PREDICTING THE SEVERITY OF DENGUE FEVER AMONG CHILDREN ADMITTED TO ANGKOR HOSPITAL FROM CLINICAL FEATURES AND LABORATORY INDICATORS: APPLICATION OF CLASSIFICATION TREE ANALYSIS

KHANSOUDAPHONE PHAKHOUNTHONG

A THEMATIC PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE (BIOMEDICAL AND HEALTH INFORMATICS) FACULTY OF GRADUATE STUDIES MAHIDOL UNIVERSITY 2014

Thematic Paper
entitled

PREDICTING THE SEVERITY OF DENGUE FEVER AMONG CHILDREN ADMITTED TO ANGKOR HOSPITAL FROM CLINICAL FEATURES AND LABORATORY INDICATORS: APPLICATION OF CLASSIFICATION TREE ANALYSIS

……..……..……..
Ms. Khansoudaphone Phakhounthong
Candidate

……..……..……
Lect. Wirichada Pan-ngum, Ph.D.
Major advisor

……..……..……
Lect. PimwadeeChaovalit, Ph.D.
Co-advisor

……..……..……
Lect. PodjaneeJittamala, M.D.
Co-advisor

……
Mr. Stuart Blacksell, Ph.D.
External advisor

……..……..……
Prof. Banchong Mahaisavariya,
M.D., Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies,
Mahidol University

……..……..……
Assoc. Prof. Pratap Singhasivanon,
Dr.P.H, M.B.B.S, DTM&H.(Bangkok)
Program Director
Master of Science in Biomedical and
Health Informatics
Faculty of Tropical Medicine
Mahidol University

Thematic Paper
entitled

PREDICTING THE SEVERITY OF DENGUE FEVER AMONG CHILDREN ADMITTED TO ANGKOR HOSPITAL FROM CLINICAL FEATURES AND LABORATORY INDICATORS: APPLICATION OF CLASSIFICATION TREE ANALYSIS

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Science (Biomedical and Health Informatics)
on
March 7, 2014

……
Ms. Khansoudaphone Phakhounthong
Candidate

……
Mr. Stuart Blacksell, Ph.D.
Chair

……
Lect. Wirichada Pan-ngum, Ph.D.
Member

……
Lect. PimwadeeChaovalit, Ph.D.
Member

……
Lect. PodjaneeJittamala, M.D.
Member

……
Prof. Banchong Mahaisavariya,
M.D., Dip Thai Board of
Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

……
Assoc. Prof. Yaowalark Sukthana,
M.D, D.V.M., D.T.M.& H.(Bangkok),
M.C.T.M.,
Dip. Thai. Board in Oto-Rhino-
Laryngology
Dean
Faculty of Tropical Medicine
Mahidol University

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor Dr. Wirichada Pan-ngum, you have been a great mentor to me, I am greatly indebted to you for your kind help, wonderful comments and kindness. I would especially like to thank my committee members, Dr. Stuart Blacksell, Dr. Pimwadee Chaovalit and Dr. Podjanee Jittamala, I deeply appreciate your advices and suggestions. I would also like to thank Dr. Michael Carter for his kind help and brilliant advices.

Furthermore I would like to express my gratitude to my parents for their tremendous support and help, without their encouragement, I would have failed in the attempt.

Khansoudaphone Phakhounthong

PREDICTING THE SEVERITY OF DENGUE FEVER AMONG CHILDREN ADMITTED TO ANGKOR HOSPITAL FROM CLINICAL FEATURES AND LABORATORY INDICATORS: APPLICATION OF CLASSIFICATION TREE ANALYSIS

KHANSOUDAPHONE PHAKHOUNTHONG 5538765 TMBI/M

M.Sc. (BIOMEDICAL AND HEALTH INFORMATICS)

THEMATIC PAPER ADVISORY COMMITTEE: WIRICHADA PAN-NGUM, Ph.D., STUART BLACKSELL, Ph.D., PIMWADEE CHAOVALIT, Ph.D., PODJANEE JITTAMALA, M.D.

ABSTRACT

Background: Dengue fever (DF) / Dengue Haemorrhagic Fever (DHF) is a viral and re-emerging disease, commonly occurring in tropical and subtropical areas. Furthermore, clinical features and laboratory abnormalities of dengue infection are also similar to other febrile illnesses. The goal of this study is to develop predictive models to characterize dengue and its severity by early clinical and laboratory measures using statistical and data mining tools.

Methods and findings: We performed a retrospective study and retrieved data from a study of febrile illness in children at Angkor Hospital for Children in Cambodia. A total of 1225 febrile episodes were in our analysis, 198 were confirmed dengue patients. Classification and regression tree (CART) and logistic regression analysis were used independently to differentiate: dengue versus non-dengue (model 1), severe dengue versus non-severe dengue (model 2) and death versus survival (model 3). Highlighting result of model 2, classifying severe dengue, the decision tree algorithm using pulse rate, hematocrit, Glasgow coma score, urine protein, creatinine and platelet count has outperformed the logistic regression model with sensitivity and specificity of 71% and 68.7%, respectively.

Conclusions: Our decision tree algorithm using simple clinical and laboratory indicators has a high classification accuracy in predicting pediatric patients who develop severe dengue. This model is potentially useful for guiding a patient monitoring plan and outpatient management of fever.

KEY WORDS: DENGUE SEVERE DENGUE / CLINICAL / LABORATORY / CLASSIFICATION AND REGRESSION TREE (CART)/ LOGISTIC REGRESSION

140 pages

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF TABLES (cont.)

LIST OF TABLES (cont.)

LIST OF FIGURE

LIST OF FIGURE (cont.)

CHAPTER I
INTRODUCTION

1.1 Background and Justification

Dengue fever (DF) / dengue haemorrhagic fever (DHF) is a viral and re-emerging disease, commonly occurred in tropical and subtropical areas, these issues have remained an important public health concern, especially in South-east Asia, Africa, the Western Pacific and the Americas [1]. Dengue virus composes of four serotypes, DENV-1, DENV-2, DENV-3 and DENV-4.Dengue transmission occurs via the bite of an infected mosquito, it is known as *Aedes aegypti* mosquito[2]. An estimated 2.5 billion people all over the world are at risk of DF/DHF, it also accounts for more than 50 million dengue infections annually, containing 500,000 hospital admissions for DHF, principally among children [3].

Dengue infection is frequently confounded with other febrile illnesses, showing with non-specific clinical symptoms. Other febrile illness (OFI) is common in tropical countries, and clinical features of dengue is analogous to OFI. During the early stage of illness, the presence of non-specific febrile illness is strikingly difficult in making precise diagnosis, inefficient treatment and possibly increased morbidity and mortality [2, 4].

Clinical manifestations of dengue such as sudden onset of high fever, headache, arthralgia, myalgia, nausea, vomiting and maculopapular rashes, which may imitate other diseases, for instance, leptospirosis and malaria are prevalent in the same zones as dengue is endemic [5].

DHF and dengue shock syndrome (DSS) represent the severe features such as severity of bleeding, severity of organ failure and severity of plasma leakage, which if not appropriately managed, would speedily lead to death, principally in children[6, 7].The presence of massive blood loss, profound shock and fluid overload are critical conditions, it can lead to death, and furthermore, the major problem causing death is related to poor prognosis during hospitalization [8]. There are difficulties in making

the diagnosis of DHF to meet WHO criteria, especially in tropical regions where dengue is endemic, resulting in a lack of hospital facilities and lacking the ability to do chest radiograph or ultrasound for detecting pleural effusion or ascites. In addition, a lack of necessary laboratory may cause difficulty to discriminate dengue infection from other febrile illnesses, particularly in rural, remote areas [9]. There is currently no vaccine for dengue infection or a dengue drug covering all four serotypes [10, 11].

In Southeast Asia ranging from year 2001 to 2010, the overall annual economic burden of dengue is US$ 950 million, composing of 12 countries and there are 2.9 million episodes of dengue infection annually, including 5,906 deaths of reported dengue cases[12].

In Cambodia, the first detected dengue virus was in 1963 and has been informed to national surveillance of dengue since 1980, from year 1980-2008, national dengue has been reported that infants less than 1 year and children aged 4-6 years were the highest age-specific incidence[6]. Dengue infection accounts for 10,000 - 40,000 cases annually reported by dengue national control program [13]. Research conducted by [14] and [6] showed that among children aged < 7 years were at higher risk for dengue infection, mainly during rainy seasons. During dengue outbreak in 2007, the most predominant serotype was DENV-3, and rural areas faced a higher risk of dengue infection than urban areas (incidence 71 versus 17/1,000 person-seasons) [14], an average annual number of dengue cases in Cambodia from 1991 to 2010, for each year is detailed in figure1.1. However, there remains under-recognition to the National Dengue Surveillance System (NDSS), especially during the large outbreak of dengue in 2007. Dengue cases reported by NDSS are identified on a clinical basis using the 1997 World Health Organization case definition and only include hospitalized cases study by Vong S et. al. conservatively estimated that there was a 4- to 30-fold degree of dengue under-recognition and underreporting to NDSS during 2006–2008 in Kampong Cham, the most populous province in Cambodia [15]. Additionally, an overall annual dengue cost was estimated to US$14,429,513 during the 2007 dengue endemic in Cambodia [16].

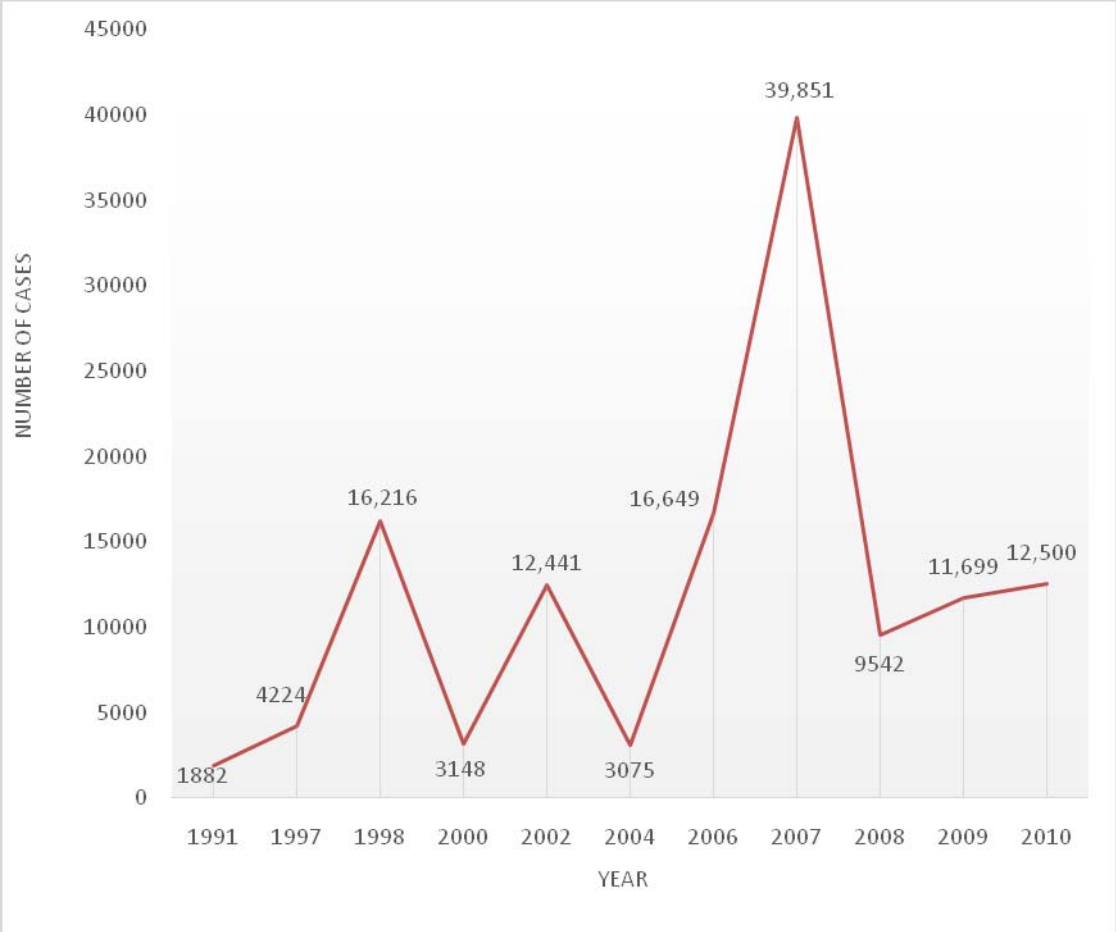


Figure 1.1 Average annual number of dengue cases in Cambodia, 1991-2010
http://hiip.wpro.who.int/hiip/DataAnalytics/GenerateTables.aspx

We aimed to apply decision tree techniques to develop a useful diagnostic algorithm, using early clinical signs and possibly laboratory testing, to discriminate severe dengue during the early stage of illness from other febrile illnesses. The data from the fever study among children admitted to Angkor hospital would be used for constructing the algorithm. This predictive tool could assist physicians in making diagnostic decision and clinical management on severe dengue and other febrile illnesses.

1.2. Objectives

1.2.1 General objective

- To characterize severe dengue by clinical and laboratory measures
- To compare clinical and laboratory characteristics of severe dengue against other febrile illnesses

1.2.2 Specific objectives

- To develop models for predicting severity of dengue at the early phase of illness using statistical and data mining techniques.
- To compare models and discuss strength and weakness of each approach

1.3 Conceptual framework

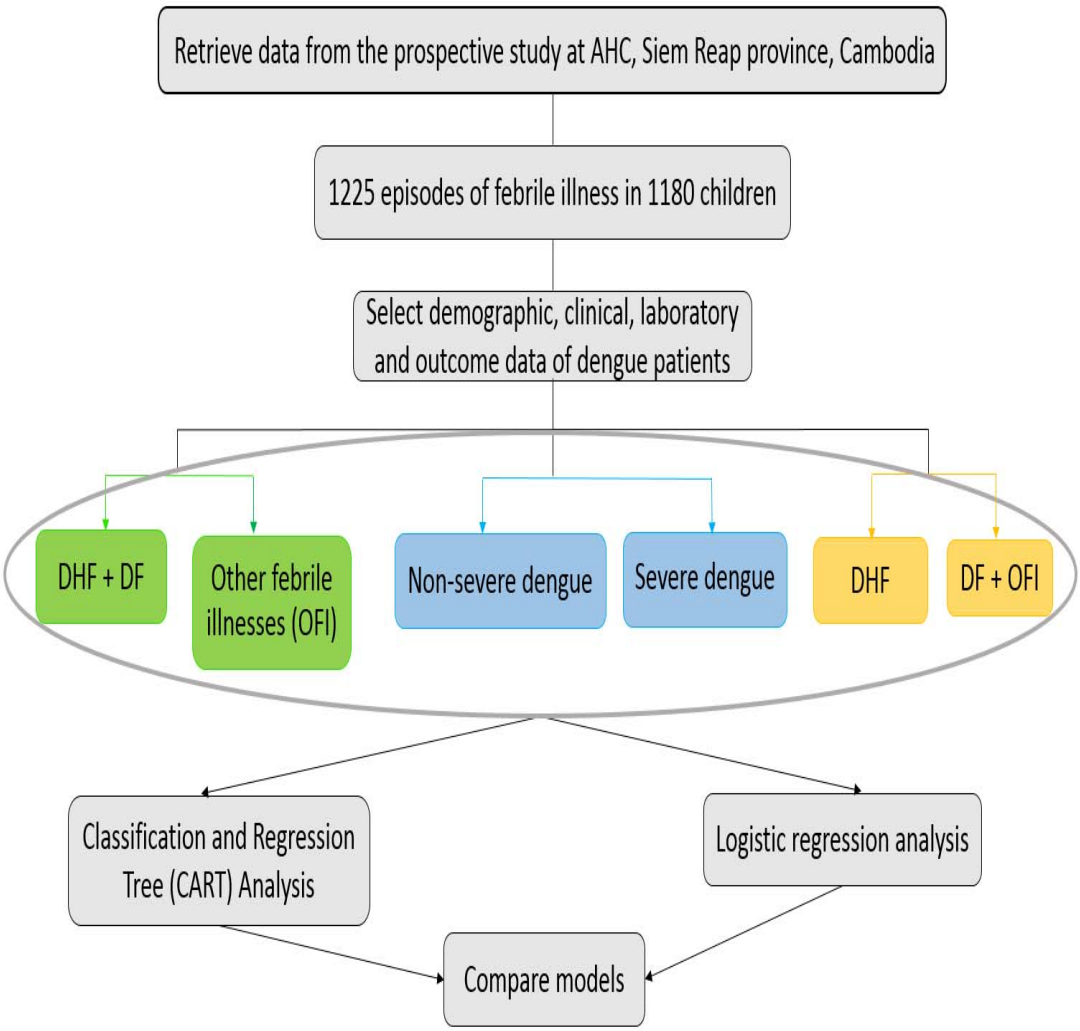


Figure 1.2 Conceptual framework

1.4 Usefulness of the study

Febrile illnesses are common in tropical countries and the symptoms of dengue is analogous to other febrile illnesses. During the early phase of illness, clinical features of dengue infection is frequently confounded with other febrile illnesses, so that is very hard to know which patients will develop a severe form of disease when presenting clinical manifestations at the hospitals. For this reason, early recognition of disease can result in making precise diagnosis, effective clinical management, saving patient's life and reducing medical expenses.

More importantly, an early prediction tool for identifying which patients are likely to undergo the severe form of dengue will help to improve the utilization of limited health facilities in poor setting where dengue is endemic. To address this we aimed to create a simple diagnostic algorithm that could aid physicians in making diagnostic decision based on using clinical signs and laboratory tests to predict severity of dengue infection.

CHAPTER II
LITERATURE REVIEW

2.1 An overview of dengue

Dengue fever is commonly widespread vector-borne diseases in tropical and subtropical regions all over the world. It impacts more than a 500 million people annually, involving more than nine million cases in the world for dengue fever and about 18,000 deaths. World-wide, an estimated 3.6 billion people are at risk for dengue virus infection. South-east Asia, Africa, the Western Pacific and the Americas are dengue-endemic areas remaining the major public health problem, and South-east Asia has the highest mortality of infection [16, 17].

2.1.1 Dengue fever virus (DENV)

It is a single stranded RNA of the *Flaviviridae* family, genus *Flavivirus*, which can be transmitted by the bite of an infected *Aedes aegypti* mosquito. All four serotypes can cause disease, which are DENV-1, DENV-2, DENV-3 and DENV-4.When getting infections by these virus serotypes, as a result, it may be asymptomatic or non-specific illness, it is known as dengue fever (DF) and can also develop severe forms of DF to DHF and DSS [18].

All four serotypes result in similar febrile illness, but DEN-2 and DEN-3 infections are more likely cause disease severity and the development of dengue haemorrhagic fever. A mild clinical manifestations of dengue fever is normally occurred after primary dengue virus infection, and in seven days it is generally cleared as a result of immune response. On the other hand, secondary infection is more severe results which is presented by high fever, an increase in vascular permeability, plasma leakage detection, the presence of haemorrhage and thrombocytopenia, which can cause either DHF or DSS and moreover, it can lead to severe illness [19].

2.1.2 Mosquito Vectors

It belongs to the *Aedes* mosquitoes, including *Aedes aegypti*, *Aedes albopictus and Aedes polynesiensis*, they are apparently responsible for dengue transmission, mainly *Aedes aegypti*. It is a day biting mosquito and commonly distributed in tropical and subtropical countries all over the world. This mosquito breeds in water-holding containers to complete its life cycle, such as stagnant water, sewage, flower vases, tree holes. The life cycle of the aedes mosquito is shown in figure 2.1. Only female mosquitoes bite humans and domesticated animals for taking a blood meal before developing eggs and laying viable eggs on the surface of the water. Their life cycle follow by eggs, larvae, pupae, and adult. A study of the relation between rainy season and aedes mosquitoes showed that during rainy season, aedes larval populations were significantly increased [20].

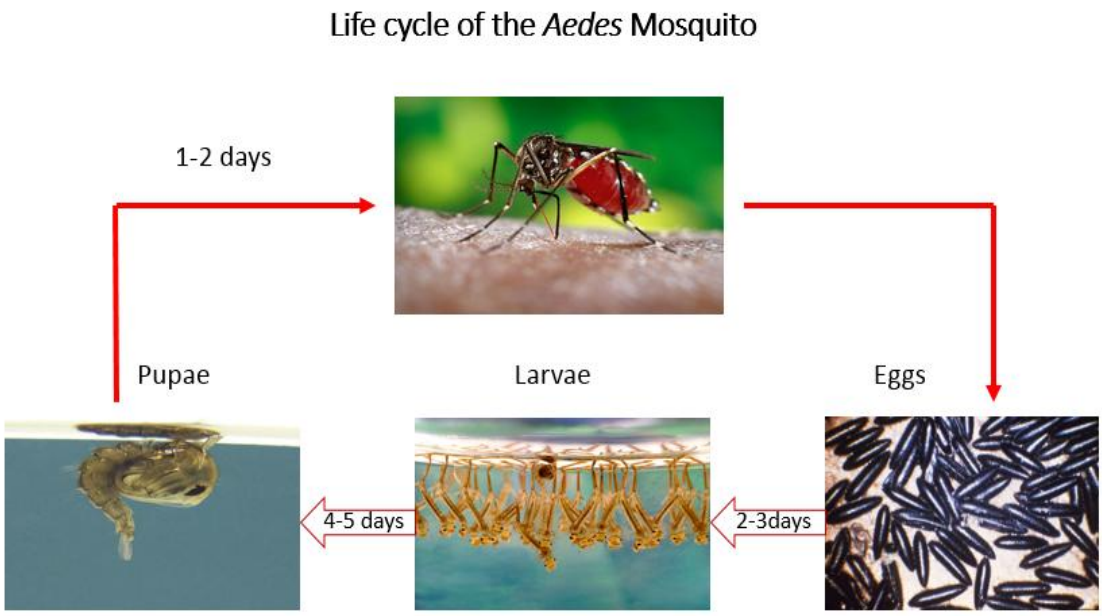


Figure 2.1 Life cycle of Aedes mosquito

2.2 Clinical manifestations of dengue infections

2.2.1 Clinical manifestations of dengue fever

The acute febrile phase generally lasts 2-7 days, accompanying by high fever, headache, nausea/vomiting, body pains and rash, which is not easy to differentiate clinical manifestations of dengue from other febrile illnesses during the initial febrile phase of illness [21].

A mild haemorrrhagic features such as petechiae and bleeding (nose and gums) may be occurred in DF, which are scarcely correlated with the severity of haemorrhagic shock [1].

2.2.2 Clinical manifestations of dengue hemorrhagic fever

As a result of distinctions in capillary permeability, the development of DHF in children may be at higher risk than adults [22]. By the presence of increased vascular permeability can lead to plasma leakage, a decrease of platelets in blood < 100, 000 mm^3 associated with bleeding and an abnormality on liver function, development of effusions and ascites (an accumulation of fluid in abdomen) [3, 23].

2.3 The Course of dengue illness

The incubation period of dengue infection is 4-7 days, ranging from 3 - 14 days, then the development of febrile illness can lead to be one of the three following conditions, and is detailed in figure 2.2.

- Undifferentiated febrile illness
- Dengue Fever
- Dengue Haemorrhagic Fever

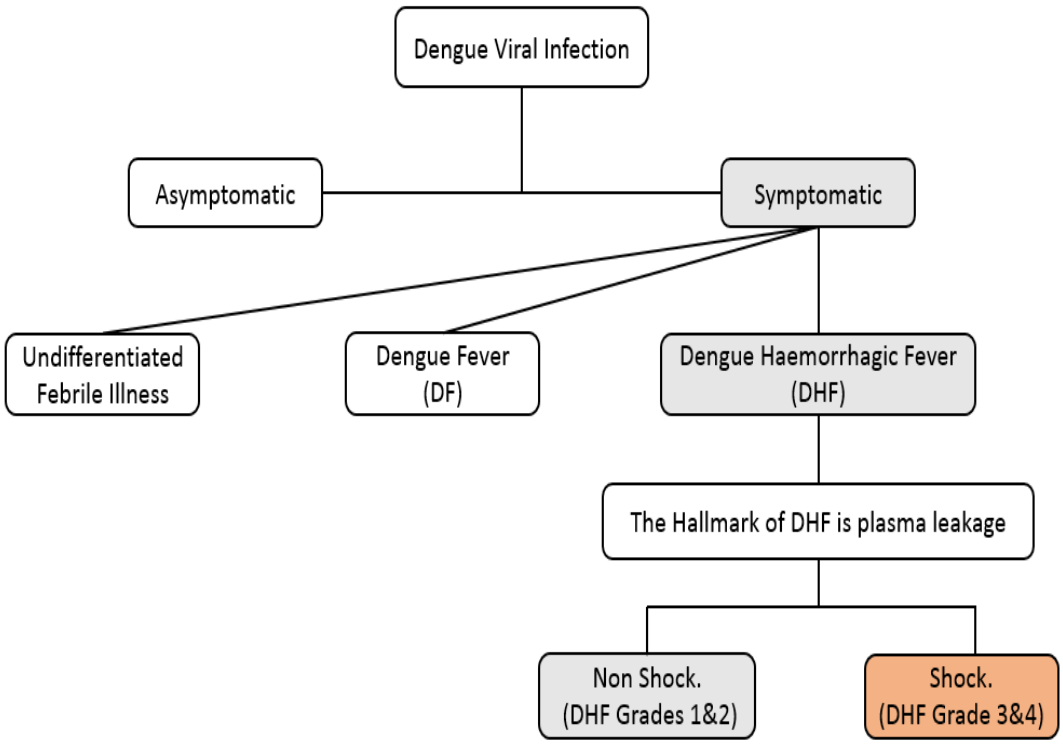


Figure 2.2 Flowchart showing undifferentiated febrile illness, DF and DHF
The figure taken from national guidelines on management of dengue fever and in adults, Sri Lanka.

