in any one of the knees, blood clotting, tongue bite, bruises) and Thalassemia (such as fatigue, weakness, yellow skin_E1, low_HB, increased heart beat, shortness of breath, and back/leg pain). The factors that have the best chance of accurately predicting the course of the disorder may be identified using the correlation function, which evaluates the direction and strength of a linear connection between two variables. Thus, using the correlation approach, a total of 13 features were chosen from the gathered dataset for Haemophilia Thalassemia, as shown in Figure 4.2.

Although, age, gender, marital status, medical history, family history, obstetric history, symptoms details associated with the proposed disorder, and partner's details (including age, gender, medical, family, obstetric, and symptom history) have been recorded as overall feature set which is acquired from the dataset of Haemophilia and Thalassemia. Later on, only the symptoms set of disorders has been decided to be utilised for the development of a diagnosis model for the identification of the status of an individual for Haemophilia or Thalassemia through classification. For this, the required 13 sub-sets of features have been selected from the symptom set of the proposed study which can also be seen in Figure

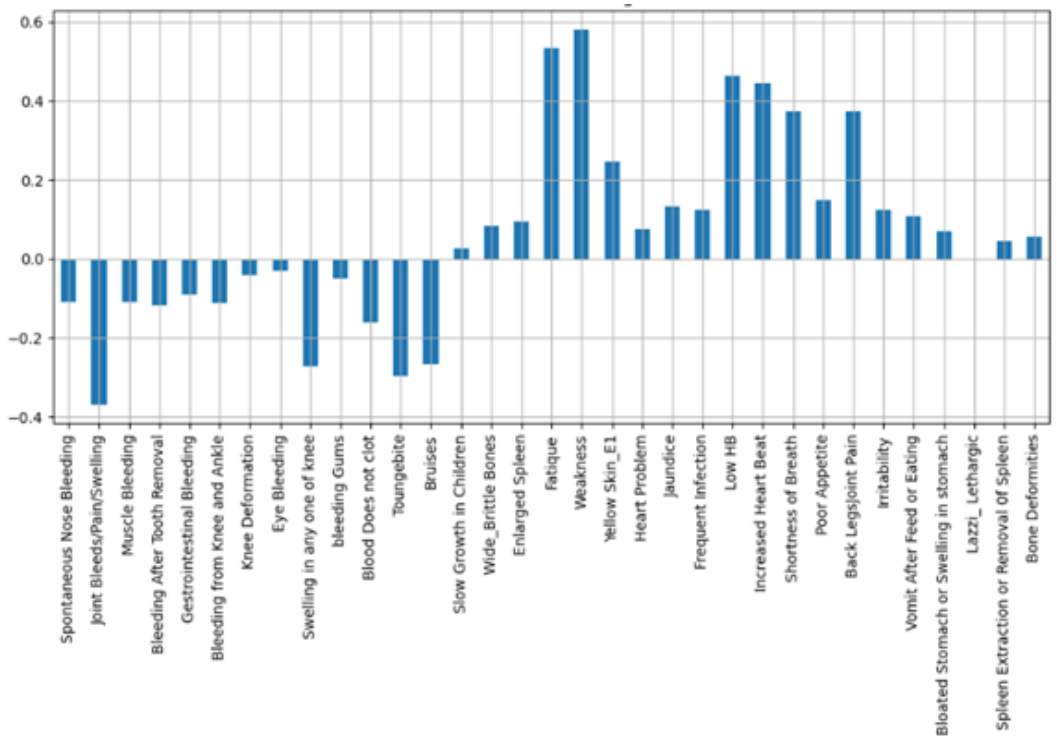
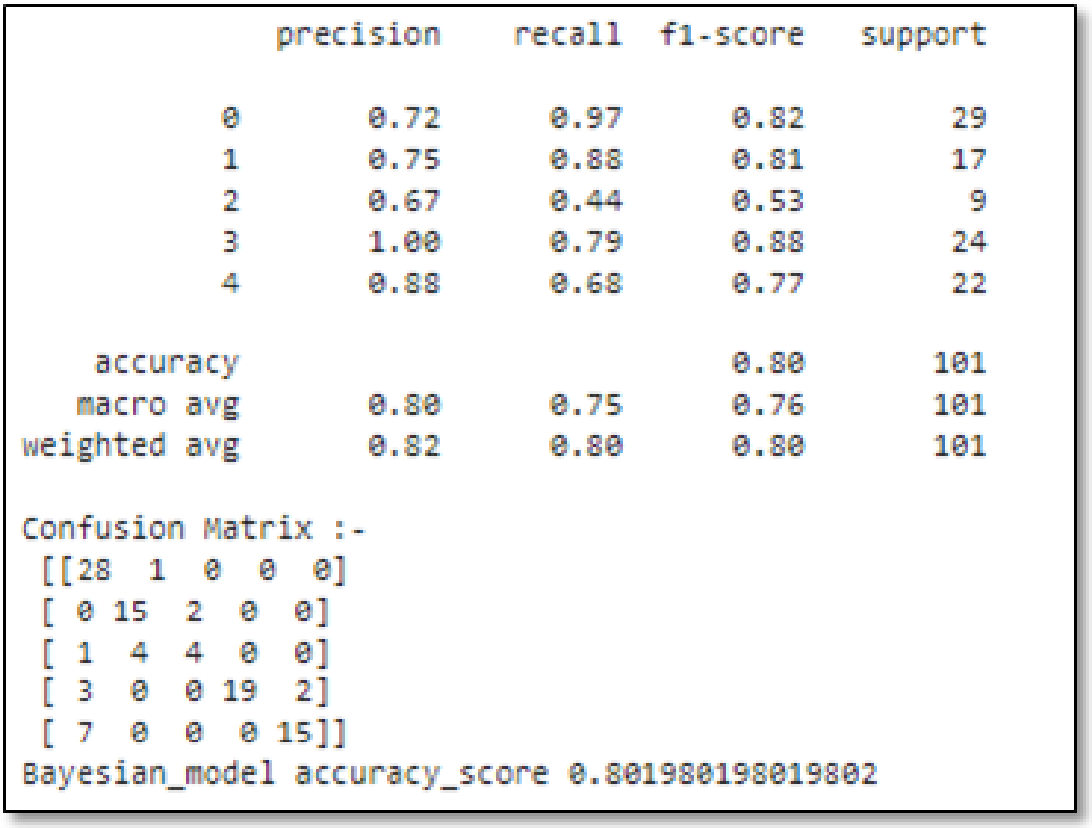


FIGURE 4.2: Feature Selection from Dataset

4.2. In this way, the current study can be streamlined and a better knowledge of how these variables relate to disorder and outcomes can be obtained by concentrating on these 13 aspects. Additionally, it might be possible to increase the precision and readability of results by reducing the number of features in the analysis. This is so that the model can be generalised to new data without becoming too complex due to over fitting, which can result from having too many variables. Overall, the study on the factors that is most likely to be predictive of illness outcomes improves our understanding of the underlying parameters of Haemophilia and Thalassemia by utilising the correlation function to choose the most pertinent characteristics or features from the dataset. The efficiency and accuracy of analysis can be enhanced by using the correlation function to pinpoint the dataset's most important features. This improved understanding will aid in the development of new treatment approaches for these inherited haematological as well as clinical decision-making.

4.2.3 Bayesian Network Based Classification Model

After feature selection, the development of a diagnosis or classification model is the next step of the current study. Hence, the Bayesian Network was initially chosen for the construction of the diagnosis model with the utilization of python; because it is also concerned with the probabilistic idea and seems to be relevant to the proposed study which aims to forecast the chance of inheritance of inherited prenatal disorders. Bayesian Network is a probabilistic graphical model used for probabilistic decision-making. It is a directed acyclic graph in which the variables are represented by the nodes, and the probabilistic relationships between the variables are represented by the edges. The characteristics or features (such as age, gender, marital status, medical history, family history, obstetric history, symptoms details associated with the proposed disorder, and partner's details (including age, gender, medical, family, obstetric, and symptom history)) that are pertinent for the diagnosis of these disorders must first be defined before using a bayesian network to identify patients with Haemophilia and Thalassemia from a dataset. Once relevant variables or characteristics have been defined, need to construct a bayesian network that represents the probabilistic relationships between these variables. The dataset is used to estimate the conditional probabilities that are required to specify the network. This has been done using techniques such as maximum likelihood estimation or bayesian inference. After constructing the network, it has been performed probabilistic inference on new patient records to determine the probability that each patient has Haemophilia or Thalassemia, based on their given symptoms. This would involve computing the posterior distribution over the relevant variables given the evidence provided by the patient record. A Bayesian network can be used to model the probabilistic relationships between variables relevant to the diagnosis of Haemophilia and Thalassemia, and to perform probabilistic inference on new patient records to identify patients with these conditions. The specific details of how to construct and use the network would depend on the particular dataset.



	precision	recall	f1-score	support
0	0.72	0.97	0.82	29
1	0.75	0.88	0.81	17
2	0.67	0.44	0.53	9
3	1.00	0.79	0.88	24
4	0.88	0.68	0.77	22
accuracy			0.80	101
macro avg	0.80	0.75	0.76	101
weighted avg	0.82	0.80	0.80	101

```
Confusion Matrix :-
 [[28  1  0  0  0]
 [ 0 15  2  0  0]
 [ 1  4  4  0  0]
 [ 3  0  0 19  2]
 [ 7  0  0  0 15]]
Bayesian_model accuracy_score 0.801980198019802
```

FIGURE 4.3: Accuracy of Bayesian Network Algorithm-Based Classification Model

A Bayesian network was applied to a dataset consisting of 501 records and 13 columns extracted as the most prominent features during the process of feature selection and achieved an accuracy of 80% as shown in Figure 4.3. The accuracy of a model is a measure of how well it predicts outcomes compared to the true values in the dataset. This is a relatively good accuracy, but it is important to consider the specific context and goals of the analysis. Overall, achieving an accuracy of 80% with a bayesian network on a dataset of 501 records and 13 columns or features suggests that the model has some predictive power and may be useful for certain applications.

```
Model training start........
Model training completed
knn Accuracy :- 0.8118811881188119
Confusion Matrix :-
 [[29  0  0  0  0]
 [ 0 15  2  0  0]
 [ 0  6  3  0  0]
 [ 1  0  0 21  2]
 [ 3  0  0  5 14]]
Classification Report :-
              precision    recall  f1-score   support

           0       0.88      1.00      0.94        29
           1       0.71      0.88      0.79        17
           2       0.60      0.33      0.43         9
           3       0.81      0.88      0.84        24
           4       0.88      0.64      0.74        22

    accuracy                           0.81       101
   macro avg       0.78      0.75      0.75       101
weighted avg       0.81      0.81      0.80       101
```

FIGURE 4.4: Accuracy of KNN Algorithm-Based Classification Model

4.2.4 K-Nearest Neighbours Algorithm-Based Classification Model

The k-nearest neighbours' algorithm, often known as KNN or k-NN, is a classifier made up of supervised learning that makes predictions or classifications about the clustering of an individual point of data using closeness. It may be used to solve classification or regression issues, but because it is based on the notion that comparable points can be discovered adjacent to one another, it is most frequently used to solve classification problems. The KNN model will utilise the features to recursively split the data into smaller subsets depending on their values, with a maximum depth or a minimum number of samples in each leaf node. Several measures, including accuracy, precision, recall, and F1 score, have been used to assess the performance of the KNN model. Figure 4.4 shows that, compared to the Bayesian Network, the KNN Algorithm delivered an accuracy of 81%. As a result of using KNN for the suggested diagnostic model, an increase of 1% has

been noted; yet requires more improvements. To further enhance, it has been determined to construct an integrated or hybrid model, which is covered in depth in section 4.2.5.

4.2.5 Hybrid Model

The proposed study initially implemented bayesian network for the development of the diagnosis of Haemophilia and Thalassemia which diagnosed the disorders with an accuracy of 80%. Because of this, another individual model has been created using the KNN method (as shown in section 4.2.4) to enhance this one. This model correctly diagnosed the disorder with an accuracy rate of 81%, however, this was insufficient. Hence, the development of a hybrid model has been decided to attain more improved results. The diagnosis of the illnesses, specifically Haemophilia, and thalassemia, has been incorporated using Decision Trees with the Gaussian Naive Bayes approach. To maximise the benefits of both approaches while minimising their drawbacks, the hybrid strategy that combines decision trees with Gaussian naive Bayes can be a potent tool for illness identification. For classification problems, decision trees are a common technique that is particularly helpful when there are complicated or nonlinear interactions between the variables. They work by recursively splitting the data into smaller and smaller groups based on the most informative feature at each step, resulting in a tree-like structure where each leaf node corresponds to a classification decision. However, decision trees can suffer from over fitting, which occurs when the model is too complex and fits the noise in the data rather than the underlying patterns.

As opposed to this, the Gaussian Naive Bayes (GNB) technique assumes that the characteristics are independent and regularly distributed. It operates by figuring out the likelihood of each class given the feature values and selecting the class with the highest likelihood as the forecast. GNB is simple, fast, and robust to

```
Model training start........
Model training completed
Hybrid_model Accuracy :- 0.8514851485148515
Classification Report :-
              precision    recall  f1-score   support

           0       0.88      0.97      0.92        29
           1       0.67      0.94      0.78        17
           2       0.67      0.22      0.33         9
           3       1.00      0.88      0.93        24
           4       0.90      0.86      0.88        22

    accuracy                           0.85       101
   macro avg       0.82      0.77      0.77       101
weighted avg       0.86      0.85      0.84       101
```

FIGURE 4.5: Accuracy of Hybrid Model

irrelevant features, but it can struggle with correlated features and non-normal distributions. Combining these two approaches allows us to divide the data into smaller subgroups based on the most useful characteristics using the Decision Tree, and then use GNB for each subgroup to calculate the class probabilities. This approach can reduce the risk of over fitting and handle correlated features by splitting the data into smaller, more homogeneous subgroups. Moreover, since GNB is less affected by irrelevant features, the decision tree can include a wider range of features without negatively impacting the performance.

As a result, the hybrid model (developed with the integration of the Decision Tree and GNB) outperformed individual models created using the Bayesian network and KNN, respectively, by achieving a greater accuracy of 85% (as shown in Figure 4.5). The Bayesian Network can model complex dependencies between features but can suffer from computational complexity, while the decision tree can be simple and interpretable but may not capture complex dependencies. By combining the decision tree and GNB, we can leverage the strengths of both methods and achieve better performance than each one separately. The 85% accuracy is a good result, but it should be noted that the performance may vary depending on the specific

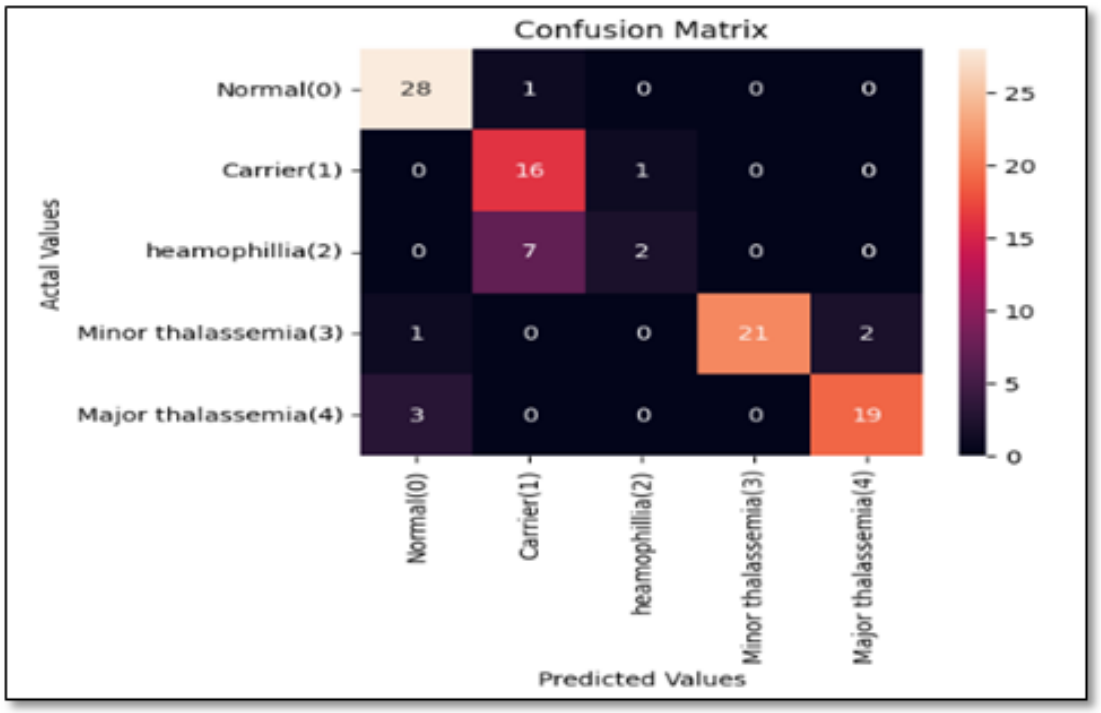


FIGURE 4.6: Confusion Matrix of All Classes

dataset and problem at hand.

4.3 Performance Evaluation of Classification Model

The current study utilized the confusion matrix for the evaluation of the newly developed classification model. A confusion matrix is a table that compares the predicted labels of a classification model against the actual labels of a set of data to assess the performance of the model. For each class, the matrix displays the number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Figure 4.6 shows that the confusion matrix contains five classes, which correspond to Normal, Carrier, Haemophilia, Minor Thalassemia, and Major Thalassemia on both the y-axis and the x-axis. The diagonal values of the confusion matrix represent the number of correct predictions made by the model for each class. In this case, the values on the diagonal are (3, 11, 5, 45, 18), which means that the model correctly predicted 3 instances of Normal, 11 instances of

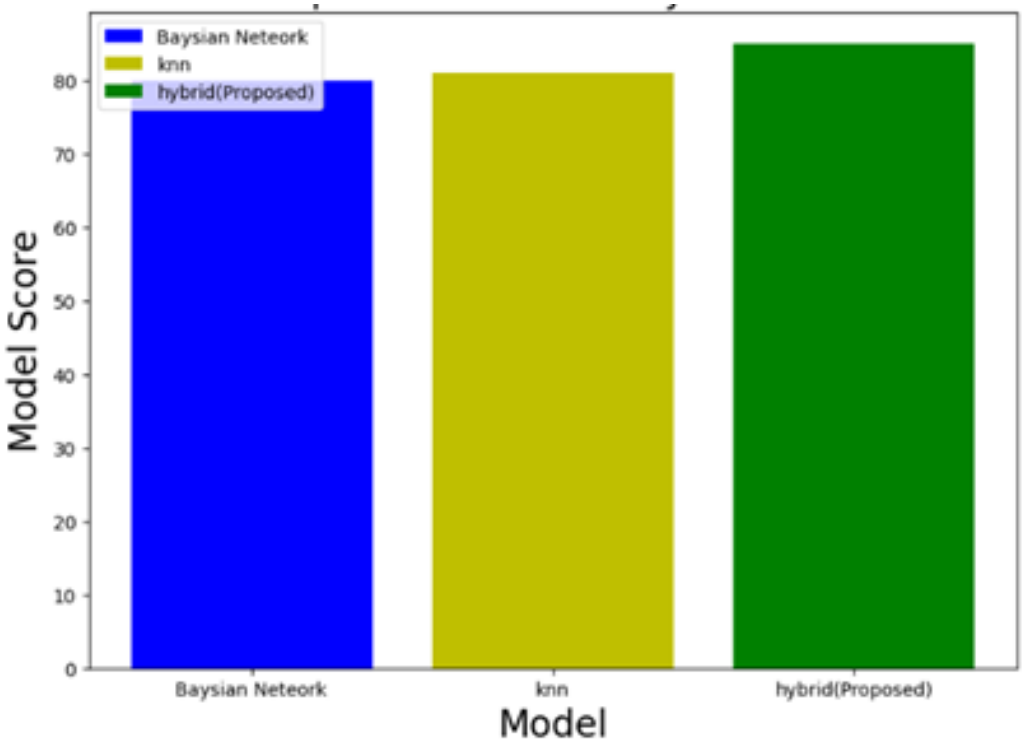


FIGURE 4.7: Confusion Matrix of All Classes

Carrier, 5 instances of Haemophilia, 45 instances of Minor Thalassemia, and 18 instances of Major Thalassemia. The off-diagonal values in the confusion matrix represent the number of incorrect predictions made by the model for each class. Hence, the total number of predictions made by the model is 101, and the number of correct predictions is 82. This means that the model achieved an overall accuracy of 82/101 = 81.2%.

4.4 Comparison of Accuracy

This section shows a comparison of the accuracy rates achieved by various individual models (such as a classification model created using a Bayesian Network and KNN, respectively), as well as a hybrid model created during the current study.

The analysis of Figure 4.7 depicted that the hybrid model has the highest accuracy among these three when comparing their respective levels of precision. In light of

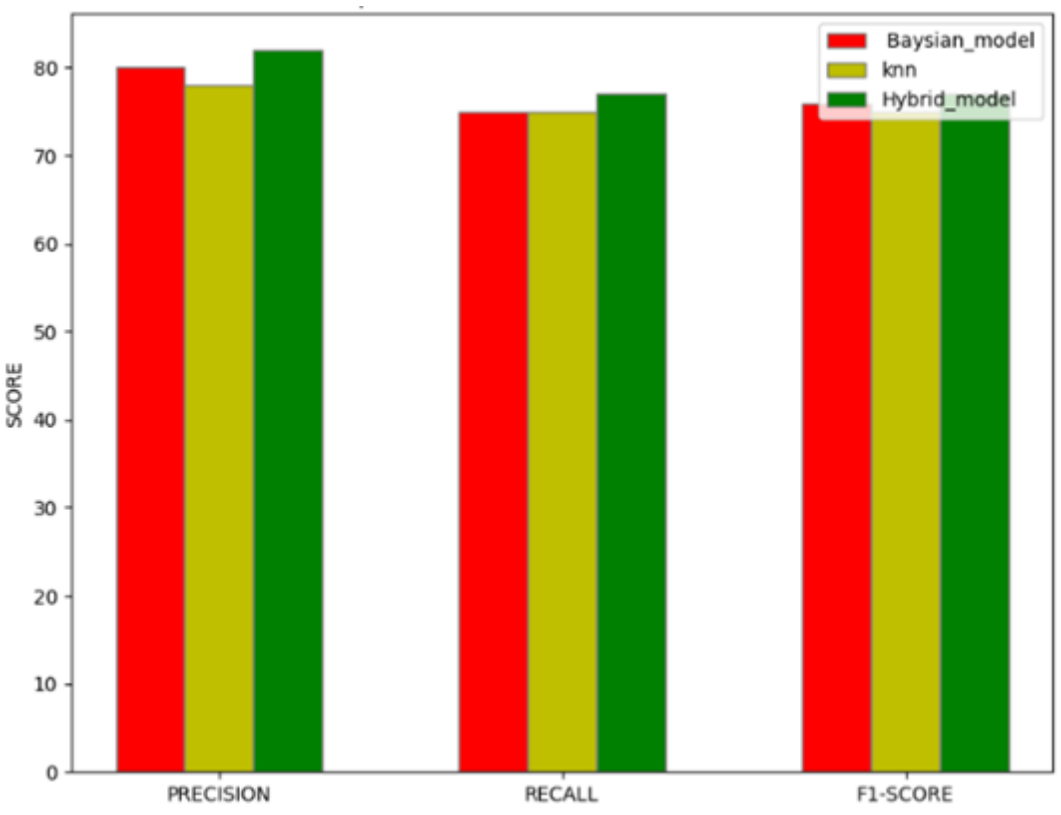


FIGURE 4.8: Comparison of Precision, Recall and F1-Score

this, it seems likely that the suggested hybrid model would be the most effective option for forecasting outcomes in the issue domain for which these models were trained. Although several measures may be utilised, accuracy is only one of them when evaluating a model's performance. Precision, recall, and F1 score are dependent on the particular problem domain and application, as shown in figure 4.8. The proposed hybrid model seems to beat both the bayesian network and KNN models based on the accuracy numbers you have supplied.