Undifferentiated febrile illness is widely known in tropical regions, its characteristic etiology is frequently unknown, making diagnosis is inaccuracy and the treatment is ineffective. The difficulty of identifying a particular pathogen is that there is no specific signs or symptoms and lack of available diagnostic tests, this problem can potentially influence morbidity and mortality [4].

Hence, all patients suspected of having DF and DHF should be observed closely, early identifying of DHF patients and treat will prevent development of plasma leakage which may occur rapidly after onset of symptoms.

The incubation period of dengue is around 4-7 days, and the natural course of the illness composes of three stages: febrile, critical and recovery phase, this course is detailed in figure 2.3.

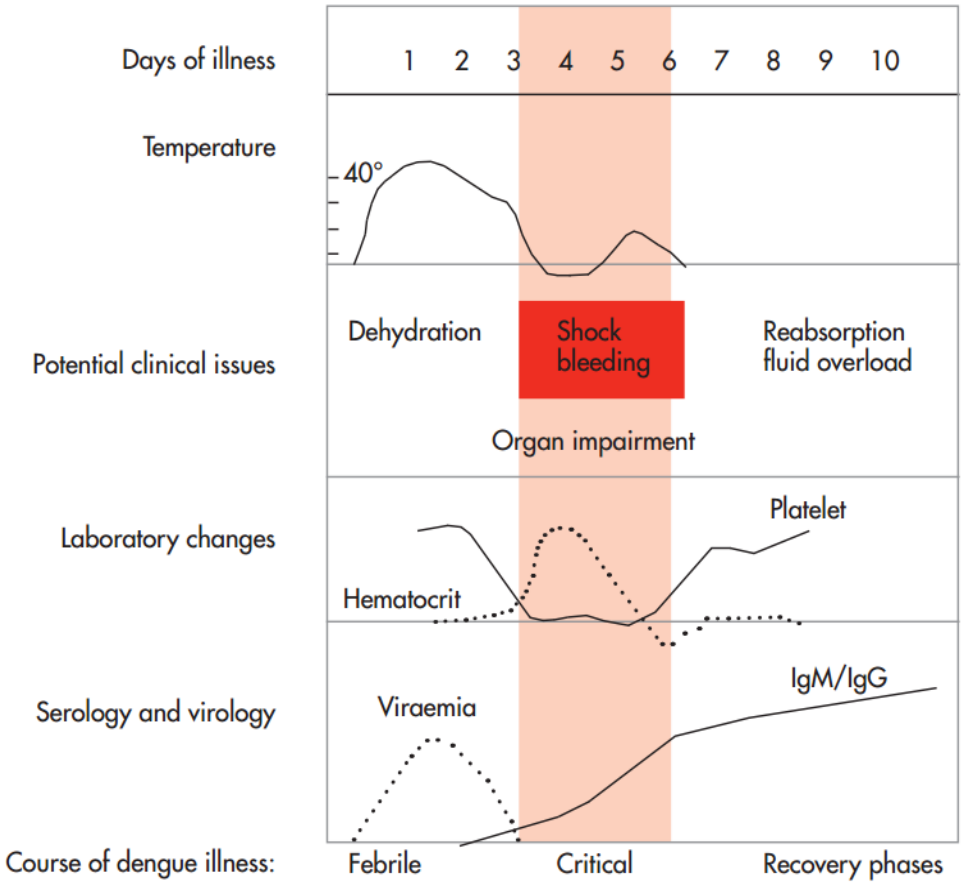


Figure 2.3 The course of dengue illness

The figure taken from WHO 2009 dengue case classification

2.3.1 Febrile phase

This phase is presented by having sudden onset of high fever, generally lasts 2-7 days, coinciding with facial flushing, headache, myalgia, arthralgia, skin erythema, nausea and vomiting. Other features may occur in some patients such as conjunctival injection, injected pharynx, sore throat and diarrhoea. In addition, mild haemorrhagic features in this phase can also happen. In the late febrile phase, leucopenia (WBC < 5000 mm^3) and mild thrombocytopenia (platelets < 150,000/mm^3) are commonly occurred. Above manifestations are normally undifferentiated between DF and DHF during this stage. Although the presence of increased liver size advocates the diagnosis of DHF. The end of this phase is usually suggested by having platelet count less than 100,000/mm^3, and may signal an entrance to the critical phase [21, 24].

2.3.2 Critical phase

The onset of plasma leakage is indicated the critical phase. It normally happens towards the late febrile phase, regularly on days 3-7 of illness.

The presence of increased capillary permeability leads to plasma leakage. The period of plasma leakage, it usually develops 24-48 hours, see figure 2.4. An increase in capillary permeability may turn out to be worse based on the degree of plasma leakage and the volume of fluid administration for patients. Pleural effusion and ascites can be happened but for pericardial effusion is unusually occurred during this phase. Therefore, helpful tools for diagnosis such as chest radiograph (chest x-ray) and ultrasound of abdomen can be clinically detected for the rapid occurrence of plasma leakage.

The severity of plasma leakage is often indicative of rising in haematocrit (HCT) 20% of the baseline value. Other confirmations of plasma leakage are a serum albumin reduction less than 3.5 g/dl and non-fasting serum cholesterol less than 100 mg/dl.

The degree of plasma leakage is variable in dengue hemorrhagic fever. Some patients may develop to be minimal manifestations while in others it can be very considerable, depending on the degree of plasma leakage.

Unusual manifestations such as severe hepatitis, encephalitis or myocarditis can occasionally develop without apparent plasma leakage.

Shock can occur owing to the lost plasma volume and it usually precedes warning signs. Prolonged shock results in progressive organ failure, disseminated intravascular coagulation (DIC) and metabolic acidosis which frequently cause significant bleeding [21, 24].

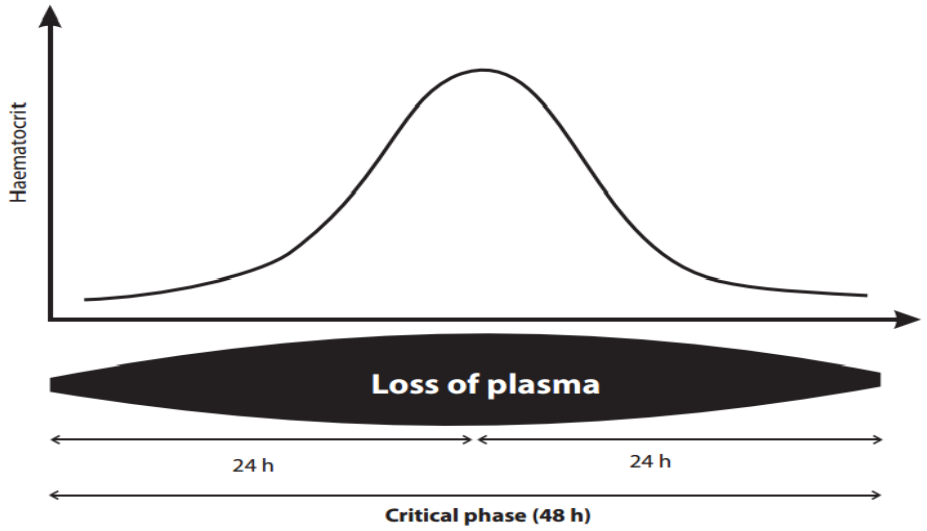


Figure 2.4 Fluid leakage in the critical phase
The figure taken from national guidelines on management of dengue fever and in adults, Sri Lanka.

2.3.3 Recovery phase

This regularly lasts 2-5 days. The reabsorption of extravasated fluid will be occurred in this phase.

The key indicators for patients reaching the recovery phase

- Improvement of general health and appetite
- Presentation of convalescent rash
- Widespread itchy (higher density in palms and soles)
- Haemodynamic stability
- Bradycardia and electrocardiographic changes
- Diuresis
- Haematocrit stability
- A rise in white cell count, followed by platelet count

Owing to intravenous fluid overload during the critical stage, respiratory distress can occur as a result of pleural effusion and ascites, and excessive fluid treatment is related to pulmonary oedema or congestive heart failure [21, 24].

2.4 Dengue classification

There has been confusion on the subject of using both WHO 1997 and 2009 dengue classification. The 1997 dengue case classification or the current WHO has been causing a lot of complaints from many physicians owing to most of the patients could not meet the WHO criteria for making diagnosis of DHF case definition, such as the presence of plasma leakage and thrombocytopenia, and furthermore many cases could not be classified by WHO classification when presenting with multi-organ involvement, such as liver, renal and heart failure [25].

2.4.1 WHO 1997 dengue classification

The WHO 1997 classification categorizes dengue into dengue fever and dengue haemorrhagic fever/dengue shock syndrome. For dengue fever requires fever with at least two or more of: headache, retro-orbital pain, myalgia, arthralgia plus positive tourniquet test, leukopenia. For diagnosing of DHF case, it is required to meet the following criteria: fever, hemorrhagic manifestations, thrombocytopenia $\leq$ 100,000/ μl and evidence of plasma leakage (haematocrit $\geq$ 20%). For a case of DSS, all four criteria of DHF case is needed to be met plus any signs of these criteria:

- Circulatory failure (weak pulse, hypotension, restlessness)
- Profound shock with undetectable blood pressure and pulse

For the spectrum of DHF is detailed in figure 2.5 and for grading dengue infection severity is shown in table 2.6.

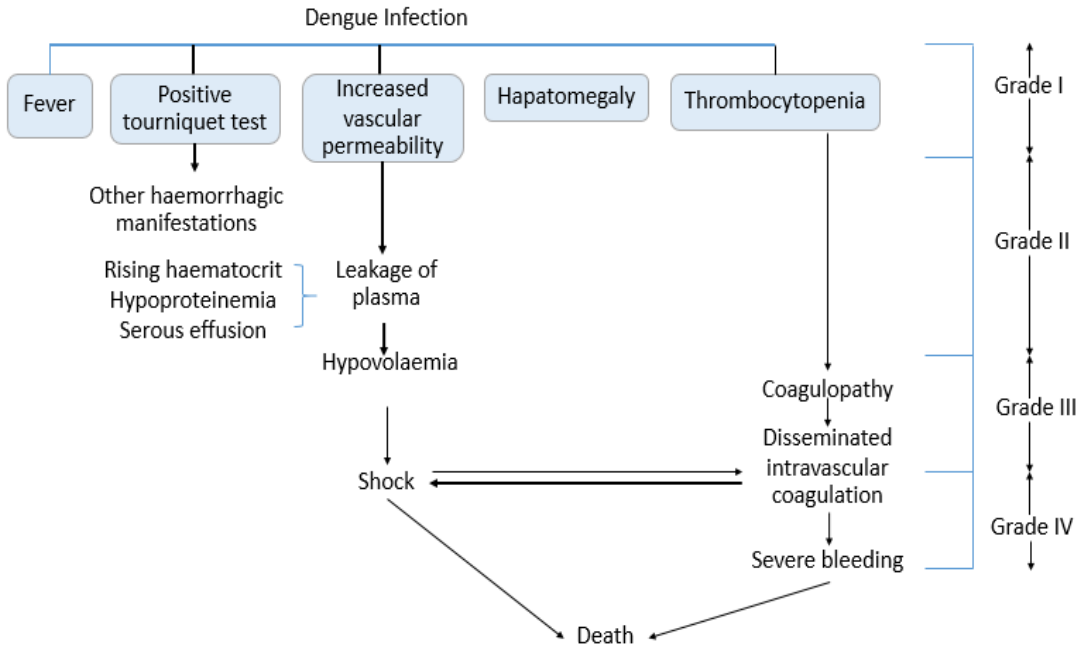


Figure 2.5 The spectrum of Dengue Haemorrhagic Fever

The figure taken from this website

http://209.61.208.233/en/Section10/Section332/Section554_2564.htm

DF/DHF	Grade	Symptoms	Laboratory
DF*		Fever with two or more of the following signs: headache, retro-orbital pain, myalgia, arthralgia plus positive tourniquet test	Leucopenia occasionally Thrombocytopenia, may be present, no evidence of plasma loss
DHF	I	Above signs plus positive tourniquet test	Thrombocytopenia ≤ 100,000/ µL, hematocrit rise ≥ 20%
DHF	II	Above signs plus spontaneous bleeding	Thrombocytopenia ≤ 100,000/ µL, hematocrit rise ≥ 20%
DHF	III**	Above signs plus circulatory failure (weak pulse, hypotension, restlessness)	Thrombocytopenia ≤ 100,000/ µL, hematocrit rise ≥ 20%
DHF	IV**	Profound shock with undetectable blood pressure and pulse	Thrombocytopenia ≤100,000/ µL, hematocrit rise ≥ 20%

DF: dengue fever; DHF: dengue hemorrhagic fever; *Modified WHO criteria; **DHF Grade III and IV are also called as dengue shock syndrome (DSS).

Figure 2.6 Grading dengue infection severity (modified WHO 1997)

The figure taken from Prognostic Indicators for Dengue Infection Severity**.**

There are many countries in Asia remain using World Health Organization classification criteria 1997, such as Cambodia [6], Thailand [8], Laos [26], Vietnam [27], Malaysia [7], Singapore [28].

2.4.2 WHO 2009 dengue classification

- The WHO 2009 criteria are categorized dengue into DF with or without warning signs and severe dengue. For dengue diagnosis, it needs to have a fever and two or more of criteria or any warning signs. For severe dengue can be diagnosed by the requirement of a diagnosis of DF plus any criteria (figure 2.7):
- Plasma leakage that may cause shock (dengue shock) and/or an accumulation of the fluid, with or without respiratory distress
- Severe bleeding
- Severe organ impairment.

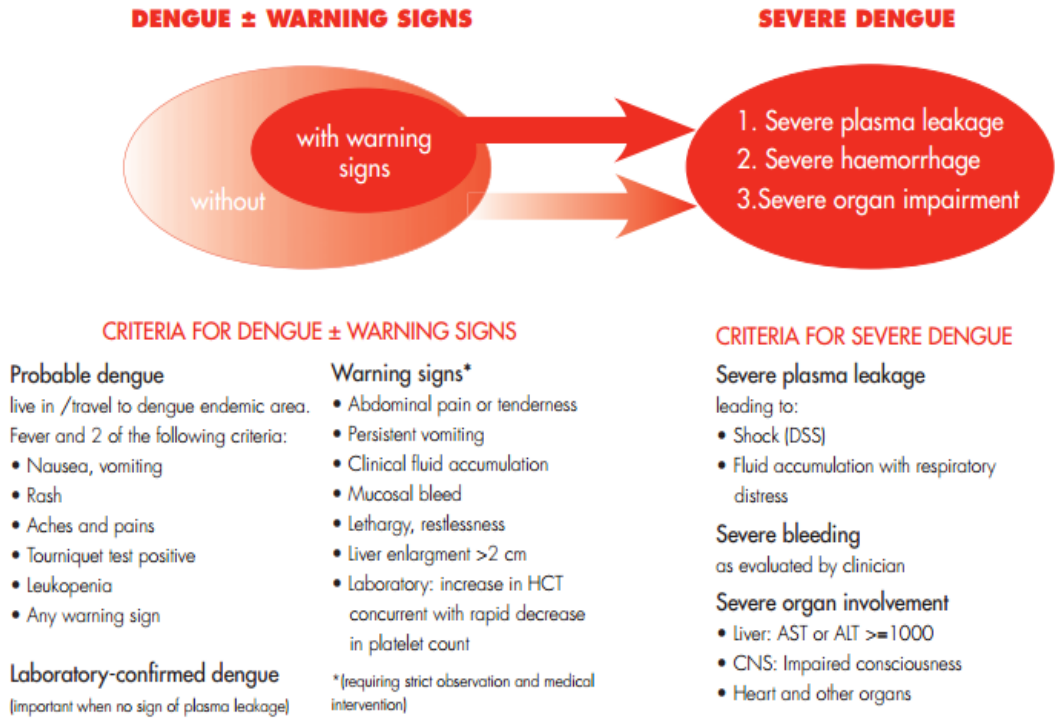


Figure 2.7 Dengue classification and levels of severity

The figure taken from dengue guidelines for diagnosis, treatment, prevention and control.

Shock is often preceded by warning signs, usually on days 3-7 of illness. During the first stage of shock, tachycardia, cold extremities and prolonged capillary refill time will occur.

If the narrowing of pulse pressure is ≤20 mm Hg, patient is considered to have shock. When massive bleeding occurs, it is nearly associated with profound shock, integrating with thrombocytopenia, acidosis and hypoxia can cause multi-organ impairment and followed by DIC.

Severe dengue should be thought about it carefully if patient comes from dengue endemic area, accompanying with fever of 2-7 days plus any of manifestations given below:

- Confirmation of plasma leakage, for instance
  - High increase in haematocrit,
  - Pleural effusions or ascites
  - Cold extremities, fast heart rate, capillary refill time ≥ 3 seconds, narrow pulse pressure or unrecordable blood pressure, weak pulse
- Massive bleeding
- Lethargy or restlessness, convulsions, coma
- An increase in abdominal pain , persistent vomiting, jaundice
- Acute liver failure, acute renal failure, encephalitis, or other unusual manifestations.

2.4.3 Simpler definition of severe dengue illness

WHO 1997 is required to be modified regarding plasma leakage for user-friendliness, clarifying definitions of severe bleeding and organ impairment [25, 29]. Clinical signs of severe dengue infection such as organ impairment is not presented in WHO 1997 criteria, many physicians have been trying to fulfill these criteria, in addition, a requirement for repeating clinical tests is also necessary but this is not easy to perform frequently, particularly for those countries which have limited resource settings [30, 31].

The study conducted by Potts (2010) [9] used a simpler definition of severe dengue where any of these four criteria must be met:

1. Patient must be diagnosed with DHF III or DHF IV (DHF grade III or IV are also known as DSS).

2. Patient requires the radiological investigation (chest x-ray) to assess the pleural effusion index (PEI > 15) from right lateral decubitus position, it is measured by the diameter of the excess fluid in the right lung (A) /the diameter of the right hemithorax (B) ×100, (see figure 2.8).

3. Patient is needed to take total fluid intervention within 24 hours (exceeded maintain of body fluid + 5% volume deficit).

4. Intravenous fluid is required when patient cannot be able to take fluids orally.

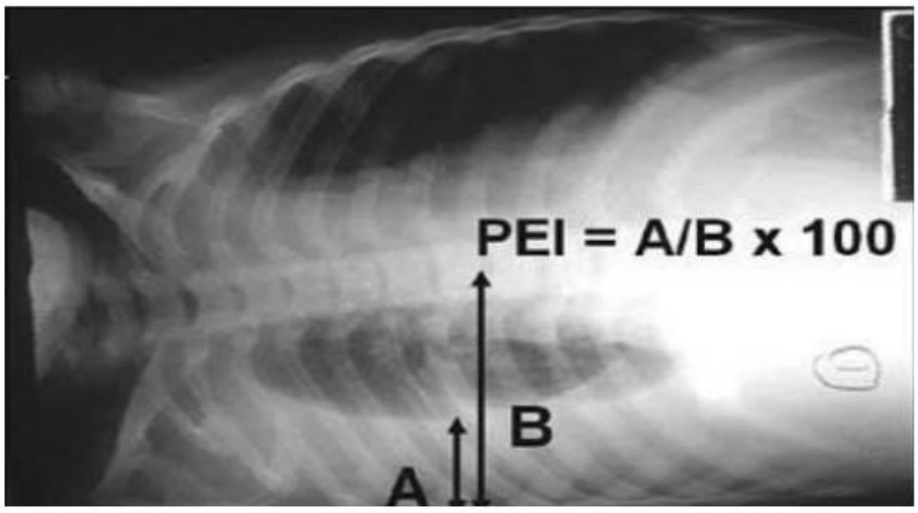


Figure 2.8 Pleural effusion index

The figure taken from the study of the relationship between PEI and DSS [32]. This simple classification of severe dengue will be applied in this project.

2.5. Laboratory diagnosis of dengue infection

As dengue infection is a wide spectrum of clinical manifestations, it can be difficult to make accurate diagnosis, progressing from mild febrile illness to severe disease. Amongst the available techniques for diagnostic dengue infection, virus isolation is the most reliable test result providing evidence of infection. However, diagnostic facilities are not always available, particularly in some rural, remote areas.

Seroconversion of IgM or IgG antibodies is the benchmark for serological confirmation of dengue infection. Detection of specific IgM or high IgG levels in serum collected from patients with suspected dengue case can consider to be a probable dengue infection [3].

2.5.1 Virus isolation

The viraemic stage, usually occurs within 3-5 days of fever. Sera collected from dengue patients are often used for virus isolation.

This is the technique in vitro isolation and requires the manipulation in the BSL3.The culture uses the C6/36 (Aedes albopictus mosquito) as a cell line to grow these virus. In addition, there are different cell lines that can apply in the culture such as Aedes pseudoscutellaris AP61, Vero cells [33] LLC- MK2 cells BHK21. The virus isolation is positive during the early phase of illness and in particular the blood should be collected before day 5 of the illness in suspected dengue infection. After the cells were infected with the virus during the incubation period. The identification of the etiology was done by using the specific monoclonal antibodies in immunofluorescence assays and follows the identification of virus by PCR [34].

In the isolation of dengue, there are different samples such as serum, whole blood, the tissue from the biopsy all of them can be used to detect the dengue infection. This technique is the most specific test for the diagnosis of dengue but requires expertise and appropriate facilities.

2.5.2 Serological testing

This is the most common test that has been used and it requires the paired serum to confirm the dengue infection. To compare with the culture and PCR this test is easiest than the isolation and it is not expensive too. In the primary dengue IgM antibody titer is higher when comparing with the secondary infection. It rises until the peak about 2 weeks after the onset of disease [35].

In the areas where there are co-circulated between Dengue virus and Japanese encephalitis virus. MAC- ELISA is the best serological test and it has shown the sensitivity of 90% and specificity of 98% when the blood was collected within 5

days of illness. However, it has some non-specific reaction where the multiple flavivirus and including the patient with malaria and leptospirosis diagnosis [36].

IgG ELISA, The IgG is used to determine that it is primary or secondary infection. It can be detected in patients after 7 days of illness. Therefore, the recommendation of IgG is used when the IgM is still negative after day 7 of IgG negative in the results of the primary test.

2.5.3 Molecular detection

PCR (Polymerase chain reaction), is a molecular method to detect the DNA of the virus. It is a reverse transcriptase polymerase chain reaction**.** This technique can be used to identify the virus of dengue infection which is less than 5 days of the illness with the sensitivity of 100% and it will decrease in day 6 of 70%**.**

2.5.4 Detection of antigen

All Flavivirus have the glycoprotein such us NS1. This antigen is necessary for the multiplication.NS1 antigen is secreted in the blood circulation. During the viraemia phase, usually lasts 4-6 days, NS1 is known to be occurred during this time. In primary dengue infection, NS1 antigen can be detected on day 5-6 by serological test. In contrast, during secondary dengue infection the presence of NS1 antigen is usually detected on day 6-12 of illness (figure 2.9).

In figure 2.9, in primary infection, IgM is the first appearance of immunoglobulin isotype, the reaction of this antibody to dengue virus presented by a slow and low titre. Anti-dengue IgG is represented by a low titre which can be detected after ending the initial week of illness, and then raising gradually. In comparison, secondary dengue infection, IgG antibody titres can be detected in the first week of illness and they increase rapidly over two weeks, IgG is a representative of previous infection.

The detection of NS1 antigen has the advantage that it can be used for early detection and investigating dengue outbreaks. However, this is not as sensitive as virus isolation [37].

The study conducted by Singh [38] has demonstrated IgM antibody with NS1 antigen in 87 samples in acute infection. It has the sensitivity of 71- 100% until

day three of fever whereas IgM cannot be detected in the early phase of viraemia stage.

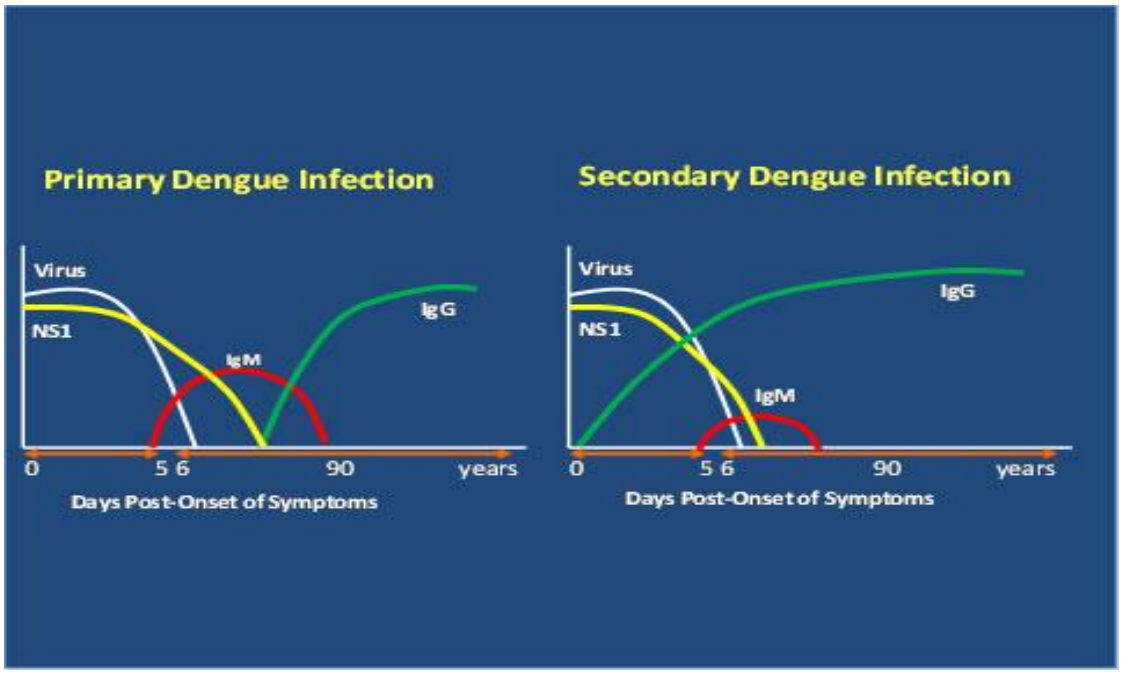


Figure 2.9 Immune response for dengue infection

The figure taken from this website:

(http://www.cdc.gov/dengue/clinicallab/laboratory.html)

2.5.5 Rapid diagnostic tests

It is an immunochromatographic test that has used to find out the non-structural protein 1 (NS1) of dengue virus, antibodies such as IgM, IgG. Rapid diagnostic tests (RDT) is shown in the pattern of a lateral flow cassette that shows the flow of the sample in horizontal vision. The interpretation of this test is the presence of NS1 antigen or IgM, IgG antibodies. This test takes only 10 – 15 minutes to perform, and no need expertise or the sophisticate equipment [39].

2.5.6 Advantages and limitations of using different dengue diagnostic tests

There are many useful diagnostic tests for dengue infection, such as virus detection, RNA detection, antigen detection, serological tests and all these tests can confirm dengue infection. For virus detection, it can identify serotypes of dengue virus but it takes more than one week for the results and cannot distinguish between primary

and secondary dengue infection. For RNA detection, it can also identify serotype and genotype of the virus, which results in 24-48 hours. However, it cannot differentiate between primary and secondary infection. For antigen detection, it is easy to perform the test and the price is cheaper than virus detection or RNA detection, but it is not sensitive as them. Serological tests can differentiate between primary and secondary dengue infection, it is also less expensive and easy to perform the test but it requires two or samples for confirmation [37]. These are great benefits and their limitations using different diagnostic tests is shown in table 2.10.

Diagnostic tests	Advantages	Limitations
Viral isolation and identification	• Confirmed infection • Specific • Identifies serotypes	• Requires acute sample (0–5 days post onset) • Requires expertise and appropriate facilities • Takes more than 1 week • Does not differentiate between primary and secondary infection • Expensive
RNA detection	• Confirmed infection • Sensitive and specific • Identifies serotype and genotype • Results in 24–48 hours	• Potential false-positives owing to contamination • Requires acute sample (0–5 days post onset) • Requires expertise and expensive laboratory equipment • Does not differentiate between primary and secondary infection
Antigen detection		
Clinical specimens (for example, using blood in an NS1 assay)	• Confirmed infection • Easy to perform • Less expensive than virus isolation or RNA detection	• Not as sensitive as virus isolation or RNA detection
Tissues from fatal cases (for immunohisto-chemistry, for example)	• Confirmed infection	• Not as sensitive as virus isolation or RNA detection • Requires expertise in pathology
Serological tests		
IgM or IgG seroconversion	• Confirmed infection • Least expensive • Easy to perform	• IgM levels can be low in secondary infections • Confirmation requires two or more serum samples • Can differentiate between primary and secondary infection*
IgM detection (single sample)	• Identifies probable dengue cases • Useful for surveillance, tracking outbreaks and monitoring effectiveness of interventions	• IgM levels can be low in secondary infections

Figure 1.10 Advantages and limitations of different dengue diagnostic tests

The table taken from evaluation of diagnostic tests: dengue [37]

2.6 Combination of NS1 antigen and IgM antibody results.

Research conducted by Stuart [40] showed that detection of IgM antibody, Panbio tests had sensitivity of 83.2% and specificity of 87.8% and detection of IgG antibody had sensitivity of 39.8% and specificity of 95.3%. In comparison, combined IgM antibody and NS1 antigen had overall sensitivity of 87.9% and specificity of 84.5%. Therefore, combining the NS1 antigen and IgM antibody testing provides both high levels of sensitivity and specificity.

Other flaviviruses, including Japanese encephalitis virus (JEV), it is a virus belonging to the family *Flaviviridae*, the same family of the dengue viruses [41]. Across-reactivity of the antibodies with JEV exists and is considered as part of dengue diagnosis i.e. Afrims' interpretation [42] and the SDB's algorithm [43].

In the study of the causes of fever in children at AHC also used the combination of NS1 and IgM antibody testing to interpret the final patient diagnosis in the following way:

The Panbio Japanese Encephalitis Dengue IgM Combo ELISA was retrospectively used for reference serology which will be used as the final diagnosis in this study. Panbio Units were calculated by multiplying the index by 10. The results were classed as negative for dengue virus and JEV if PanBio units were <9, equivocal if 9-11 and positive if >11. If both anti-dengue virus and anti-JEV IgM results were positive, the anti-JEV result was divided by the anti-dengue virus result to give a ratio, with >1 indicating JEV infection and <1 indicating dengue virus infection. All ELISAs were repeated if the positive, negative or calibrator samples were out of range, or the results equivocal (repeated equivocal results counted as negative). NS1 antigen ELISA (Standard Diagnostics, Korea; Lot # 224007) was used to detect dengue NS1 antigen [13]. To interpret the final patient results, the results of NS1 antigen and anti-dengue/JEV IgM ELISAs were combined for children that had paired serum collections and those that only had one serum collected (figure 2.11).

	Diagnosis	DENV NS1	Paired sera DENV IgM	JEV IgM	Single serum DENV IgM	JEV IgM	JEV/ DENV ratio
DENV	Acute DENV	Positive	Negative, static, falling	Rising[a], falling[b], static[c], negative[d]			<1
	Acute DENV	Negative, positive	Rising	Rising, falling, static, negative			<1
	Acute/recent DENV	Negative	Static	Static, negative			<1
	Acute/recent indeterminate flavivirus	Negative	Falling	Static, negative			<1
	Acute/recent indeterminate flavivirus	Negative	Rising	Negative, static, falling			>1
	Acute/recent DENV				Positive	Positive, negative	<1
JEV	Acute JEV	Negative	Negative, static, falling	Rising			>1
	Acute/recent JEV	Negative	Negative, static	Static			>1
	Acute/recent indeterminate flavivirus	Negative	Negative, static	Falling			>1
	Acute/recent indeterminate flavivirus	Negative	Negative, static, falling	Rising			<1
	Acute/recent JEV				Negative, positive	Positive	>1
Mixed Dengue/ JEV	Acute/recent indeterminate flavivirus	Negative	Falling	Falling			>1 or <1

Figure 2.11 The result interpretation of combined Panbio and DEN NS1 antigen

This table taken from the study of the causes of fever in children at AHC

2.7 Management of dengue infections

To present, there are no specific antiviral drugs to treat dengue fever. The management of dengue depends on clinical manifestations and the presentation of a patient at an emergency department, some patients may be sent home after the physical examination and laboratory investigation, and some patients may be hospitalized for close monitoring based on the degree of disease severity. The use of fluid administration has become a cornerstone in dengue management depending on the degree of disease severity. Mild dengue infection is required bed rest, adequate oral fluid replacement and antipyretics, it is not necessary for hospital admission. On the other hand, in cases of patients with DHF or DSS, fluid administration or intravenous volume replacement should be cautiously used under close monitoring in order to reduce resultant complications and saving patients' lives.

2.8 Prevention and control of dengue infections

Due to the fact that there is no cure for dengue infection, so an appropriate strategy of dengue is prevention and control in order to reduce the risk of illness, mortality rate and save human lives. Health education on sufficient knowledge of dengue, its transmission, severity and prevention is very important sustaining prevention and control at personal and community levels, especially in remote and rural areas. There have been several KAP, which is a study survey assessing the level of knowledge, attitude and practices on the particular disease among the community. A study conducted in urban and rural communities where the knowledge of dengue infection were often obtained from media sources such as television and radio [44]. In Laos, in peri-urban district, people in the community they heard about dengue infection but their in-depth perception of dengue knowledge is not sufficient for preventing themselves [45].

2.8.1 Prevention of mosquito bites

Wearing long dresses and covering the upper and lower limbs during the daytime is a one way of personal protection in order to reduce contact or the bites from mosquitoes. Apart from wearing long dresses, the use of mosquito repellent is also a

protective way, especially in children and elderly people. Furthermore, using bed nets is a suggestive way for people who want to relax or rest during the daytime. Using curtains to cover the windows and doorways for preventing mosquitoes come inside the house.

2.8.2 Prevention of multiplication of mosquitoes

Elimination of mosquito breeding sites such as wastewater, stagnant water, water-holding containers, tanks, flower vases, coconut shells, plastic bags, and tree holes by draining out the unused water or water containers should be covered all the time after using in order to prevent mosquitoes laying eggs on the surface of the stagnant water.

2.8.3 Early detection and outbreak control

National surveillance data on dengue have often been shown to significantly underestimate true incidence and burden for a number of diseases. This is also the case in Cambodia [46]. Limitations of case recognition included weak diagnostic facilities (Griffiths 2013) and clinical diagnosis is mainly used to confirmed dengue without any laboratory results. In case of outbreak response, many countries face with a lack of budget and poor coordination to follow some recommendations such as spraying on a sufficiently large scale early in an epidemic (WHO guideline 2009).

2.9 Dengue vaccine

Presently, there is no vaccine to protect against dengue virus. There are many dengue vaccine development for preventing all four dengue serotypes, until now none are sufficient to prevent the disease [47]. So the best way to avoid getting the infection is personal protection and elimination of mosquito breeding sites.

2.10 Comparison of WHO 1997 and 2009 criteria for dengue classifications

The revised dengue guidelines or the WHO 2009 guideline, it highlights the severity of dengue combining at least one of three groups: severe plasma leakage that may cause shock, severe bleeding or severe organ impairment, but on the contrary, the WHO 1997 determines a DSS case to coincide with bleeding, thrombocytopenia, plasma leakage, circulatory failure, undetectable blood pressure and pulse [29].

Victor (2013) found that both WHO 1997 and WHO 2009 dengue classification were inconsistent in identifying severe illness in 2004 and 2007 dengue endemic in Singapore, 15% of DSS patients were categorized as dengue without warning signs and there were 10.8% of severe dengue patients did not meet DHF criteria. Using WHO 2009 dengue classification, the duration of illness in hospital was increased, and it would be up to 51.3% in both severe dengue and dengue with warning signs, but on the other hand it was only 17.0% of hospitalization of DHF and DSS cases based on WHO 1997 dengue classification [29].

There are some recommendations for current WHO (WHO 1997) and the revised WHO (WHO 2009). Siripen (2011) suggested that the current WHO classification should continue using, owing to the revised WHO classification highlights in dengue with or without warning signs, this may increase the amount of work to healthcare workers, and a number of cases with suspected dengue infection. Furthermore, the WHO 1997 classification should be adjusted to make it easier to use or understand in simple way. Another recommendation is to solve the plasma leakage problem, the presence of plasma leakage should be considered as the first concern or major criteria [25].

2.11 Dengue infection in Cambodia

Cambodia is classified as low-income country. A country is situated in South-East Asia, an estimated population 14.9 million in 2012. It covers 181,035 km^2 dividing into 24 provinces and 180 districts (see figure 2.12).

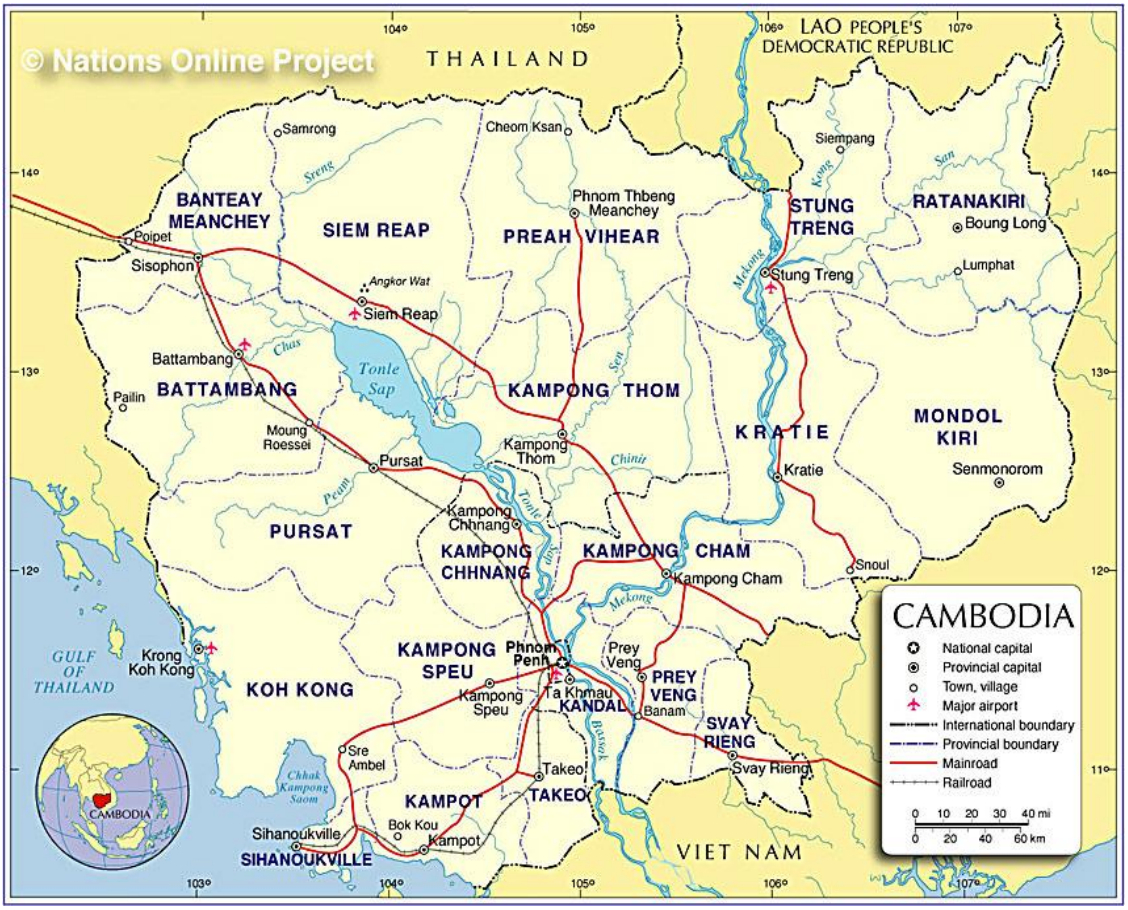


Figure 2.12 Map of Cambodia

The map of Cambodia taken from this website:

(http://www.nationsonline.org/oneworld/map/cambodia_map.htm)

The first detection of dengue virus infection in Cambodia occurred in 1963 and has been informed to national surveillance of dengue since 1980 [6]. Overall, from 1980 to 2008, there were 194,726 dengue cases, which reported by the National Dengue Control Program (NDCP). In the table below (table 2.1), the NDCP reported the percentage of cases of DF, age-specific of DHF, DSS and virus serotypes from 2000 to 2008. A number of cases were highest during the large endemic in 2007 and infants aged < 1 year were the highest age-specific incidence of DHF and DSS. Moreover, the predominant serotype of dengue virus in 2007 dengue outbreak in Cambodia was DENV-3 [6].

Table 2.1 Cases of DF, DHF and DSS reported by the National Dengue Control Programme, Cambodia, 2000-2008

Parameter	Year of surveillance								
	2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
DF cases, no.	3145	10 266	12 441	12 099	9991	9006	16 635	39 618	9546
DHF, %	not determined	not determined	27.8	20.5	41.5	34.6	54.0	51.2	42.9
DSS, %	not determined	not determined	5.4	7.0	8.7	8.4	6.6	3.0	3.6
DHF and DSS, %	not determined	not determined	33.2	27.5	50.2	43.0	60.6	54.2	46.5
DF case facility rate, %	not determined	not determined	1.2	1.6	0.9	1.7	0.9	1.0	0.7
Age-specific incidence of DHF and DSS (per 1000 population)									
< 1 year	not determined	not determined	0.45	0.25	0.59	0.61	2.43	5.49	1.01
1-4 year	not determined	not determined	0.58	0.71	1.00	0.72	2.15	4.07	0.77
5-9 year	not determined	not determined	1.22	0.77	1.14	0.91	2.21	4.41	0.86
10-14 year	not determined	not determined	0.54	0.44	0.65	0.42	0.98	2.30	0.54
15-19 year	not determined	not determined	0.00	0.01	0.01	0.03	0.05	0.27	0.07
Specific virus serotype, %									
DENV-1	5.6	23.3	21.0	10.4	3.3	5.5	5.7	4.3	10.8
DENV-2	24.4	20.1	41.0	61.2	74.1	45.3	9.2	9.1	44.1
DENV-3	58.9	45.0	18.0	15.5	16.7	39.4	82.2	83.6	19.7
DENV-4	11.1	11.6	20.0	12.9	5.9	9.7	2.9	3.1	25.4

During dengue endemic in 2007, as a result, rural areas were more likely to face a higher risk of dengue infection than urban areas (incidence 71 versus 17/1,000 person-seasons) [14]. There remained under-recognition to the National Dengue Surveillance System (NDSS), most often for hospitalized dengue cases, especially during the huge outbreak of dengue in 2007, most of the cases were reported as cases with complications. Another issue is due to Cambodia still remains using WHO 1997 dengue classification, only dengue case definition could be hospitalized, and many cases that are severe than those DHF/DSS likely to be missed by the National Dengue Surveillance System, so under-reported cases of dengue in 2007 dengue outbreak were higher when comparing to other years [15]. An overall annual dengue cost in 2007 was US$14,429,513 which was higher than annual cost for dengue illness in year 2006 and 2008, as shown in table 2.2 [16].

Table 2.2 Cost of dengue illness by year ($US), 2006-2008

	2006	2007	2008
Cost of illness (US$)	5,771,079	14,429,513	3,327,284

This table taken from the study of cost and disease burden of Dengue in Cambodia

In the study of the causes of febrile children admitted to Angkor hospital, dengue virus was commonly microbiological diagnoses, it accounted for 16.2% among children aged 5-16 years, and five cases of dengue virus infection died during the study [43].

2.12 Data mining

It is the procedure of discovering pattern interestingness in large data sets. It generally includes data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation.

Interesting patterns substitute for knowledge. It can be proceeded on any type of data. It has combined technologies from many other domains such as statistics, machine learning, database and data warehouse systems, and information retrieval.

Popular techniques in data mining compose of: the first technique, decision trees which are used for classification and prediction. Classification tree is a subset of decision trees which can be used to classify by building a decision tree to provide better understanding of data set. It can be used both in medical and other sectors, such as marketing, manufacturing and finance. The second technique, clustering is the technique to separate data set into smaller subset. This can be used to organize the data into group which represent similarity or dissimilarity. The third technique, association rules: finding frequent patterns, associations, correlations among sets of items. As for the purpose of gaining a predictive algorithm which can practically be used to guide dengue diagnosis, we will focus on the decision tree technique.

2.12.1 Classification and Regression Tree (CART) Analysis

A recursive partitioning technique, is a tool for predicting both categorical predictor variables (classification tree) and continuous dependent variables (regression tree)**.** CART is very popular algorithm for building decision trees and commonly used in data mining techniques.

A classification tree has three types of nodes. The top node is called the root node and contains all the observation in the sample. The second type of node consists of terminal nodes that are the end nodes assigned to one of the classes. The other nodes are called as non-terminal nodes. CART analysis composes of four procedures: the first procedure, tree building, the method of building tree is recursively split nodes, from root node to child nodes, every node is given to predict an outcome. The second procedure, stopping tree building after the process of building tree is maximum, over fitting, and incapable of continuing the tree. The third procedure, tree

pruning, creating a set of simpler trees sequentially. The fourth procedure, optimal tree selection, is the process of reducing the over fitting and giving the best fit for data set.

In this study, we used CART to build a classification tree to predict the severity of dengue fever among children admitted to Angkor hospital in Siem Reap province, Cambodia. For this, clinical features and laboratory indicators of children admitted between October 12th 2009 and October 12th 2010 using data obtained from a prospective study of the causes of febrile illness requiring hospitalization in children in Cambodia.

2.12.2 Tree pruning

Tree pruning is the reduction of the size of the decision trees, preventing over-fitting in noisy data. The purpose of pruning is reduced complexity of terminal nodes and provides predictive accuracy.

2.12.3 Review of studies using Data mining techniques

Research conducted by Lee and Lye (2009) suggested that a simple decision tree algorithm in predicting adult patients with DHF was efficient in the clinical setting. The decision tree algorithm for dengue had only tree branches, including a history of bleeding, the level of urea and the total protein levels. Using this decision tree would increase classification of DHF and would reduce 43.9% of dengue hospitalizations. There were some limitations in this study, such as the number of patients with DHF were quite small and during early stage of illness, patients were hardly indicative of having dengue infection, they were mostly identified as dengue owing to constant fever, low platelet counts and petechiae [28].

James and Robert (2010) showed that developing diagnostic algorithm using CART was applicable for identifying patients who were at increased risk of developing dengue infection severity and those were at low risk using early clinical laboratory indicators within the first three days of illness for discrimination. The tree had a sensitivity of 97% to specify patients who would develop DSS, and correctly classified 48% of mild dengue or other febrile illness patients. The limitations of this study, firstly, only pediatric patients and two hospitals were included. Secondly, this

study recruited patients only 72 hours after illness. Therefore, this diagnostic decision tree may inadequately indicate clinical practice outside [10].

2.12.4 Advantages and Disadvantages of CART

CART analysis has many advantages over other alternative methods, containing multivariate logistic regression. Firstly, it is non-parametric, it can manipulate numerical data which are highly skewed or multi-modal, and categorical predictors with ordinal or non-ordinal structure. Secondly, CART has experienced methods for dealing with the problems of missing variables, so beneficial CART trees can be created even when significant predictor variables are not recognized for all patients. Another benefit of CART analysis is quite automatic machine learning technique, comparing to the complexity of the analysis, little input is needed from the analyst. Ultimately, CART trees are easy for non-statisticians to comprehend.

For disadvantages of CART, first, CART analysis is new and unfamiliar, therefore there may be some refusals to accept it by conventional statisticians. Furthermore, there is some reasonable skepticism concerning tree methodologies from unreliable claims and poor performance of previous methods. Owing to its novelty, it is hard to find expert statisticians in CART.

2.13 Logistic regression analysis

It is type of predictive model, which the outcome variable is categorical. A variable is referred to a binary or dichotomous.

Logistic regression model is a statistical model, a linear model in the log-odds scale, describing the association between two dichotomous data and providing odds ratio of interest by calculating the probability of being a case over the probability of non-case.

The logistic function:

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x,$$

$\pi(x)$= the probability (the dependent variable = a case)

β_0 = intercept

$\beta_1 x$= regression coefficient

- Interpreting the odds ratio (OR)

If OR = 1 the exposure is not related to the disease

If OR > 1 the exposure is positively related to the disease

If OR < 1 the exposure is negatively relates to the disease

2.13.1 Review of studies using logistic regression

David (2005) showed that the outcomes of logistic regression analysis using the clinical presentation and laboratory tests for predicting the likelihood of disease, rash and total white blood cell count were more likely to be significant and associated with dengue fever [48].

The study conducted by [49] showed that the revised dengue classification or WHO guidelines 2009 was more likely to detect severe dengue compared to the WHO guidelines 1997 by using binary logistic regression in two classification systems.

2.13.2 Advantages and Disadvantages of using logistic regression

Logistic regression analysis has advantages that it can be used in many study designs, such as cohort, case-control and cross-sectional study. Another benefit is when the outcome is uncommon or unusual, the Odds Ratio is a good estimate of the Relative Risk. Furthermore, logistic regression is easy to perform by using many statistical software. Lastly, it provides multivariate analysis. For the disadvantages of using logistic regression are when the outcome is common, the Odds Ratio overestimates the Relative Risk and the interpretation of the Odds Ratio as Relative Risk causing its possible exaggeration. For non-statistician, the outcomes of logistic regression as log odds may seem more difficult to interpret when compare with the discrete scoring system in CART.