4.5 Statistical Analysis

The diagnostic model created during the current study effort has been statistically analysed using the Receiver Operating Characteristic (ROC) Curve; which is an analytical method used to assess the effectiveness of a diagnostic model in graphical

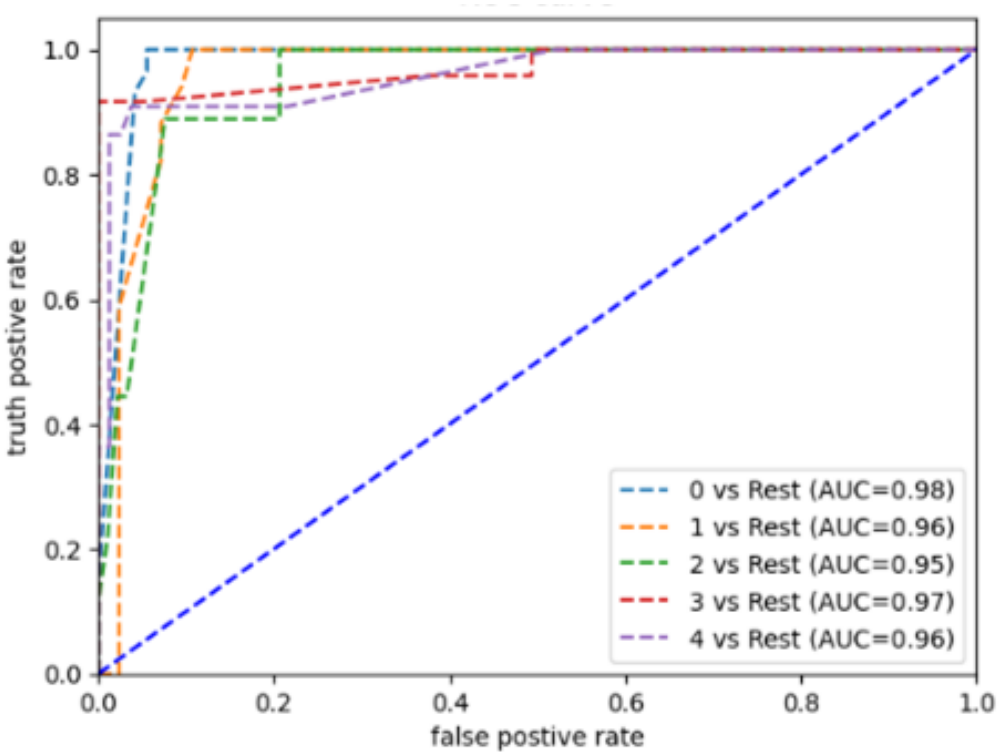


FIGURE 4.9: Comparison of Precision, Recall and F1-Score

form. Hence, the ROC (Receiver Operating Characteristic) curve visually depicts a binary classifier's performance at various discriminating levels. On the ROC curve, the True Positive Rate (TPR) and False Positive Rate (FPR) are shown for various threshold values. The area under the ROC curve (AUC) is typically used to describe the classifier's overall performance. The ROC curve and AUC may be used to evaluate the trade-off between TPR and FPR. While reducing false positives could be more important in some applications than it might be in others, the ROC curve and AUC offer a method to assess how well various classifiers perform in terms of these choices. Hence, understanding the trade-off between true positive rate and false positive rate at various discriminating thresholds may be done with the use of the ROC curve and AUC, which are useful tools for assessing the effectiveness of binary classifiers.

Figure 4.9 displays the ROC curves for five various binary classifiers. These classifiers each make a distinction between a specific class (0, 1, 2, 3, or 4) and the

other classes. The AUC (Area under the Curve) measures the classifier's overall efficacy. As the AUC goes up, the classifier's performance gets better.

- For class 0, the AUC is 0.98. This means that the classifier performs reasonably well in distinguishing class 0 from the rest, but there is room for improvement.
- For class 1, the AUC is 0.96. This indicates that the classifier performs very well in distinguishing class 1 from the rest.
- For class 2, the AUC is 0.95. This indicates that the classifier performs perfectly in distinguishing class 2 from the rest.
- For class 3, the AUC is 0.97. This indicates that the classifier performs well in distinguishing class 3 from the rest, but there is some room for improvement.
- For the class 4, the AUC is 0.96. This indicates that the classifier performs very well in distinguishing class 4 from the rest.

Hence, it can be seen that the classifiers perform differently in separating their respective classes, according to the ROC curves and AUC values. Some classifiers perform better than others, and some classifiers still have room for development.

4.6 Hypothesis Testing

To calculate the effectiveness of machine learning models, hypothesis testing offers statistical support for conclusions to be drawn from datasets that are executed with the use of cross-validation. Thus, the hybrid model and KNN diagnosis model have been evaluated using repeated stratified k-fold cross-validation, a popular technique for determining the effectiveness of classification models. The hybrid model is an integration of two or more techniques to improve the accuracy of the model. Hence, the hybrid model developed during the current study is being

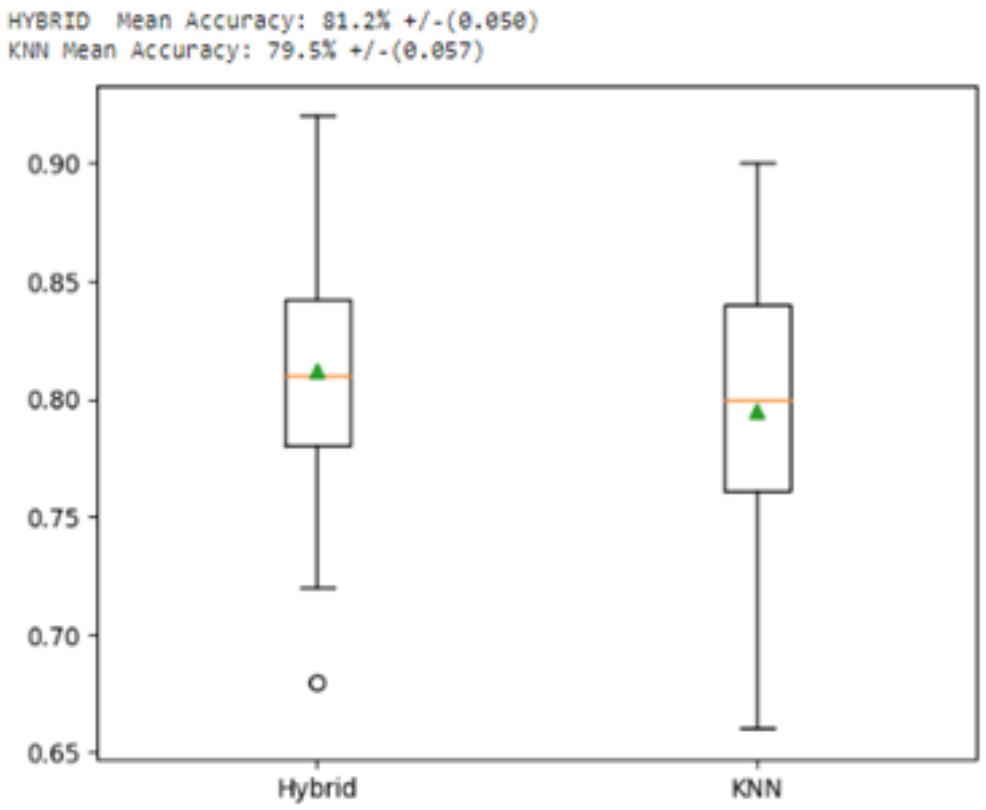


FIGURE 4.10: Mean Accuracy of Hybrid Approach vs KNN

compared to the KNN diagnosis model based on k-mean accuracy. A higher k-mean accuracy indicates better performance because it indicates how accurately the model identified the data.

According to the results of the repeated stratified k-fold cross-validation method, the k-mean accuracy of the hybrid approach was 81.2 %, which is higher than the k-mean accuracy of the KNN model, which was 79.1%, as shown in Figure 4.10. This shows that a better model for this dataset would be the hybrid technique, which has a higher degree of data classification accuracy.

4.7 Execution and Outcome of Newly Developed Hybrid Diagnosis Model

This section depicts the working of a newly developed hybrid diagnosis model which diagnoses the disorders (namely Haemophilia and Thalassemia) by taking the symptoms as input from an individual or patient. As we can see in Figure 4.11 the model asks for details of patients in the form of symptoms such as joint bleeds/pain/swelling, gastrointestinal bleeding, swelling in any one of the knees, blood clotting, tongue bite, bruises, fatigue, weakness, yellow skin˙E1, low HB, increased heartbeat, shortness of breath and back legs joint pain and then patient or user will enter the responses against these symptoms by entering numerical value such as 1 or 0; where 1 represents to the presence of symptoms and 0 represents to the absence of symptom. After entering the details of symptoms, the result will be generated by the model and presented as below in figure 4.11 and 4.12.

The same thing is depicted in Figure 4.11, where the patient has chosen joint bleeds/pain/swelling, gastrointestinal bleeding, swelling in any one of the knees, blood that does not clot, tounge bite, bruises, and shortness of breath as input options for a newly developed hybrid diagnosis model. This model will take the inputs and predict the outcome that will inform the patient that he or she has Haemophilia. Similarly, in Figure 4.12 newly developed hybrid diagnosis model predicts the result of having major Thalassemia in a patient based on the selection of symptoms. In this way, it can be said that the newly developed hybrid model takes the input (in the form of symptoms) as input from a new patient to forecast Haemophilia or Thalassemia which blends machine learning techniques and Haematology specialist expertise. Although, this model diagnoses the presence or absence of bleeding disorders such as Haemophilia and Thalassemia; but does not predict the probability of inheriting the genetic bleeding disorder by offspring at the prenatal stage. To accomplish the primary objective of the current study, it has been chosen to proceed with the construction of the prediction model (which

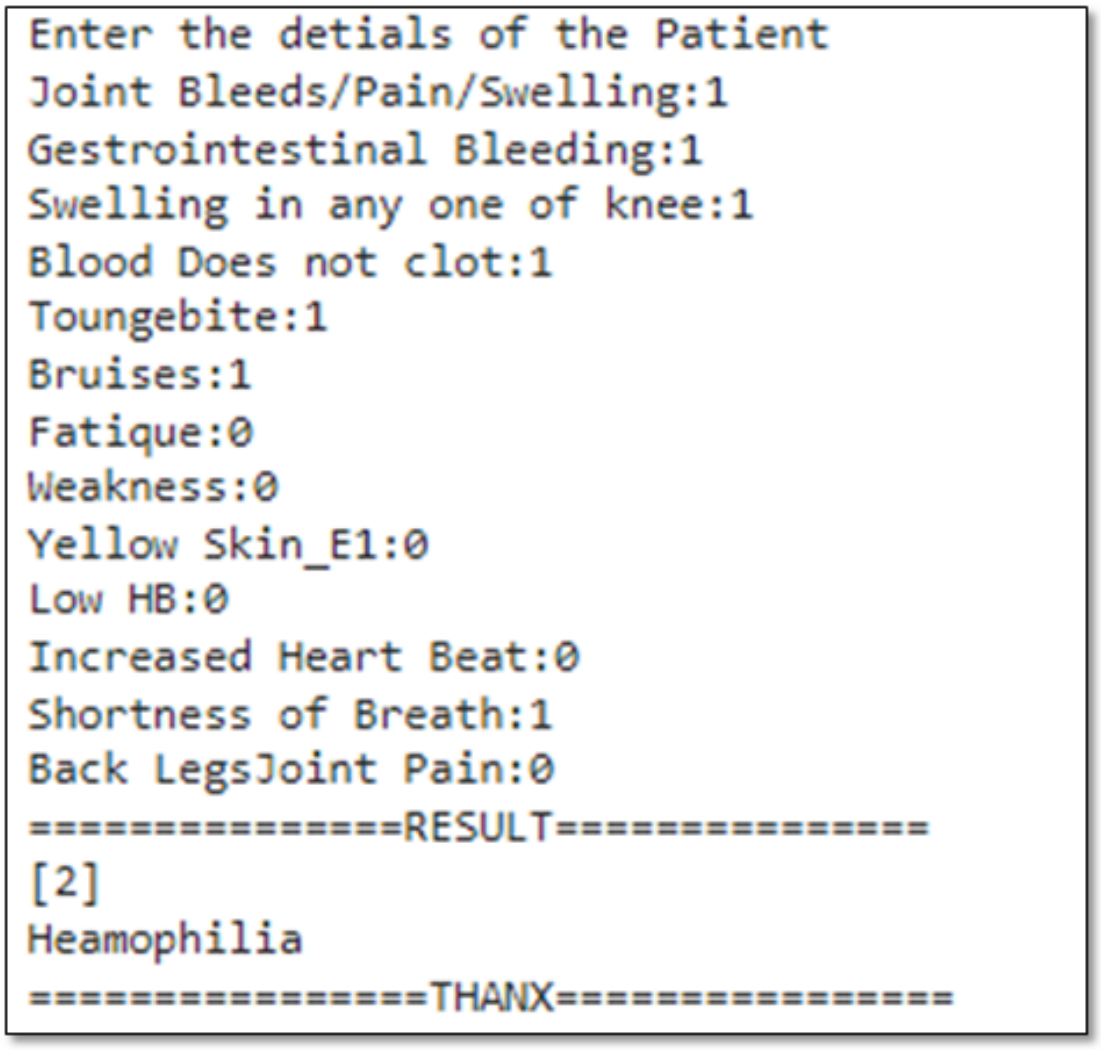


FIGURE 4.11: Prediction of Haemophilia based on Symptoms

will forecast the chance of having a kid with Haemophilia and Thalassemia at the prenatal stage).

For this, further analysis has been conducted over the dataset and as a result, a rule based prediction model has been developed. The model utilizes the symptom base of the above hybrid diagnosis model along with the medical, family, and obstetric history of individuals or patients to predict the possibility of inheriting Haemophilia and Thalassemia by offspring. Besides this, the newly developed rule set discussed in previous chapter 3 has been utilised for the development of

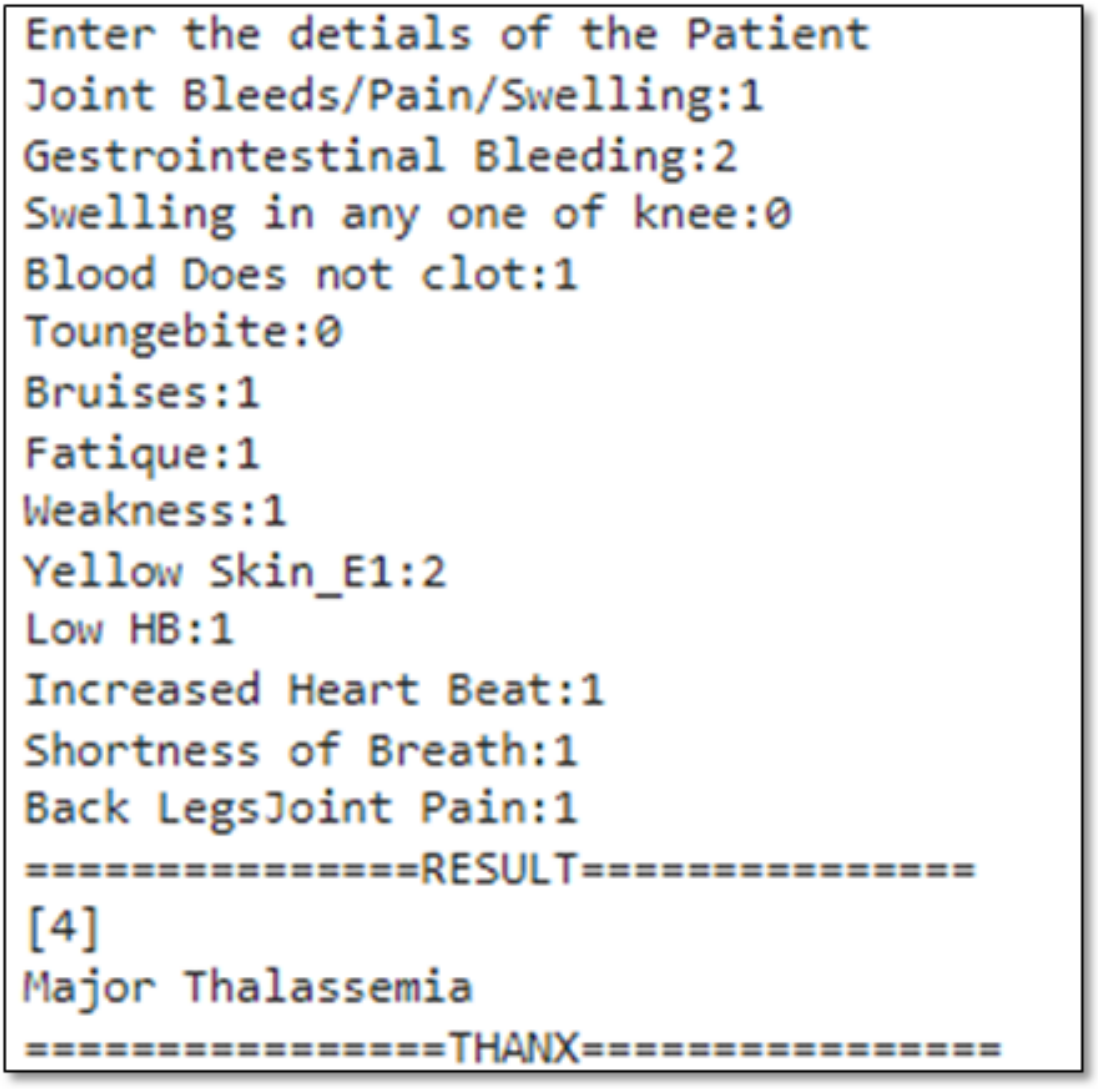


FIGURE 4.12: Prediction of Hybrid Model based on Symptoms

this model. The detail development process of this model has been described in Chapter 5.

4.8 Conclusion

One of the most challenging tasks a medical specialist has ever undertaken is illness diagnosis and prediction since even a minor error could compromise a patient's life. Utilisation of computer-aided or an electronic healthcare technique

has brought major transformation in the prediction of diseases or disorders. As a result, ideas like e-Health are becoming more and more popular in the medical field, which is a wonderful encouragement for the creation of more e-Health systems that will revolutionise the healthcare industry. Hence, the chapter presented the details about the dataset, the development of various models (individual and integrated/hybrid models) throughout the current research project using the Gaussian Naive Bayes approach, K-Nearest Neighbours, Decision Trees, and Bayesian Network techniques and an analysis of the model's output. It also discusses the investigation's findings and their implications for diagnosis. Implementation outcomes proved that the hybrid model developed for diagnosis was preferable to other individual diagnostic models implemented in the chapter. Hence, the hybrid model that combines Decision Tree and Gaussian Naive Bayes has been developed to classify bleeding disorders into five categories: Normal, Haemophilia Carrier, Haemophilia Diseases or Patient, Thalassemia Major, and Thalassemia Minor class. Before that, individual machine learning algorithms like Bayesian Network and KNN were used to construct individual classification models for the diagnosis of Haemophilia and Thalassemia. However, these models were eventually abandoned in favour of the hybrid model due to their poor accuracy rates. According to the experiment's results, the hybrid model outperformed individual models in terms of prediction accuracy. Not only did the model yield remarkable accuracy and true positive rates, but it also increased accuracy rates with a percentage of 85%. This method employs a classification system to identify the illness using the symptoms as input; however, it is unable to predict the likelihood of inheriting a bleeding issue during the prenatal stage. To do this, a Bayesian network, a rule set, and an existing inheritance pattern have been combined to build a new rule based prediction model, which addresses the issue of people's lack of knowledge of their families' medical histories, which were not taken into account when determining their risk for certain diseases [67] [68] [73]. Then, a mobile-based application with the name of IPDP (inherited prenatal disorder predictor) has been developed for

the intended e-Health system that will estimate the possibility that haematological disorders like Haemophilia and Thalassemia will be inherited from parents to their offspring at the prenatal stage. The detail presentation of these two models has given in chapter 5 and 6 of this thesis.

Chapter 5

A Rules-Based Prediction Model

The structure of this chapter is as follows: The chapter is introduced in Section 5.1. Section 5.2 describes the rules based prediction model, its workings, and its outcome in detail. Section 5.3 concluded the findings.

5.1 Introduction

To achieve the set goal of the proposed study, this section discusses in detail the rules based prediction model for inherited prenatal disorders that were created during the conduct of the proposed study. This rules based prediction model predicts the probability of having a child with Haemophilia and Thalassemia at a prenatal stage which helps individuals or couples in making informed decisions in the context of family planning. For this, a relational dataset that contains an individual's and his or her partner's medical history, family history, physiological history, and obstetric history has been used. This dataset is further divided into two categories, such as the Haemophilia relation and the Thalassemia relation as depicted in figure 3.2 and 3.3. In short, the development of the model utilizes the methodology depicted in Figures 3.2 and 3.3 of chapter 3. Following that, a newly developed rule set which was created based on an existing inheritance pattern that was acquired from the literature and discussed in chapter 3 has been utilized for

the development of this model. The implementation of this model has been done systematically with the utilisation of a relational dataset (which includes medical history, family history, physiological symptoms, and obstetric history of the couple), the newly developed rule set along with existing inheritance patterns, and a Bayesian Network. For then implementation python platform has been utilized. In this way, the current model covers not only known cases of disorder but also cases where couples are not sure about the status of disorders. Additionally, the model includes family members as participants in addition to potential partners, taking into account their medical and symptom histories for the prediction. The next part includes a visual representation of all the rule sets that were created after the implementation and put into use throughout the current research, together with the existing inheritance pattern. The graphical representation of inheritance rules is shown in the following sections and includes participants (such as husband or wife), medical status about Haemophilia and Thalassemia (such as normal, carrier, Haemophilia, not sure, Thalassemia major and Thalassemia minor), family history, obstetric history, physical examination results, and genetic testing results. Python is used to implement all the rules for producing outcomes, which are further illustrated in the form of cases for better understanding.

5.2 Working of a Rule Based Prediction Model

This part graphically represented all the rules, including existing patterns, with the prediction outcome expressed in Python as cases. In addition, the section explains the operation of the model, which uses an if-then or nested-if-then pattern for the execution of rules, outcome prediction, and graphical representation. The next sections discuss several predictions made using a rule-based model for different cases related to Haemophilia and Thalassemia. As a result, the rule set used by the current model is divided into two main categories, such as rule set for

Thalassaemia and rule set for Haemophilia, which are further subdivided into categories like rules for the cases where couple already knows their medical status, rules for the cases where anyone from husband and wife is not sure about their medical status.

This part graphically represented all the rules, including existing patterns, with the prediction outcome expressed in Python as cases. In addition, the section explains the operation of the model, which uses an if-then or nested-if-then pattern for the execution of rules, outcome prediction, and graphical representation. The next sections discuss several predictions made using a rule-based model for different cases related to Haemophilia and thalassemia. As a result, the rule set used by the current model is divided into two main categories, such as the rule set for Thalassemia and the rule set for Haemophilia, which are further subdivided into categories like rules for the cases where the couple already knows their medical status, rules for the cases where anyone from husband and wife is not sure about their medical status.

5.2.1 Prediction Results of Model in Context of Thalassemia

All the rules developed for the early diagnosis of inheritance of Thalassemia among offspring have been presented in this segment with a graphical demonstration of execution. All these rules cover not only known cases of existing inheritance patterns but also unknown cases.

Prediction results for the cases where the couple (Husband and wife or father and mother) is sure about the medical history of Thalassemia

This section depicts the prediction outcome of inheritance rules acquired after the implementation on python and covers the cases where participants already know their medical history.

Case 1: Husband is normal and wife is carrier.

This is the case where husband and wife utilises their medical history as an input

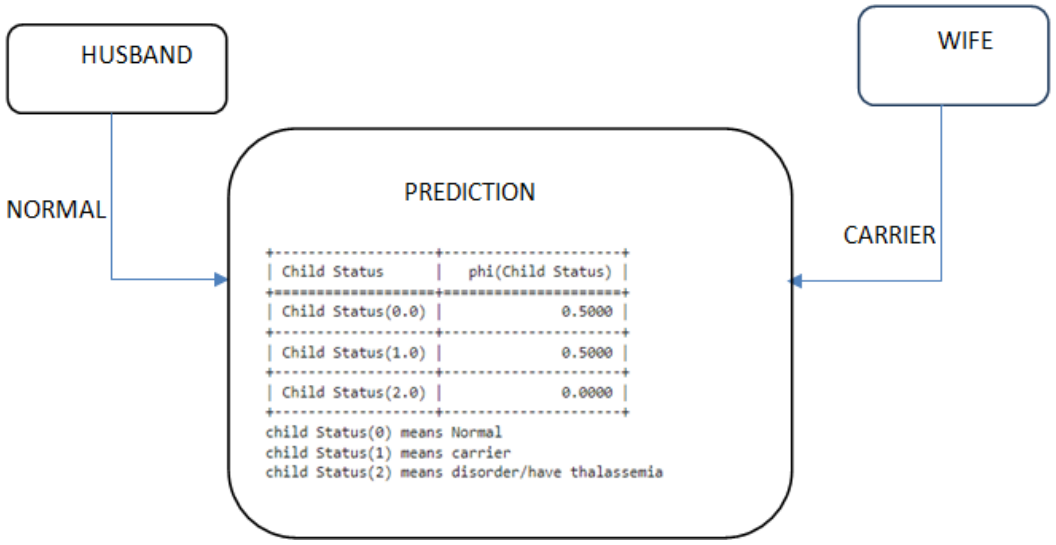


FIGURE 5.1: Probability of inheriting Thalassemia by offspring for Case 1

and feed it to the model; which will further predict the result. In Figure 5.1, husband and wife are participants; whereas normal and carriers are the medical status of disorders of the participant which is provided as input to the prediction model named Inherited Prenatal Disorders Predictor (IPDP). Under the heading of child status, which indicates the status or severity of a condition, a prediction result has been created based on the input. For example, child status (0.0) denotes normal, child status (1.0) denotes a carrier, and child status (2.0) denotes Thalassemia and phi (child status), which indicates the likelihood of inheriting the condition and how severe it will be.

As shown in figure 5.1, the parameter with the heading "child status" (0.0) refers to the status of offspring in the normal range, and the value of phi (child status) is 0.5000, or 50%, towards that parameter. This illustrates that there is a 50% chance of having a kid without thalassemia. Therefore, the probability of producing a normal kid is 50% if the husband is normal and the woman is a Thalassemia carrier, as shown in figure 5.1 Similar probabilities of having a kid with mild Thalassemia or a child who contains the Thalassemia gene are 50% and 0%, respectively.

Case 2: Husband and Wife both are carriers.

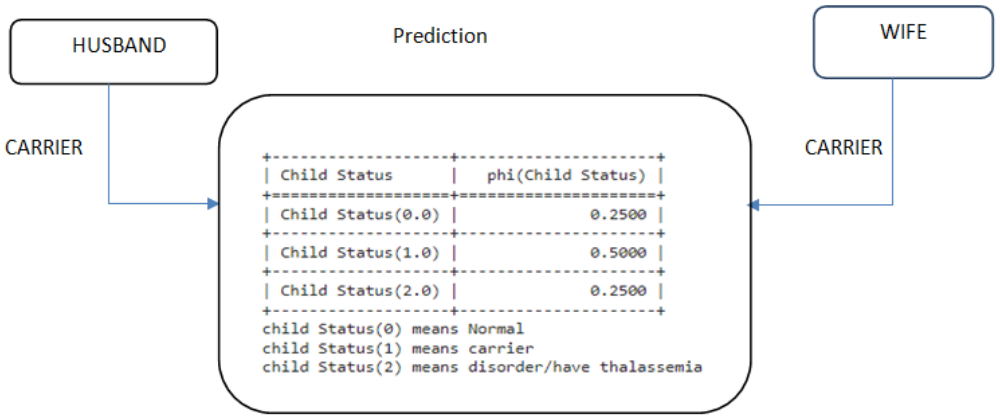


FIGURE 5.2: Probability of Inheriting Thalassemia by Offspring for Case 2

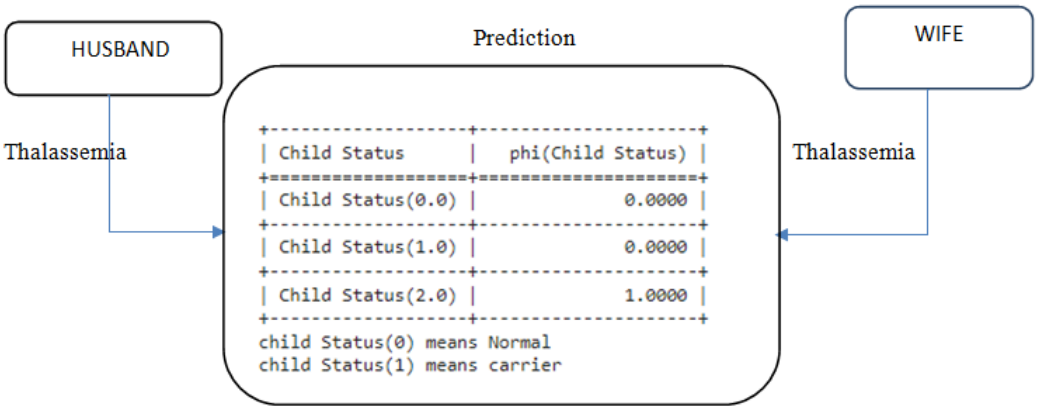


FIGURE 5.3: Probability of inheriting Thalassemia by Offspring for Case 3

There is a 25% probability of producing a normal kid in this situation when both the husband and wife are carriers. Similar to this, there is a 50% probability of having a kid with mild Thalassemia, a 25% chance of having a carrier child, or a child who has the gene. Figure 5.2 illustrates this point.

Case 3: Husband and Wife both are patients of Thalassemia or Major Thalassemia

When both parents have thalassemia, as shown in figure 5.3, there is a 100% risk that the kid will have significant Thalassemia or Thalassemia.

Case 4: Husband is Carrier and Wife is Normal.

Figure 5.4 depicts that if the husband is a carrier of Thalassemia and the wife is

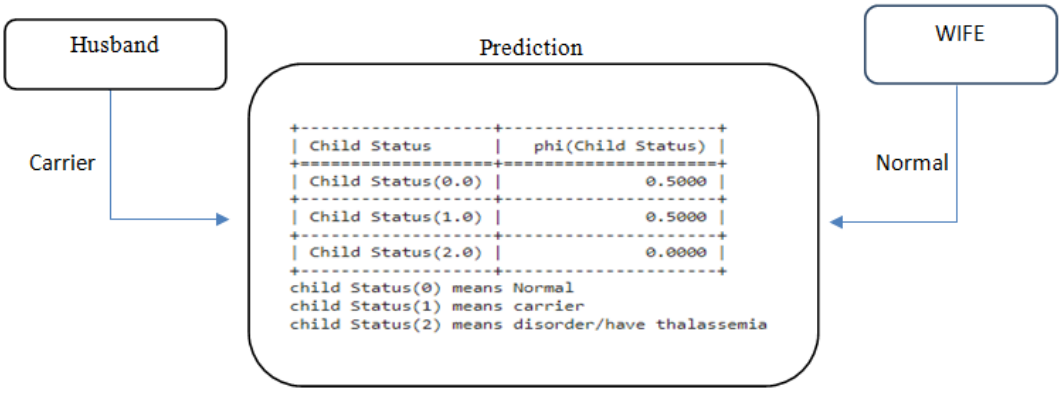


FIGURE 5.4: Probability of inheriting Thalassemia by Offspring for Case 4

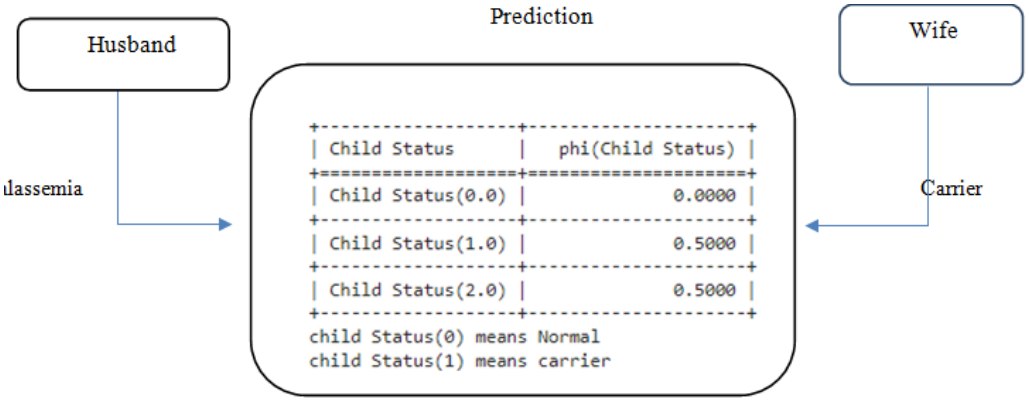


FIGURE 5.5: Probability of inheriting Thalassemia by Offspring for Case 5

normal then the chances of having a normal child are 50%.

Similarly, there is a 50% probability that a kid will have mild Thalassemia or have the Thalassemia gene and there is no risk of having a child who has thalassemia.

Case 5: Husband has Thalassemia and Wife is Carrier.
Figure 5.5 shows that there is a 50% probability of having a carrier child and a 0% chance of having a normal child.

Similarly to this, if the wife is a carrier and the husband has Thalassemia, there is a 50% probability that the kid will have major Thalassemia or Thalassemia.

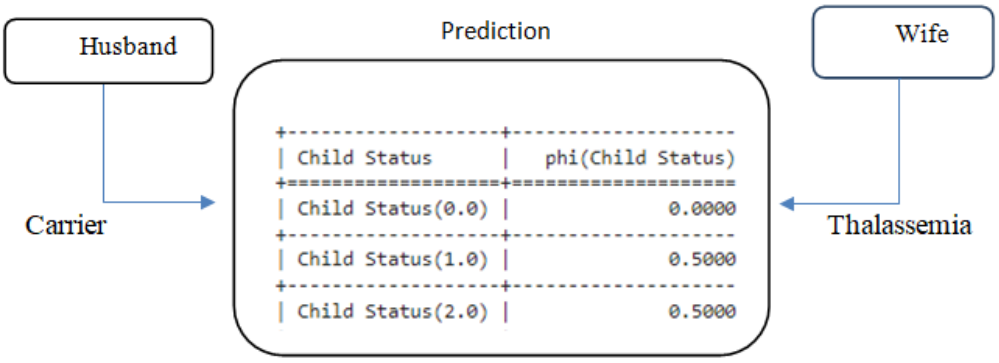


FIGURE 5.6: Probability of inheriting Thalassemia by Offspring for Case 6

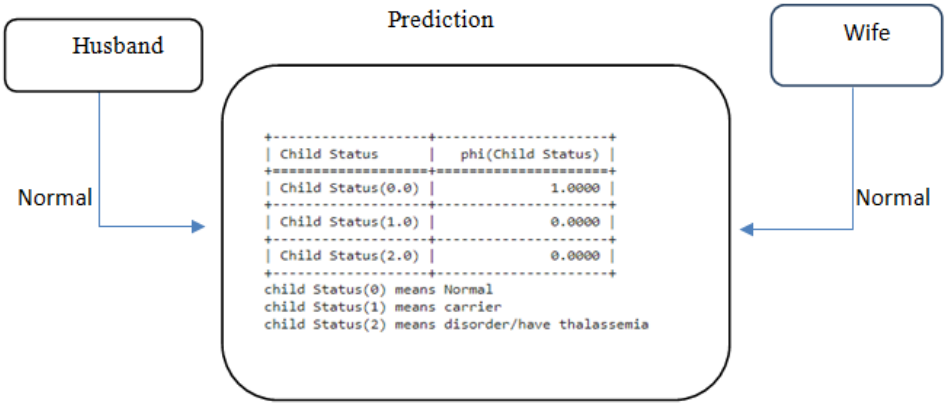


FIGURE 5.7: Probability of inheriting Thalassemia by Offspring for Case 7

Case 6: Husband is carrier and wife has Thalassemia.

Figure 5.6 depicts that there will be a 50% probability of having a carrier child and a 0% chance of having a normal child.

Similarly to this, there is a 50% chance that the child will develop Thalassemia or major Thalassemia if the husband is a carrier and the mom is a patient.

Case 7: Husband and wife both are normal.

Figure 5.7 shows that there is a 100% possibility that the child will be normal and won't inherit the condition if both parents (husband and wife) are normal.

Case 8: The husband is a patient of Thalassemia and the wife is normal.

In this case, a 100% chance is that the child will have Thalassemia trait or minor

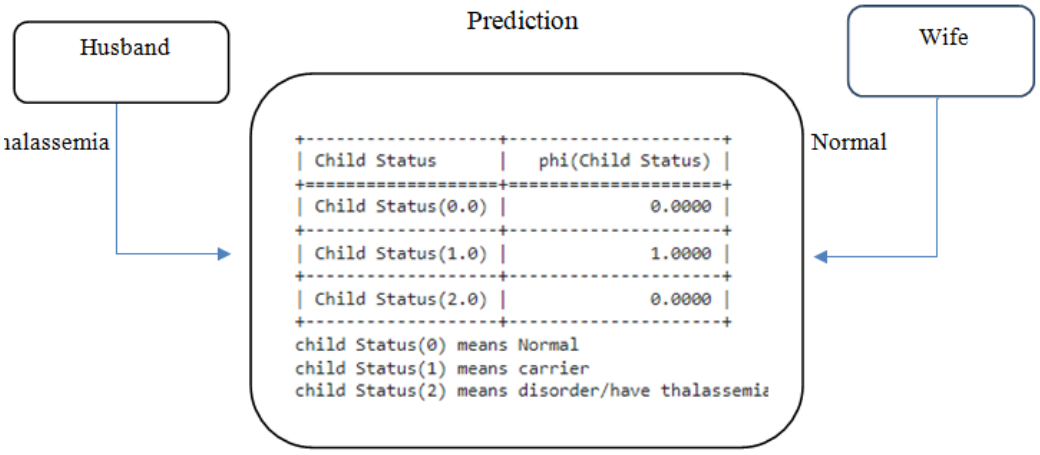


FIGURE 5.8: Probability of inheriting Thalassemia by Offspring for Case 8

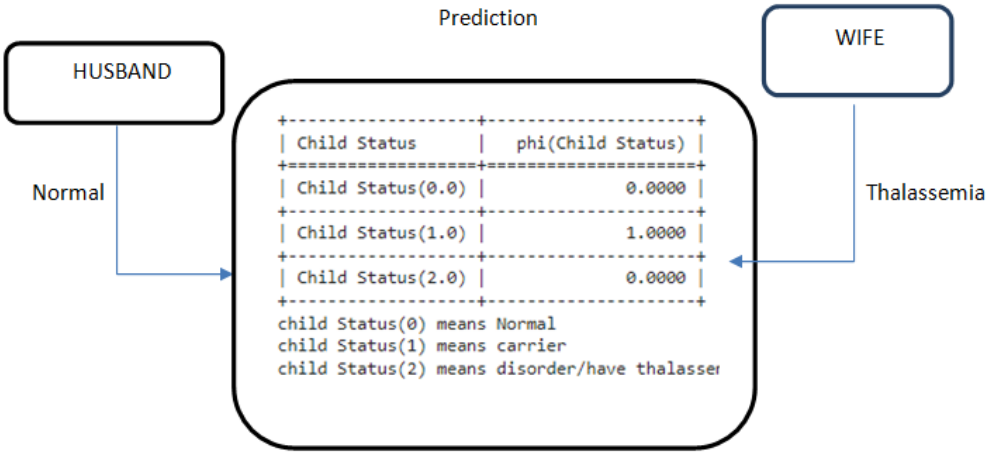


FIGURE 5.9: Probability of inheriting Thalassemia by Offspring for Case 9

Thalassemia as the husband is a patient of Thalassemia and the wife is normal. The same can be seen in figure 5.8.

Case 9: Husband is normal and wife is a patient of Thalassemia
In this case, the model predicted a 100% chance of inheriting the Thalassemia trait by offspring from parents as the husband is normal and the wife is a patient of Thalassemia. The same can be seen in figure 5.9.

Prediction results for the cases where any one of a couple (Husband or wife) is not sure about their medical history.

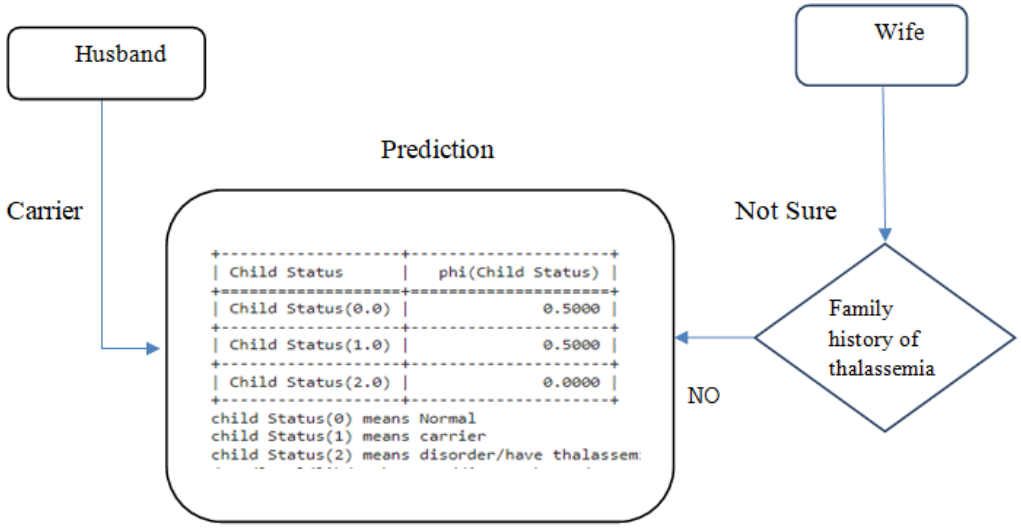


FIGURE 5.10: Probability of inheriting Thalassemia by Offspring for Case 10

This section depicts the prediction outcome of inheritance rules acquired after the implementation in Python and covers the cases where any one of the participants is not sure about medical history; which is depicted below:

Case 10: Husband is a carrier and wife is not sure about her medical history
In this situation, when the woman is unsure of her medical history and the husband is a carrier of Thalassemia, the IPDP model queries the wife about her family history to determine if she has Thalassemia or not. After that, the wife selected "NO" as an input for the model; hence, the matching forecast was produced. According to figure 5.10, there is a 50% probability that a kid will be normal and a 50% chance that the child will have mild Thalassemia or Thalassemia trait. However, since the husband is a carrier of Thalassemia and the woman is unsure of his medical condition with Thalassemia, there is no likelihood of conceiving a kid with Thalassemia major. However, when she enquired about the history of Thalassemia, she was able to establish that no one in his family had the disease. She must thus be considered normal.

Case 11: The husband is the carrier of Thalassemia and the wife did not sure about the disorder. But, she has a family history of Thalassemia.

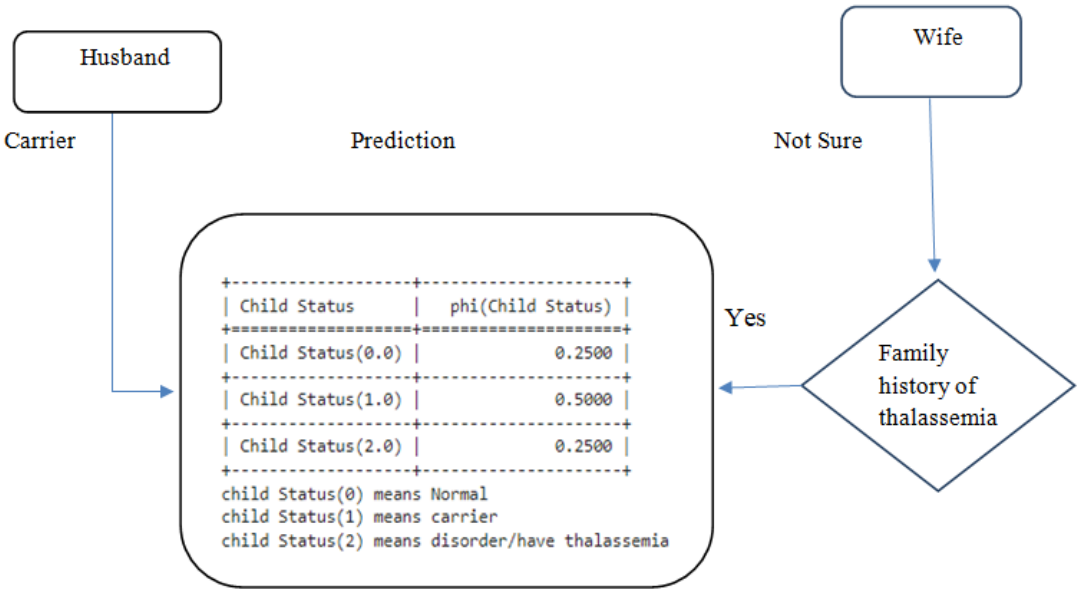


FIGURE 5.11: Probability of inheriting Thalassemia by Offspring for Case 11

In this case, when the husband is a carrier and the wife did not sure about the status of Thalassemia (which means she does not know whether she has Thalassemia or not?). However, she confirmed that she has a family history of the disorder (any family member suffering from Thalassemia). For this, the rule set-based model assumed that she may carry the genes of Thalassemia as Thalassemia is an inherited disorder and generated the prediction result. This suggests that there is a 25% chance the infant will be healthy. There is a 50% risk of having a kid with Thalassemia trait or mild Thalassemia and a 25% chance the child will inherit the condition. The same can be seen in Figure 5.11.

Case 12: The husband is normal and the wife is not sure about the medical history and family history of Thalassemia. However, she has physiological symptoms history in the family.

Figure 5.12 shows the 50% likelihood of having a carrier kid and the 50% likelihood of having a normal child. The likelihood of inheriting Thalassemia, however, is zero because the husband is healthy and the wife is unsure of her family's medical history or history of Thalassemia. However, her family has a history of

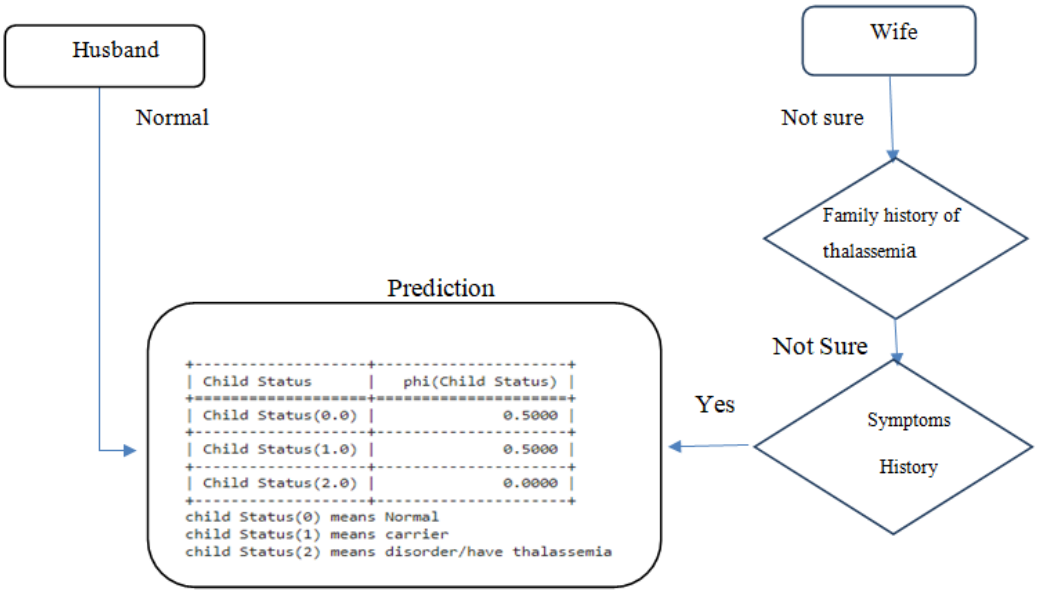


FIGURE 5.12: Probability of inheriting Thalassemia by Offspring for Case 12

physiological symptoms.

Case 13: The husband is a patient of Thalassemia and the wife did not sure about the status of Thalassemia and the family history of Thalassemia. Besides this, a history of physiological symptoms has also been found negative in the family.
Figure 5.13 depicts a couple whose husband has Thalassemia while the wife is unsure of her own health status and family history. She was thus questioned if she or any family members or relatives had ever experienced physical symptoms, to which she gave a negative response. Therefore, a newly created rule-based model indicated that there would be no chance of having a normal child. There is a 50% probability that the kid will be a Thalassemia carrier. In the same vein, there is a 50% risk of conceiving a child with major thalassemia.