CHAPTER III
MATERIAL AND METHODS

3.1 Study site and population

Siem Reap province is located in northwest Cambodia near the famous temple of Angkor Wat. It covers 10,299km^2, composing of 12 districts, 100 communes and 875 villages. The province capital is Siem Reap (see figure 3.1). The population is about 903,030 people in 2007.

Figure 3.1 Map of Siem Reap province, Cambodia

The figure taken from **(**http://www.travelfish.org/country_map/cambodia**)**

Figure 3.2 Angkor Hospital for Children (AHC), Siem Reap province, Cambodia (http://frank.itlab.us/photo_essays/wrapper.php?sep_18_2010_clinic_sojourn.html)

The study site is at Angkor hospital for children (see figure 14). It is a 50-bed children's hospital in Siem Reap province in Cambodia, providing free health care to children less than 16 years of age, including inpatient and outpatient care. Additionally, it also repays travel expenses, provide food for free and materials for sleeping to families during hospitalization. This children's hospital provides emergency, surgical, intensive care treatment, dental and ophthalmologic care and antiretroviral HIV therapy.

Febrile children admitted to Angkor hospital for children in Siem Reap province, Cambodia were included for enrollment. There were 1225 episodes of febrile illness in 1180 children, including 198 serologically confirmed dengue episodes, 941 non-dengue episodes and 86 episodes were no samples, thus, we excluded it from our analysis.

3.2 Study design

The study will be conducted a retrospective cohort study aiming to discriminate severe dengue during the early phase of illness from other febrile illnesses among children admitted to Angkor hospital for children (AHC) in Siem Reap province, Cambodia, using data taken from a prospective study of the causes of febrile illness requiring hospitalization in children in Cambodia. Data were collected between 12th October 2009 and 12th October 2010 from patients who admitted to AHC. These data include demographic, clinical, laboratory and outcome data. The Case Record Form (CRF) used in this fever study is attached as an appendix.

3.3 Inclusion and exclusion criteria

3.3.1 Inclusion criteria

The following criteria are the eligibility criteria

- Patients aged less than 16 years
- Documented axillary temperature $\geq$ 38.0 ◦c within 48h of admission
- Patients admitted to Angkor hospital for children
- Informed consent by parent/caregiver

Our study all patients had to have fever with two or more of the following signs: headache, retro-orbital pain, vomiting, rash, myalgia, arthralgia, positive tourniquet test, leukopaenia for classifying as DF. For DHF classification, requiring criteria of fever, thrombocytopenia $\leq$ 100,000/ μL, haematocrit rise $\geq$ 20%, spontaneous bleeding, evidence of plasma loss or DSS criteria using WHO guidelines 1997. For severe dengue, requiring any of the criteria such as severe plasma leakage, severe bleeding and severe organ involvement using WHO dengue classification 2009. For positive dengue serology, the combination of NS1 antigen and the dengue IgM ELISA (Panbio, Australia) are used for patients who had the results of admission and discharge specimens (paired serum) and those who had single serum.

3.3.2 Exclusion criteria

- Febrile post-surgical patients were excluded from the study.

3.4 Study variables

Demographic, history/symptoms, clinical, laboratory data will be used for constructing the algorithm and analyzing the data, for each data are detailed in table 3.1, 3.2, 3.3, 3.4 and 3.5.

Table 3.1 Demographic, history/symptoms, clinical and laboratory features for dengue patients

Demographic data	**History / Symptoms**	**Clinical parameters**	**Laboratory parameters**
Gender Age groups	Number of days of fever Vomiting Abdominal pain Headache or retro-orbital pain	Rash Temperature ◦c Pulse /minute Capillary refill time Respiratory rate Liver enlargement Glasgow coma score Positive tourniquet test	Haematocrit, % Platelet, x$10^3/\mu l$ White blood cell, x$10^3/\mu l$ Neutrophil Lymphocyte Urea, mmol/L Creatinine, μmol/l Alanine transaminase, IU/l Urinary protein mg/dL Urinary red blood cells SDB DEN/JEV interpretation

Table 3.2 Demographic data for dengue patients

Demographic data					
Name	**Data explanation**	**Data variable**	**Missing**	**Data type**	**Code**
Gender	Male, Female	gender	0 (0%)	binary	Male = 1, female = 2
Age groups	< 28 days 28 days to < 1 year ≥ 1 year to < 5 years ≥ 5 years to < 16 years	agegroups	0 (0%)	categorical	< 28 days = 1 28 days to 1 year = 2 1-5 years = 3 Over 5 years = 4

Table 3.3 History/symptoms for dengue patients

History / Symptoms				
Data Variable	**Data explanation**	**Missing**	**Data type**	**Code**
afeverdays	Number of days of fever	18 (1%)	continuous	
hvomiting	Vomiting	0 (0%)	categorical	No = 1, Yes = 2, Not assessable = 3
kabdominalpain	Abdominal pain	0 (0%)	categorical	No = 1, Yes = 2, Not assessable = 3
mnheadacheorretroorbitalpain	Headache or retro-orbital pain	0 (0%)	categorical	No = 1, Yes = 2, Not assessable = 3

Table 3.4 Clinical features for dengue patients

Clinical parameters							
Data variable	**Data explanation**	**Range (min-max)**	**Mean**	**Standard deviation**	**Missing**	**Data type**	**Code**
rrash	Rash				0 (0%)	categorical	No = 1, Yes = 2, Not assessable = 3
ctemppres	Maximum temperature before enrollment	36 – 42	38.684	0.588	0 (0%)	continuous	
epulsepres	Pulse per minute	0 – 220	139.378	27.744	1 (0%)	continuous	
gcrtpres	Capillary refill time	2 – 6	2.094	0.376	13 (1%)	continuous	
hrrpres	Respiratory rate	0 - 88	43.028	14.876	4 (0%)	continuous	
ciiihepatomegaly	Liver enlargement				0 (0%)	categorical	No = 1, Yes = 2, Not assessable = 3
igcstotalpres	Glasgow coma score	3 - 20	14.458	2.079	16 (1%)	continuous	
vitourniquettest	Positive tourniquet test				85%	categorical	No = 1, Yes = 2, Not assessable = 3

Table 3.5 Laboratory features for dengue patients

Laboratory parameters							
Data variable	**Data explanation**	**Range (min-max)**	**Mean**	**Standard deviation**	**Missing**	**Data type**	**Code**
haemhc	Haematocrit	6 – 78	30.937	7.389	71 (6%)	continuous	
haemplts	first recorded platelet result	0 - 1288	351.182	206.087	71 (6%)	continuous	
haemwbc	first recorded white blood cell result	0.5 - 328	15.015	21.44	72 (6%)	continuous	
haemneuts	first recorded neutrophil result	0 – 93	8.406	7.027	75 (6%)	continuous	
haemlymph	first recorded lymphocyte result	0 – 88.5	4.072	4.787	77 (6%)	continuous	
biochemurea	first recorded Urea result	0.9 - 68	4.782	5.16	109 (9%)	continuous	
biochemcreat	first recorded Creatinine result	21 - 710	66.135	36.081	118 (10%)	continuous	
biochemalt	first recorded ALT result	3.6 - 1757	66.395	142.822	101 (8%)	continuous	
urineprotein	The presence of urinary protein	0 - 300	17.785	49.933	160 (13%)	continuous	
urinerbc	The presence of urinary red blood cells	0 - 100	5.463	18.552	175 (14%)	continuous	
Sdbdenjevinterpretation	SDB DEN/JEV interpretation				0 (0%)	categorical	Non-dengue = 0 Dengue = 1

3.5 Statistical analyses

3.5.1 Descriptive statistics

They are used to describe the basic features of our data including demographic data, history /symptoms, clinical and laboratory parameters, as shown in table 3.2, 3.3, 3.4 and 3.5.

3.5.2 Analytical statistics

3.5.2.1 Model 1 (non-dengue vs. dengue)

Model 1 is used for dengue prediction, a patient had dengue infection or non-dengue based on clinical features and laboratory indicators.

We used the attribute "Sdbdenjevinterpretation" or SDB DEN/JEV interpretation as a dependent variable (see table 3.6), whereas gender, age groups, number of days of fever, vomiting, blood or mucous in stool, abdominal pain, headache or retro-orbital pain, rash, temperature, pulse, capillary refill time, respiratory rate, Glasgow coma score, hematocrit, platelets, white blood cells, neutrophil, lymphocyte, urea, creatinine, ALT, urine red blood cells and urine protein are independent variables.

Table3.6 Dependent variable of model 1

Attribute "Sdbdenjevinterpretation"	**Code**
Non-dengue	0
Dengue	1

3.5.2.2 Model 2 (non-severe vs. severe dengue)

A total of 198 dengue episodes, we thoroughly categorized patients into two groups: non-severe dengue and severe dengue based on ICU admission and WHO 2009 dengue classification. In our data set, grading severity of dengue could be difficult due to clinical presentation of dengue and laboratory abnormalities, especially if we only had the first recorded hematocrit, platelet, white

blood cell, neutrophil, lymphocyte, urea, creatinine and ALT result, including the presence of urinary protein and red blood cells. Note that these were not the results of routine laboratory investigations, but they were recorded at first presentation to hospital. Thus, in our case the attribute ICU admission had priority over WHO 2009 classification, besides, our data set did not have information on the results of chest x-ray to evaluate pleural effusion and the results of abdominal ultrasound to detect peritoneal fluid (ascites). Owing to we only had the attribute "blood or mucous in stool" in our data set, so we used this attribute for evaluating bleeding.

Out of 198 dengue episodes, we found that 43 episodes were required intensive care unit admission, but only 29 episodes were categorized as severe dengue based on their clinical signs of severity such as low hematocrit and platelets, liver enlargement, ALT $\geq$ 1000 and other laboratory abnormalities. For low hematocrit, if patients had comorbidities: anaemia or hookworm, these could explained by low hematocrit and were classified as non-severe dengue, meaning that they were severe from other confirmed diagnosis. For low platelet count, this suggests that dengue was severe. For liver enlargement, this could be signs of fluid shift. For low Glasgow coma score, if patients had co-morbidities such as seizure, clinical stroke or meningitis, they were categorized as non-severe dengue not convincing for severe dengue.

After categorizing all 198 dengue cases, we conclude that 160 patients were classified as non-severe dengue due to they had other confirmed diagnosis and no real signs of dengue severity, only 38 episodes were classified as severe dengue. Grading severity of dengue was classified using both ICU admission and WHO 2009 dengue classification as the first opinion, and a pediatrician's diagnosis was the second opinion to help with classifying dengue severity.

Model 2 is used for predicting the severity of dengue, the patient had severe dengue or non-severe dengue based on clinical features and laboratory indicators.

The variable "grading severity" is used for the dependent variable (see table 3.7), whereas gender, age groups, number of days of fever, vomiting, blood or mucous in stool, abdominal pain, headache or retro-orbital pain, rash, temperature, pulse, capillary refill time, respiratory rate, Glasgow coma score,

hematocrit, platelets, white blood cells, neutrophil, lymphocyte, urea, creatinine, ALT, urine red blood cells and urine protein are independent variables.

Table 3.7 Dependent variable of model 2

Attribute "grading severity"	Code
Non-severe dengue	0
Severe dengue	1

3.5.2.3 Model 3 (survived vs. died)

Model 3 is used for predicting survival and death in dengue cases based on clinical features and laboratory indicators.

The variable "outcome" is used for the dependent variable (see table 3.8), whereas gender, age groups, number of days of fever, vomiting, blood or mucous in stool, abdominal pain, headache or retro-orbital pain, rash, temperature, pulse, capillary refill time, respiratory rate, Glasgow coma score, hematocrit, platelets, white blood cells, neutrophil, lymphocyte, urea, creatinine, ALT, urine red blood cells and urine protein are independent variables.

Table 3.8 Dependent variable of model 3

Attribute "outcome"	Code
Survived	0
Died	1

3.5.3 Statistical techniques

3.5.3.1 CART

CART is used for constructing diagnostic algorithm based on both clinical and laboratory features.

In our models, we used 10-fold cross validation (number of folds used in cross validation), it would split data set into 10 partitions, 9 partitions for training and 1 partition for testing, as shown in figure 3.3.

Figure 3.3 Cross-validation technique

Then we built decision trees with the J48 algorithm changing these parameters:

- ***Confidence factor (CF)***

It is an original technique used for pruning based on confidence intervals. A default value is set to 0.25, lowering the confidence values result in more pruning and general models.

- ***The minimal number of objects (miNumObj)***

It is the minimum number of instances per leaf, a default value is set to 2. Higher values result in smaller trees. A particular node containing the data would be stopped for further partition when the class or leaf node has equal to or less than the "miNumObj values.

We tested the J48 algorithm with Confidence Factor (CF) values ranging from 0.01 to 0.5 (0.01, 0.025, 0.05, 0.1, 0.25 and 0.5) and the miNumObj values ranging from 2 to 15 (2, 5, 10 and 15) of J48 algorithm in order to generate the decision trees to find the most appropriate value. We also applied pruning method to our decision tree for reducing the tree size in order to avoid over-fitting and complexity of our data set. There are two techniques of applying pruning. Firstly, selecting the option “unpruned” to false and then choosing the confidence factor value, secondly, setting the option “reduced error pruning “ to true (see figure 3.4). In our model we only applied the traditional technique, option “unpruned”.

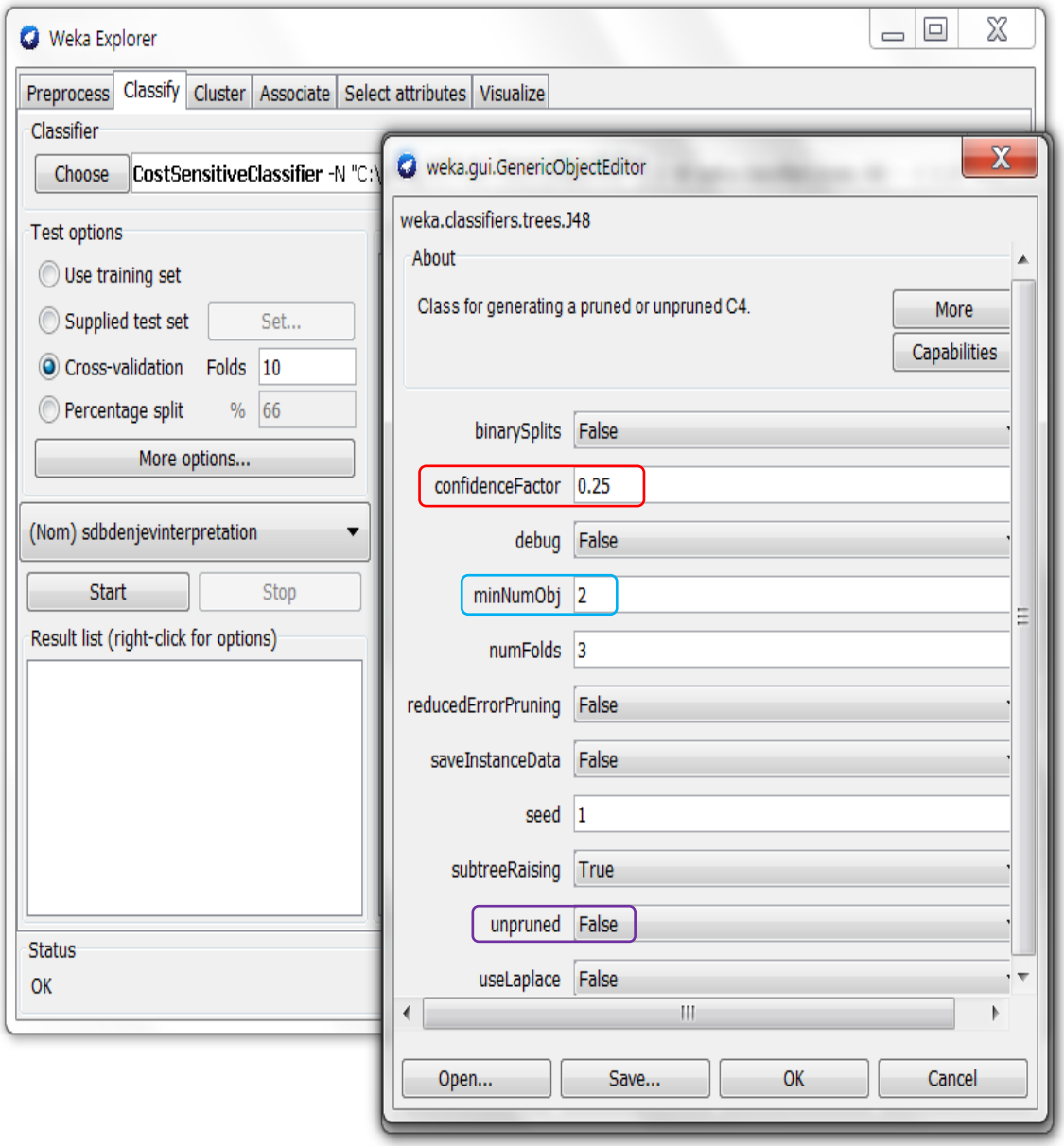


Figure 3.4 Setting the classifier parameters of J48 decision

Besides, we made decisions based on a cost matrix too (see figure 3.5).

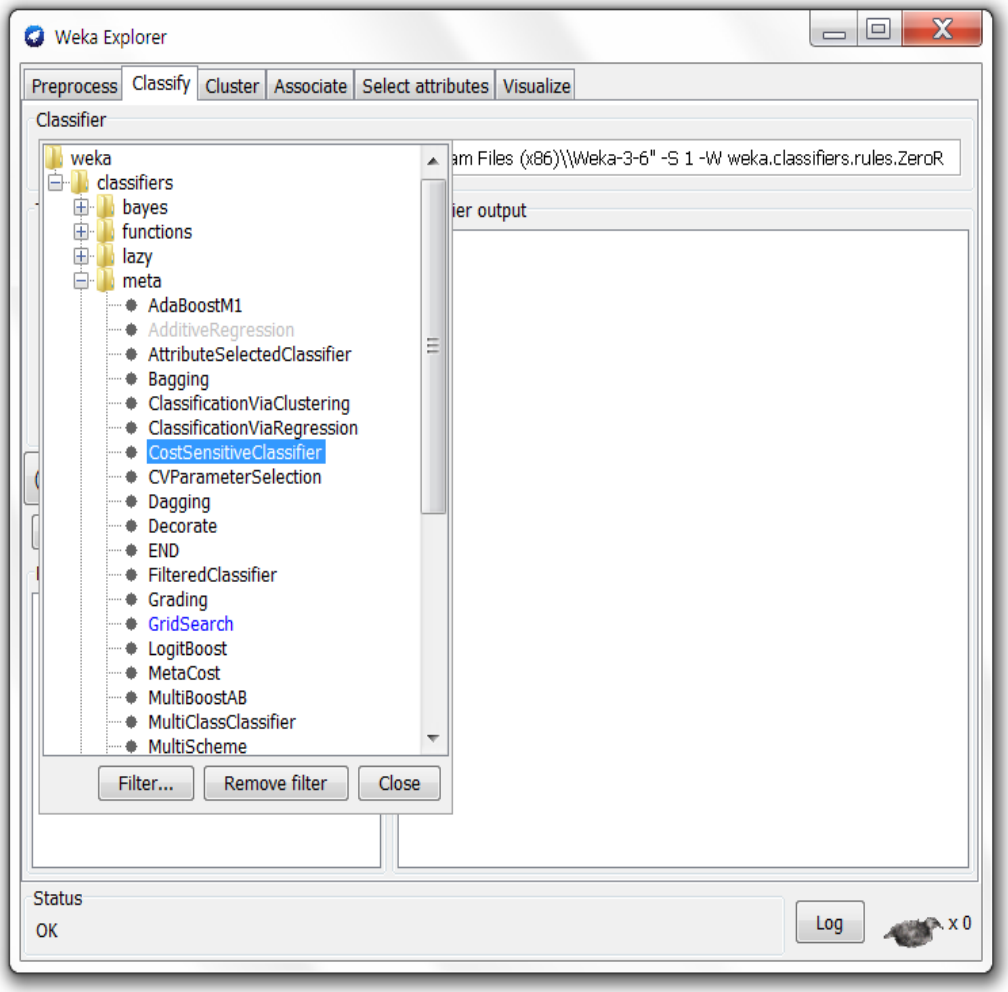


Figure 3.5 Cost sensitive classifier

The cost matrix rows refer to actual classes, while the columns refer to predicted classes, as shown in table 3.9.

Table 3.9 Cost matrix

	Actual class	**Actual class**
Predicted class	true positive (TP)	false negative (FN)
Predicted class	false positive (FP)	true negative (TN)

- **Model 1**

The cost matrix rows referred to actual non-dengue / dengue class, while the columns referred to predicted non-dengue / dengue class. If a dengue patient is diagnosed with non-dengue fever and non-dengue patient is diagnosed with dengue, the cost is assigned to 1, as presented in table 3.10.

Table 3.10 Cost matrix of model 1

		Actual class	
		Dengue	Non-dengue
Predicted class	Dengue	0	1
	Non-dengue	1	0

- **Model 2**

The cost matrix rows referred to actual classes (non-severe and severe dengue), while the columns referred to predicted classes (non-severe and severe dengue). If a patient with severe dengue is diagnosed with non-severe dengue, the cost is assigned to 5. If a patient with non-severe dengue is diagnosed with severe dengue, the cost is assigned to 1 in order to predict the classes with minimum misclassification cost or to minimize the problems of misclassifying severe dengue patients as non-severe dengue patients (see table 3.11).

Table 3.11 Cost matrix of model 2

		Actual class	
		Severe dengue	Non-severe dengue
Predicted class	Sever dengue	0	1
	Non-severe dengue	5	0

- **Model 3**

The cost matrix rows referred to actual classes (survived dengue and died from dengue), while the columns referred to predicted classes (survived dengue and died from dengue). If a patient is confirmed to survive but dies, the cost is assigned to 5. If a patient is confirmed to die but survives, the cost is assigned to 1 in order to minimize the problems of misclassifying died from dengue as survived dengue (see table 3.12).

Table 3.12 Cost matrix of model 3

		Actual class	
		Died from dengue	Survived dengue
Predicted class	Died from dengue	0	1
	Survived dengue	5	0

3.5.3.2 Logistic regression analysis

Binary logistic regression analysis (see section 2.13, page 33) was performed in three models using the attribute "Sdbdenjevinterpretation", "grading severity" and "outcome" as a dependent variable for model 1, 2 and 3, respectively, whereas gender, age groups, number of days of fever, vomiting, blood or mucous in stool, abdominal pain, headache or retro-orbital pain, rash, temperature, pulse, capillary refill time, respiratory rate, Glasgow coma score, hematocrit, platelets, white blood cells, neutrophil, lymphocyte, urea, creatinine, ALT, urine red blood cells and urine protein are independent variables.

For model 3, Owing to a very small number of our cases, maximum likelihood estimation of logistic model could not be able to handle it since there were only eight deaths out of 198 episodes. Trying to fit ordinary logistic regression will cause extremely large standard errors, while at other times the software may report an error or a warning. There are some approaches which can handle this problem. One is using exact logistic regression. Exact logistic regression is complex and may require prohibitive computational resources. Here we show how to use a penalized likelihood method originally proposed by Firth (1993 Biometrika 80:27-38) which can reduce small-sample bias in maximum likelihood estimation. (http://www.cytel.com/software/logxact).

- **Classification table**

It is another technique for evaluating the predictive accuracy of the logistic regression model. The observed values for dependent outcome and the predicted values, our cut-off value used is 0.5 by default (see table 3.13).

Classification Table[a]

Table 3.13 Classification table with the cut-off value 0.5

Observed			Predicted		
			gradingseverity		Percentage Correct
			Severe dengue	Non-severe dengue	
Step 19	gradingseverity	Severe dengue	27	50	50
		Non-severe dengue	11	110	95.6
Overall Percentage					86.9

a. The cut value is .500

- **Odds ratios**

$$\text{Odds} = \frac{\text{probability that the person will have the disease}}{\text{Probability that the person will not have the disease}} = \frac{p}{1-p}$$

Interpreting the odds ratio (OR)

- If OR = 1 the person is not related to the disease
- If OR > 1 the person is positively related to the disease
- If OR < 1 the person is negatively related to the disease

Our models were constructed by applying a method of Backward Likelihood Ratio which removal testing is based on the probability of the likelihood-ratio statistic based on the maximum partial likelihood estimates. Likelihood ratios (LR) are used for assessing a diagnostic test using both sensitivity and specificity to estimate the likelihood of disease in individual patients.

In table 3.14, 3.15 and 3.16 show dependent variable encoding.

Table 3.14 Dependent variable encoding of model 1

Dependent Variable Encoding

Original Value	Internal Value
Non-dengue	0
Dengue	1

Table 3.15 Dependent variable encoding of model 2

Dependent Variable Encoding

Original Value	Internal Value
Non-severe dengue	0
Severe dengue	1

Table 3.16 Dependent variable encoding of model 3

Dependent Variable Encoding

Original Value	Internal Value
Survived	0
Died	1

In our 198 dengue episodes, we observed that there were no febrile infants less than 28 days were confirmed as dengue. Thus, we created new age groups by combining group 1 with goup2, as presented in table 3.17.