Case 14: Husband is normal and wife is not sure about medical, family, and physiological symptoms and history of Thalassemia. Her obstetric history is also found negative means she did not have any child with Thalassemia.
Since the husband is healthy and the woman is unsure about her medical history, there is no likelihood of producing a kid with Thalassemia in this situation, and

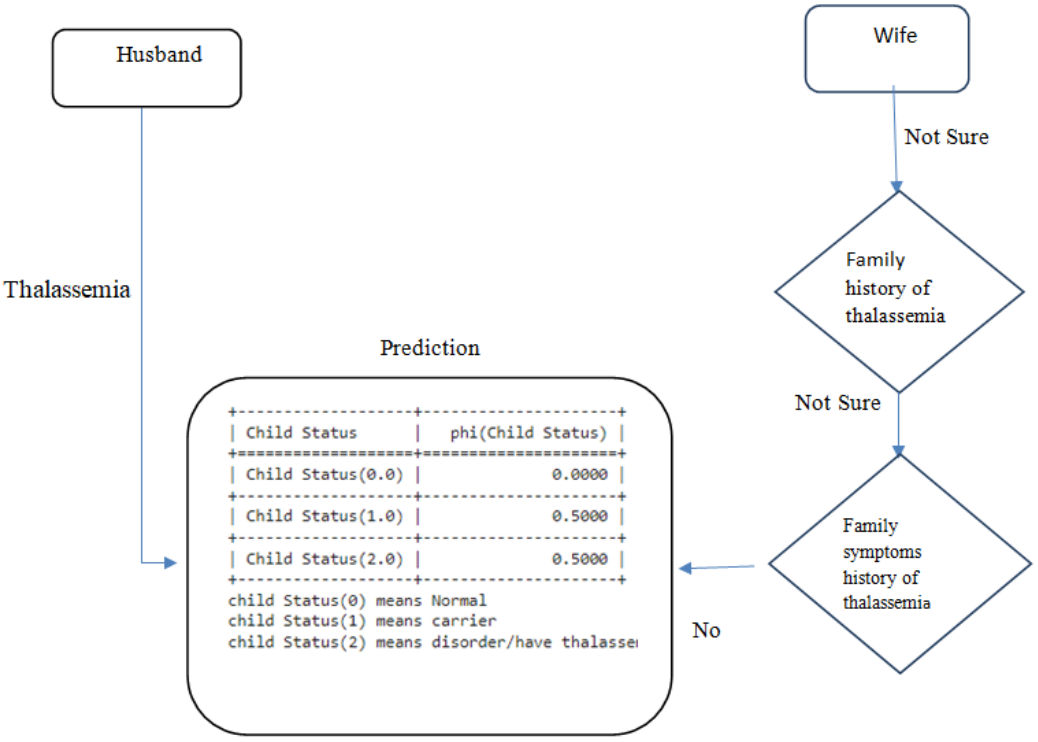


FIGURE 5.13: Probability of inheriting Thalassemia by Offspring for Case 13

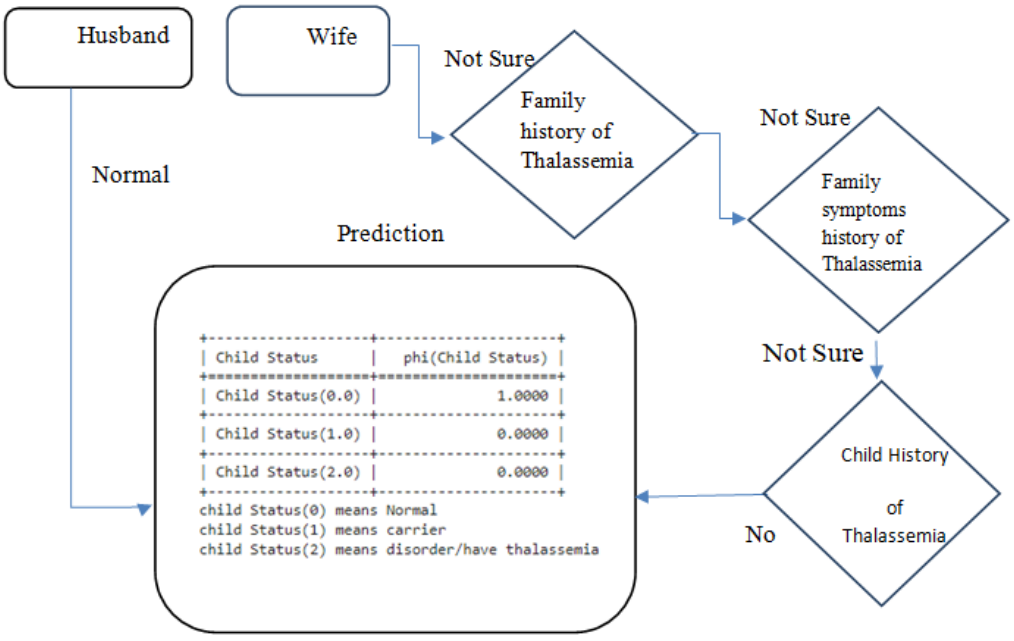


FIGURE 5.14: Probability of inheriting Thalassemia by Offspring for Case 14

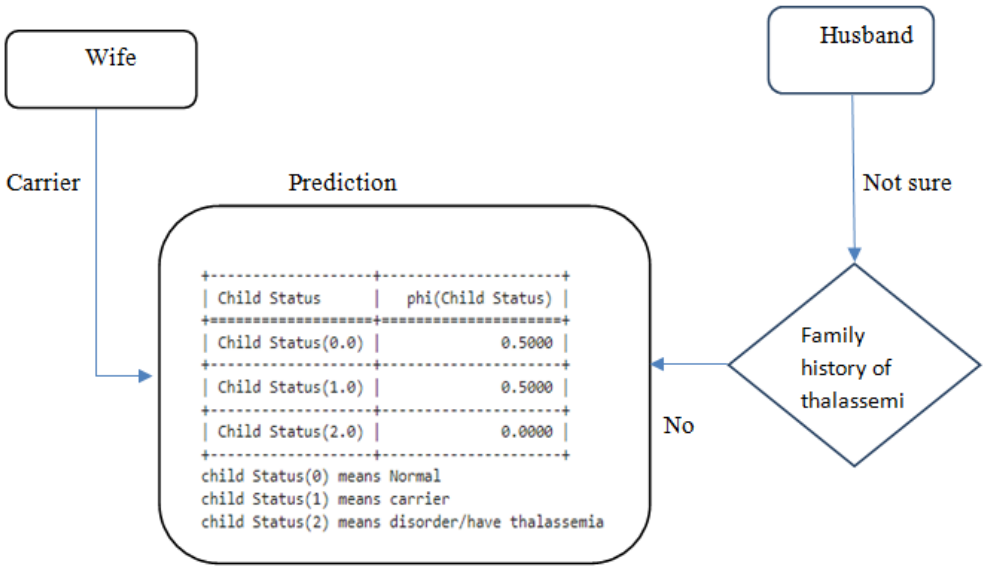


FIGURE 5.15: Probability of inheriting Thalassemia by Offspring for Case 15

there is a 100% chance that the child will be normal. However, there is no Thalassemia history in the family. She is therefore a regular person. The same can be seen in 5.14.

Case 15: Wife is Carrier and husband is not sure about the disorder. But, he did not have a family history of Thalassemia.
Since the husband is healthy and the woman is unsure about her medical history, there is no likelihood of producing a kid with Thalassemia in this situation, and there is a 100% chance that the child will be normal. However, there is no Thalassemia history in the family. She is therefore a regular person.

Figure 5.15 shows that the woman carries the Thalassemia gene whereas the husband is unsure of his healthcare background. However, his family history is negative, thus he has no relatives who have thalassemia. So, it's safe to say that the husband is normal. In this manner, the model predicted that there would be a 50% chance of producing a normal child. The likelihood of conceiving a kid with mild or carrier Thalassemia is 50%. Similar to this, there is no probability of conceiving a child who has significant thalassemia.

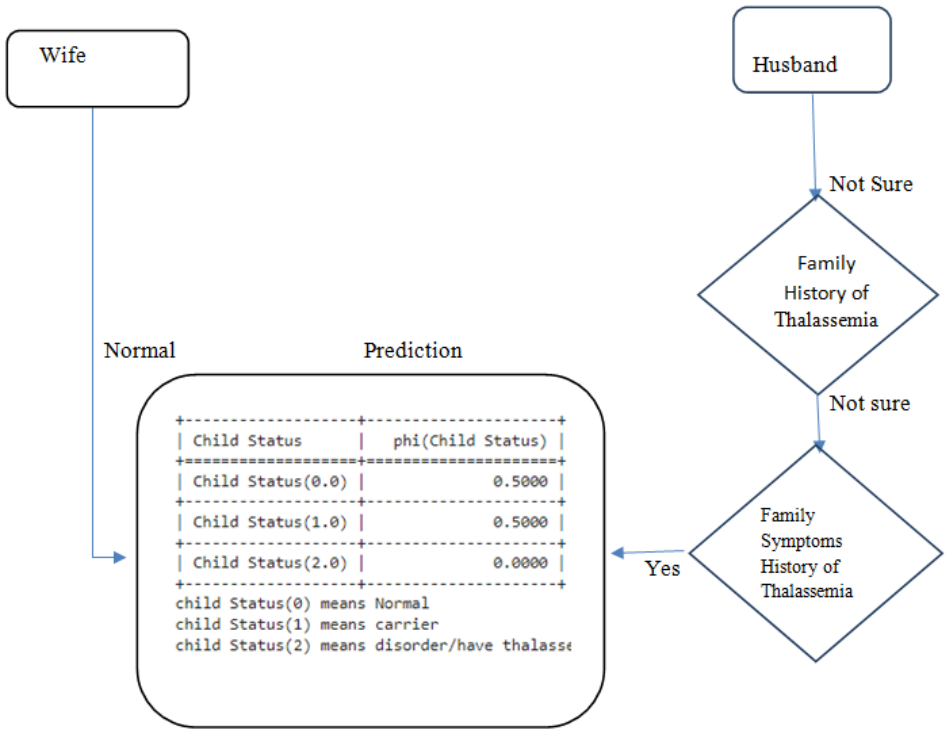


FIGURE 5.16: Probability of inheriting Thalassemia by Offspring for Case 16

Case 16: The wife is normal and the husband is not sure about his medical history and family history of Thalassemia. But, physiological symptoms of Thalassemia have been observed in his family.

Figure 5.16 shows that the wife is normal and that the husband is unsure about his medical history and his family's Thalassemia history. However, his family has experienced the physiological signs of Thalassemia.

Therefore, it is reasonable to infer that he may inherit Thalassemia, and if he is a Thalassemia patient or has major Thalassemia, then the couple has odds of having a normal kid are 50%. The likelihood of conceiving a kid with mild or carrier Thalassemia is 50%. Similar to this, there is no probability of conceiving a child who has major thalassemia.

Case 17: The wife is a patient of Thalassemia and the husband is not sure about his medical, family history, and history of physiological symptoms of Thalassemia. But, the obstetric history of the couple has been found negative which means the

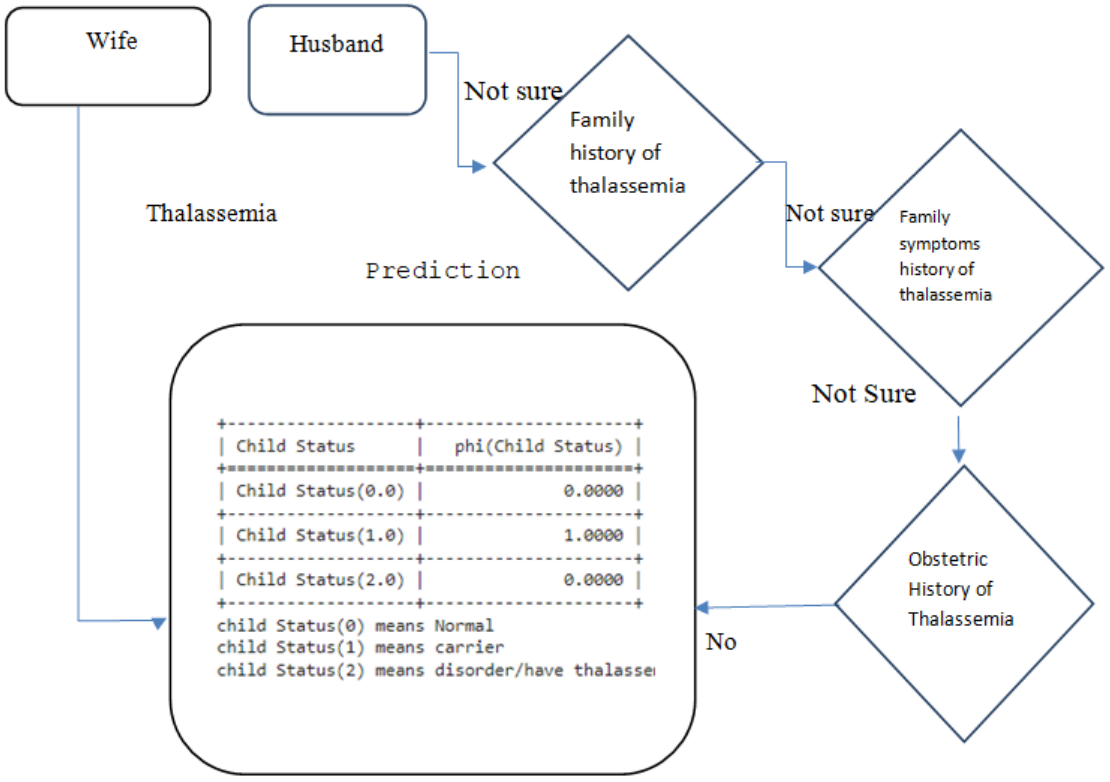


FIGURE 5.17: Probability of inheriting Thalassemia by Offspring for Case 17

couple does not have any children with Thalassemia.

Figure 5.17 shows a situation where the wife has Thalassemia and the husband is unsure about his health history, family history, or previous experiences with the physiological signs of Thalassemia. However, the couple's obstetric history was shown to be negative, indicating that they had never had a kid with thalassemia. There is a 100% possibility that the couple will give birth to a kid who has mild Thalassemia or Thalassemia trait in this situation. The likelihood of having a healthy kid or a patient with Thalassemia or serious Thalassemia is 0%.

5.2.2 Prediction Results of Model in Context of Haemophilia

All the rules developed for the early diagnosis of inheritance of haemophilla among offspring have been presented in this segment with a graphical demonstration of execution. All these rules cover not only known cases of existing inheritance

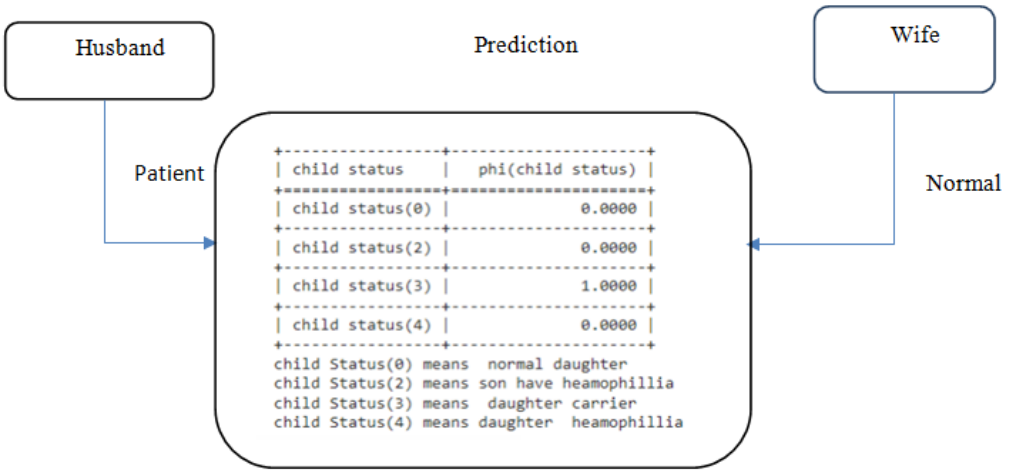


FIGURE 5.18: Probability of inheriting Haemophilia by Offspring for Case 1

patterns but also unknown cases.

Prediction outcomes for couples (husband and wife) who are certain of their Haemophilia history.

This section depicts the prediction outcome of inheritance rules acquired after the implementation in Python and covers the cases where participants already know their medical history.

Case 1: The husband is a patient of Haemophilia and the wife is normal.
In this situation, there is a 100% chance that the child will be born with a carrier daughter because the father has Haemophilia and is responsible for passing the gene to the daughter. All of the couple's sons will be normal because the wife is healthy and passes the normal X chromosome, which is what causes Haemophilia to occur. The same can be seen in figure 5.18.

Additionally, there is zero possibility of conceiving a son or daughter with Haemophilia. The husband and wife, who are already aware of their history, provided input to the prediction model, which then generated the outcome based on the pattern and relational dataset that previously existed. The likelihood that one kid out of every four will inherit Haemophilia has been computed as a percentage.

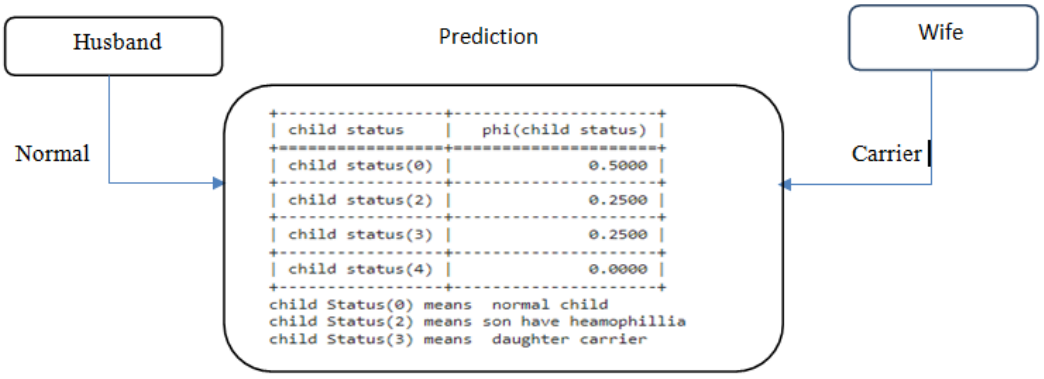


FIGURE 5.19: Probability of inheriting Haemophilia by Offspring for Case 2

Case 2: When the husband is normal and the wife is a Carrier of Haemophilia. Given that the mother is a carrier of Haemophilia and has just one defective X chromosome out of two, there is a 25% probability that the pair will produce a carrier daughter (1/4 offspring) and a 25% chance that they will have a boy with Haemophilia. Similar to this, there is no probability of having a daughter with Haemophilia and a 50% chance of having a normal child. Figure 5.19 makes the execution sequence very obvious.

Case 3: When the husband has Haemophilia and the wife is a carrier of Haemophilia. Figure 5.20 shows that if the husband has Haemophilia and the woman is a carrier, there is a 25% (1/4) chance of having a normal kid and a 25% chance of having a boy who has Haemophilia. Similarly, there is a 25% probability that the couple will have a daughter who is a carrier and a 25% risk that their daughter will develop Haemophilia.

Case 4: Both the husband and the wife are normal.
Figure 5.21 states that there is a 100% likelihood that a normal kid will be born to a normal couple.The likelihood of having a kid with Haemophilia is 0%. Additionally, there is zero possibility of conceiving a son or daughter with Haemophilia. The husband and wife, who are already aware of their history, provided input to the prediction model, which then generated the outcome based on the pattern and

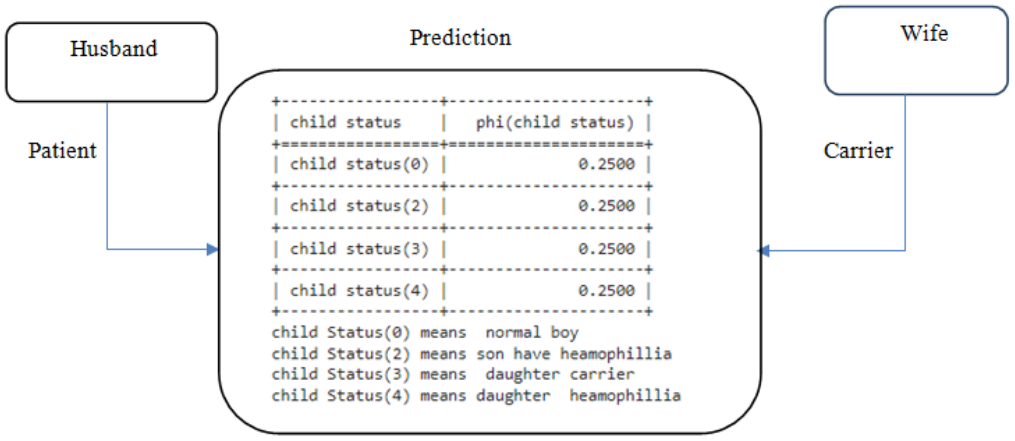


FIGURE 5.20: Probability of inheriting Haemophilia by Offspring for Case 3

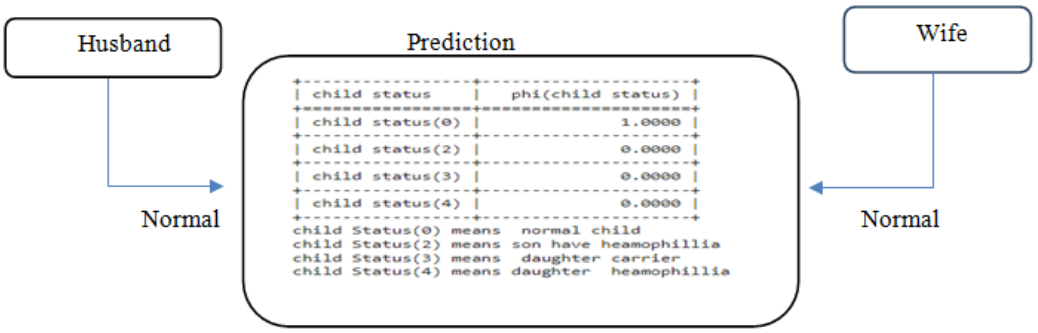


FIGURE 5.21: Probability of inheriting Haemophilia by Offspring for Case 4

relational dataset that previously existed. The likelihood that one kid out of every four will inherit Haemophilia has been computed as a percentage.

Results of the prediction in situations when either the husband or the wife is unsure about their medical history.

This section addresses the situations where any one of the participants is unsure about their medical history and illustrates the prediction outcome of inheritance rules gained after their implementation on Python as follows:

Case 5: The husband is a patient of Heamophilia and the wife is not sure about the medical history of Haemophilia; but she has a family history of Heamophilia. Figure 5.22 states that there is a 100% likelihood that a normal kid will be born to a normal couple.The likelihood of having a kid with Haemophilia is 0%.

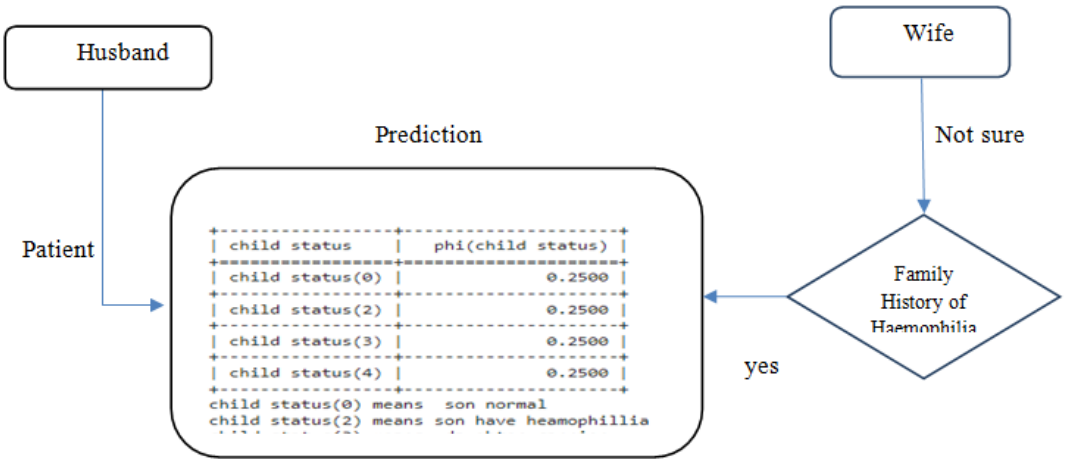


FIGURE 5.22: Probability of inheriting Haemophilia by Offspring for Case 5

Figure 5.22 shows that if the wife is unsure of her medical history and the husband has Haemophilia, the wife's family history has been inquired about, which is positive in the context of the wife. As a result, the model predicted that the probability of having a normal son would be 25% and the chance of having a son with Haemophilia would be 25%. Similar to this, there is a 25% chance that the couple will have a carrier daughter and a 25% risk that their daughter will develop Haemophilia.

Case 6: The husband is normal and the wife is not sure about her medical history and family history of Haemophilia; but she has a history of physiological symptoms in the family.

Figure 5.23 shows that if the husband is healthy and the wife is unsure of her medical history, the model will ask the wife whether there have been any physiological symptoms in the family by providing a list of symptoms for the choice. If the wife has confirmed the presence of physiological symptoms in the family, she may be a carrier of Haemophilia.

As a consequence, the model estimated that there is a 50% chance of having a normal child and a 25% risk of having a boy with Haemophilia. Similar to this, there is a 25% possibility that the couple will have a daughter who is a carrier and a 0% chance that their daughter will develop Haemophilia.

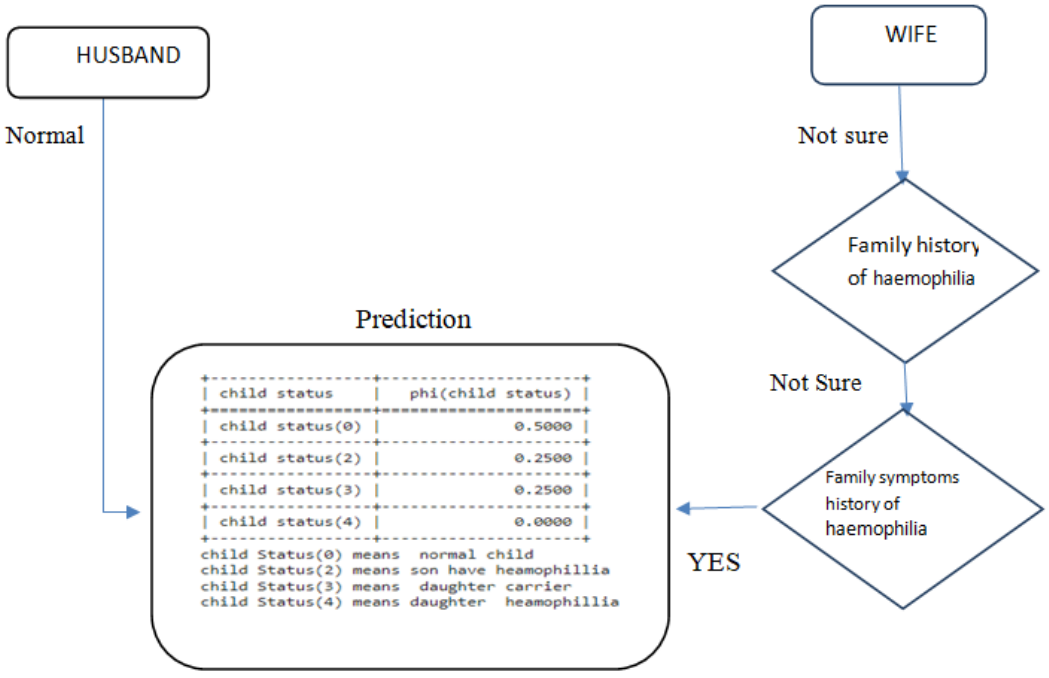


FIGURE 5.23: Probability of inheriting Haemophilia by Offspring for Case 6

Case 7: The husband is a patient of Haemophilia and the wife is not sure about her medical and family history of Haemophilia. But she did not find any symptoms or history of Haemophilia.