Table 3.17 Combination age group 1 and 2 in logistic regression analysis

Old age groups		New age groups	
< 28 days	= group 1	< 28 days to 1 year	= group 1
28days to 1 year	= group 2	1-5 years	= group 2
1 – 5 years	= group 3	Over 5 years	= group 3
Over 5 years	= group 4		

3.6 Performance measuring tools

Sensitivity, specificity and accuracy will be used to assess performance of the two approaches.

- **Sensitivity**

The sensitivity of a test is the proportion of people that are known to have the disease who test positive for it.

- Sensitivity = true positives / (true positive + false negative)
- Calculating sensitivity using 2x2 table (see table 3.18), sensitivity = A/ (A+C)

- **Specificity**

- Specificity relates to the test's ability to identify negative results or the probability of a negative test given that the patient is well.
- Specificity = true negatives / (true negative + false positives)
- Calculating specificity using 2x2 table (see table 3.18), Specificity = D / (B+D)

- **Accuracy**

- The accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's actual (true) value.

- Accuracy = (true positives + true negatives) / (true positives + false positives + false negatives + true negatives)

- Calculating accuracy using 2x2 table (see table 3.18), Accuracy = (A+D) / (A+B+C+D)

Table 3.18 Calculating sensitivity, specificity and accuracy

	Disease (Yes) (as determined by "Gold standard")	Disease (No)	Total
Detected (Yes)	True + (A)	False + (B)	A+B
Detected (No)	False – (C)	True – (D)	C+D
	A+C	B+D	A+B+C+D

- **Kappa statistic**

It measures the agreement of prediction with the true class. A kappa of 1 indicates perfect agreement.

- **TP Rate (True positive rate)**

The proportion of the positive instances correctly classified as positive.

- **FP Rate (False positive rate)**

The proportion of the negative instances correctly classified as positive.

- **Precision**

The number of correctly classified instances divided by the whole classified instances number.

3.7 Software

These software will be used in our study, such as Excel, SPSS and Weka.

3.7.1 Excel

It is used for data format and filter feature to view different values in a variable (a column). It can also be used for storing, organizing and manipulating data.

3.7.2 SPSS

It is used for recoding data, calculating descriptive statistics such as mean, range, standard deviation, variance, minimum and maximum, and measures of kurtosis and skewness. This software also provides graphic display including barchart, histogram, line chart, scatterplot, boxplot, pie chart. Furthermore, it can be used for performing logistic regression analysis providing odds ratio of interest. SPSS for Windows, version 21 (SPSS, Chicago, IL, USA) will be used in the analysis.

3.7.3 Weka

Weka or Waikato Environment for Knowledge Analysis. It is an open source software. It is a tool for data mining and machine learning algorithms created by researchers at University of Waikato, New Zealand. In figure 3.6 shows when opening the software, pre-processing functions is shown in figure 3.7 providing different data mining algorithms, such as classification, clustering, association rules. In figure 3.8 shows 10-fold cross-validation. Weka version 3.6.9 will be used in the analysis.

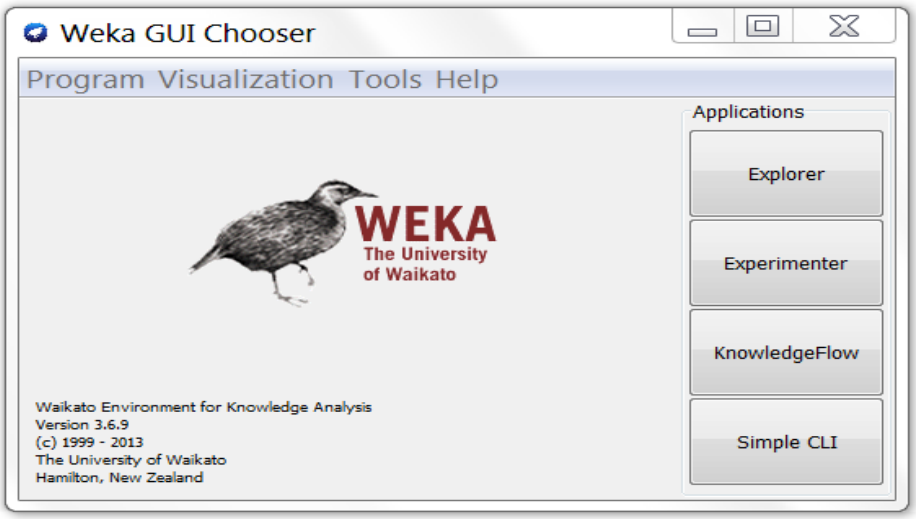


Figure 3.6 Weka starup screen

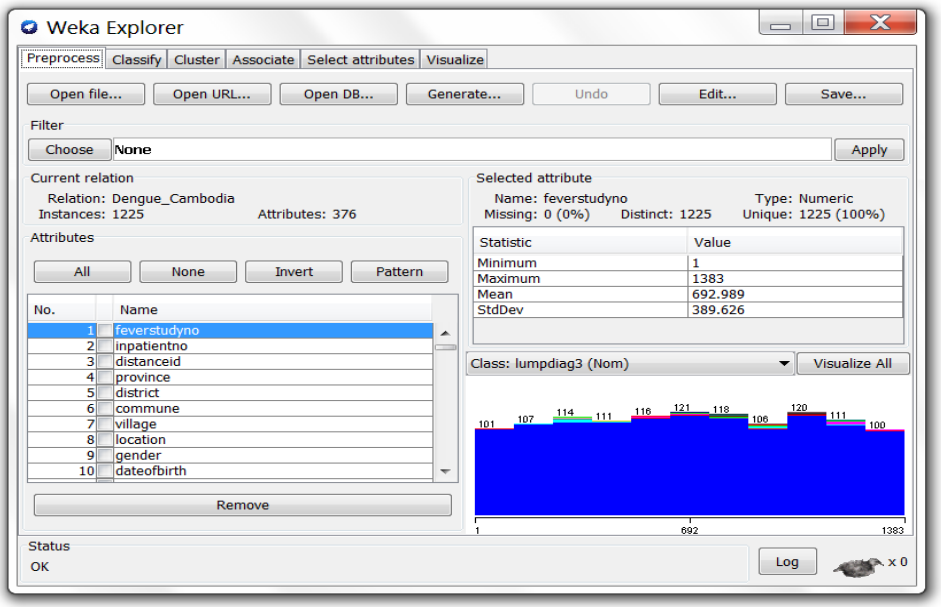


Figure 3.7 Pre-processing functions

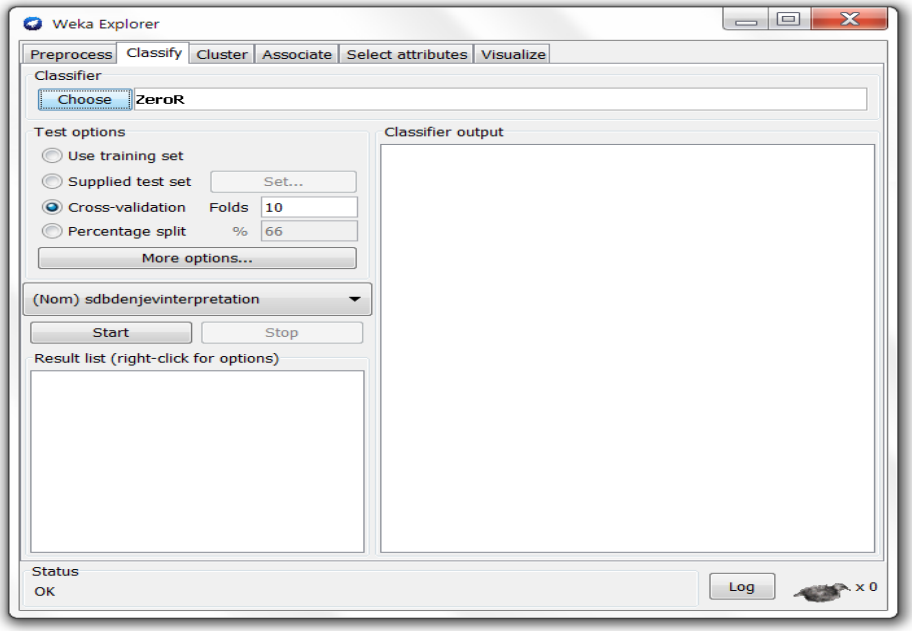


Figure 3.8 Using 10-fold cross-validation

3.8 Data pre-processing

Missing values were replaced with their mean value in case of they were continuous variables. If their mean value were abnormal we would replace missing

values by the median value of the variable, as presented in table 3.19, 3.20 and 3.21. For the attribute tourniquet test, it accounted for 85 % of the clinical record having missing values, so it was considered to be excluded from our analysis.

Table 3.19 Replacing missing values with mean values

Continuous attributes	**Mean values**
afeverdays	18 values – replaced with 4.63
gcrtpres	13 values – replaced with 2.09
igcstotalpres	9 values – replaced with 14.46
biochemurea	77 values – replaced with 4.7819

Table3.20 Replacing missing values with median values

Continuous attributes	**Median values**
haemhc	66 values – replaced with 31
urineprotein	144 values – replaced with 0.1
urinerbc	159 values – replaced with 0
haemwbc	67 values – replaced with 11.6

Table3.20 Replacing missing values with median values (cont.)

Continuous attributes	Median values
hrrpres(28 days to 1 year)	1 value – replaced with 40
biochemalt	69 values – replaced with 28
biochemcreat	86 values – replaced with 60
haemplts	66 values – replaced with 329
haemneuts	70 values – replaced with 6.4
haemlymph	72 values – replaced with3

Table 3.21 Replacing missing values with normal values

Continuous attributes	Normal values
epulsepres (over 5 years)	1 value – replaced with 75
hrrpres(1 – 5 years)	1 value – replaced with 20

CHAPTER IV
RESULTS

4.1 Demographic data

Demographics of the study of the causes of febrile children at Angkor Hospital for Children in Cambodia is shown in table 4.1. Males contained 668 (54.4 %) of episodes and the median age of this study was 2 years. Their study had four different age groups which included febrile infants < 28 days, 28 days to 1 year, ≥ 1 year to < 5 years, ≥ 5 years to <16 years, it is observed that, the number of episodes of aged ≥ 1 year to < 5 years was higher than other age groups, and it also accounted for 39.8 % of all episodes.

Table 4.1 Demographic data of 1225 febrile episodes admitted to Angkor hospital in Cambodia

Demographics	**Number or median (IQR[a], range)**
Male	668 (54.5 %)
Median age	2.0 (0.8-6.4, 0.0-16.0)
Numbers in each age range:	
< 28 days	32 (2.6 %)
28 days to 1 year	330 (26.9 %)
≥ 1 year to < 5 years	488 (39.8 %)
≥ 5 years to < 16 years	375 (30.6 %)

[a] IQR interquartile range

4.2 Causes of febrile illness among the children admitted to Angkor Hospital for Children in Cambodia

The results of microbiological diagnoses were found predominantly in dengue virus, which was 198 of 1225 (16.2%) episodes, followed by 96 of 1225 (7.8%) episodes of orientia tsutsugamushi, 71 of 1225 (5.8%) episodes of Japanese encephalitis virus, 65 of 1225 (5.3%) episodes of indeterminate flavivirus, and 37 of 1225 (3.0%) of episodes of staphylococcus aureus. Top five microbiological diagnoses is shown in table 4.2.

There were 1333 clinical syndromes in 1225 febrile episodes were diagnosed, the most common clinical syndromes were lower respiratory tract infection (38.3 %), followed by 312 of 1225 (25.5%) of episodes of undifferentiated fever, 239 of 1225 (19.5%) of episodes of diarrhoeal disease (see table 4.3). There were 69 deaths (5.6%) during this study which top three causes of death were clinically diagnosed pneumonia (19), dengue virus (5) and melioidosis (4) (see table 4.4).

Table 4.2 Top five microbiology diagnoses

Microbiological diagnoses	**Total (n = 1225)**
Dengue virus	198 (16.2%)
Orientia tsutsugamushi	96 (7.8%)
Japanese encephalitis virus	71 (5.8%)
Indeterminate flavivirus	65 (5.3%)
Staphylococcus aureus	37 (3.0%)

Table 4.3 Presenting clinical syndromes diagnosed for 1225 episodes of febrile illness

Clinical syndrome	Total (n = 1225 episodes)
Lower respiratory tract infection	469 (38.3%)
Undifferentiated fever	312 (25.5%)
Diarrhoeal disease	239 (19.5%)
Skin/soft tissue/bone/joint infection	88 (7.2%)
Upper respiratory tract infection	64 (5.2%)
CNS infection	56 (4.6%)
Genitourinary	41 (3.4%)
Abdominal disease/surgical abdomen	21 (1.7%)
Non-infectious cause of fever	43 (3.5%)
Total clinical diagnoses	1333
Acute mortality	69 (5.6%)

Table 4.4 Primary diagnosis for 69 children who died during the study

Primary diagnosis	Number of deaths (n = 69)
Pneumonia	19 (27.5%)
Dengue virus	5 (7.2%)
Melioidosis	4 (5.8%)

4.3 Dengue infection among the children admitted to Angkor Hospital for Children in Cambodia

There were 198 episodes of dengue infections confirmed by SDB DEN/JEV interpretation. By using WHO 2009 dengue classification, we could category dengue into dengue (n = 55 or 27.7%), dengue with warning signs (n = 105 or 53%), and severe dengue (n = 38 or 19.19 %). For dengue infection, children aged between 28 days and 5 years were commonly infected, a large number of children aged over 5 years was classified as having dengue with warning signs and aged 28 days to 1 year were most prevalent for severe dengue (figure 4.1).

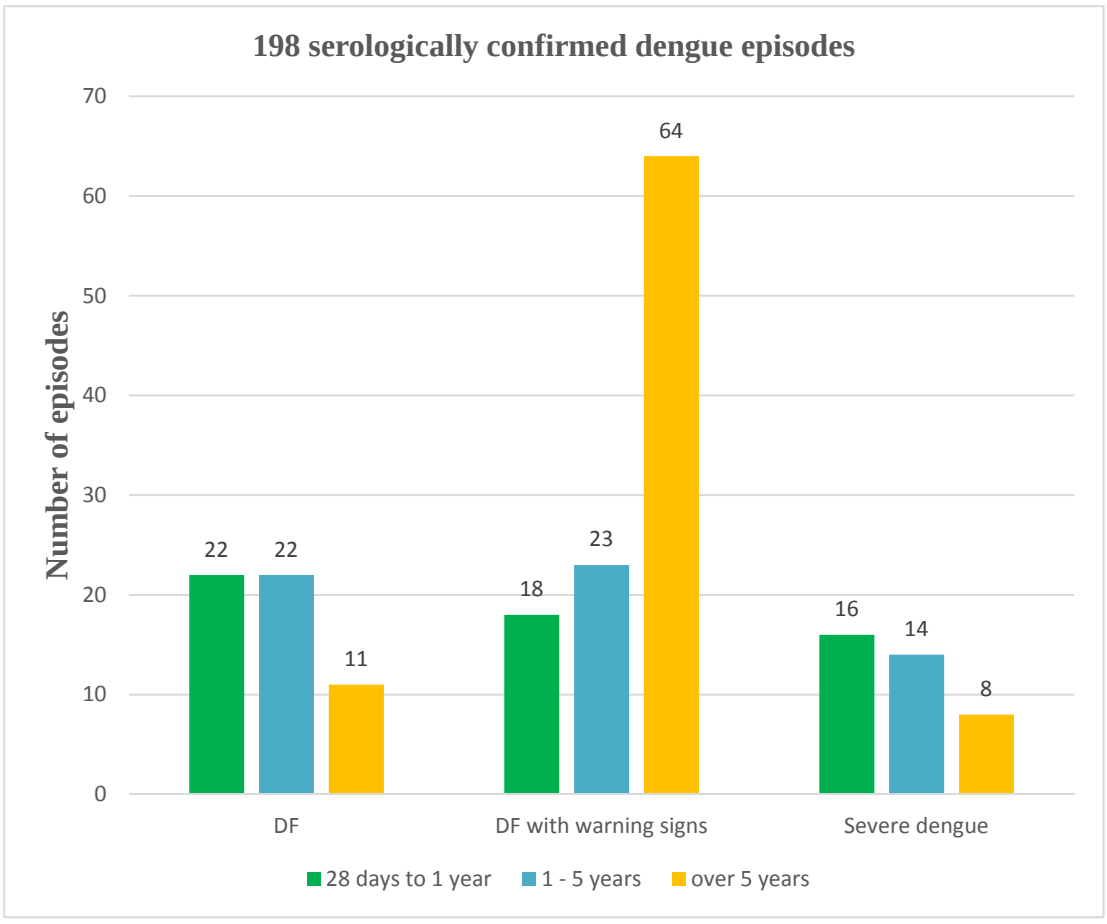


Figure 4.1 Age groups in DF, DF with warning signs and severe dengue

4.4 Classification Analytic Regression Trees (CART) and logistic regression analysis.

There are three following research questions, each will be answered by a specific model (Model 1 – 3) using two different predicting techniques, CART model and logistics regression model.

For CART model, we select the most appropriate model by looking at the sensitivity and specificity together with the classification tree obtained. The model with the highest sensitivity and specificity will be preferable, however, the tree which has more than 6 predicting nodes will be classified as being too complicated and will not be favorable. The researcher made final decision on choosing the most appropriate model. Details of the model selection process are shown in Appendix C.

For logistic regression model, stepwise selection of variables was applied in the model selection. The model with the maximum likelihood estimate will be chosen as the final model.

4.4.1 Model 1 (DF vs. OFI)

Research question: Can we differentiate dengue infections from other febrile illnesses using some early signs and symptoms? We used SDB interpretation as described in section 2.6, page 24 to confirm dengue. The final output in this model is dengue (n = 198) and non-dengue (n = 941).

4.4.1.1 Classification tree

The decision tree is shown in figure 4.2, the first splitting variable is a neutrophil count of 4.2, followed by hematocrit, respiratory rates and platelet count of 154,000/mm^3. The leaf nodes indicates which class would be assigned to a particular leaf (non-dengue or dengue). The numbers in brackets indicate the number of instances assigned to that leaf and represents as decimal number, this number is just an artifact of the algorithm. The number of leaves and size of the our tree were 5 and 9, respectively.

The first splitting " neutrophil or haemneuts " indicates that if neutrophil is greater than 4.2, then the class (leaf node) is assigned to non-dengue.

There were 838 instances reached this “non-dengue” leaf, but 92 instances were not non-dengue (incorrectly classified as non-dengue)

If the neutrophil is less than or equal to 4.2 and hematocrit value is less than or equal to 35 then the class label is also assigned to non-dengue, 224/61 meaning that 224 instances were classified as non-dengue, while 61 instances were classified incorrectly.

If hematocrit value (haemhc) is > 35, and respiratory rate (hrrpres) also ≤ 34 then it is dengue, 47 instances were dengue, but 10 instances were mis-classified.

If respiratory rates (hrrpres) is > 34 and platelet count also ≤ 154, it is dengue, otherwise it is non-dengue.

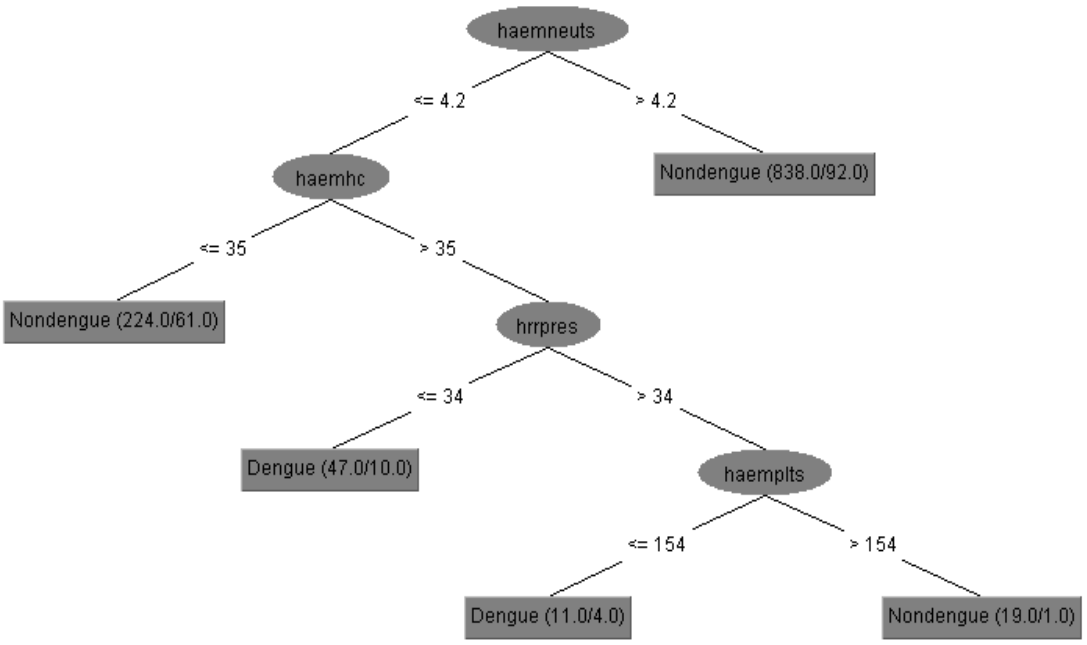


Figure 4.2 Decision tree algorithm for model 1

After assigning confidence factor value and minimum number of instances per leaf in order to prune and reduce the size of the tree, we found an appropriate model with a sensitivity of 16.6 %, specificity of 97.9% and overall accuracy of the

tree was 83.8% (see table 4.5), with a Kappa of 0.206 indicated a slight agreement of prediction with the actual class, followed by precision of non-dengue and dengue.

Table 4.5 Stratified cross-validation of model 1

	Model 1
Sensitivity	16.6%
Specificity	97.9 %
Kappa statistic	0.206
Precision	Non-dengue = 0.848
	Dengue = 0.635

This confusion matrix shows the total of dengue and non-dengue instances. For class “dengue” 33 instances were correctly classified but 165 instances were assigned to class “non-dengue”. For class “non-dengue” 922 instances were correctly classified, 19 were assigned to class “dengue”, as presented in table 4.6.

Table 4.6 Confusion matrix of model 1

		Final confirmation using SDB algorithm	
		Dengue	Non-dengue
Model prediction	Dengue	33	19
	Non-dengue	165	922

4.4.1.2 Logistic regression analysis

The classification table is use to evaluate the predictive accuracy of the logistic regression model is shown in table 4.7. Using the probability cut-off value of 0.5 (default setting), this model had a sensitivity of 14.6%, specificity of 98.2% and overall accuracy was 83.7%.

Table 4.7 Classification table of model 1 by logistic regression model with the cut-off value of 0.5

		Final confirmation using SDB algorithm		
		Dengue	Non-dengue	Percentage correct
Model	Dengue	29	17	83.7 %
	Non-dengue	169	924	
	Overall percentage	14.6 %	98.2 %	

The clinical significance of model 1 is shown in table 4.8, with the odds ratio calculated. Febrile infants aged less than 28 days to 1 year, abdominal pain, increasing hematocrit, low platelet count (below 100,000) were factors which increased the probability of being diagnosed as having dengue fever. The significant odds ratio ranged from 1.030 to 3.895. The presence of low platelet count, infants less than 28 days to 1 year, abdominal pain, hematocrit, white blood cell count, ALT had been shown to have clinical significance in diagnosis of dengue fever [21].

Table 4.8 Final results of model 1 using logistic regression analysis

Variables	**OR**	**95 % CI**		***p* value**
Age over 5 years	1			
- Age group (< 28 days to 1 year)	2.125	1.267	3.564	.004
- Age group (1-5 years)	0.881	0.571	1.359	.566
Abdominal pain	2.718	1.780	4.150	.000
Hematocrit, %	1.030	1.005	1.055	.019
White blood cell count, $x10^3/\mu l$	0.953	0.928	0.980	.001
ALT, U/L	1.001	1.000	1.002	.019
Normal platelet count (100,000 and over)	1			
- Platelet count (19,000)	1.117	0.351	3.555	.852
- Platelet count (20,000 – 40,000)	2.484	1.149	5.374	.021
- Platelet count (41,000 – 99,000)	3.895	2.012	7.543	.000

OR = odds ratio, CI = confidence interval

4.4.1.3 Comparison of model 1 using classification tree and logistic regression Analysis

Observing table 4.10, the variables contribute to the model using classification tree are neutrophil, hematocrit, respiratory rate and platelet count, whereas age, abdominal pain, hematocrit, white blood cell count, ALT and platelet

count contribute significantly to the model when using logistic regression analysis. Hematocrit and platelet count are found in both models.

Both models poorly perform the dengue prediction (sensitivity of 16.6% and 14.6%) where both can predict non-dengue well (specificity of 97.9% and 98.2%), as presented in table 4.9. The sets of initial symptoms identified by these two models may be able to apply when ruling out the possibility of dengue. In the situation where mortality caused by dengue is comparably low, this algorithm can be used to confirm the treatment decision and/or patient management focusing on other causes of fever.