Figure 5.24 shows what would happen if the husband had Haemophilia and the woman did not know her medical history or any Haemophilia in the family. She then confirmed that there had never been a family history of physical problems. According to the model's results, there is a 25% chance of having a son without Haemophilia and a 25% risk of having a son with the condition. Additionally, the couple has a 25% chance of having a daughter who is haemophilic and a 25% chance of having a daughter who is a carrier.

Case 8: The husband is normal and the wife is not sure about her medical history of Haemophilia. No Family history of Haemophilia has been found in the case of the wife. But, she has an obstetric history (history of child related to Haemophilia). Figure 5.25 states that if the husband is healthy and the woman is unsure of her

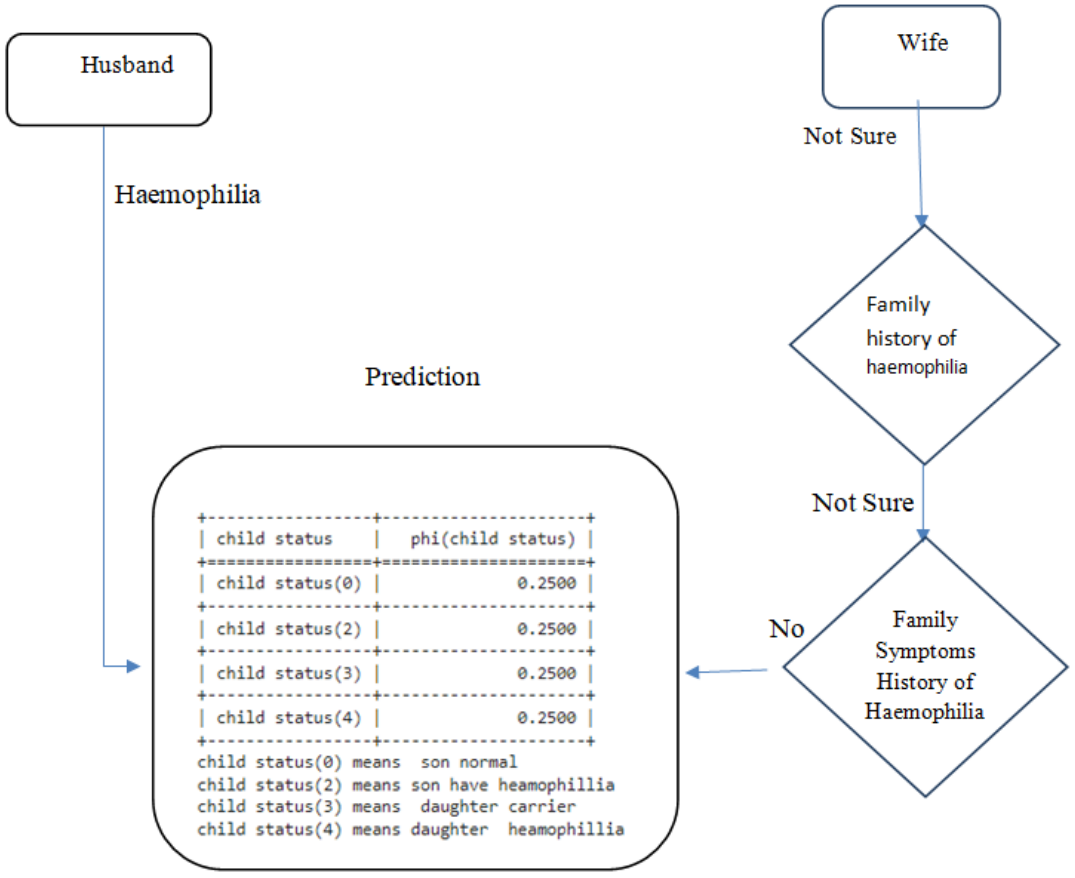


FIGURE 5.24: Probability of inheriting Haemophilia by Offspring for Case 7

medical history, she may have Haemophilia. She then affirmed that she had neither a family history of physical complaints nor a prior experience of them. But she had no obstetric history of Haemophilia, meaning she had no children who had the condition. As a result, the model predicted that the probability of having a normal child would be 50% and the chances of having a son with Haemophilia would be 25%. Similarly, a 25% probability is that the couple will have a 25% carrier daughter and a 0% chance of having a daughter with Haemophilia.

Case 9: The wife is a carrier of Haemophilia and the husband is not sure about his medical history of Haemophilia. But, he did not have a family history of Haemophilia.

If the wife is a carrier and the husband is unsure of his medical history for

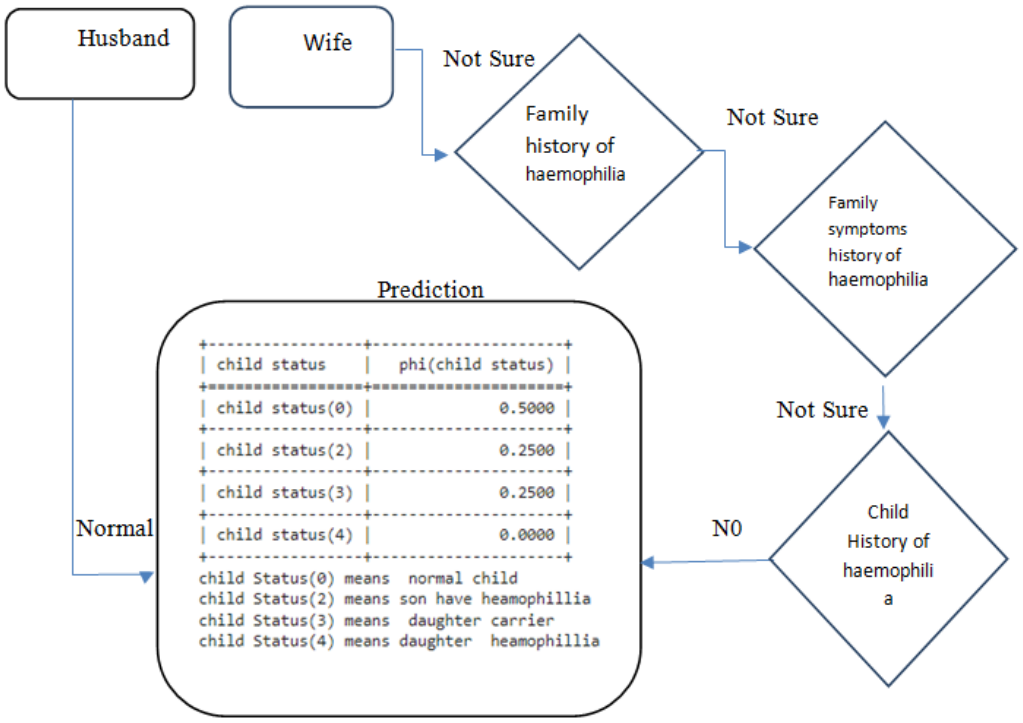


FIGURE 5.25: Probability of inheriting Haemophilia by Offspring for Case 8

Haemophilia, this is shown in figure 5.26. He then confirmed that he had no family history and that he was regarded as normal. As a consequence, the model estimated that there would be a 50% chance of having a normal child and a 25% risk of having a boy with Haemophilia. Similarly, there is a 25% chance that the couple will have a daughter who is a carrier and a 0% chance that their daughter will have Haemophilia.

Case 10: The wife is Normal and the husband is not sure about his medical history and family history of Haemophilia. But, he had a history of physiological symptoms in the family.

Figure 5.27 shows a normal-appearing wife and a husband who is unsure about his medical history and his family's history of Haemophilia. But his family has a history of physical issues. The model thus indicated that there was a 100% likelihood of having a carrier daughter. There is no likelihood that the couple will

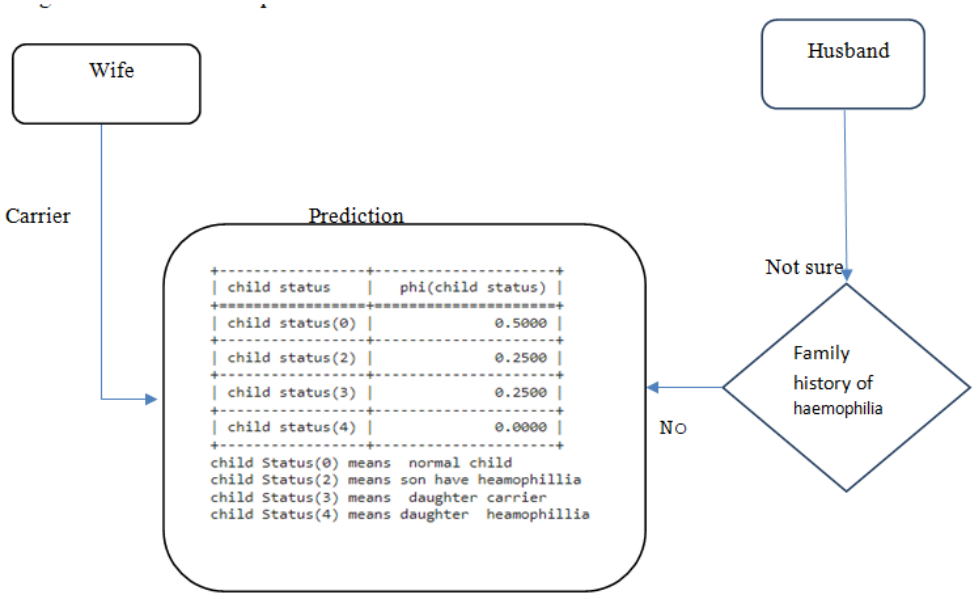


FIGURE 5.26: Probability of inheriting Haemophilia by Offspring for Case 9

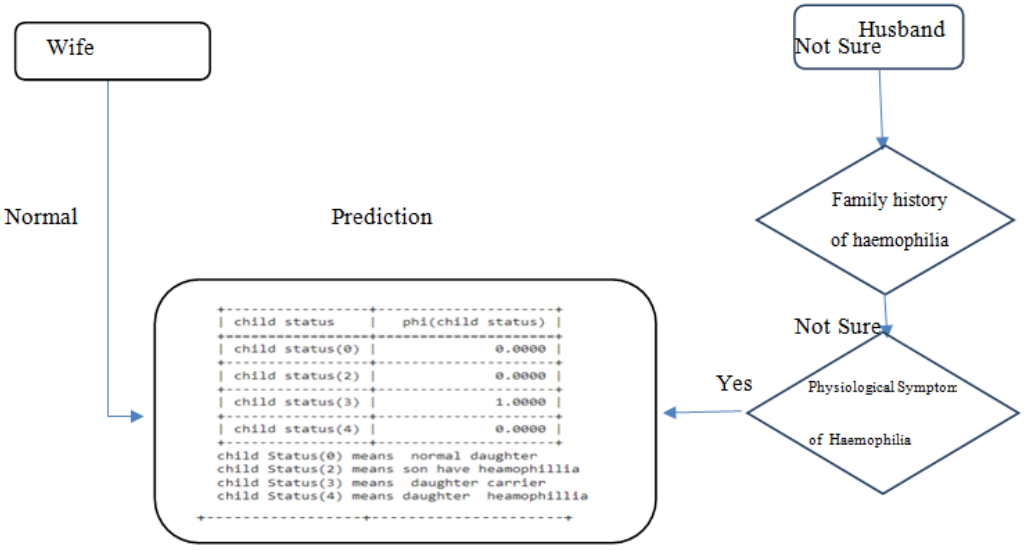


FIGURE 5.27: Probability of inheriting Haemophilia by Offspring for Case 10

produce a son or daughter with Haemophilia.0%. Similarly, there is zero likelihood that the couple will produce a healthy girl.

Case 11: The wife is a carrier and the husband is not sure about the medical, family, and symptom history of Haemophilia. But, his obstetric history of Haemophilia

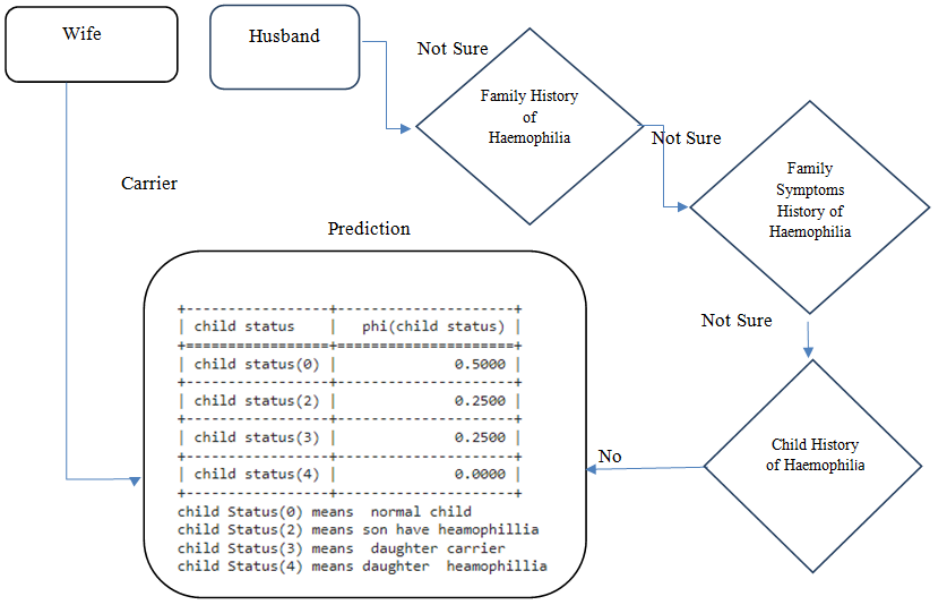


FIGURE 5.28: Probability of inheriting Haemophilia by Offspring for Case 11

was found negative.

Figure 5.28 show that the husband does not know his medical history, family history, or physiological symptom history of Haemophilia, whereas the wife is a carrier. However, the obstetric history was negative; due to this husband will be regarded as normal.

As a consequence, the model estimated that there was a 50% probability of having a normal kid and a 25% risk of having a boy with Haemophilia. Similarly, the couple has a 25% probability of having a daughter who is a carrier and a 0% chance of having a daughter who has Haemophilia. Therefore, the model offers a solid foundation for the creation of a mobile application-based eHealth system that is intended to forecast the likelihood of having a kid with Haemophilia and thalassemia.

5.3 Conclusion

Prediction of probability of inheritance of haemotological disorder at prenatal stage is complex task and requires enough time and detailed medical background related to disorders. Any single mistake could leads to wrong result. Moreover, medical resources associate to this such as genetic counselling and genetic counsellors are limited. Above of all, awareness and attitude of public towards such kind of disorders and also for diagnosis services or options are less. Hence, to resolve this problem a rules based prediction model has been developed during the current research work; which has been presented in this chapter. This model helps in predicting the probability of inheritance of haemotological disorder such as Haemophillia and Thalassaemia in offspring at prenatal stage. Thus, the suggested model has filled in the entire identified gap with the development of a mobile application-based e-Health system using rule-based prediction models and a hybrid model (created by combining a decision tree and a naive bayes algorithm); which is discussed in detail in chapter 6. The publications of the articles that are connected to this chapter are listed below.

Sharma, P., Sharma, A., & Kakkar,S. (2024), "A multi-layered framework for analysis and prediction of Haemophilia using machine learning techniques based e-health system.", Tuijin Jishu/Journal of Propulsion Technology, 45 (3), 2295-2307.

Chapter 6

Inherited Prenatal Disorder Predictor (IPDP) Application

The structure of this chapter is as follows: The chapter is introduced in Section 6.1. Section 6.2 describes the rules based prediction model, its workings, and its outcome in detail. The whole process of creating the mobile-based e-Health system known as the Inherited Prenatal Disorder Predictor (IPDP) is covered in Section 6.3. Additionally, this part includes information about the application's design and implementation. Evaluation of application performed by doctors has been discussed in section 6.4. Section 6.5 concluded the findings.

6.1 Introduction

A smartphone application called Inherited Prenatal Disorder Predictor (IPDP) uses information and communication technologies to deliver e-Health services to remote users. This application predicts the probability of inheritance of haematological disorders from parents to offspring. Inherited Prenatal Disorder Predictor (IPDP) application mainly covers major inherited haematological disorders namely Haemophilia and Thalassemia. By providing their medical or family history, people can find out their likelihood of having a child with a disorder like Haemophilia or thalassemia. If they are unsure of the history of the disorder, the application

offers them additional options (like a list of symptoms associated with the disorder and previous obstetric history) to confirm the situation. This application provides prenatal, genetic counselling, and premarital services like facilities in the context of Haemophilia and Thalassemia to couples so that they can take informed decisions regarding the pregnancy or marriage. The dashboard of the application provides two main modules namely Haemophilia and Thalassemia for the prediction of the probability of inheriting haematological disorders by the offspring of couples. Additionally, a report module is also provided to show all the details along with results to the individual using the application. Besides this, a list of a few expert doctors from different hospitals has been provided for the further confirmation of prediction results. In this way, this application disseminate awareness regarding inherited disorders and provides prenatal diagnosis/premarital or genetic counselling electronically, so that the lack of knowledge regarding the mentioned services and the shortage of staff associated (such as genetic counsellors) among the underserved public can be overcome. The Inherited Prenatal Disorder Predictor (IPDP) application has utilized a hybrid diagnosis model and rules based prediction model for its development. For this, Kotlin programming Language on the Android studio platform has been used to implement above mentioned models. For all this, the design flow depicted in figure 3.4 of chapter 3 has been utilized. The application consists of Haemophilia, Thalassemia, connect to the doctor, and report module. Each module performs a different task as Haemophilia provides prediction regarding the chance of inheritance of Haemophilia in offspring based on the input given by the user. Similarly, Thalassemia provides prediction regarding the chance of inheritance of Thalassemia in offspring based on the input given by the user. Connect to Doctor option provides the list of doctors associated with the disease to the user so that he/she can confirm the prediction result provided by the application and take further consultancy. Report section provides the details regarding the user along with the partner, prediction results, and risk factors related to pregnancy or fertility issues based on medical, family, physiological history, and

obstetric history in the context of disorders (Haemophilia and Thalassemia).

6.2 Implementation of Inherited Prenatal Disorder Predictor (IPDP) Application

The Inherited Prenatal Disorder Predictor (IPDP) application has utilized the design depicted as flow chart in figure 3.3 of chapter 3. Besides this, a rules based prediction model has also utilized for the final implementation and development of the application. Hence, the depicted design in figure 3.4 and rules based prediction model have been implemented through Kotlin programming on the android studio platform. This platform requires an i5 processor, 8GB of RAM, and a minimum of 256 GB SSD. After the development of the application, apk file of the programming code of the proposed application was installed on the mobile. So that, the output generated by the application can be checked. The Inherited Prenatal Disorder Predictor (IPDP) application consists of an introductory page and dashboard with different options such as Haemophilia, Thalassemia, connection to the doctor, and report. Each option consists of an application that performs different operations; which are elaborated through various screenshots of the Inherited Prenatal Disorder Predictor (IPDP) application depicted in section 6.3 of this chapter.

6.3 Working of Inherited Prenatal Disorder Predictor Application

This section uses screenshots to illustrate how the Inherited Prenatal Disorder Predictor (IPDP) Application functions and the results that are produced after choosing specific input choices or buttons. After installing the application, the very first pages that appear on the screen provide introductory information about the target bleeding disorders such as Haemophilia and Thalassemia. In this case,

FIGURE 6.1: Inputs for Haemophilia when the medical status of the couple is known

what should the user do if they are unaware of what Haemophilia and Thalassemia are? Then, he or she can get knowledge about it from the working of the application. Hence, it can be said that the application disseminates awareness about the Haemophilia and Thalassemia among the general public. The next page prompts for registration of users; where users who want to use the application can register themselves by providing their details along with their partner's details.

These details will be merged with prediction results, later on when a report will be generated. In this phase, username and password will also be generated for

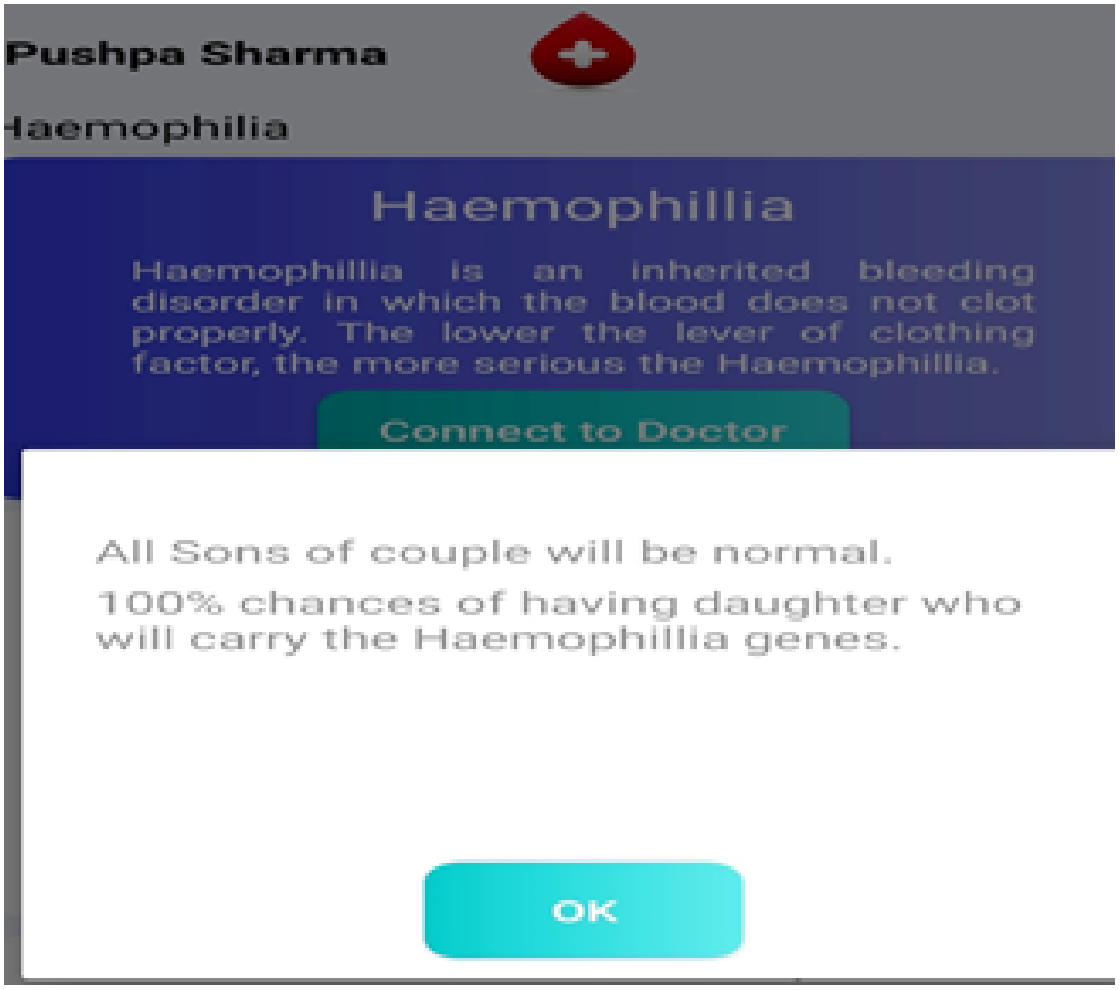


FIGURE 6.2: Prediction Result for Haemophilia

registered users for future login processes. After signup, the user has to login for the utilization of the application. As the user login, a dashboard is prompted on the screen with the option Haemophilia, Thalassemia, report, and logout option. Hence, a dashboard is the place from where an actual task has started. As this application provides two option buttons, namely Haemophilia, and Thalassemia, to provide prediction related to the inheritance of these two in offspring of the user, who is using the app, this is the place where different options work together to predict the probability for the occurrence of inherited or genetic disorders in offspring. To do this, choose the relevant radio button. After that, a screen with a Haemophilia page and other input choices will display. A similar thing can be seen in Figure 6.1,

FIGURE 6.3: When a Patient's Medical History Is Uncertain but Their Family History Is Known

where the user is requested to indicate whether or not the couple has Haemophilia by checking a yes or no box. In this manner, after supplying the known status of the illness in the form of yes or no, the application will forecast the likelihood that the expected kid or offspring would inherit a bleeding problem based on the medical history of the couple, as shown in Figure 6.2.