Table 4.9 Sensitivity, specificity and accuracy of two different methods in model 1

Technique	**Sensitivity**	**Specificity**	**Accuracy**
Classification tree	16.6 %	97.9 %	83.8 %
Logistic regression analysis	14.6 %	98.2 %	83.7 %

Table 4.10 Variables contrubute to model 1 using two different methods

Technique	**Neutrophil**	**Hematocrit**	**Respiratory rate**	**Platelets**	**Age**	**Abdominal pain**	**White blood cell count**	**ALT**
Classification tree	√	√	√	√				
Logistic regression analysis		√		√	√	√	√	√

4.4.2 Model 2 (Non-severe vs. Severe dengue)

Research question: Can we differentiate severe dengue from non-severe dengue using some early signs and symptoms?

The final output in this model is based on both ICU admission and WHO 2009 criteria as the first opinion, a pediatrician's diagnosis as the second opinion to help with classifying dengue severity. Finally, the confirmed dengue patients were graded as being severe dengue (n = 38) and non-severe dengue (n = 160).

4.4.2.1 Classification tree

The first splitting variable is pulse rate of 111 per minute, followed by hematocrit, Glasgow coma score, urine protein, creatinine and platelet count. The number of leaves and size of the our tree were 7 and 13, respectively (see figure 4.3).

The first split " pulse ", if pulse is less than or equal to 111, then the class is assigned to non-severe dengue. There were 25.46 instances reached this "non-severe dengue" leaf, but 15.84 instances were not non-severe dengue (incorrectly classified as non-severe dengue)

If the pulse is greater than 111 and hematocrit value is less than or equal to 28 then it is severe dengue, 78.07/15.84 meaning that 78.07 instances were classified as severe dengue, while 15.84 instances were classified incorrectly.

If hematocrit value is > 28, and glasgow coma score also ≤ 11 then it is severe dengue, 18.1 instances were severe dengue, but 1.13 of the instances was mis-classified as severe dengue.

If urine protein is > 0, and creatinine also > 84 then it is severe dengue, 15.27 instances were classifide as severe dengue, but 1.13 of the instances was not severe dengue.

If creatinine is ≤ 84 and platelet also ≤ 146, it is severe dengue, otherwise it is non-severe dengue.

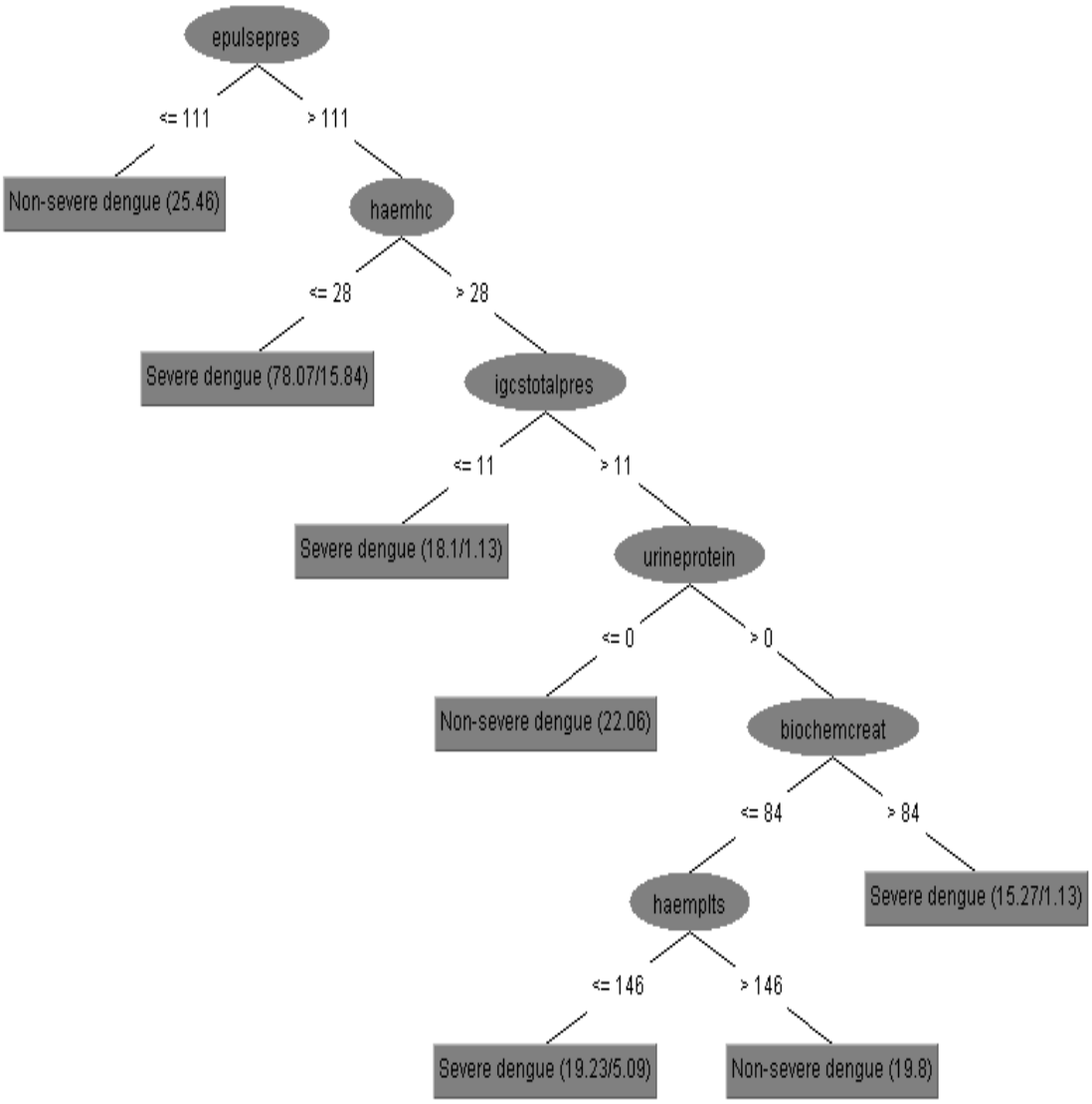


Figure 4.3 Decision tree algorithm for model 2

After assigning confidence factor value and minimum number of instances per leaf in order to prune and make the tree smaller, we found an appropriate model with a sensitivity of 71%, specificity of 68.7% and overall accuracy of the tree was 69.1% (see table 4.11), with a Kappa of 0.286 indicated a fair agreement of prediction with the actual class, followed by precision of non-severe dengue and severe dengue.

Table 4.11 Stratified cross-validation of model 2

	Model 2
Sensitivity	71%
Specificity	68.7%
Kappa statistic	0.286
Precision	Non-dengue = 0.909
	Dengue = 0.351

This confusion matrix shows the total of severe dengue and non-severe dengue instances. For class "severe dengue" 27 instances were correctly classified but 11 instances were assigned to class "non-severe dengue". For class "non-severe dengue" 110 instances were correctly classified but 50 were assigned to class "severe dengue", as presented in table 4.12.

Table 4.12 Confusion matrix of model 2

		Final confirmation using SDB algorithm	
		Sever dengue	Non-severe dengue
Model prediction	Severe dengue	27	50
	Non-severe dengue	11	110

4.4.2.2 Logistic regression analysis

The classification table is shown in table 4.13. This model had a sensitivity of 50%, specificity of 95.6% and overall accuracy was 86.9%.

Table 4.13 Classification table of model 2 by logistic regression analysis with the cut-off value of 0.5

		Final confirmation using SDB algorithm		
		Severe dengue	Non-severe dengue	Overall Percentage
Model	Severe dengue	19	7	86.9%
	Non-severe dengue	19	153	
	Percentage correct	50%	95.6%	

The clinical significance of model 2 is shown in table 4.14, with the odds ratio calculated. Pulse rate, lymphocyte, ALT, urine red blood cells, urine protein and low platelet count increased the probability of a diagnosis of severe dengue, significant odds ratio ranging from 1.003 to 63.449.

Table 4.14 Final results of model 2 using logistic regression analysis

Variable	**OR**	**95 % CI**		***p* value**
Rash	0.063	0.003	1.299	.073
Pulse rate	1.026	1.009	1.043	.002
Glasgow Coma Score	0.816	0.685	0.972	.022
Hematocrit	0.929	0.873	0.989	.021
Lymphocyte	1.107	0.981	1.249	.098

Table 4.14 Final results of model 2 using logistic regression analysis (cont.)

Variable	OR	95 % CI		*p* value
ALT, IU/l	1.003	1.000	1.005	.020
Urine red blood cells, /mm^3	1.030	1.006	1.055	.013
Urine protein	1.008	0.999	1.017	.080
Normal platelet count (100,000 and over)	1			
- Platelet count (19,000)	63.449	3.696	1089.258	.004
- Platelet count (20,000 – 40,000)	3.518	0.783	15.796	.101
- Platelet count (41,000 – 99,000)	0.930	0.159	5.448	.936

OR = odds ratio, CI = confidence interval

4.4.2.3 Comparison of model 2 using classification tree and logistic regression Analysis

Observing table 4.16, variables contribute to the model using classification tree are pulse rate, hematocrit, Glasgow coma score, urine protein, creatinine and platelet count, whereas rash, lymphocyte, pulse rate, Glasgow coma score, hematocrit, ALT, urine protein, urine red blood cells and platelet count contribute significantly to the model when using logistic regression analysis. Pulse rate, hematocrit, Glasgow coma score, urine protein, platelet count are found in both models.

The decision tree model showed a high sensitivity of 71% in distinguishing severe dengue from non-severe dengue with sets of six initial variables. The high sensitivity of the algorithm performed by CART is promising. If a patient is

confirmed to have dengue infection, by following the set of these early signs we may be able to predict whether the patient may become severe or not. This will be very useful for guiding treatment decision and patient management.

The logistic regression model had the performance improved when compared with model 1, however, it performed less well compared with CART. It had a sensitivity of 50% and very high specificity of 95.6 % (see table 4.15). Again, it may be able to apply when ruling out the possibility of severe dengue. In the situation where a patient has been diagnosed and confirmed to have dengue infection, this algorithm can highly be capable of ruling out the possibility of severe infection among the majority. Dengue confirmed cases with some warning signs should, however, be closely monitored since there was 50% chance that their disease severity may have been overlooked (50% false negative rate). If the algorithm predicted a severe infection, we would be highly confident that it was a severe case since false positive rate was low.

Table 4.15 Sensitivity, specificity and accuracy of two different methods in model 2

Technique	**Sensitivity**	**Specificity**	**Accuracy**
Classification tree	71 %	68.7 %	69.1 %
Logistic regression analysis	50 %	95.6 %	86.9 %

Table 4.16 Variables contribute to model 2 using two different methods

Technique	**Rash**	**Lymphocyte**	**Pulse rate**	**Hematocrit**	**Glasgow coma score**	**Urine protein**	**Creatinine**	**Platelets**	**ALT**	**Urine red blood cells**
Classification tree			√	√	√	√	√	√		
Logistic regression analysis	√	√	√	√	√	√		√	√	√

4.4.3 Model 3 (Survived vs. Died)

Owing to we did not have information on blood pressure (narrow pulse pressure) for considering or indicating that the patient had shock and restlessness to evaluate circulatory failure, this could not be qualified as DHF or DSS using WHO 1997 criteria. Thus, we predicted survival and death in dengue cases instead of predicting DHF and DF+OFI (see figure 4.4).

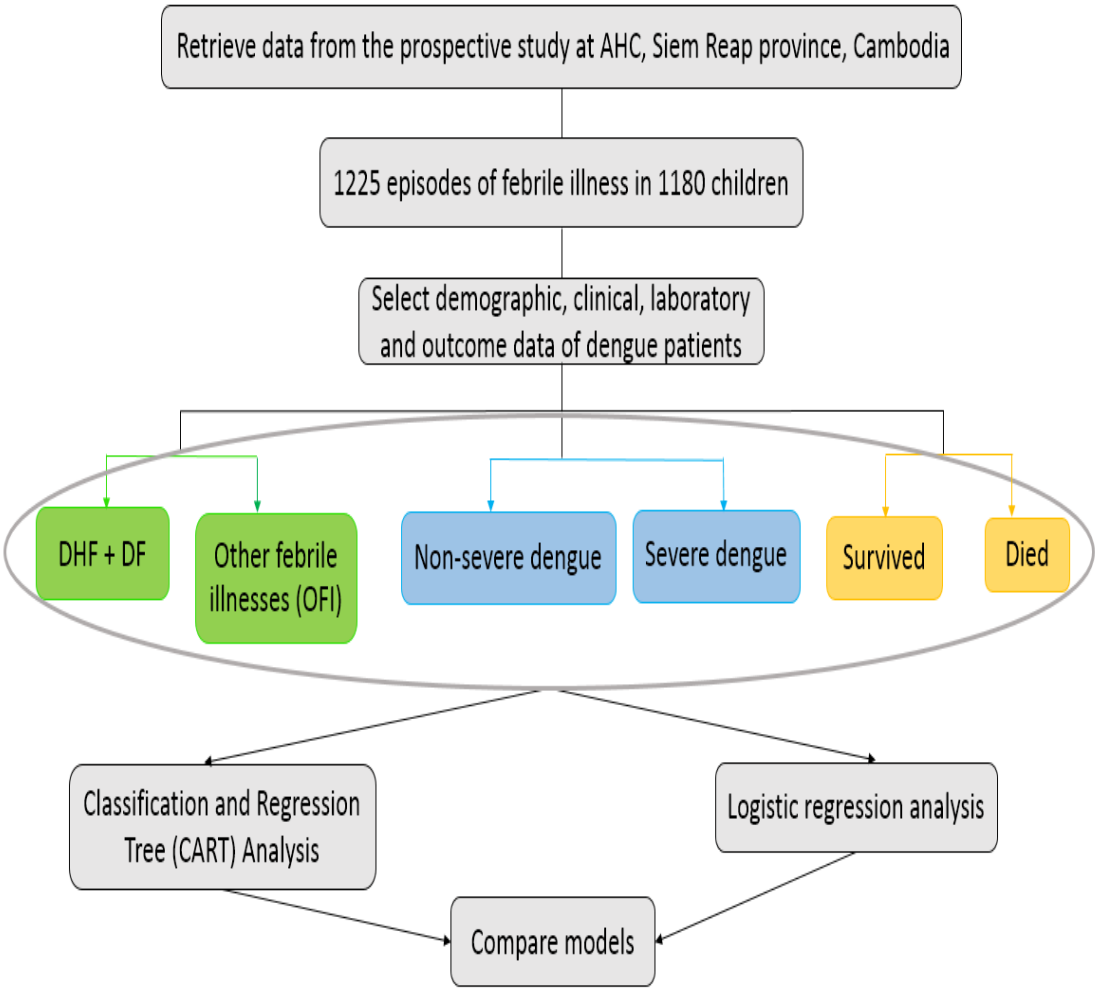
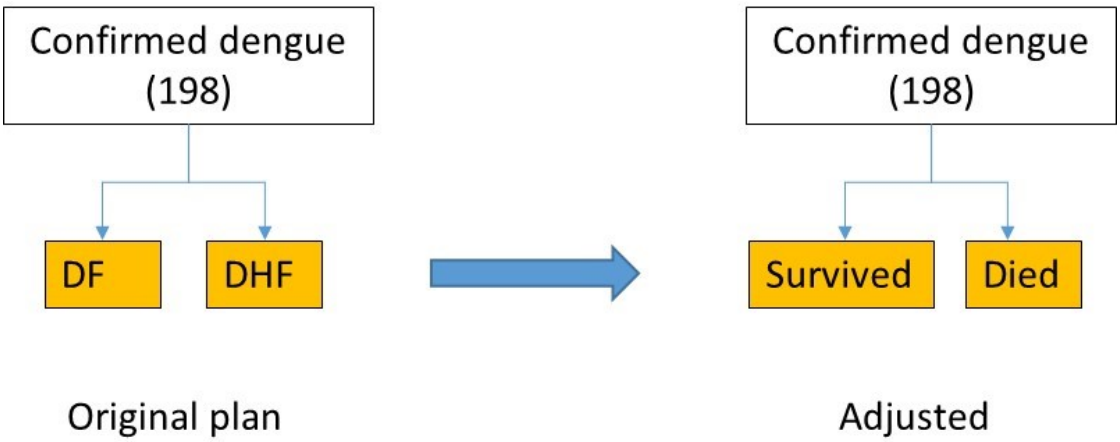


Figure 4.4 Adjusted model 3 for predicting clinical outcome, survival and death in dengue confirmed cases (the original figure, see figure 1.2, page 5)



Research question: Can we differentiate a patient die from dengue or survive dengue fever using some early signs and symptoms?

The final output in this model is "died" (n = 8) and "survived" (n= 190). In order to keep sufficient numbers of outcomes required to run the model we included all deaths where dengue infection was confirmed though cautions must be taken as dengue may not have been the main cause of death in some cases.

4.4.3.1 Classification tree

The first splitting variable is urea of 7, followed by Glasgow coma score. The number of leaves and size of the our tree were 3 and 5, respectively (see figure 4.5).

The first split " urea ", if urea is greater than 7, then it is death (died from dengue). There were 24.97 instances reached this "died" leaf, but 7.75 instances were dead (incorrectly classified as died from dengue)

If urea is less than or equal to 7 and igcstotalpres (Glasgow coma score) also ≤ 10, then he/she dies from dengue, 16.36/7.75 meaning that 16.36 instances were classified as died from dengue, but 7.75 instances were misclassified. If Glasgow coma score is greater than 10, then he/she survives dengue, the numbers in brancket meaning 156.68 instances survived dengue, but 8.61 of instances were incorrectly classified (see figure 4.5).

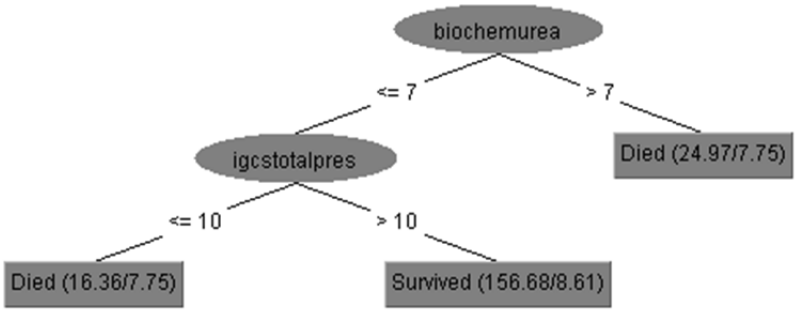


Figure 4.5 Decision tree algorithm for model 3

After assigning confidence factor value and minimum number of instances per leaf in order to prune and make the tree smaller, we found an appropriate model with a sensitivity of 37.5 %, specificity of 90% and overall accuracy of the tree was 83.8% (see table 4.17), with a Kappa of 0.149 indicated a slight agreement of prediction with the actual class, followed by precision of survival and death in dengue episodes.

Table 4.17 Stratified cross-validation of model 3

	Model 3
Sensitivity	37.5%
Specificity	90%
Kappa statistic	0.149
Precision	Survived = 0.972
	Died = 0.136

This confusion matrix shows the total of “died” and “survived” instances. For class “died” 3 instances were correctly classified but 5 instances were assigned to class “survived”. For class “survived” 171 instances were correctly classified, but 19 were assigned to class “died”, as presented in table 4.18.

Table 4.18 Confusion matrix of model 3

		Final confirmation using SDB algorithm	
		Died	Survived
Model prediction	Died	3	19
	Survived	5	171

4.4.3.2 Logistic regression analysis

Owing to a very small number of our eight death cases, maximum likelihood estimation of logistic model could not be able to handle it since there were only eight deaths out of 198 dengue episodes. Thus, we used another method, known as penalized likelihood which can handle it in reducing small-sample bias.

The clinical significance of model 3 is shown in table 4.19 with the odds ratio calculated. Abdominal pain, urea and urine protein increased the odds that the patient died, significant odds ratio ranging from 1.013 to 6.088.

Table 4.19 Results of model 3 using penalized likelihood

Variable	**OR**	**95 % CI**		***p* value**
Abdominal pain	6.088	0.1324	3.48	0.034
Urea, mmol/L	1.156	-0.000104	0.2895	0.050
Urine protein, mg/dL	1.013	0.000278	0.02509	0.045

4.4.3.3 Comparison of model 3 using classification tree and penalized likelihood

Observing table 4.20, variables contribute significantly to the model using classification tree are urea and Glasgow coma score, whereas abdominal pain, urea and urine protein contribute to the model when using penalized likelihood. Urea is found in both classification tree and penalized likelihood.

Table 4.20 Variables contribute to model 3 using different methods

Technique	Urea	Glasgow coma score	Urine protein	Abdominal pain
Classification tree	√	√		
Penalized likelihood	√		√	√

4.5 Comparison of three models using classification tree and logistic regression analysis

4.5.1 Sensitivity, specificity and accuracy of three models

As shown in table 4.21, using classification tree, it is observed that model 2 had the highest sensitivity (71 %), followed by model 3 (37.5 %) and model 1 (16.6 %). In contrast, model 1 had the highest specificity both in classification and logistic regression analysis (97.9 % and 98.2 %, respectively).

Table 4.21 Sensitivity, specificity and accuracy for three models

Models	Classification trees	Logistic regression analysis
Model 1 (Non-DF vs. DF)	• Sensitivity : 16.6 % • Specificity : 97.9 % • Accuracy : 83.8 %	• Sensitivity : 14.6 % • Specificity : 98.2 % • Accuracy : 83.7 %
Model 2 (Non-sever vs. Severe)	• Sensitivity : 71 % • Specificity : 68.7 % • Accuracy : 69.1 %	• Sensitivity : 50 % • Specificity : 95.6 % • Accuracy : 86.9 %
Model 3 (Survived vs. Died)	• Sensitivity : 37.5 % • Specificity : 90 % • Accuracy : 87.8 %	

4.5.2 Variations of three models predicting dengue, severe dengue and died from dengue in classification trees and logistic regression analysis

From table 4.22, using classification tree, it is observed that, model 1 and 2, hematocrit and platelets are found in predicting dengue and severe dengue. Glasgow coma score is found in predicting severity (model 2) and death in dengue cases (model 3).

Using logistic regression analysis, it can be seen that hematocrit, platelet count and serum ALT are found in both model 1 and 2. Abdominal pain is found in both model 1 and 3. Urine protein is found in both model 2 and 3 (see table 4.23).

Table 4.22 Variations of three models predicting dengue, severe dengue and died from dengue by classification trees

Model	**Variables**								
	Neutrophil	**Hematocrit**	**Respiratory rates**	**Platelets**	**Pulse**	**GCS**	**Urine protein**	**Creatinine**	**Urea**
Model 1 (Non-DF vs. DF)	√	√	√	√					
Model 2 (Non-severe vs. Severe)		√		√	√	√	√	√	
Model 3 (Survived vs. Died)						√			√

Table 4.23 Variations of three models predicting dengue, severe dengue and died from dengue by logistic regression analysis and penalized likelihood

Model	Variables												
	Rash	**Lymphocyte**	**Age group**	**Abdominal pain**	**Hematocrit**	**GCS**	**Platelets**	**White blood cells**	**Pulse**	**Urine red blood cells**	**ALT**	**Urea**	**Urine protein**
Model 1			√	√	√		√	√			√		
Model 2	√	√			√	√	√		√	√	√		√
Model 3				√								√	√

CHAPTER V
DISCUSSION

Dengue fever (DF) / dengue haemorrhagic fever (DHF) is a viral and re-emerging disease, commonly occurred in tropical and subtropical areas, these issues have remained an important public health concern, especially in South-east Asia, Africa, the Western Pacific and the Americas [1]. Dengue infection is frequently confounded with other febrile illnesses, showing with non-specific clinical symptoms. Other febrile illness is common in tropical countries, and clinical features of dengue is analogous to OFI. During the early stage of illness, the presence of non-specific febrile illness is strikingly difficult in making precise diagnosis, inefficient treatment and possibly increased morbidity and mortality [2, 4]. The first detection of dengue virus infection in Cambodia occurred in 1963 and has been informed to national surveillance of dengue since 1980, from 1980 to 2008, there were 194,726 dengue cases, which reported by the National Dengue Control Program [6].

In our study, we developed three models using both simple clinical manifestations and laboratory indicators to predict severity of dengue at the early phase of illness. Three different research questions were set in relations to dengue identification at each stage from being differentiable from other febrile illnesses (Model 1), predicting severity (Model 2) and finally the clinical outcome (Model 3). In order to build decision trees, we used ten-fold cross validation to estimate the accuracy. The J48 algorithm has been used for generating decision trees since they are able to handle nominal, categorical, numeric data, and missing values. Optimized parameters for J48 algorithm provide equivalent or higher classification accuracy. Reducing the size of decision trees in comparison with using the default parameters, for a particular data set, tuning parameters of algorithm is found to be better than applying the default parameters [50-52]. For model performance comparison, the conventional approach, logistic regression, was applied. In terms of classification accuracy, using different methods, model 1 of both techniques were nearly equal, but

model 2, a higher classification accuracy was found in application of decision algorithm. In terms of sensitivity, using decision tree models, both model 1 and 2 were found to be higher sensitivity, compared to logistic regression models.

The difficulties of classifying patients with DHF or DSS in our data set were due to lack of information on clinical bleeding sites, we only had a little information on blood or mucous in stool which could refer to gastrointestinal bleeding. In addition, lack of information on blood pressure or narrow pulse pressure for considering or indicating that the patient had shock [21] and restlessness to evaluate circulatory failure [8]. Besides, our data set did not have information on the results of chest x-ray to evaluate pleural effusion and also abdominal ultrasound to detect peritoneal fluid (ascites), these signs play an important role in identifying plasma leakage, these could not be qualified or fulfilled either as DHF or as DSS using WHO 1997 criteria, consistent with previous studies that WHO 1997 was required to be modified regarding plasma leakage for user-friendliness, clarifying definitions of severe bleeding and organ impairment [25, 29]. Other previous studies had also encountered the WHO classification issue, for example Barniol et al showed that 13.7% of dengue cases could not be classified using WHO 1997 dengue classification by expert reviewers while using the WHO 2009 criteria only accounted for 1.6% [53]. Clinical signs of severe dengue infection such as organ impairment is not presented in WHO 1997 criteria, many physicians have been trying to fulfill these criteria. In addition, a requirement for repeating clinical tests is also necessary but this is not easy to perform frequently, particularly for those countries which have limited resource settings [30, 31]. Basuki and colleagues showed that, in order to detect the severity of dengue the application of revised dengue classification (WHO 2009 criteria) was found to be better than the WHO 1997 dengue classification, with a high sensitivity and specificity [49]. However, there were other studies stated the problems using WHO 2009 e.g. Hadinegoro and Kalayanarooj stated the issues of its applicability such as classifying dengue patient with warning signs requiring special criteria to fulfill and changes in this system also increased the amount of work for health care workers [25, 31].

In our study, by using WHO 2009 dengue criteria, we could classified 198 dengue episodes into dengue fever (n = 55), dengue with warning signs (n = 105) and

severe dengue (n = 38), in terms of classifying severe dengue, not only WHO 2009 criteria was used but also the information on ICU admission and two pediatricians' opinions were considered.

Our decision tree algorithm for dengue prediction in the early stage of illness had sets of initial variables which include neutrophil, hematocrit, respiratory rates and low platelets (figure 4.2). For node hematocrit and platelet count had clinical significant, and high hematocrit increased the probability of dengue diagnosis (table 4.8). In the previous study by Tanner and colleagues, their decision tree for the diagnosis of dengue had similar sets of initial variables as ours such as platelet count, neutrophil, hematocrit with a high sensitivity and specificity of 71.2% and 90.1%, respectively [2]. Although, the tree had very low sensitivity of 16.6% and a very high specificity of 97.9%, this algorithm may be useful for ruling out those people who do not have dengue fever from other febrile illness which can lead to severe consequences, in contrast with a study of predicting dengue and chikungunya in adults using classification trees showed a very high sensitivity of 99% and specificity of 52% [54].

The decision algorithm for predicting severe dengue (figure 4.3), the first split is pulse rate, followed by hematocrit as the second split if pulse rate is greater than 111 per minute, Glasgow coma score as the third split, urine protein as the fourth split, creatinine as the fifth split and the final split uses platelet count less than 146,000 mm3 blood. In the previous study Potts, Tanner and colleagues, their decision algorithm had some similar sets of initial variables as ours such as low hematocrit and platelet count [2, 10]. A study of hospitalized children with DHF in Thailand showed that close observation in DHF patients was important in order to avoid developing shock and only laboratory data on hospital admission was not be able to predict shock in DHF patients [55].

The decision tree algorithm for prognostication, death or survival uses the urea as the first split, followed by Glasgow coma score as the final split for those with urea less than or equal to 7 mmol/L. For urea node of the decision tree has clinically significant odds ratio 1.156 (table 4.19), high urea increased the odds that the patient died. Owing to a very small number of our eight death cases, maximum likelihood estimation of logistic model could not be able to handle it since there were only eight

deaths out of 198 dengue episodes. Thus, we used another method, known as penalized likelihood which could handle it in reducing small-sample bias. For decision tree, only eight death outcomes, the performance could not be so good when running this through the decision algorithm and not be able to provide enough information since the sample of death was not large.

The results of logistic regression analysis for model 1, infants aged < 28days to < 1year, abdominal pain, increasing hematocrit, serum ALT, low platelet count (below 100,000) were factors which increased the probability of dengue diagnosis with significant odds ratio ranging from 1.001 to 3.895 and they also had clinical significance in diagnosis of dengue fever. National dengue surveillance in Cambodia, 1980-2008 has been reported that infants less than 1 year and children aged 4-6 years were the highest age-specific incidence of dengue fever [6]. The study of dengue severity in infants, children and adults in Nicaragua also showed that the burden of disease infants aged 4-9 months and children 5-9 years were mostly affected [56]. One possible explanation for this result is due to insufficient maternal antibodies to protect dengue particularly in infants less than 1 year [56, 57]. In model 2, final outcome shows pulse rate, lymphocyte, serum ALT, urine red blood cells, urine protein, platelet count (19,000) and platelet count (20,000 – 40,000) increased the possibility of a diagnosis of severe dengue with significant odds ratio ranging from 1.003 to 63.449. Khan and colleagues showed that increased ALT was associated with developing severe dengue while the presence of urine red blood cells related to develop DHF [19]. The variables in model 3, abdominal pain, urea and urine protein were more likely to increase the likelihood of death in dengue cases. There has not been such a model for predicting the final outcome in the past to compare with.

The WHO 1997 and WHO 2009 dengue guidelines include a tourniquet test as a diagnostic test for dengue and it is found in the early febrile phrase. Previous studies have been shown that tourniquet test had low sensitivity of dengue diagnosis, and negative result of this test did not exclude dengue infection [56, 58, 59]. Unfortunately, tourniquet test in our data set accounted for 85% of the clinical record had missing values and thus was initially omitted from the analysis. The following predictive factors will be discussed in detail since they have been identified, by the current and previous studies, to potentially be for early predicting of dengue severity.

Pulse

The presence of rapid pulse (tachycardia) had clinical significant and increased the likelihood of severe dengue diagnosis, it is a sign of circulatory failure. Severe bleeding could lead to shock depending on the volume of blood loss and resulting in pulse pressure fall and heart rate is creased [21]. Pulse rate is also different in age groups, our study showed severe dengue was mostly found in infants aged 28 days to 1 year, thus, their normal pulse rate would be 120-160 beats per minute [60]. The study of fatal DHF in Singapore showed that only tachycardia on hospital admission was related to dengue deaths [61]. In our tree, pulse rate had clinical significant.

Glasgow coma score (GCS)

The GCS is used for measuring levels of consciousness (mental alteration) [62]. In our results, the node tree with GCS $\leq$ 11 for model 2, and $\leq$ 10 for model 3, this considered to be moderate GCS. Rao and colleagues showed that 2 out of 24 patients died in dengue encephalitis had low GCS, seizure, headache and gastrointestinal bleeding [63].

Urine protein

The presence of urine protein might indicate the severity of dengue infection due to plasma leakage. Having urine protein was found to be associated with developing DHF or DSS [19, 64].

Creatinine

An increased serum creatinine level indicates kidney dysfunction. In patients with DHF, a mild increase of serum creatinine is common in contrast with acute kidney injury (AKI). Our model showed the serum creatinine > 84 mmol/l (4.6 mg/dl) led to severe dengue, this value is similar to the study of pediatric Thai patients with DHF that had mean serum creatinine was 4.9 mg/dl, their analysis also showed that twenty-four of 25 AKI patients had DSS as a final diagnosis, and 64% of AKI patients died due to development of profound shock together with other conditions such as liver failure, respiratory failure and a large amount of bleeding [65].

Platelet count and Hematocrit

Understanding the mechanism of thrombocytopenia caused by dengue virus is very difficult to understand [66]. Previous studies suggest that dengue virus probably contributes to bone marrow suppression and results in platelet destruction [67, 68]. To fulfill the WHO guideline in classifying patients with DHF, thrombocytopenia (platelet count ≤100,000) is required. Srikiatkhachorn and colleagues demonstrated that patients with thrombocytopenia were related to dengue severity, and not all severe cases have met WHO criteria in terms of classifying patients with DHF [69]. Recent study showed that most of patients with spontaneous bleeding had low platelet count of < 20,000. A platelet count of 20,000-40,000 was found in patients with petichae/purpura [70]. Research conducted by Makroo and Raina, they classified patients with having platelet count into high, moderate, low and no risk, patients with high risk had platelet count < 20,000/cumm were more likely to receive platelet transfusion, those with moderate risk of platelet transfusion had platelet count 21,000-40,000/ cumm, those with low risk had platelet count > 40,000 - < 100,000/cumm and those with no risk had platelet count > 100,000/cumm [71]. In a study of 167 Vietnamese patients with DSS, they categorized severity of bleeding into mild, moderate and severe bleeding, patients with mild bleeding had spontaneous petechiae or bruising at many sites of puncturing of a vein, those with moderate bleeding had mucosal bleeding and should not any have influence on hematocrit, and those with severe bleeding had epistaxis or gastrointestinal bleeding resulting in decreased hematocrit or blood transfusion [72]. The cut-off value of platelet count in our tree was ≤ 146,000 indicating patients developed severe dengue [10]. The WHO 1997 definition requires low plate count ≤ 100,000 together with an increased hematocrit ≥ 20% above baseline value would be indicative of having plasma leakage. In a study by Potts (2010) showed that some patients with DHF were diagnosed by physicians did not had a low platelet count ≤ 100,000, a majority of DHF patients had plasma leakage with the presence of pleural effusion instead of evaluating hematocrit ≥ 20%. Although low platelets suggest dengue infection is severe, malaria and scrub typhus are also common in differential diagnosis for dengue infection [48]. A study in Thailand stated that patients who had internal bleeding such as gastrointestinal tract may not manifest hemoconcentration or rising hematocrit 20% of the baseline value

[66], consistent with our study that we only had blood in stool for hemorrhagic manifestation.

Limitations of this study

Our study has several limitations such as the smaller number of patients with dead episodes, there were eight of fatal episodes in 198 dengue but who truly died from dengue were five cases according to grading severity of dengue and the opinion of pediatrician's diagnosis. There were co-infections in our three cases were SLE (Systemic lupus erythematosus), pneumonia (empyema) and meliodosis. Our model 1, patients presented at a median of 3 days of illness, and a median of 4 days of illness for both model 2 and 3, these decision algorithms may not be appropriate for applying if patients present their illness earlier to a doctor. We did not have pediatric normal laboratory values from Angkor hospital for children which leads to inaccurate laboratory results because age-related changes in hematocrit and laboratory values. We studied only in Angkor hospital for children, not in different settings or other populations, however in Cambodia, dengue infection is predominantly found in children.

In conclusion we have shown our three decision algorithms which may be able to distinguish other causes of febrile illness from dengue infection to rule out the possibility of dengue fever where mortality is relatively low. In addition, our decision tree algorithm using simple clinical and laboratory indicators has a high classification accuracy in predicting pediatric patients who develop severe dengue. This model is potentially useful for guiding patient monitoring plan and outpatient management of fever.

ADVISOR'S COMMENTS

Dengue virus is endemic in tropical and sub-tropical resource-poor countries including Thailand and Cambodia. A wide range of symptoms related to dengue infections, from nonspecific febrile illness to severe forms such as severe bleeding or circulatory failure, making it difficult in clinical management and resources allocation. This research project focuses on developing predictive tools to characterize severity of dengue infection using early clinical and laboratory indicators. Secondary data used in this project are from the study of febrile illnesses among children admitted to Angkor Hospital for Children (AHC) in Siem Reap province, Cambodia. The study focuses on using classification tree to derive diagnostic algorithms which aim to distinguish between patients at risk of developing a severe infection and those at low risk. This relatively large data set with a fair proportion of dengue infections allows us to well apply data mining and statistical techniques. In addition, Khansoudaphone has suitable clinical background which is essential for data analysis and interpreting outcomes of the study.

This study demonstrates the potential usefulness of some specific clinical and laboratory indicators for predicting severity of dengue infections.

……

Wirichada Pan-ngum,

Ph.D. (Infectious Disease)

TIMETABLE

Tasks	September /2013	October /2013	November /2013	December /2013	January /2014
Proposal defense					
Data management					
Data analysis					
Writing thematic paper					
Final defense					

REFERENCES

1. Tantawichien, T., Dengue fever and dengue haemorrhagic fever in adolescents and adults. Paediatr Int Child Health, 2012. 32 Suppl 1: p. 22-7.
2. Tanner, L., et al., Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. PLoS Negl Trop Dis, 2008. 2(3): p. e196.
3. Guzman, M.G., et al., Dengue: a continuing global threat. Nat Rev Microbiol, 2010. 8(12 Suppl): p. S7-16.
4. Manock, S.R., et al., Etiology of acute undifferentiated febrile illness in the Amazon basin of Ecuador. Am J Trop Med Hyg, 2009. 81(1): p. 146-51.
5. Yong, L.S. and K.C. Koh, A case of mixed infections in a patient presenting with acute febrile illness in the tropics. Case Rep Infect Dis, 2013. 2013: p. 562175.
6. Huy, R., et al., National dengue surveillance in Cambodia 1980-2008: epidemiological and virological trends and the impact of vector control. Bull World Health Organ, 2010. 88(9): p. 650-7.
7. Sam, S.S., et al., Review of Dengue hemorrhagic fever fatal cases seen among adults: a retrospective study. PLoS Negl Trop Dis, 2013. 7(5): p. e2194.
8. Surangrat, P., Prognostic Indicators for Dengue Infection Severity 2013, Elmer press. p. 12-18.
9. Potts, J.A., et al., Classification of dengue illness based on readily available laboratory data. Am J Trop Med Hyg, 2010. 83(4): p. 781-8.
10. Potts, J.A., et al., Prediction of dengue disease severity among pediatric Thai patients using early clinical laboratory indicators. PLoS Negl Trop Dis, 2010. 4(8): p. e769.
11. Gibbons, R.V. and D.W. Vaughn, Dengue: an escalating problem. BMJ, 2002. 324(7353): p. 1563-6.

12. Shepard, D.S., E.A. Undurraga, and Y.A. Halasa, Economic and disease burden of dengue in Southeast Asia. PLoS Negl Trop Dis, 2013. 7(2): p. e2055.
13. Duong, V., et al., Complex dynamic of dengue virus serotypes 2 and 3 in Cambodia following series of climate disasters. Infect Genet Evol, 2013. 15: p. 77-86.
14. Vong, S., et al., Dengue incidence in urban and rural Cambodia: results from population-based active fever surveillance, 2006-2008. PLoS Negl Trop Dis, 2010. 4(11): p. e903.
15. Vong, S., et al., Under-recognition and reporting of dengue in Cambodia: a capture-recapture analysis of the National Dengue Surveillance System. Epidemiol Infect, 2012. 140(3): p. 491-9.
16. Beaute, J. and S. Vong, Cost and disease burden of dengue in Cambodia. BMC Public Health, 2010. 10: p. 521.
17. Tapia-Conyer, R., M. Betancourt-Cravioto, and J. Mendez-Galvan, Dengue: an escalating public health problem in Latin America. Paediatr Int Child Health, 2012. 32 Suppl 1: p. 14-7.
18. Matheus, S., et al., Discrimination between primary and secondary dengue virus infection by an immunoglobulin G avidity test using a single acute-phase serum sample. J Clin Microbiol, 2005. 43(6): p. 2793-7.
19. Khan, M.I., et al., Factors predicting severe dengue in patients with dengue Fever. Mediterr J Hematol Infect Dis, 2013. 5(1): p. e2013014.
20. Wee, L.K., Relationship between rainfall and aedes larval population at two insular sites in Pulau Ketam, Selangor, Malaysia. Southeast Asian J Trop Med Public Health, 2013. 44(2): p. 9.
21. Organization, W.H., Dengue guidelines for diagnosis, treatment, prevention and control, in Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition. 2009: Geneva.
22. Wichmann, O., et al., Risk factors and clinical features associated with severe dengue infection in adults and children during the 2001 epidemic in Chonburi, Thailand. Trop Med Int Health, 2004. 9(9): p. 1022-9.

23. Rigau-Perez, J.G., et al., Dengue: A Literature Review and Case Study of Travelers from the United States, 1986-1994. J Travel Med, 1997. 4(2): p. 65-71.
24. unit, E., Guidelines on Management of Dengue Fever & Dengue Haemorrhagic Fever In Adults. 2010, Epidemiology Unit, Ministry of Health 2005.
25. Kalayanarooj, S., Dengue classification: current WHO vs. the newly suggested classification for better clinical application? J Med Assoc Thai, 2011. 94 Suppl 3: p. S74-84.
26. Blacksell, S.D., et al., Prospective study to determine accuracy of rapid serological assays for diagnosis of acute dengue virus infection in Laos. Clin Vaccine Immunol, 2007. 14(11): p. 1458-64.
27. Tsuzuki, A., et al., Can daytime use of bed nets not treated with insecticide reduce the risk of dengue hemorrhagic fever among children in Vietnam? Am J Trop Med Hyg, 2010. 82(6): p. 1157-9.
28. Lee, V.J., et al., Decision tree algorithm in deciding hospitalization for adult patients with dengue haemorrhagic fever in Singapore. Trop Med Int Health, 2009. 14(9): p. 1154-9.
29. Gan, V.C., et al., Implications of discordance in world health organization 1997 and 2009 dengue classifications in adult dengue. PLoS One, 2013. 8(4): p. e60946.
30. Bandyopadhyay, S., L.C. Lum, and A. Kroeger, Classifying dengue: a review of the difficulties in using the WHO case classification for dengue haemorrhagic fever. Trop Med Int Health, 2006. 11(8): p. 1238-55.
31. Hadinegoro, S.R., The revised WHO dengue case classification: does the system need to be modified? Paediatr Int Child Health, 2012. 32 Suppl 1: p. 33-8.
32. Novrianti Hawarini, M.S.K., The relationship between pleural effusion index and mortality in children with dengue shock syndrome. 2012. 52: p. 3.
33. Armstrong, P.M., et al., Detection of infectious virus from field-collected mosquitoes by vero cell culture assay. J Vis Exp, 2011(52).
34. Limkittikul, K., et al., Clinical differences among PCR-proven dengue serotype infections. Southeast Asian J Trop Med Public Health, 2005. 36(6): p. 1432-8.

35. Moi, M.L., et al., Detection of dengue virus nonstructural protein 1 (NS1) by using ELISA as a useful laboratory diagnostic method for dengue virus infection of international travelers. J Travel Med, 2013. 20(3): p. 185-93.
36. Rathakrishnan, A. and S.D. Sekaran, New development in the diagnosis of dengue infections. Expert Opin Med Diagn, 2013. 7(1): p. 99-112.
37. Peeling, R.W., et al., Evaluation of diagnostic tests: dengue. Nat Rev Microbiol, 2010. 8(12 Suppl): p. S30-8.
38. Singh, M.P., et al., NS1 antigen as an early diagnostic marker in dengue: report from India. Diagn Microbiol Infect Dis, 2010. 68(1): p. 50-4.
39. Blacksell, S.D., Commercial dengue rapid diagnostic tests for point-of-care application: recent evaluations and future needs? J Biomed Biotechnol, 2012. 2012: p. 151967.
40. Blacksell, S.D., et al., Comparison of seven commercial antigen and antibody enzyme-linked immunosorbent assays for detection of acute dengue infection. Clin Vaccine Immunol, 2012. 19(5): p. 804-10.
41. Cardosa, M.J., et al., Antibodies against prM protein distinguish between previous infection with dengue and Japanese encephalitis viruses. BMC Microbiol, 2002. 2: p. 9.
42. B. L. INNIS, A.N., An enzyme-linked immunosorbent assay to characterize dengue infections where dengue and Japanese encephalitis co-circulate. Am. J. Trop. Med. Hyg, 1989: p. 418-427.
43. Chheng, K., et al., A prospective study of the causes of febrile illness requiring hospitalization in children in Cambodia. PLoS One, 2013. 8(4): p. e60634.
44. Hairi, F., et al., A knowledge, attitude and practices (KAP) study on dengue among selected rural communities in the Kuala Kangsar district. Asia Pac J Public Health, 2003. 15(1): p. 37-43.
45. Mayxay, M., et al., Dengue in peri-urban Pak-Ngum district, Vientiane capital of Laos: a community survey on knowledge, attitudes and practices. BMC Public Health, 2013. 13: p. 434.
46. Griffiths, K., et al., Public health responses to a dengue outbreak in a fragile state: a case study of Nepal. J Trop Med, 2013. 2013: p. 158462.

47. Thomas, S.J., The necessity and quandaries of dengue vaccine development. J Infect Dis, 2011. 203(3): p. 299-303.
48. Chadwick, D., et al., Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis. J Clin Virol, 2006. 35(2): p. 147-53.
49. Basuki, P.S., et al., Application of revised dengue classification criteria as a severity marker of dengue viral infection in Indonesia. Southeast Asian J Trop Med Public Health, 2010. 41(5): p. 1088-94.
50. Koblar, V., Optimizing parameters of machine learning algorithms. 2012.
51. Kohavi, R., Automatic parameter selection by minimizing estimated error. 1995: p. 304-312.
52. Tusar, T., Optimizing accuracy and size of decision trees.
53. Barniol, J., et al., Usefulness and applicability of the revised dengue case classification by disease: multi-centre study in 18 countries. BMC Infect Dis, 2011. 11: p. 106.
54. Lee, V.J., et al., Simple clinical and laboratory predictors of Chikungunya versus dengue infections in adults. PLoS Negl Trop Dis, 2012. 6(9): p. e1786.
55. Wiwanitkit, V., Can hematocrit and platelet determination on admission predict shock in hospitalized children with dengue hemorrhagic fever? A clinical observation from a small outbreak. . 2004.
56. Hammond, S.N., et al., Differences in dengue severity in infants, children, and adults in a 3-year hospital-based study in Nicaragua. Am J Trop Med Hyg, 2005. 73(6): p. 1063-70.
57. Hongsiriwon, S., Dengue hemorrhagic fever in infants. Southeast Asian J Trop Med Public Health, 2002. 33(1): p. 49-55.
58. Cao, X.T., et al., Evaluation of the World Health Organization standard tourniquet test and a modified tourniquet test in the diagnosis of dengue infection in Viet Nam. Trop Med Int Health, 2002. 7(2): p. 125-32.
59. Mayxay, M., et al., Predictive diagnostic value of the tourniquet test for the diagnosis of dengue infection in adults. Trop Med Int Health, 2011. 16(1): p. 127-33.
60. West, B., Vital signs. 2012. p. 297-321.

61. Ong, A., et al., Fatal dengue hemorrhagic fever in adults during a dengue epidemic in Singapore. Int J Infect Dis, 2007. 11(3): p. 263-7.
62. The, T.D., Clinical features of Dengue in a large Vietnamese cohort: Instrinsically lower platelet counts and greater risk for bleeding in adults than children. 2012. 6(6): p. 11.
63. Rao, M., Dengue encephalitis - clinical spectrum and outcome. 2013.
64. Vasanwala, F.F., Could peak proteinuria determine whether patient with dengue fever develop dengue hemorrhagic/dengue shock syndrome? - A prospective cohort study. 2011: p. 2-6.
65. Laoprasopwattana, K., et al., Outcome of dengue hemorrhagic fever-caused acute kidney injury in Thai children. J Pediatr, 2010. 157(2): p. 303-9.
66. Chuansumrit, A., Pathophysiology and management of dengue hemorrhagic fever. 2006: p. 3-11.
67. Lei, H.Y., et al., Immunopathogenesis of dengue virus infection. J Biomed Sci, 2001. 8(5): p. 377-88.
68. Hottz, E., Platelets in dengue infection. 2011. 8: p. 33-38.
69. Srikiatkhachorn, A., et al., Dengue hemorrhagic fever: the sensitivity and specificity of the world health organization definition for identification of severe cases of dengue in Thailand, 1994-2005. Clin Infect Dis, 2010. 50(8): p. 1135-43.
70. Pruthvi, D., Evaluation of platelet count in dengue fever along with seasonal variation of dengue infection. 2012. 3(4): p. 1-4.
71. Makroo, R.N., et al., Role of platelet transfusion in the management of dengue patients in a tertiary care hospital. Asian J Transfus Sci, 2007. 1(1): p. 4-7.
72. Wills, B.A., et al., Coagulation abnormalities in dengue hemorrhagic Fever: serial investigations in 167 Vietnamese children with Dengue shock syndrome. Clin Infect Dis, 2002. 35(3): p. 277-85.

APPENDICES

APPENDIX A

Case Report Form (CRF) used in the fever study at the Angkor Hospital for Children (AHC) in Cambodia.