The application also provides the provision where any one of a couple is not sure about the medical history of Haemophilia. In this situation, the user will pick the not sure option under the relevant heading, as shown in Figure 6.2. For example, if the husband is unsure, the not sure option under the heading spouse is selected; otherwise, the not sure option shown under the wife's Haemophilia status will be selected; This provision is absent from earlier versions of the inheritance pattern. The application will inquire about a family history of disorder as soon as the

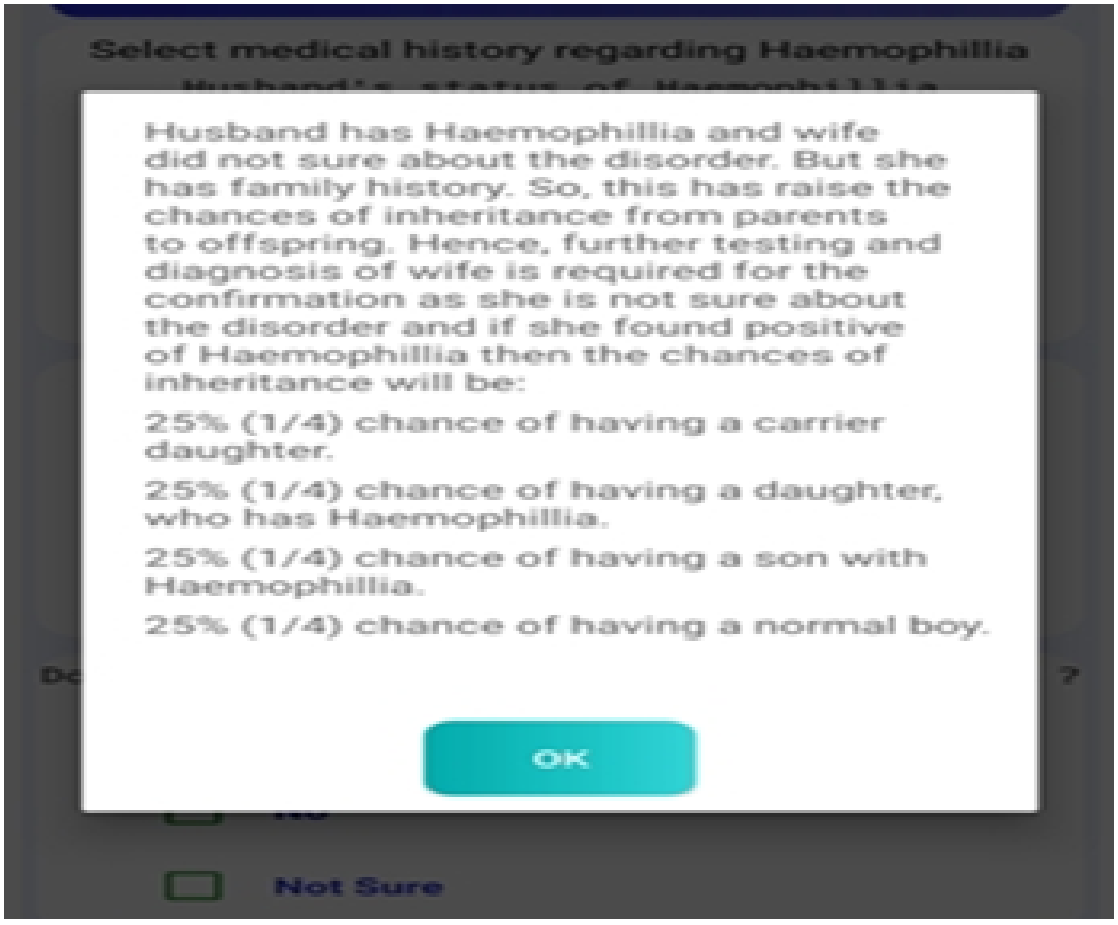


FIGURE 6.4: Prediction Result Based On Medical and Family History of User

user picks the "not sure" option in any area titled as husband's/ Wife's Status of Heamophilia; which is also depicted in Figure 6.3. This implies that whether or not a family member has a bleeding issue. The tool forecasts the outcome based on the user's selection and the couple's family history, as shown in Figure 6.4.

Now, again if there is a situation where the user is also not sure about family history then again they can select the not sure option under this section which will open the list of symptoms for further selection the can be seen in Figure 6.5.

Users can select any one two or all the symptoms from the list as per the family record in this context. Otherwise, the user can select no one option from the list as seen in Figure 6.6 .If he/she thinks that the user or the family member did not have any symptoms. Similarly, if a user is not sure then he/she can select the not

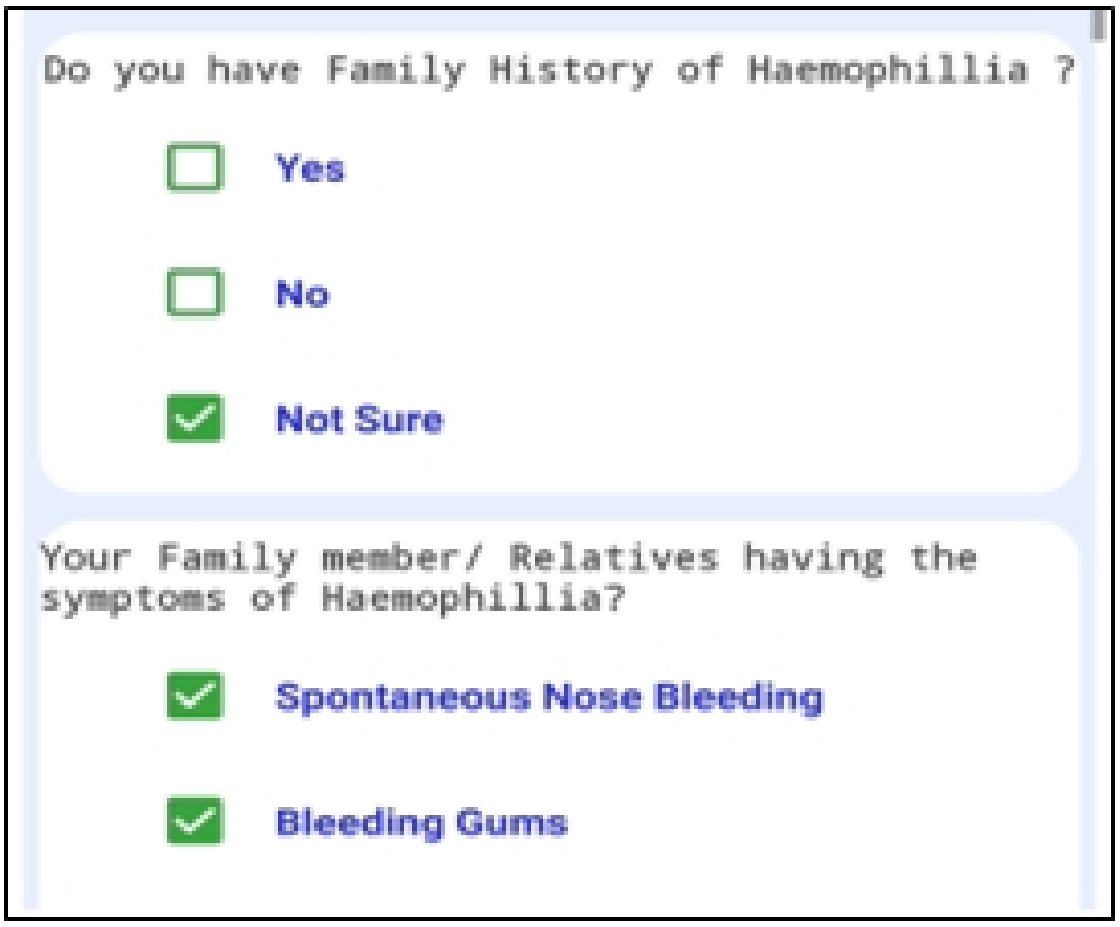


FIGURE 6.5: Physiological Symptoms as Inputs

sure option from the list; which directs the user toward an obstetric history of a bleeding disorder or Haemophilia as depicted in 6.6.

Obstetric history verifies whether the user has any children who have bleeding disorders or who exhibit symptoms related to them. After this response is prompted on the screen, as shown in Figure 6.6, the user might respond with a yes or no and which will predict the result as depicted in Figure 6.7.

When the user clicks on the submit button of the result page then it redirects the user toward the report section which depicts all the details along with prediction results and a list of expert doctors for further consultancy to confirm the application result as this application does not replace the role of medical professionals instead it provides awareness and assistance to the general public regarding

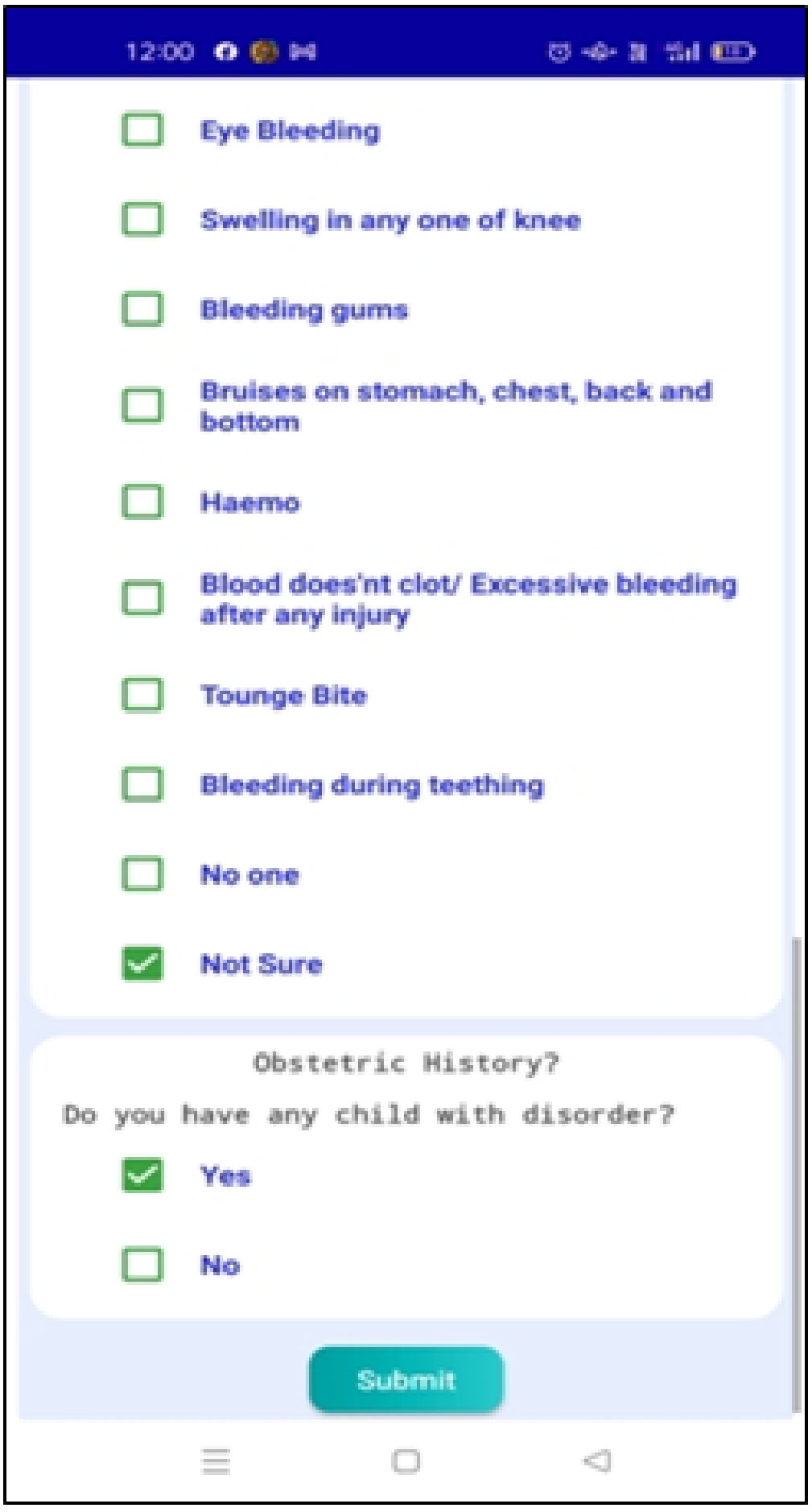


FIGURE 6.6: Obstetric History

FIGURE 6.7: Prediction Results after inputting Obstetric History

inherited bleeding disorder by providing e-Health services.

Similar to the first module, the second module, Thalasemia, asks the user for information about their medical history as well as their partner's medical history, family history (if any of the users is unsure about their medical history), physiological symptoms history (if the user's family history is uncertain), and obstetric history (if the user's physiological symptoms are uncertain or no symptoms of Thalassemia have been noticed in themselves or their family members/relatives) with the name normal, carrier, patients and not sure. The application generates a prediction after receiving these inputs and prompts the user on the screen. As in the Haemophilia module, a report of the prediction result will also be generated

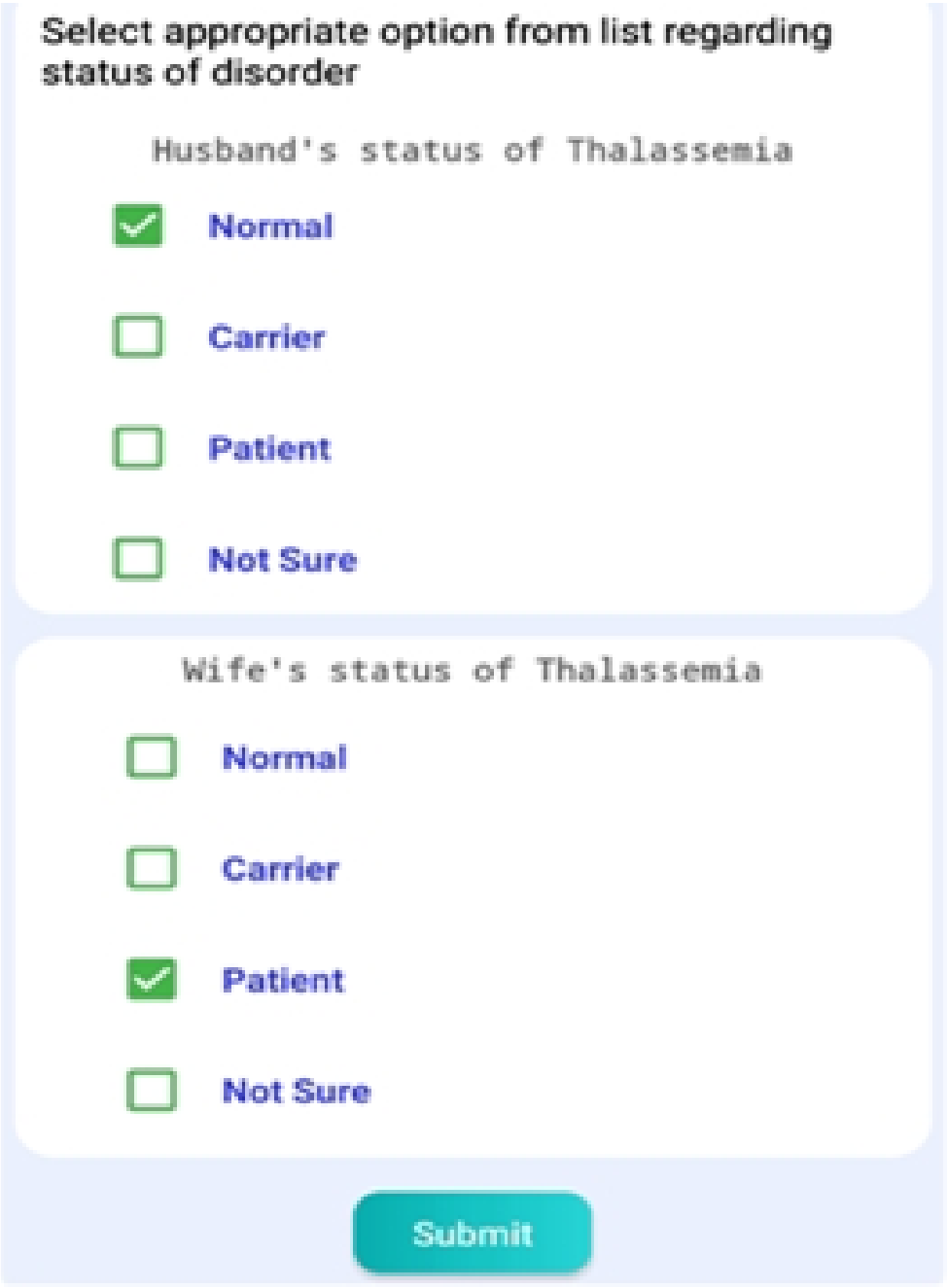


FIGURE 6.8: Input for Thalassemia

in addition to this. For further confirmation of the prediction and the administration of associated care, a list of doctors is also provided. The screenshot related to Thalassemia can be seen in Figure 6.8. In which the user chooses the option labelled as normal and the patient's present state of thalassemia. As soon as the user clicks the submit button after providing their medical history, the programme will then anticipate the outcome.

Thus, the application uses a rules-based prediction model, which makes use of the rule set developed on the inheritance pattern of bleeding disorders during the current research work, to predict the likelihood of inheriting Haemophilia and Thalassemia from parents to offspring. In addition, patient medical, family, physiological, and obstetric information was used. Throughout the whole process of development and post-development activities, patients with specific bleeding problems, gynaecologists, and haematologists have offered guidance and advice.

6.4 Performance Evaluation

Following the creation of the Inherited Prenatal Disorder Predictor (IPDP), an evaluation task was carried out while considering the results. To further execute this, an application has been sent to qualified doctors, enabling the results to be verified. Following review, medical professionals have given their approval for this novel application to be used in the real time treatment protocol. Physicians will treat haematological problems using the recently created e-Health system, Inherited Prenatal Disorder Predictor, which is based on a cell phone. Physicians request that patients use the technique to determine the likelihood that they may have children affected by Thalassaemia and Haemophilia. Expert physicians will later cross-check the obtained results to determine the need for additional actions. It would help to save time and effort during the diagnosing process in this way. The application will be made available on websites like "google play" in order to accomplish this.

6.5 Conclusion

Hence, the suggested e-Health system may be used to educate the general public about disorders and medical resources, serve as a model for training medical students, and aid medical professionals and couples planning kids by serving as a predictor of hereditary prenatal disease. The system will also assist medical professionals in analysing complex cases (where patients lacked confidence in their medical history and family history of disorders) so that inheritance of haematological disorders can be predicted and appropriate decisions can be made regarding pregnancy (such as whether to continue the pregnancy or not), as these disorders are challenging to treat and manage. The user interface of developed system is very easy to use. Beside this, it is very fast and less complex to know the prediction results. By estimating the likelihood that a child would inherit a bleeding illness from their parents at a prenatal stage, this application functions like an e-Health system that offers healthcare services to remote public in the context of hereditary prenatal disorders (such as Haemophilia and Thalassaemia). In this way, it also offers the general public services like genetic counselling, premarital counselling, and preconception counselling to those who are less aware of and concerned about inherited diseases (like Haemophilia and thalassemia) and are unable to use these services due to a lack of facilities (especially in remote areas) or personnel (such as genetic counsellors) associated with them. Medical science does, however, offer certain inheritance patterns that may be used to determine if a bleeding problem is passed down from parents to children or not depending on the expecting couple's medical histories. However, they don't have any provisions in place for situations where one of them is unsure whether a bleeding condition exists. Hence, the application covers such kind of cases by extending the existing inheritance pattern. This application can be used by medical professional to speed their task as the task of calculating the probability of inheritance of haematological disorders from parents to offspring's takes too much time. Secondly, this application can

be utilized by general public as genetic/ pre-marital/preconception counsellors as it help the expected couple to take informed decision regarding the pregnancy or even regarding the marriage based of their medical, family and obstetric history. It can also be used in teaching learning process of medical domain.

Chapter 7

Conclusion and Future Scope

The structure of this chapter is as follows: Section 7.1 summarises the work covered in the thesis. Section 7.2 then highlights some potential topics for future study.

7.1 Summary of Deductions

The purpose of this thesis was to investigate and develop a mobile application-based e-Health system for the early diagnosis of inherited prenatal disorders using a newly developed hybrid diagnosis model and a rule-based prediction model that was developed during the conduct of the current research study. The process of creating a hybrid diagnostic model includes creating an individual model (such as a Bayesian network model or a KNN model), as well as a hybrid model that combines a decision tree and Gaussian Naive Bayes to identify hereditary bleeding disorders; whereas rule set based prediction model utilizes different rules developed during the current study based on existing inheritance pattern to predict the probability of inheriting the bleeding disorder such as Haemophilia and Thalassemia by offspring. A significant change in the discipline of healthcare has resulted from the implementation of the proposed e-Health system with the use of a hybrid diagnosis model and a rule-based predictor model created during ongoing research. This is because it helps determine the likelihood of having a child with

Thalassemia and Haemophilia, which take a long time to diagnose. Besides this, it also provides electronic services to the remote public for their self-diagnosis regarding inherited bleeding disorders; so that they can take informed decisions under the supervision of an expert doctor; if they find any kind of probability of inheritance of Haemophilia and Thalassemia after the utilization of the mobile-based e-Health system. In this research work, the collection of data set (associated with Haemophilia and Thalassemia), development of the hybrid model, development of inherited prenatal disorders predictor model, development of new rules based on existing inheritance patterns and development of mobile-based application has been conducted for the diagnosis of Haemophilia and Thalassemia as well as to assess the chances of inheritance of bleeding disorders such as Haemophilia and Thalassemia by offspring's at prenatal stage; which explained in form of chapters. The fundamental study related to e-Health system, inherited disorders, Haemophilia, and Thalassemia has been discussed in detail in Chapter 1.

Chapter 2 presented a comprehensive literature review on e-Health systems developed and associated with different domains of the healthcare sector. Different types of inherited disorders identified during the literature review have been mentioned in this chapter. Besides this, names of the selected inherited bleeding disorder such as Haemophilia and Thalassemia have been specified. Furthermore, the research gap identified after the review of the literature, the problem statement formulated, objectives, and proposed methodology related to current research work have been mentioned in this chapter.

Chapter 3 discusses the proposed research methodology in detail; which comprises of research approach, research strategy, time horizon, Sampling Strategy, data collection methods, data analysis methods, research process, and proposed work.

Chapter 4 discusses the development and implementation of individual models, a hybrid model, and performance evaluation of these models in detail. The chapter started with brief introduction of proposed disorders and e-Health system. After this, a description of the methodology has been given along with a graphical flow

chart which is utilized for the development of proposed models. A description of the data set and feature selection criteria has also been discussed in this chapter. Besides this, the development of two individual models (implemented through classification algorithms such as the Bayesian network algorithm and KNN algorithm for the diagnosis of Haemophilia and Thalassemia) has been discussed in this chapter. At the start of implementation, a Bayesian algorithm was utilized for the development of the first diagnosis model. This model estimates the conditional probabilities required to specify the network with the utilization of the collected dataset and maximum likelihood estimation or bayesian inference technique. After constructing the network, it has been performed probabilistic inference on new patient records to determine the probability that each patient has Haemophilia or Thalassemia, based on their given symptoms. This would involve computing the posterior distribution over the relevant variables given the evidence provided by the patient record. A Bayesian network was applied to a dataset consisting of 501 records and 13 columns extracted as the most prominent features during the process of feature selection and achieved an accuracy of 80%. To improve the accuracy rate, another model has been developed with the utilization of the KNN algorithm and produced an accuracy of 81%, which is higher than that of the Bayesian network. In this way, an increase of 1% has been observed after the utilization of KNN for the proposed diagnosis model. But require more improvements. Hence, the development of an integrated or hybrid model has been done for further improvement. For this, decision trees with the Gaussian Naive Bayes technique have been integrated for the diagnosis of the disorders namely Haemophilia and Thalassemia. The hybrid approach that combines decision trees with Gaussian Naive Bayes can be a powerful technique for disease detection, as it can take advantage of the strengths of both methods while mitigating their weaknesses. Hence, the hybrid model utilizes the decision tree to split the data into smaller subgroups based on the most informative features and then apply GNB to each subgroup to calculate the class probabilities. This approach can reduce

the risk of over fitting and handle correlated features by splitting the data into smaller, more homogeneous subgroups. Moreover, since GNB is less affected by irrelevant features, the Decision tree can include a wider range of features without negatively impacting the performance. In this way, the hybrid approach achieved a higher accuracy rate of 85% as compared to the Bayesian network and KNN. The current study utilizes the confusion matrix for the evaluation of a newly developed classification model.

Furthermore, an evaluation of the performance of classification models has been depicted in the chapter. For this, the Confusion matrix has been used to assess the effectiveness of models developed during the current research work by comparing its predicted labels against the true labels of the data set. The matrix shows the number of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) for each class. Besides this, the chapter also depicted the comparison of the accuracy of individual models and hybrid models; and depicted that the hybrid model has the highest accuracy among the models developed during the current study; when comparing their respective levels of precision. Additionally, during the current research effort and as stated in this chapter, precision, recall, and F1 score have been used for the comparison of produced models. Statistical analysis with the utilization of the ROC (Receiver Operating Characteristic) curve has also been presented in this chapter. The efficiency of models has been assessed through the cross-validation method and discussed under the heading of hypothesis testing of this chapter; which depicted that the k-mean accuracy of the hybrid model was 72.3%, according to the findings of the repeated stratified k-fold cross-validation procedure, which is greater than the k-mean accuracy of the KNN model, which was 56.1%. After this, the execution and outcome of the newly developed hybrid diagnosis model have been presented. This section depicts the working of a newly developed hybrid diagnosis model that diagnosis the disorders (namely Haemophilia and Thalassemia) by taking the symptoms as input from an

individual or patient. Lastly, a Rule Set Based Inherited Prenatal Disorder Predictor (IPDP) Model has been discussed in detail along with the implementation and outcome of all the rules developed during the current research. Besides this, the existing inheritance pattern has been depicted in the chapter.
Chater 5 discusses the rules based prediction model in detail along with the implementation and outcome of all the rules developed during the current research. Chapter 6 also discusses the design and implementation details of the proposed mobile application developed during current research work along with screen shots of prediction results generated after taking input from users. The chapter also discuss the evaluation of innovative mobile phone based e-Health system and the process of including this system in real life treatment protocol.
Chapter 7 discusses the conclusion and finding of the current thesis. There are also some new directions for in investigation, which are depicted in this chapter briefly.

In conclusion, it can be said that each e-Health system has been utilised to serve remote patients in the hour of emergency based on the type of disease and domain. By overcoming resource constraints, the development and implementation of e-Health systems using ICT increases the efficiency and effectiveness of healthcare services. There are several systems have been developed for the diagnosis and management of different diseases. In this context, current research work provides a model and a mobile application-based e-Health system as the utility of research work which predicts the probability of inheriting the bleeding disorders such as Haemophilia and Thalassemia at the prenatal stage, electronically. Any remote patients can check the chance of having a child with Haemophilia and Thalassemia based on their medical, family, physiological symptoms, and obstetric history at a fundamental level and provide the list of expert doctors for further consultation. In this way, current work will help in controlling the prevalence of Haemophilia and Thalassemia by providing early diagnosis methods. Additionally, the system will help the consultant doctor in predicting the likelihood that Haemophilia and

Thalassemia would be inherited by offering an automatic calculation approach since a human calculation would take too long to make the same prediction.

7.2 Future Scope of Work

To assist the underserved population in distant areas, the e-Health system in healthcare is an active research topic. Using information and communication technology and machine learning techniques, dynamic e-Health systems have been continuously created for the diagnosis, prognosis, and treatment of diseases. In light of this, the present study created an e-Health system based on mobile applications for the identification of hereditary prenatal illnesses that forecast the likelihood of having a kid with Haemophilia and thalassemia. Likewise, the current proposed systems in this work can also be extended for further research work. The different significant future directions are discussed in the following section:

1. The current e-Health system discussed in this work can also be extended for the prediction of other bleeding disorders such as sickle cell anemia

2. The proposed system can also be extended for the management or treatment of proposed inherited bleeding disorders with large databases to enhance the system's learning capability and robustness, which will improve the system's resilience and capacity for learning.

3. In addition to this, the current study might be utilized to create a premarital health compatibility analyzer e-Health system that would forecast the amount of medical compatibility between couples to limit the likelihood of having children with inherited illnesses.

4. The current system predicts the likelihood of situations where a couple has prior knowledge of an inherited disorder's medical history and situations

where one partner is unsure of their medical history while the other is aware of it. However, it does not account for situations where both partners are unsure of their medical history of purposeful disorders, which will be the subject of future research.

REFERENCES

[1] G. Eysenbach, A. R. Jadad *et al.*, "Evidence-based patient choice and consumer health informatics in the internet age," *Journal of medical Internet research*, vol. 3, no. 2, p. e841, 2001.

[2] J. Flower and P. Guillaume, "E-health: eight power factors, three scenarios." in *Health Forum Journal*, vol. 44, no. 1, 2001, pp. 12–6.

[3] P. Churi, A. Pawar, and A.-J. Moreno-Guerrero, "A comprehensive survey on data utility and privacy: Taking indian healthcare system as a potential case study," *Inventions*, vol. 6, no. 3, p. 45, 2021.

[4] K. Kuhn *et al.*, "E-healthcare in india: critical success factors for sustainable health systems," in *MEDINFO 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics*, vol. 129. IOS Press, 2007, p. 257.

[5] M. A. Curran and K. E. Curran, "The e-health revolution: Competitive options for nurse practitioners as local providers," *Journal of the American Academy of Nurse Practitioners*, vol. 17, no. 12, pp. 495–498, 2005.

[6] A. Banbury, A. Roots, and S. Nancarrow, "Rapid review of applications of e-health and remote monitoring for rural residents," *Australian Journal of Rural Health*, vol. 22, no. 5, pp. 211–222, 2014.

[7] T. Spil and R. W. Schuring, *E-Health Systems Diffusion and Use: The Innovation, the User and the Use IT Model: The Innovation, the User and the Use IT Model*. IGI Global, 2005.

[8] A. T. Simpson, C. Doarn, and S. Garber, "A brief history of nasa's contributions to telemedicine," *NASA https://www. nasa. gov/content/a-brief-history-of-nasa-s-contributions-totelemedicine/# _edn1 (Accessed December 20 2021)*, 2013.

[9] N. F. Güler and E. D. Übeyli, "Theory and applications of telemedicine," *Journal of medical systems*, vol. 26, pp. 199–220, 2002.

[10] N. R. Armfield, S. K. Edirippulige, N. Bradford, and A. C. Smith, "Telemedicine–is the cart being put before the horse," *Med J Aust*, vol. 200, no. 9, pp. 530–533, 2014.

[11] W. R. Hersh, D. H. Hickam, S. M. Severance, T. L. Dana, K. P. Krages, and M. Helfand, "Diagnosis, access and outcomes: Update of a systematic review of telemedicine services," *Journal of telemedicine and telecare*, vol. 12, no. 2_suppl, pp. 3–31, 2006.

[12] N. Van den Berg, M. Schumann, K. Kraft, and W. Hoffmann, "Telemedicine and telecare for older patients—a systematic review," *Maturitas*, vol. 73, no. 2, pp. 94–114, 2012.

[13] A. C. Smith and L. C. Gray, "Telemedicine across the ages," *Medical journal of Australia*, vol. 190, no. 1, pp. 15–19, 2009.

[14] A. K. Sharma, "Telemedicine: A technology in waiting," *Indian Journal of Public Health*, vol. 55, no. 4, pp. 286–288, 2011.

[15] V.-n. Ryan, S. Stathis, A. C. Smith, D. Best, and R. Wootton, "Telemedicine for rural and remote child and youth mental health services," *Journal of Telemedicine and Telecare*, vol. 11, no. 2_suppl, pp. 76–78, 2005.

[16] I. Morrison, "The rise and fall and rise of e-health," in *Health Forum Journal*, vol. 44, no. 1. Health Forum, 2001, pp. 48–48.

[17] D. J. Antonacci, R. M. Bloch, S. A. Saeed, Y. Yildirim, and J. Talley, "Empirical evidence on the use and effectiveness of telepsychiatry via videoconferencing: implications for forensic and correctional psychiatry," *Behavioral sciences & the law*, vol. 26, no. 3, pp. 253–269, 2008.

[18] G. Elliott, A. C. Smith, M. E. Bensink, C. Brown, C. Stewart, C. Perry, and P. Scuffham, "The feasibility of a community-based mobile telehealth screening service for aboriginal and torres strait islander children in australia," *Telemedicine and e-Health*, vol. 16, no. 9, pp. 950–956, 2010.

[19] S. K. Mishra, L. Kapoor, and I. P. Singh, "Telemedicine in india: current scenario and the future," *Telemedicine and e-Health*, vol. 15, no. 6, pp. 568–575, 2009.

[20] C. Cartwright, R. Wade, and K. Shaw, *The Impact of Telehealth and Telecare on Clients of the Transition Care Program (TCP).* Southern Cross University-Aged Services Learning & Research Collaboration, 2011.

[21] R. K. Das, M. R. Patra, and S. C. Mahapatra, "e-grama: a tool for bridging the digital divide in rural india," in *Proceedings of the 2nd international conference on Theory and practice of electronic governance*, 2008, pp. 361–366.

[22] M. Dongier, R. Tempier, M. Lalinec-Michaud, and D. Meunier, "Telepsychiatry: psychiatric consultation through two-way television. a controlled study." *The Canadian Journal of Psychiatry*, vol. 31, no. 1, pp. 32–34, 1986.

[23] J. Greenwood and R. Williams, "Continuing professional development for australian rural psychiatrists by videoconference," *Australasian Psychiatry*, vol. 16, no. 4, pp. 273–276, 2008.

[24] E. Saurman, D. Perkins, R. Roberts, A. Roberts, M. Patfield, and D. Lyle, "Responding to mental health emergencies: implementation of an innovative telehealth service in rural and remote new south wales, australia," *Journal of emergency nursing*, vol. 37, no. 5, pp. 453–459, 2011.

[25] A. C. Smith, N. R. Armfield, W.-I. Wu, C. A. Brown, and C. Perry, "A mobile telemedicine-enabled ear screening service for indigenous children in queensland: activity and outcomes in the first three years," *Journal of Telemedicine and Telecare*, vol. 18, no. 8, pp. 485–489, 2012.

[26] S. Soegijoko, "Application specific e-health & telemedicine systems: Implementation experience for community healthcare and systematic review of disaster publications," in *2011 2nd International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*. IEEE, 2011, pp. 415–416.

[27] M. T. Karthikeya, M. N. Reddy, M. S. Kumar, and M. H. Yogendra, "e-sanjeevani-telemedicine service," 2022.

[28] M. Heitner and B. Fotsch, "Is there a doctor on the web? mark heitner and bill fotsch, medem creating a far-reaching e-health network can improve physician-patient communications," *Pharmaceutical Executive*, vol. 21, no. 6; SUPP, pp. 18–21, 2001.

[29] H. Larkin, "Build your own information highway," *Medical Economics*, vol. 77, no. 15, pp. 65–65, 2000.

[30] J. Everett and D. Kerr, "Telehealth as adjunctive therapy in insulin pump treated patients: a pilot study," *Practical Diabetes International*, vol. 27, no. 1, pp. 9–10i, 2010.

[31] N. Swartz, "Electronic medical records' risks feared," *Information Management Journal*, vol. 39, no. 3, pp. 9–10, 2005.

[32] D. W. Bates, "Physicians and ambulatory electronic health records," *Health affairs*, vol. 24, no. 5, pp. 1180–1189, 2005.

[33] R. Carol, "Ehrs, the doctor will see you now." *Journal of AHIMA*, vol. 76, no. 4, pp. 24–28, 2005.

[34] L. B. Villa and I. Cabezas, "A review on usability features for designing electronic health records," in *2014 IEEE 16th international conference on e-health networking, applications and services (Healthcom)*. IEEE, 2014, pp. 49–54.

[35] I. de la Torre, B. Martínez, and M. López-Coronado, "Analyzing open-source and commercial ehr solutions from an international perspective," in *2013 IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*. IEEE, 2013, pp. 399–403.

[36] D. B. Hier, A. Rothschild, A. LeMaistre, and J. Keeler, "Differing faculty and housestaff acceptance of an electronic health record," *International journal of medical informatics*, vol. 74, no. 7-8, pp. 657–662, 2005.

[37] B. Kelly, "Crafting a web site game plan." *Health Data Management*, vol. 8, no. 9, pp. 42–4, 2000.

[38] J. Li, L. P. W. Land, P. Ray, and S. Chattopadhyaya, "E-health readiness framework from electronic health records perspective," *International Journal of Internet and Enterprise Management*, vol. 6, no. 4, pp. 326–348, 2010.

[39] R. H. Miller, C. West, T. M. Brown, I. Sim, and C. Ganchoff, "The value of electronic health records in solo or small group practices," *Health Affairs*, vol. 24, no. 5, pp. 1127–1137, 2005.

[40] M. Thakkar and D. C. Davis, "Risks, barriers, and benefits of ehr systems: a comparative study based on size of hospital," *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, vol. 3, 2006.

[41] D. Grandinetti, "Patients are putting their records online," *Medical Economics*, vol. 77, no. 7, pp. 93–93, 2000.

[42] R. J. Cline and K. M. Haynes, "Consumer health information seeking on the internet: the state of the art," *Health education research*, vol. 16, no. 6, pp. 671–692, 2001.

[43] N. Szecket, H. J. Wong, R. C. Wu, H. D. Berman, and D. Morra, "Implementation of a continuous admission model reduces the length of stay of patients on an internal medicine clinical teaching unit," *Journal of Hospital Medicine*, vol. 7, no. 1, pp. 55–59, 2012.

[44] A. H. Association *et al.*, "Forward momentum: hospital use of health information technology," 2005.

[45] J. DeLuca and O. Doyle, *Health Care Information Systems: An Executive's Guide for Successful Management*. American Hospital Publishing, Incorporated, 1991.

[46] P. G. Goldschmidt, "Hit and mis: implications of health information technology and medical information systems," *Communications of the ACM*, vol. 48, no. 10, pp. 68–74, 2005.

[47] R. Feijo and L. Costa, "Virtual health library prevention and cancer control: providing reliable information," in *ASIA-PACIFIC JOURNAL OF CLINICAL ONCOLOGY*, vol. 10. WILEY-BLACKWELL 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2014, pp. 1–1.

[48] T. W. Burkholder, J. W. Bellows, and R. A. King, "Free open access medical education (foam) in emergency medicine: the global distribution of users in 2016," *Western Journal of Emergency Medicine*, vol. 19, no. 3, p. 600, 2018.

[49] J. Ross, C. Chapman, C. Murray, M. Stevenson, D. Natin, and K. Rogstad, "How much interest is the internet to patients?" *Sexually Transmitted Infections*, vol. 76, no. 5, p. 393, 2000.

[50] M. J. Dutta, R. Pfister, and C. Kosmoski, "Consumer evaluation of genetic information online: The role of quality on attitude and behavioral intentions," *Journal of Computer-Mediated Communication*, vol. 15, no. 4, pp. 592–605, 2010.

[51] J. Avalon, P. Hayward, V. Lane, and J. Snaith, "E-health modelling for decision support: Can technology turn us all into experts?" in *2010 IEEE 23rd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2010, pp. 432–437.

[52] H. Ranjbartabar, A. Maddah, and M. Kavakli, "emedicalhelp: A customized medical diagnostic application: Is a single questionnaire enough to measure stress? in 2015 ieee/acis 14th international conference on computer and information science (icis)(pp. 367–372)," 2015.

[53] P. Sharma and A. Sharma, "Collection and analysis of patients dataset to develop an e health system for the prediction of inherited prenatal disorders," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 12, pp. 5118–5121, 2019.

[54] P. Sharma, T. R. Shivaram, and A. Sharma, "A methodical review of e-health systems developed for indian healthcare sector," *Indian Journal of Science and Technology*, vol. 9, no. 44, pp. 1–6, 2016.

[55] A. Deorari, "Standard treatment protocol (stp)," https://play.google.com/store/apps/details?id=who.searo.stpapps, 2014, accessed on 15-Jan-2016.

[56] P. Sharma and A. Sharma, "e-health systems for mother and child care domain: A systematic review," in *Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1*, 2022, pp. 439–447.

[57] A. I. I. of Medical Sciences, "Neonatology," https://play.google.com/store/apps/details?id=com.itdcs.neonatology, accessed on 19-Jan-2016.

[58] Newbornwhoc. Accessed on 3-Sep-2015. [Online]. Available: https://www.newbornwhocc.org/smart_phon.html

[59] Inherited diseases, hereditary genetics: Current research. Accessed on 23-Mar-2023. [Online]. Available: https://www.walshmedicalmedia.com/scholarly/inherited-diseases-journals-articles-ppts-list-385.html

[60] A. Muhammad, "Ob insulin," https://itunes.apple.com/us/app/wusm-ob-insulin/id829933660?mt=8, accessed on 6-Oct-2015.

[61] Inherited disorder. Accessed on 12-Mar-2021. [Online]. Available: https://medical_dictionary.freedictionary.com

[62] E. Marchesini, M. Morfini, and L. Valentino, "Recent advances in the treatment of hemophilia: A review," *Biologics: Targets and Therapy*, pp. 221–235, 2021.

[63] E. Twain, "Genetic disorder haemophilia: Its diagnosis and management," *Hereditary Genet.*, vol. 11, p. 229, 2022.

[64] S. Sarmiento Doncel, G. A. Díaz Mosquera, J. M. Cortes, C. Agudelo Rico, F. J. Meza Cadavid, and R. G. Peláez, "Haemophilia a: A review of clinical manifestations, treatment, mutations, and the development of inhibitors," *Hematol. Rep.*, vol. 15, pp. 130–150, 2023.

[65] R. C. Ljung, "Prenatal diagnosis of haemophilia," *Baillière's Clinical Haematology*, vol. 9, no. 2, pp. 243–257, 1996.

[66] Diagnosis of haemophilia. Accessed on 16-May-2021. [Online]. Available: https://www.cdc.gov/ncbddd/hemophilia/diagnosis.html

[67] How haemophilia is inherited? Accessed on 17-May-2021. [Online]. Available: https://www.hog.org/handbook/section/2/how-hemophilia-is-inherited

[68] How haemophilia is inherited? Accessed on 17-May-2021. [Online]. Available: https://www.cdc.gov/ncbddd/hemophilia/course/Hemophilia_Patterns_v3.pdf

[69] E. Higgs, K. E. Wain, J. Wynn, M. T. Cho, S. Higgins, D. Blaisdell, D. Dugan, S. Valek, and S. Cohen, "Measuring quality and value in genetic counseling: The current landscape and future directions," *Journal of Genetic Counseling*, 2022.

[70] Genetic counselling. Accessed on 05-Apr-2023. [Online]. Available: https://nbiacure.org/learn/genetic-counseling/

[71] M. Abacan, L. Alsubaie, K. Barlow-Stewart, B. Caanen, C. Cordier, E. Courtney, E. Davoine, J. Edwards, N. J. Elackatt, K. Gardiner, and Y. Guan, "The global state of the genetic counseling profession," *European Journal of Human Genetics*, vol. 27, no. 2, pp. 183–197, 2019.

[72] K. A. Schneider, A. Chittenden, and K. M. Shannon, *Counseling about Cancer: Strategies for Genetic Counseling.* John Wiley and Sons, 2023.

[73] Z. Qasim, N. Asif, and K. Hassan, "Pattern of bleeding in haemophilia patients," *J Islam Med Dent Coll*, vol. 3, pp. 3–6, 2014.

[74] M. S. Hossain, M. M. Hasan, E. Raheem, and et al., "Lack of knowledge and misperceptions about thalassemia among college students in bangladesh: A cross-sectional baseline study," *Orphanet J Rare Dis*, vol. 15, p. 54, 2020.

[75] R. Gupta and G. BM, "Global thalassemia research: A scientometric assessment of publications output during 2008-17," *Open Acc Blood Res Trans J*, vol. 2, no. 4, p. 555594, 2018.

[76] F. Shafique, S. Ali, T. Almansouri, F. Van Eeden, N. Shafi, M. Khalid, S. Khawaja, and S. Andleeb, "Thalassemia, a human blood disorder," *Brazilian Journal of Biology*, vol. 83, 2021.

[77] Genetics of thalassemia. Accessed on 07-Apr-2023. [Online]. Available: https://thalassemia.com/geneticsinheritance.aspx#gsc.tab=0

[78] Symptoms and diagnosis. Accessed on 07-Apr-2020. [Online]. Available: https://thalassemia.com/diagnosis.aspx#diagnosis&gsc.tab=0

[79] N. Mobarra, M. Shanaki, H. Ehteram, H. Nasiri, M. Sahmani, M. Saeidi, and M. Azad, "A review on iron chelators in treatment of iron overload syndromes," in *International Journal of Hematology-Oncology and Stem Cell Research*, 2016, vol. 10, no. 4, p. 239.

[80] H. L. Muncie Jr and J. S. Campbell, "Alpha and beta thalassemia," *American Family Physician*, vol. 80, no. 4, pp. 339–344, 2009.

[81] Thalassemia genetic prevalence. Accessed on 07-Apr-2023. [Online]. Available: https://www.newsmedical.net/health/Thalassemia-Genetic-Prevalence.aspx

[82] S. Russell and P. Norvig, "Ai: A modern approach," *Learning*, vol. 2, no. 3, p. 4, 2005.

[83] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[84] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[85] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. Springer, 2006, vol. 4, no. 4.

[86] Z. Obermeyer and E. J. Emanuel, "Predicting the future—big data, machine learning, and clinical medicine," *N Engl J Med*, vol. 375, no. 13, pp. 1216–1219, 2016.

[87] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.

[88] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.

[89] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G. Z. Yang, "Deep learning for health informatics," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[90] G. Briganti, M. Scutari, and R. J. McNally, "A tutorial on bayesian networks for psychopathology researchers," *Psychological Methods*, 2022.

[91] W. Xing and Y. Bei, "Medical health big data classification based on knn classification algorithm," *IEEE Access*, vol. 8, pp. 28 808–28 819, 2019.

[92] B. N. Lakshmi, T. S. Indumathi, and N. Ravi, "An hybrid approach for prediction based health monitoring in pregnant women," *Procedia Technology*, vol. 24, pp. 1635–1642, 2016.

[93] E-health. Accessed on 4-Sep-2015. [Online]. Available: https://www.who.int/topics/ehealth/en/

[94] J. W. Creswell, "Research designs. qualitative, quantitative, and mixed methods approaches," 2009.

[95] A. Raza, F. Rustam, H. U. R. Siddiqui, I. d. l. T. Diez, B. Garcia-Zapirain, E. Lee, and I. Ashraf, "Predicting genetic disorder and types of disorder using chain classifier approach," *Genes*, vol. 14, no. 1, p. 71, 2022.

[96] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," *Artificial Intelligence Review*, vol. 26, pp. 159–190, 2006.

[97] M. A. Curran and K. E. Curran, "The e-health revolution: competitive options for nurse practitioners as local providers," *Journal of the American Association of Nurse Practitioners*, vol. 17, no. 12, pp. 495–498, 2005.

[98] I. T. Network. (2015) Teleconsultation. Accessed on 15-Oct-2015. [Online]. Available: http://stelemedindia.org/teleconsultation.html

[99] (2013) e-sanjeevani. Center for Development of Advanced Computing (CDAC). Accessed on 15-09-2015. [Online]. Available: https://www.esanjeevani.in/aboutus.aspx

[100] Accenture, "Delivering e-healthcare in india," in *National Conference on e-Governance*, Kochi, 2014, pp. 1–28.

[101] Online registration system (ors). National Informatics Centre, Government of India. Accessed on 1-Jan-2016. [Online]. Available: https://ors.gov.in

[102] Healthyyou. MediIT Health Solutions. Accessed on 6-Oct-2015. [Online]. Available: https://play.google.com/store/apps/details?id=in.mediit.android.healthyyou&hl=en

[103] (2015) Practo. Accessed on 6-Oct-2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.practo.fabric&hl=en

[104] M. Thakkar and D. C. Davis, "Risks, barriers, and benefits of ehr systems: A comparative study based on size of hospital," *Perspectives in Health Information Management*, vol. 3, no. 5, pp. 1–19, 2006.

[105] (2008) 21st century's healthnet in goa. Express Healthcare. Accessed on 09-Sep-2015. [Online]. Available: https://www.21ci.com/pdfs/published_news/ExpressHealthcare-Jan2008.pdf

[106] M. R. Institute, "Medical records institute's seventh annual survey of electronic health record trends and usage for 2005," Tech. Rep., 2005.

[107] Microsoft. (2015) Health vault. Accessed on 5-Jan-2016. [Online]. Available: https://www.healthvault.com

[108] (2015) Health memo. Notch Media. Accessed on 6-Oct-2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.notchmag.healthmemo

[109] (2015) Health pie. mTatva mHealth. Accessed on 6-Oct-2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.mtatva.healthpie

[110] R. Sengupta, R. Foundation, and M. Arogya. (2014) Kolkata medical emergency (kmes). Accessed on 3-Oct-2014. [Online]. Available: https://play.google.com/store/apps/details?id=in.kmes.androidapp

[111] (2014) Malaria early epidemic detection system (meeds). National Health Portal. Accessed on 3-Oct-2015. [Online]. Available: https://www.nhp.gov.in/mobile-apps

[112] P. Tandon, S. Maheshwari, G. Agarwal, and A. Jain. (2015) 1mg. Accessed on 6-Oct-2015. [Online]. Available: https://www.1mg.com/aboutUs

[113] (2015) Webmd. Accessed on 6-Oct-2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.webmd.android&hl=en

[114] (2015) 5minuteconsult. Wolters Kluwer Health. Accessed on 6-Oct-2015. [Online]. Available: https://5minuteconsult.com/collectioncontent/3-197794/patient-handouts/abdominal-hernia-repair-discharge-instructions

[115] (2015) Nhp-health directory services. National Health Portal. Accessed on 3-Oct-2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.nhp.bloodbank&hl=en

[116] (2015) Healthcentral.com. Remedy Health Media. Accessed on 7-Jan-2016. [Online]. Available: https://www.healthcentral.com

[117] (2015) Symptom checker. Isabel Health Care. Retrieved on 12-Jan-2016. [Online]. Available: https://symptomchecker.isabelhealthcare.com/

[118] (2015) Veegilo. Innovative Support to Emergencies Diseases and Disasters (InSTEDD). Accessed on 15-Jan-2016. [Online]. Available: https://instedd.org/technologies/veegilo/

[119] Medscape. (2016) emedicine. Accessed on 8-Jan-2016. [Online]. Available: https://emedicine.medscape.com/

[120] (2011) Geo chat. Innovative Support to Emergencies Diseases and Disasters (InSTEDD). Accessed on 6-Oct-2015. [Online]. Available: https://instedd.org/technologies/geochat/

[121] (2015) Mobile alliance for maternal action (mama). Washington D.C., USA. Accessed on 3-Oct-2015. [Online]. Available: https://www.mobilemamaalliance.org/

[122] A. Deorari, "Essential newborn care," https://play.google.com/store/apps/details?id=drdeorari.aiims.enc, accessed on 15-Jan-2016.

[123] A. I. I. of Medical Sciences and A. Deorari, "Aiims-who cc enbc," https://play.google.com/store/apps/details?id=drdeorari.aiims.enc, accessed on 3-Oct-2015.

[124] (2016) Kangaroo mother care initiative (kmc). Accessed on 02-Feb-2016. [Online]. Available: http://www.kmcindia.in/healthcare/fact_immediate.html

[125] J. Kaewkungwal, P. Singhasivanon, A. Khamsiriwatchara, S. Sawang, P. Meankaew, and A. Wechsart, "Application of smart phone in "better border healthcare program : a module for mother and child care," *BMC medical informatics and decision making*, vol. 10, no. 1, p. 1, 2010.

[126] (2011) Verboise. Innovative Support to Emergencies Diseases and Disasters (InSTEDD). Accessed on 3-Oct-2015. [Online]. Available: https://verboice.instedd.org/

[127] (2023) How to write the methodology. Accessed on 13-April-2023. [Online]. Available: https://gradcoach.com/how-to-write-the-methodology-chapter/

[128] G. Roumeliotis, "Athens as an international tourism destination: An empirical investigation to the city's imagery and the role of local dmo's," Ph.D. dissertation, Athens University and Business, 2014.

[129] D. Okaygoun, D. D. Oliveira, S. Soman, and R. Williams, "Advances in the management of haemophilia: Emerging treatments and their mechanisms," *Journal of Biomedical Science*, vol. 28, no. 1, pp. 1–13, 2021.

[130] (2022) Hemophilia carrier. Accessed on 16-05-2022. [Online]. Available: https://www.nationwidechildrens.org/conditions/hemophiliacarrier

Publications and Annexures

Journals

- Sharma, P., & Sharma, A. (2019), "Collection and Analysis of Patients Dataset to Develop an E Health System for the Prediction of Inherited Prenatal Disorders," Journal of Computational and Theoretical Nanoscience, 16(12), 5118-5121.

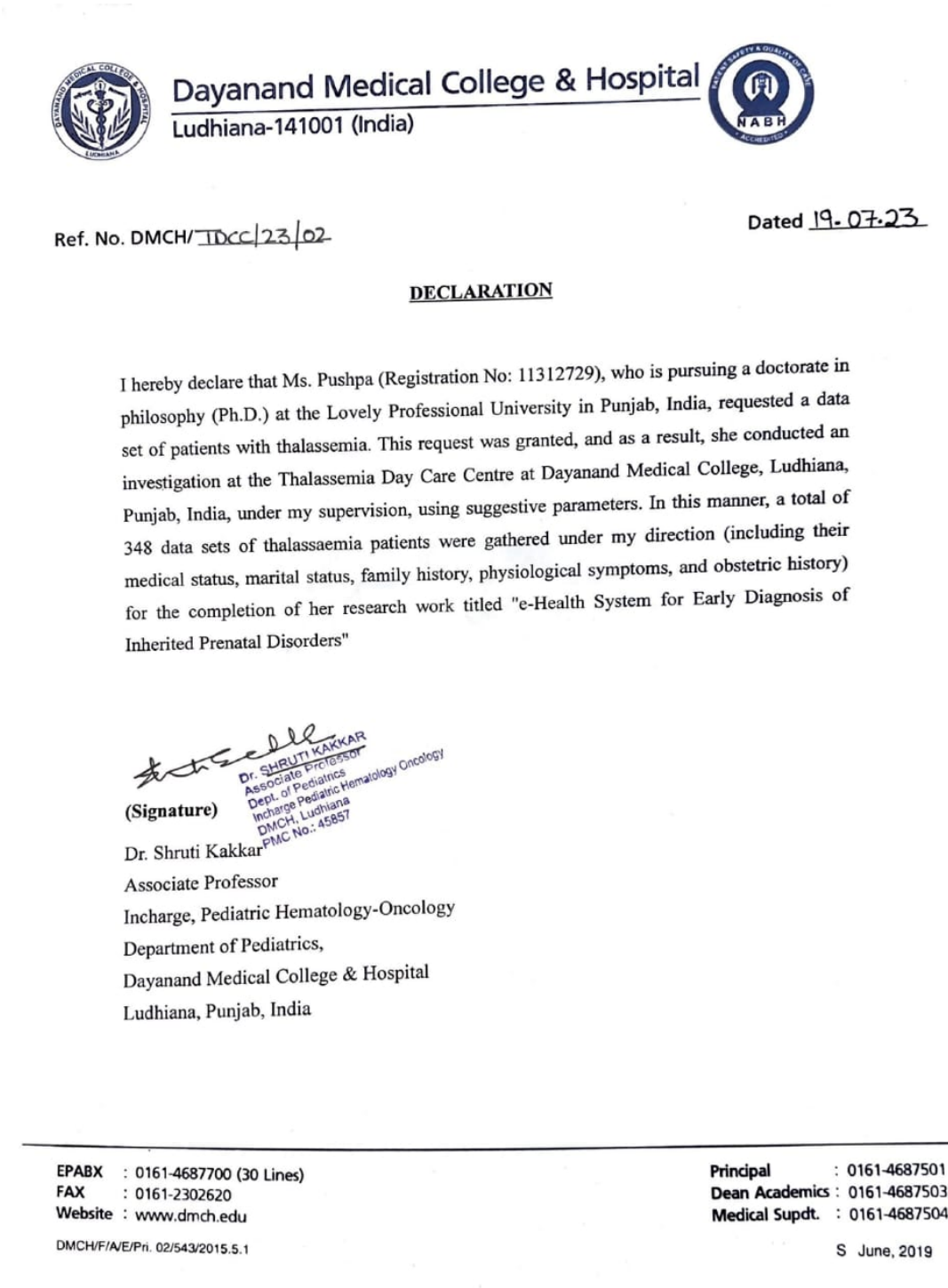
- Sharma, P., Sharma, A., & Kakkar,S. (2024), "A multi-layered framework for analysis and prediction of Haemophilia using machine learning techniques based e-health system.", Tuijin Jishu/Journal of Propulsion Technology, 45 (3), 2295-2307.

Conferences

- Sharma, P., Shivaram, T. R., & Sharma, A. (2016), "A methodical review of e-Health systems developed for Indian Healthcare Sector," Indian Journal of Science and Technology, 9(44), 1-6.

- Sharma, P., & Sharma, A. (2022), "e-Health Systems for Mother and Child Care Domain: A Systematic Review," Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1, 439-447.

Certificates



Dayanand Medical College & Hospital
Ludhiana-141001 (India)

Ref. No. DMCH/TDCC/23/02

Dated 19.07.23

DECLARATION

I hereby declare that Ms. Pushpa (Registration No: 11312729), who is pursuing a doctorate in philosophy (Ph.D.) at the Lovely Professional University in Punjab, India, requested a data set of patients with thalassemia. This request was granted, and as a result, she conducted an investigation at the Thalassemia Day Care Centre at Dayanand Medical College, Ludhiana, Punjab, India, under my supervision, using suggestive parameters. In this manner, a total of 348 data sets of thalassaemia patients were gathered under my direction (including their medical status, marital status, family history, physiological symptoms, and obstetric history) for the completion of her research work titled "e-Health System for Early Diagnosis of Inherited Prenatal Disorders"

Dr. SHRUTI KAKKAR
Associate Professor
Dept. of Pediatrics
Incharge Pediatric Hematology Oncology
DMCH, Ludhiana
PMC No.: 45857

(Signature)

Dr. Shruti Kakkar
Associate Professor
Incharge, Pediatric Hematology-Oncology
Department of Pediatrics,
Dayanand Medical College & Hospital
Ludhiana, Punjab, India

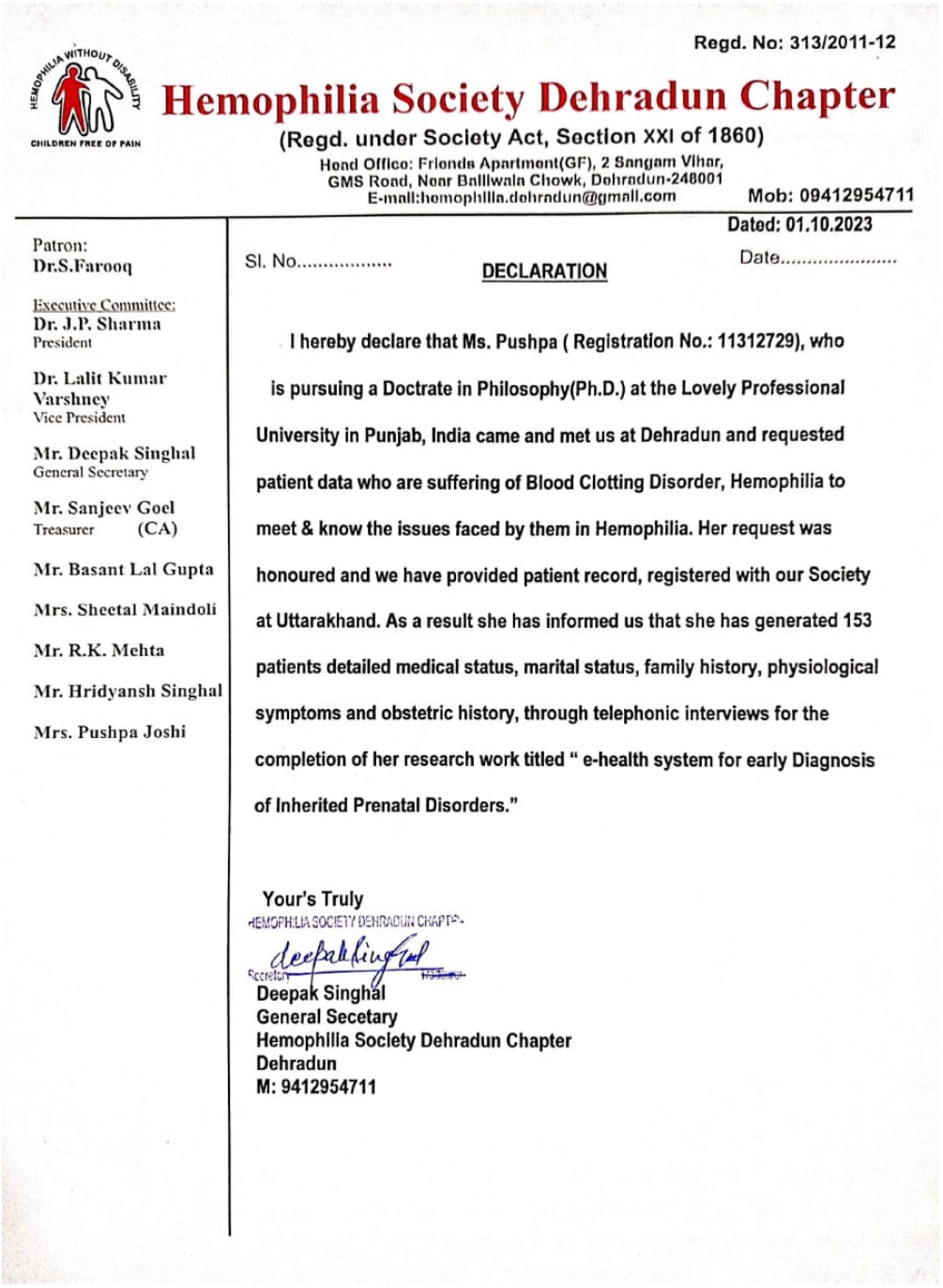
EPABX : 0161-4687700 (30 Lines)
FAX : 0161-2302620
Website : www.dmch.edu

Principal : 0161-4687501
Dean Academics : 0161-4687503
Medical Supdt. : 0161-4687504

DMCH/F/A/E/Pri. 02/543/2015.5.1

S June, 2019

FIGURE 7.1: Annexure 1



Regd. No: 313/2011-12

Hemophilia Society Dehradun Chapter

(Regd. under Society Act, Section XXI of 1860)

Head Office: Friends Apartment(GF), 2 Sangam Vihar,
GMS Road, Near Ballliwala Chowk, Dehradun-248001
E-mail:hemophilia.dehradun@gmail.com Mob: 09412954711

Dated: 01.10.2023

Patron:
Dr.S.Farooq

Executive Committee:
Dr. J.P. Sharma
President

Dr. Lalit Kumar Varshney
Vice President

Mr. Deepak Singhal
General Secretary

Mr. Sanjeev Goel
Treasurer (CA)

Mr. Basant Lal Gupta

Mrs. Sheetal Maindoli

Mr. R.K. Mehta

Mr. Hridyansh Singhal

Mrs. Pushpa Joshi

Sl. No.....

Date.....

DECLARATION

I hereby declare that Ms. Pushpa (Registration No.: 11312729), who is pursuing a Doctrate in Philosophy(Ph.D.) at the Lovely Professional University in Punjab, India came and met us at Dehradun and requested patient data who are suffering of Blood Clotting Disorder, Hemophilia to meet & know the issues faced by them in Hemophilia. Her request was honoured and we have provided patient record, registered with our Society at Uttarakhand. As a result she has informed us that she has generated 153 patients detailed medical status, marital status, family history, physiological symptoms and obstetric history, through telephonic interviews for the completion of her research work titled " e-health system for early Diagnosis of Inherited Prenatal Disorders."

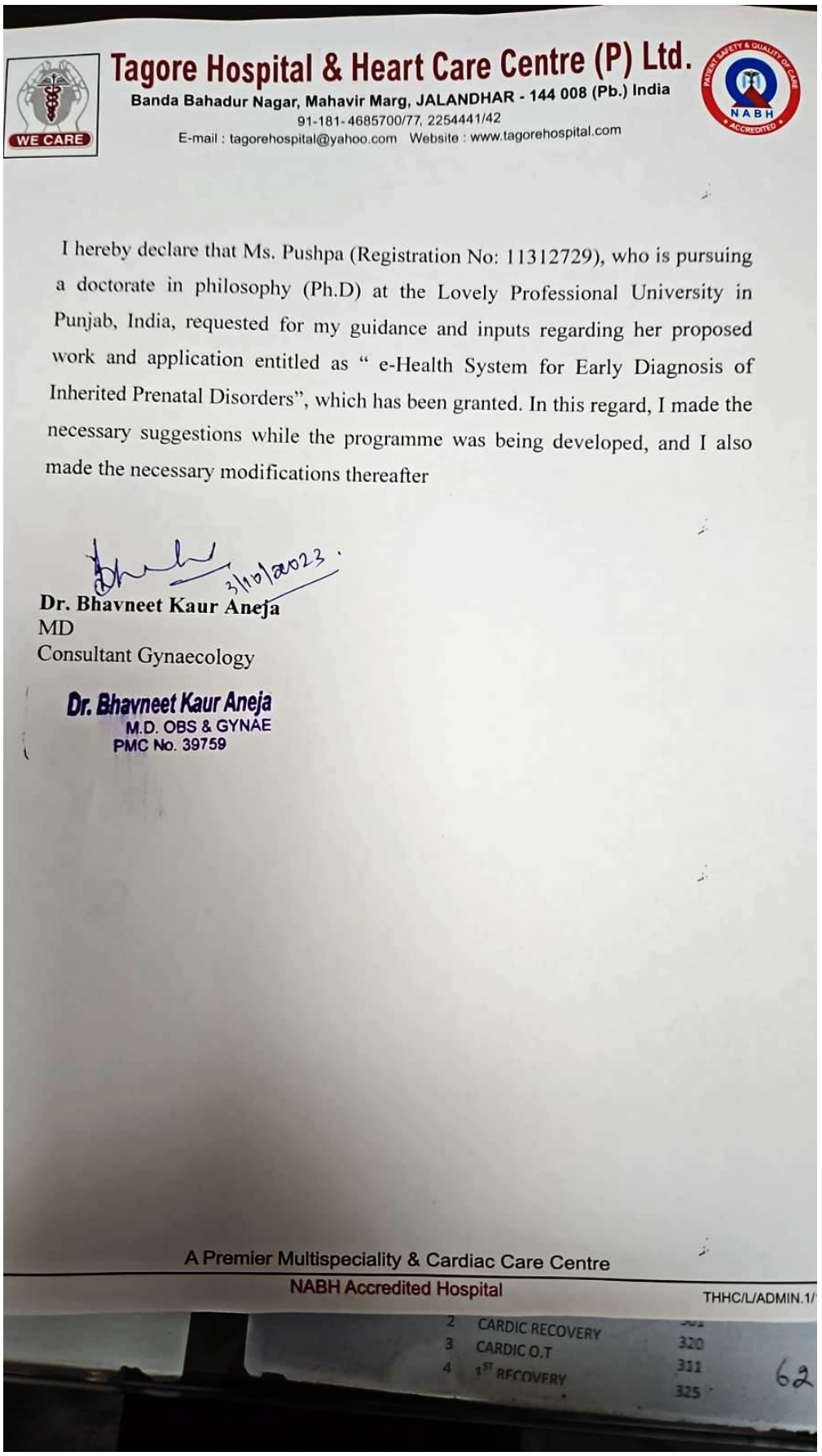
Your's Truly

HEMOPHILIA SOCIETY DEHRADUN CHAPTER

Secretary

Deepak Singhal
General Secetary
Hemophilia Society Dehradun Chapter
Dehradun
M: 9412954711

FIGURE 7.2: Annexure 2



Tagore Hospital & Heart Care Centre (P) Ltd.
Banda Bahadur Nagar, Mahavir Marg, JALANDHAR - 144 008 (Pb.) India
91-181-4685700/77, 2254441/42
E-mail : tagorehospital@yahoo.com Website : www.tagorehospital.com

I hereby declare that Ms. Pushpa (Registration No: 11312729), who is pursuing a doctorate in philosophy (Ph.D) at the Lovely Professional University in Punjab, India, requested for my guidance and inputs regarding her proposed work and application entitled as " e-Health System for Early Diagnosis of Inherited Prenatal Disorders", which has been granted. In this regard, I made the necessary suggestions while the programme was being developed, and I also made the necessary modifications thereafter

3/10/2023

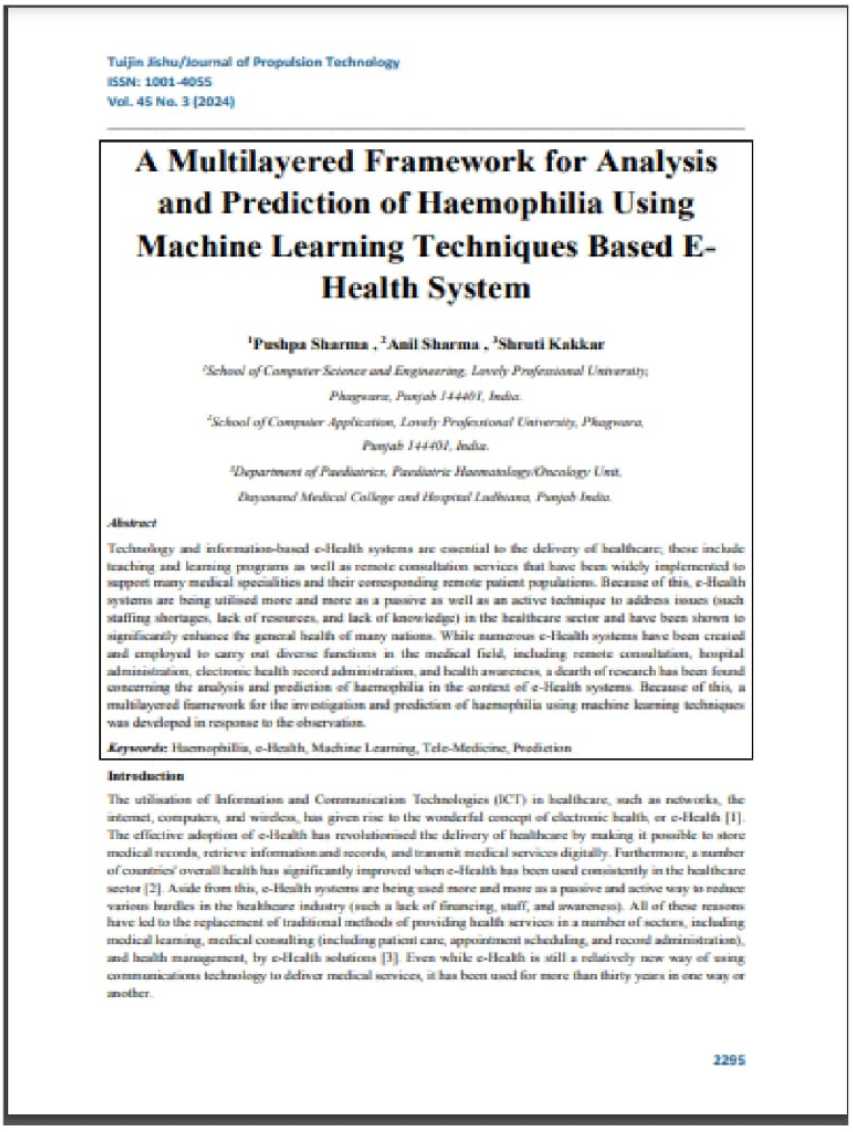
Dr. Bhavneet Kaur Aneja
MD
Consultant Gynaecology

Dr. Bhavneet Kaur Aneja
M.D. OBS & GYNAE
PMC No. 39759

A Premier Multispeciality & Cardiac Care Centre
NABH Accredited Hospital
THHC/L/ADMIN.1/

2 CARDIC RECOVERY 320
3 CARDIC O.T 311
4 1ST RECOVERY 325
62

FIGURE 7.3: Annexure 3



Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 3 (2024)

A Multilayered Framework for Analysis and Prediction of Haemophilia Using Machine Learning Techniques Based E-Health System

[1]Pushpa Sharma , [2]Anil Sharma , [3]Shruti Kakkar

[1]School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab 144401, India.

[2]School of Computer Application, Lovely Professional University, Phagwara, Punjab 144401, India.

[3]Department of Paediatrics, Paediatric Haematology/Oncology Unit, Dayanand Medical College and Hospital Ludhiana, Punjab India.

Abstract

Technology and information-based e-Health systems are essential to the delivery of healthcare; these include teaching and learning programs as well as remote consultation services that have been widely implemented to support many medical specialities and their corresponding remote patient populations. Because of this, e-Health systems are being utilised more and more as a passive as well as an active technique to address issues (such staffing shortages, lack of resources, and lack of knowledge) in the healthcare sector and have been shown to significantly enhance the general health of many nations. While numerous e-Health systems have been created and employed to carry out diverse functions in the medical field, including remote consultation, hospital administration, electronic health record administration, and health awareness, a dearth of research has been found concerning the analysis and prediction of haemophilia in the context of e-Health systems. Because of this, a multilayered framework for the investigation and prediction of haemophilia using machine learning techniques was developed in response to the observation.

Keywords: Haemophilia, e-Health, Machine Learning, Tele-Medicine, Prediction

Introduction

The utilisation of Information and Communication Technologies (ICT) in healthcare, such as networks, the internet, computers, and wireless, has given rise to the wonderful concept of electronic health, or e-Health [1]. The effective adoption of e-Health has revolutionised the delivery of healthcare by making it possible to store medical records, retrieve information and records, and transmit medical services digitally. Furthermore, a number of countries' overall health has significantly improved when e-Health has been used consistently in the healthcare sector [2]. Aside from this, e-Health systems are being used more and more as a passive and active way to reduce various hurdles in the healthcare industry (such a lack of financing, staff, and awareness). All of these reasons have led to the replacement of traditional methods of providing health services in a number of sectors, including medical learning, medical consulting (including patient care, appointment scheduling, and record administration), and health management, by e-Health solutions [3]. Even while e-Health is still a relatively new way of using communications technology to deliver medical services, it has been used for more than thirty years in one way or another.

FIGURE 7.4: Annexure 4