AHC Fever Study CRF

Patient Label	Enrollment Date (dd/mm/yyyy): |__|__|/|__|__|/|__|__|__|__|

1. Inclusion Criteria (tick closed boxes in appropriate column)	**NO**	**YES**
a. Age less than 16 years	☐	☐
b. Documented fever of ≥ 38°C at admission, or within either 48 hours before admission or 48 hours after admission	☐	☐
c. Admitted for treatment to LAU, IPD, ICU or surgical wards	☐	☐
d. Informed consent was signed (by parent or caretaker)	☐	☐
If any of the above are answered NO, the subject CANNOT enter the study		

2. Exclusion Criteria	**NO**	**YES**
a. NOT admitted	☐	☐
b. Consent NOT given	☐	☐
If any of the above are answered YES, the subject CANNOT enter the study		

3. Eligibility Summary	**NO**	**YES**
a. Is participant eligible for study?	☐	☐
b. Date study informed consent was signed (dd/mm/yyyy):	|__|__|/|__|__|/|__|__|__|__|	
c. Was Information Sheet given to participant?	☐	☐
If NO, specify reason: _____		

4. Enrollment Summary	**NO**	**YES**
a. i. Is participant being enrolled?	☐	☐
ii. If NO, Reason for not enrolled (write number in open box) |__| 1 = Did not meet eligibility criteria		
2 = Met eligibility criteria but refused to participate		
3 = Met eligibility criteria, gave consent, but not enrolled for other reason (e.g. LAMA)		
4 = Missed by admitting team and investigators		

Study Checklist
1. Consent, including caretaker and witness signatures
2. Patient labels on top of CRF and consent form
3. Order fever study bloods (see green form)
4. Take fever study respiratory samples if suspected respiratory infection
5. Take urine for urine antibiogram and urinalysis on all patients
6. Fill out CRF (use examination findings from admission note where possible).
7. Order discharge bloods when / if appropriate

Name of clinician / researcher:_____ Signature: _____ Date: ____/____/____

AHC Fever Study CRF

7. Associated Symptoms	NO	YES	Not Assessable
ii. If YES, specify from what site: _____			
u. Jaundice	☐	☐	☐
v. Dysuria	☐	☐	☐
w. Chills or rigors	☐	☐	☐
x. Skin infection	☐	☐	☐
y. Other	☐	☐	☐
ii. If YES, specify: _____			

8. Drug Allergies and Drug History	NO	YES	Not Assessable
a. i.Does the participant have any known allergy to any drugs?	☐	☐	☐
ii. If YES, specify the drugs name: _____			
b. i. Has the subject taken any antibiotic, antimalarial and/or steroid in the past 7 days?	☐	☐	☐
ii. If YES, specify the drug's name(s), dose(s) and length of treatment: _____			

9. Environmental History	NO	YES	Not Assessable
a. i. History of animal (not insect) bite?	☐	☐	☐
ii. If YES, specify what animal: _____			
b. Lives in a village with malaria	☐	☐	☐
c. Has been to the forest within the last month	☐	☐	☐
d. Has been into pond / river / ditch water within the last month	☐	☐	☐
e. Has been in wet paddy within the last month	☐	☐	☐
f. Has been exposed to pigs or chickens	☐	☐	☐
g. i. Has been exposed to dead animals	☐	☐	☐
ii. If YES, specify what animal: _____			
h. Recent minor trauma to lower limbs	☐	☐	☐

10. Past Medical and Family History	NO	YES	Not Assessable
a. i. Previous admissions to any hospital?	☐	☐	☐
ii. if YES, specify diagnoses: _____			
b. Admission to any healthcare facility (e.g. hospital, OPD) in last month?	☐	☐	☐
c. i. Any chronic disease?	☐	☐	☐
ii. If YES, please specify: _____			
d. If < 2 months old, was the mother ill with fever during pregnancy?	☐	☐	☐
e. History of contact with tuberculosis	☐	☐	☐
f. i. Family history of recent febrile illness	☐	☐	☐
ii. If YES, specify what illness: _____			
g. i. Family history of chronic illness	☐	☐	☐
ii. If YES, specify what illness: _____			

AHC Fever Study CRF

12. Systems Examination (USE ADMISSION EXAM IF POSSIBLE)	NO	YES	Not Assessable
c. Abdominal			
i. Severe wasting?	☐	☐	☐
ii. Signs of ascites (e.g. distended abdomen)?	☐	☐	☐
iii. Increased liver size or other liver abnormality?	☐	☐	☐
iv. If abnormal, please specify: _____			
v. Increased spleen size or other splenic abnormality?	☐	☐	☐
vi. If abnormal, please specify: _____			
vii. Other abdominal mass on examination (not USS)	☐	☐	☐
viii. If other mass, where?:	|___| 1 = left upper 2 = epigastric 3 = right upper 4 = left flank 5 = umbilical 6 = right flank 7 = RIF 8 = supra-pubic 9 = LIF		
ix. generalised abdominal pain with guarding	☐	☐	☐
x. Localised abdominal pain	☐	☐	☐
xi. If localised pain, where?	|___| 1 = left upper 2 = epigastric 3 = right upper 4 = left flank 5 = umbilical 6 = right flank 7 = RIF 8 = supra-pubic 9 = LIF		
xii. Specify any other abnormality: _____			
d. Skin			
i. Jaundice?	☐	☐	☐
ii. Lymphadenopathy?	☐	☐	☐
iii. If lymphadenopathy, please specify site:	|___| 1 = neck 2 = groin 3 = both 4 = generalised		
iv. Rash?	☐	☐	☐
v. If rash, please specify type:	|___| 1= erythema, 2 = petechiae, 3 = maculopapular 4 = purpura 5 = other		
vi. If rash, please specify where?	|___| 1= face or head 2 = chest or abdo 3 = limb(s) 4 = widespread 5 = other		
vi. Positive tourniquet test (leave empty if not done)?	☐	☐	☐
vii. Eschar present?	☐	☐	☐
viii. If eschar, please specify site: _____			
ix. Specify any other abnormality: _____			
e. Eyes / Ears / Mouth / Throat			
i. Conjunctival pallor?	☐	☐	☐
ii. Conjunctivitis / eye discharge?	☐	☐	☐
iii. Subconjunctival haemorrhage?	☐	☐	☐
iv. Inflammation, dullness of eardrum, or pus from ear?	☐	☐	☐
iv. Abnormal oral mucosa?	☐	☐	☐
v. If YES, please specify type:	|___| 1 = Koplik's spots 2 = petechiae 3 = oral candida 4 = other		
vi. Inflamed pharynx?	☐	☐	☐
vii. Exudate on tonsils?	☐	☐	☐
viii. Specify and other abnormality: _____			

AHC Fever Study CRF

14. Drug Therapy Commenced in AHC (KC, MC, VK or NCP to complete)

a. Tick if no drug therapy during hospitalisation ☐

b. Drug therapy commenced in AHC (do not complete length until finished treatment)

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

i. Drug name / start date / length: _____ / _____ / _____

c. According to the records, was the patient given a blood transfusion before the **admission** fever serology was taken? **NO** ☐ **YES** ☐

d. According to the records, was the patient given a blood transfusion before the **discharge** fever serology was taken? **NO** ☐ **YES** ☐

APPENDIX B

1. Vital signs of various age groups

Table 19-5 Respiratory Rates of Various Age Groups

	Average Respiratory Range, breaths/min	Respiratory Average, breaths/min
Infant (birth to 1 yr)	30-40	35
Toddler (1-3 yr)	23-35	30
Preschool child (3-6 yr)	20-30	25
School-age child (6-12 yr)	18-26	22
Adolescent (12-18 yr)	12-20	16
Adult (after 18th yr)	12-20	16

From Bonewit-West K: *Clinical procedures for medical assistants*, ed 8, St Louis, 2011, Saunders.

Table 19-4 Pulse Rates of Various Age Groups

Age Group	Pulse Range (beats/min)	Average Pulse (beats/min)
Infant (birth to 1 yr)	120-160	140
Toddler (1-3 yr)	90-140	115
Preschool child (3-6 yr)	80-110	95
School-age child (6-12 yr)	75-105	90
Adolescent (12-18 yr)	60-100	80
Adult (after 18th yr)	60-100	80
Adult (after 60th yr)	67-80	74
Well-trained athletes	40-60	50

From Bonewit-West K: *Clinical procedures for medical assistants*, ed 8, St Louis, 2011, Saunders.

2. Pediatric normal laboratory values

	White blood count	Platelets
Adults	5,000 – 10,000 cell/mm3	150,000 – 450,000 cell/mm3
Children	5,000 – 10,000 cell/mm3	200,000 – 473,000 cell/mm3
Infants	9,000 – 30,000 cell/mm3	140,000 – 300,000 cell/mm3

Source: Laboratory testing and nursing, Khonkaen University

Test	Reference Range	Age
Blood: Electrolytes, Urea & Creatinine		
Sodium	133 – 143 mmol/L	All ages
Potassium	4.5 – 6.5 mmol/L 3.8 – 6.0 mmol/L 3.5 – 5.5 mmol/L	<1 wk 1 wk – 1 month >1 month
Chloride	95 – 110 mmol/L	All ages
Bicarbonate	18 – 24 mmol/L 22 – 26 mmol/L	<2 yrs >2 yrs
Urea	1.0 – 6.0 mmol/L	All ages
Creatinine	1 – 100 umol/L 20 – 50 umol/L 30 – 75 umol/L 55 – 120 umol/L	<1 wk 2 months – 5 yrs 5 – 15 yrs >15 yrs
Total protein	45 – 70 g/L 50 – 71 g/L 55 – 80 g/L	<1 month 1 month – 1 yr >1 yr
Blood: Liver Function Tests		
Total Bilirubin	1 – 100 umol/L 1 – 30 umol/L <10 umol/L	<2 wks 2 – 4 wks >4 wks
Direct Bilirubin	1 – 25 umol/L <10 umol/L	<2 wks >2 wks
ALT (Alanine transaminase)	0 – 105 U/L 10 – 50 U/L	<1 month >1 month
GGT (Gamma glutamyl transpeptidase)	8 – 154 U/L 7 – 103 U/L 9 – 76 U/L 6 – 52 U/L 5 – 40 U/L 5 – 24 U/L <45 U/L	0 – 1 wk 1 – 2 wks 2 – 4 wks 4 – 6 wks 6 – 7 wks 7 wks – 16 yrs >16 yrs
ALP (Alkaline phosphatase)	60 – 320 U/L 40 – 300 U/L 15 – 125 U/L	<6 months 6 months – 15 yrs adult

Site: http://mededucation.bjmu.edu.cn/reportreference/reference/Heatomology.htm

APPENDIX C

The decision tree model selection process

Table A1, Table B1 and Table C1 show the outcomes in terms of number of leaves, size of the trees, root nodes, prediction nodes, correctly and incorrectly classified instances when varied minNumObj (2,5,10 and 15) and confidence (0.01, 0.025, 0.05, 0.1, 0.25 and 0.5).

We selected the most appropriate model by looking at the sensitivity and specificity together with the classification tree obtained. The model with the highest sensitivity and specificity would be preferable. However, we also considered number of nodes in the decision tree. If it has more than six prediction nodes, it would be classified as being too complicated and would not be favorable (see figure A1). We focused on simple tree, easy to understand and the prediction nodes of the selected tree should be relevant to the clinical manifestations and laboratory indicators, and moreover should be practical uses.

Table A 1 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor

minNumObj	Confidence Factor					
	0.01		**0.025**		**0.05**	
2	L: 8 T: 15 RN: neut N1: hct N2: rr . .	C: 82.35 % (938) IC: 17.64 % (201) a b 923 18 a 183 15 b	L: 8 T: 15 RN: neut N1: hct N2: rr . .	C: 82.87 % (944) IC: 17.12 % (195) a b 907 34 a 161 37 b	L: 8 T: 15 RN: neut N1: hct N2: rr . .	C: 82.61 % (941) IC: 17.38 % (198) a b 905 36 a 162 36 b

Table A 2 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

<table>
<tr><th rowspan="2">minNumObj</th><th colspan="6">Confidence Factor</th></tr>
<tr><th colspan="3">0.1</th><th colspan="2">0.25</th><th>0.5</th></tr>
<tr><td rowspan="2">2</td><td>L: 8
T: 15
RN: neut
N1: hct
N2: rr
.
.</td><td>C: 82.79 %
(943)
IC: 17.20 %
(196)
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>902</td><td>39</td><td>a</td></tr><tr><td>157</td><td>41</td><td>b</td></tr></table></td><td>L: 49
T: 85
RN: neut
N1: hct
N2: le
.
.</td><td>C: 81.21 %
(925)
IC: 18.78 %
(214)
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>877</td><td>64</td><td>a</td></tr><tr><td>150</td><td>48</td><td>b</td></tr></table></td><td>L: 140
T: 242
RN: neut
N1: hct& alt
.
.
.</td><td>C: 78.05 %
(889)
IC: 21.94 %
(250)
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>827</td><td>114</td><td>a</td></tr><tr><td>136</td><td>62</td><td>b</td></tr></table></td></tr>
</table>

Table A 3 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		**0.025**		**0.05**	
5	L: 4 T: 7 RN: neut N1: hct N2: rr . .	C: 82.35 % (938) IC: 17.64 % (201) a b 927 14 a 187 11 b	L: 5 T: 9 RN: neut N1: hct N2: rr . .	C: 82.96 % (945) IC: 17.03 % (194) a b 913 28 a 166 32 b	L: 9 T: 16 RN: neut N1: hct N2: gcs . .	C: 82.87 % (944) IC: 17.12 % (195) a b 906 35 a 160 38 b

Table A 4 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

<table>
<tr><th rowspan="2">minNumObj</th><th colspan="6">Confidence Factor</th></tr>
<tr><th colspan="2">0.01</th><th colspan="2">0.025</th><th colspan="2">0.05</th></tr>
<tr><td rowspan="2">5</td><td>L: 9
T: 16
RN: neut
N1: hct
N2: gcs
.
.</td><td>C: 82.61 %
(941)
IC: 17.38 %
(198)
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>900</td><td>41</td><td>a</td></tr><tr><td>157</td><td>41</td><td>b</td></tr></table></td><td>L: 9
T: 16
RN: neut
N1: hct
N2: gcs
.
.</td><td>C: 81.82 %
(932)
IC: 18.17 %
(207)
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>889</td><td>52</td><td>a</td></tr><tr><td>155</td><td>43</td><td>b</td></tr></table></td><td>L: 77
T: 128
RN: neut
N1: hct& alt
.
.
.</td><td>C: 78.92 %
(899)
IC: 21.07 %
(240)
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>842</td><td>99</td><td>a</td></tr><tr><td>141</td><td>57</td><td>b</td></tr></table></td></tr>
</table>

Table A 5 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
10	L: 5 T: 9 RN: neut N1: hct N2: rr . .	C: 83.49 % (951) IC: 16.50 % (188) a b 929 12 a 176 22 b	L: 5 T: 9 RN: neut N1: hct N2: rr . .	C: 83.66 % (953) IC: 16.33 % (186) a b 929 12 a 174 24 b	L: 5 T: 9 RN: neut N1: hct N2: rr . .	C: 83.66% (953) IC: 16.33 % (186) a b 924 17 a 169 29 b

Table A 6 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
10	L: 5 T: 9 RN: neut N1: hct N2: rr . . .	C: 83.84 % (955) IC: 16.15 % (184) a b 922 19 a 165 33 b	L: 8 T: 15 RN: neut N1: hct N2: rr . . .	C: 83.05 % (946) IC: 16.94 % (193) a b 904 37 a 156 42 b	L: 24 T: 42 RN: neut N1: hct & alt	C: 80.68 % (919) IC: 19.31 % (220) a b 878 63 a 157 41 b

Table A 7 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
15	L: 4 T: 7 RN: neut N1: hct N2: rr . . .	C: 83.40 % (950) IC: 16.59 % (189) a b 929 12 a 177 21 b	L: 4 T: 7 RN: neut N1:hct N2: rr . .	C: 83.58 % (952) IC: 16.41 % (187) a b 929 12 a 175 23 b	L: 4 T: 7 RN: neut N1: hct N2: rr . .	C: 83.84 % (955) IC: 16.15 % (184) a b 929 12 a 172 26 b

Table A 8 Model 1 (Non-dengue vs. dengue) when varying the minimal number of objects (minNumObj) and Confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
15	L: 4 T: 7 RN: neut N1: hct N2: rr . . .	C: 84.02 % (957) IC: 15.97 % (182) a b 927 14 a 168 30 b	L: 7 T: 12 RN: neut N1: hct N2: le & rr . . .	C: 84.10 % (958) IC: 15.89 % (181) a b 921 20 a 161 37 b	L: 7 T: 12 RN: neut N1: hct N2: le & rr . . .	C: 83.93 % (956) IC: 16.06 % (183) a b 918 23 a 160 38 b

Model 1 (Non-dengue vs. dengue) continue

Model 1 (Non-dengue vs. dengue)

L: number of leaves

T: size of the tree

RN: root node

N1: node 1

N2: node 2

neut: neutrophil

hct: haematocrit

rr: respiratory rate

le: liver enlargement

a: non-dengue

b: dengue

C: correctly classified instances

IC: incorrectly classified instances

Model prediction			
a (Non-dengue)	**b (Dengue)**		
True negative	False negative	**a (Non-dengue)**	**Final confirmation using SDB algorithm**
False positive	True positive	**b (Dengue)**	

Table B 1 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor

minNumObj	**Confidence Factor**					
	0.01		**0.025**		**0.05**	
2	L: 9 T: 16 RN: pulse N1: hct N2: lym	C: 69.19 % (137) IC: 30.80 % (61) a b 115 45 a 16 22 b	L: 13 T: 22 RN: pulse N1: rash N2: hct	C: 68.68 % (136) IC: 31.31 % (62) a b 117 43 a 19 19 b	L: 13 T: 22 RN: pulse N1: rash N2: hct	C: 69.19 % (137) IC: 30.80 % (61) a b 118 42 a 19 19 b

Table B 2 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.1		0.25		0.5	
2	L: 17 T: 29 RN: pulse N1: rash N2: hct	C: 70.20 % (139) IC: 29.79 % (59) a b 121 39 a 20 18 b	L: 24 T: 42 RN: pulse N1: rash N2: hct	C: 71.21 % (141) IC: 28.78 % (57) a b 124 36 a 21 17 b	L: 24 T: 42 RN: pulse N1: rash N2: hct	C: 71.21 % (141) IC: 28.78 % (57) a b 124 36 a 21 17 b

Table B 3 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
5	L: 8 T: 14 RN: pulse N1: hct N2: lym	C: 68.18 % (135) IC: 31.81 % (63) a b 122 48 a 15 23 b	L: 8 T: 14 RN: pulse N1: hct N2: lym	C: 66.16 % (131) IC: 33.83 % (67) a b 111 49 a 18 20 b	L: 8 T: 14 RN: pulse N1: hct N2: lym	C: 66.16 % (131) IC: 33.83 % (67) a b 111 49 a 18 20 b

Table B 4 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.1		**0.25**		**0.5**	
5	L: 10 T: 17 RN: pulse N1: rash N2: hct	C: 66.66 % (132) IC: 33.33 % (66) a b 112 48 a 18 20 b	L: 10 T: 17 RN: pulse N1: rash N2: hct	C: 67.67 % (134) IC: 32.32 % (64) a b 115 45 a 19 19 b	L: 10 T: 17 RN: pulse N1: rash N2: hct	C: 67.17 % (133) IC: 32.82 % (65) a b 115 45 a 20 18 b

Table B 5 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
10	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 69.19 % (137) IC: 30.80 % (61) a b 110 50 a 11 27 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 68.18 % (135) IC: 31.81 % (63) a b 107 53 a 10 28 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 66.66 % (132) IC: 33.33 % (66) a b 108 52 a 14 24 b

Table B 6 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.1		0.25		0.5	
10	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 68.18 % (135) IC: 31.81 % (63) a b 112 48 a 15 23 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 68.18 % (135) IC: 31.81 % (63) a b 112 48 a 15 23 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 68.18 % (135) IC: 31.81 % (63) a b 112 48 a 15 23 b

Table B 7 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	**Confidence Factor**					
	0.01		**0.025**		**0.05**	
15	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 69.19 % (137) IC: 30.80 % (61) a b 113 47 a 14 24 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 69.19 % (137) IC: 30.80 % (61) a b 113 47 a 14 24 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 69.19 % (137) IC: 30.80 % (61) a b 113 47 a 14 24 b

Table B 8 Model 2 (Non-severe dengue vs. severe dengue) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.1		0.25		0.5	
15	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 69.19 % (137) IC: 30.80 % (61) a b 113 47 a 14 24 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 70.20 % (139) IC: 29.79 % (59) a b 117 43 a 16 22 b	L: 7 T: 13 RN: pulse N1: hct N2: gcs	C: 70.70 % (140) IC: 29.29 % (58) a b 118 42 a 16 22 b

Model 2 (Non-severe dengue vs. severe dengue)

L: number of leaves

T: size of the tree

RN: root node

N1: node 1

N2: node 2

lym: lymphocyte

hct: haematocrit

gcs: glascow coma score

a: non-severe dengue

b: severe dengue

C: correctly classified instances

IC: incorrectly classified instances

Model prediction			
a (Non-severe dengue)	**b (Severe dengue)**		
True negative	False negative	**a (Non-severe dengue)**	**Final confirmation using SDB algorithm**
False positive	True positive	**b (Severe dengue)**	

Table C 1 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor

minNumObj	Confidence Factor					
	0.01		**0.025**		**0.05**	
2	L: 10 T: 18 RN: urea N1: gcs & vomiting	C: 89.89 % (178) IC: 10.10 % (20) a b 177 13 a 7 1 b	L: 10 T: 18 RN: urea N1: gcs & vomiting	C: 89.89 % (178) IC: 10.10 % (20) a b 177 13 a 7 1 b	L: 10 T: 18 RN: urea N1: gcs & vomiting	C: 89.89 % (178) IC: 10.10 % (20) a b 177 13 a 7 1 b

Table C 2 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.1		0.25		0.5	
2	L: 10 T: 18 RN: urea N1: gcs & vomiting	C: 89.89 % (178) IC: 10.10 % (20) a b 177 13 a 7 1 b	L: 10 T: 18 RN: urea N1: gcs & vomiting	C: 89.89 % (178) IC: 10.10 % (20) a b 177 13 a 7 1 b	L: 12 T: 21 RN: urea N1: gcs & vomiting	C: 90.40 % (179) IC: 9.59 % (19) a b 178 12 a 7 1 b

Table C 3 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		**0.025**		**0.05**	
5	L: 8 T: 14 RN: urea N1: gcs & plt	C: 83.83 % (166) IC: 16.16 % (32) a b 164 26 a 6 2 b	L: 8 T: 14 RN: urea N1: gcs & plt	C: 84.34 % (167) IC: 15.65 % (31) a b 167 23 a 8 0 b	L: 8 T: 14 RN: urea N1: gcs & plt	C: 84.34 % (167) IC: 15.65 % (31) a b 167 23 a 8 0 b

Table C 4 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	**Confidence Factor**					
	0.1		**0.25**		**0.5**	
5	L: 8 T: 14 RN: urea N1: gcs & plt	C: 84.34 % (167) IC: 15.65 % (31) a b 167 23 a 8 0 b	L: 8 T: 14 RN: urea N1: gcs & plt	C: 84.34 % (167) IC: 15.65 % (31) a b 167 23 a 8 0 b	L: 8 T: 14 RN: urea N1: gcs & plt	C: 84.34 % (167) IC: 15.65 % (31) a b 167 23 a 8 0 b

Table C 5 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		0.025		0.05	
10	L: 2 T: 3 RN: N1: urea	C: 84.84 % (168) IC: 15.15 % (30) a b 166 24 a 6 2 b	L: 6 T: 10 RN: urea N1: gcs N2: bloodinstool	C: 84.84 % (168) IC: 15.15 % (30) a b 166 24 a 6 2 b	L: 6 T: 10 RN: urea N1: gcs N2: bloodinstool	C: 83.33 % (165) IC: 16.66 % (33) a b 163 27 a 6 2 b

Table C 6 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	**Confidence Factor**					
	0.1		**0.25**		**0.5**	
10	L: 6 T: 10 RN: urea N1: gcs N2: bloodinstool	C: 83.33 % (165) IC: 16.66 % (33) a b 163 27 a 6 2 b	L: 6 T: 10 RN: urea N1: gcs N2: bloodinstool	C: 83.33 % (165) IC: 16.66 % (33) a b 163 27 a 6 2 b	L: 6 T: 10 RN: urea N1: gcs N2: bloodinstool	C: 83.33 % (165) IC: 16.66 % (33) a b 163 27 a 6 2 b

Table C 7 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.01		**0.025**		**0.05**	
15	L: 2 T: 3 RN: urea	C: 91.91 % (182) IC: 8.08 % (16) a b 179 11 a 5 3 b	L: 2 T: 3 RN: urea	C: 91.91 % (182) IC: 8.08 % (16) a b 179 11 a 5 3 b	L: 2 T: 3 RN: urea	C: 90.90 % (180) IC: 9.09 % (18) a b 177 13 a 5 3 b

Table C 8 Model 3 (Survived vs. Died) when varying minNumObj and confidence factor (cont.)

minNumObj	Confidence Factor					
	0.1		0.25		0.5	
15	L: 2 T: 3 RN: urea	C: 90.90 % (180) IC: 9.09 % (18) a b 177 13 a 5 3 b	L: 2 T: 3 RN: urea	C: 87.87 % (174) IC: 12.12 % (24) a b 179 19 a 5 3 b	L: 3 T: 5 RN: urea N1: gcs	C: 87.87 % (174) IC: 12.12 % (24) a b 171 19 a 5 3 b

Model 3 (Survived vs. Died)

L: number of leaves

T: size of the tree

RN: root node

N1: node 1

N2: node 2

gcs: Glasgow coma score

plt: platelet count

a: survived

b: died

C: correctly classified instances

IC: incorrectly classified instances

Model prediction			
A (Survived)	**b (Died)**		
True negative	False negative	**a (Survived)**	**Final confirmation using SDB algorithm**
False positive	True positive	**b (Died)**	

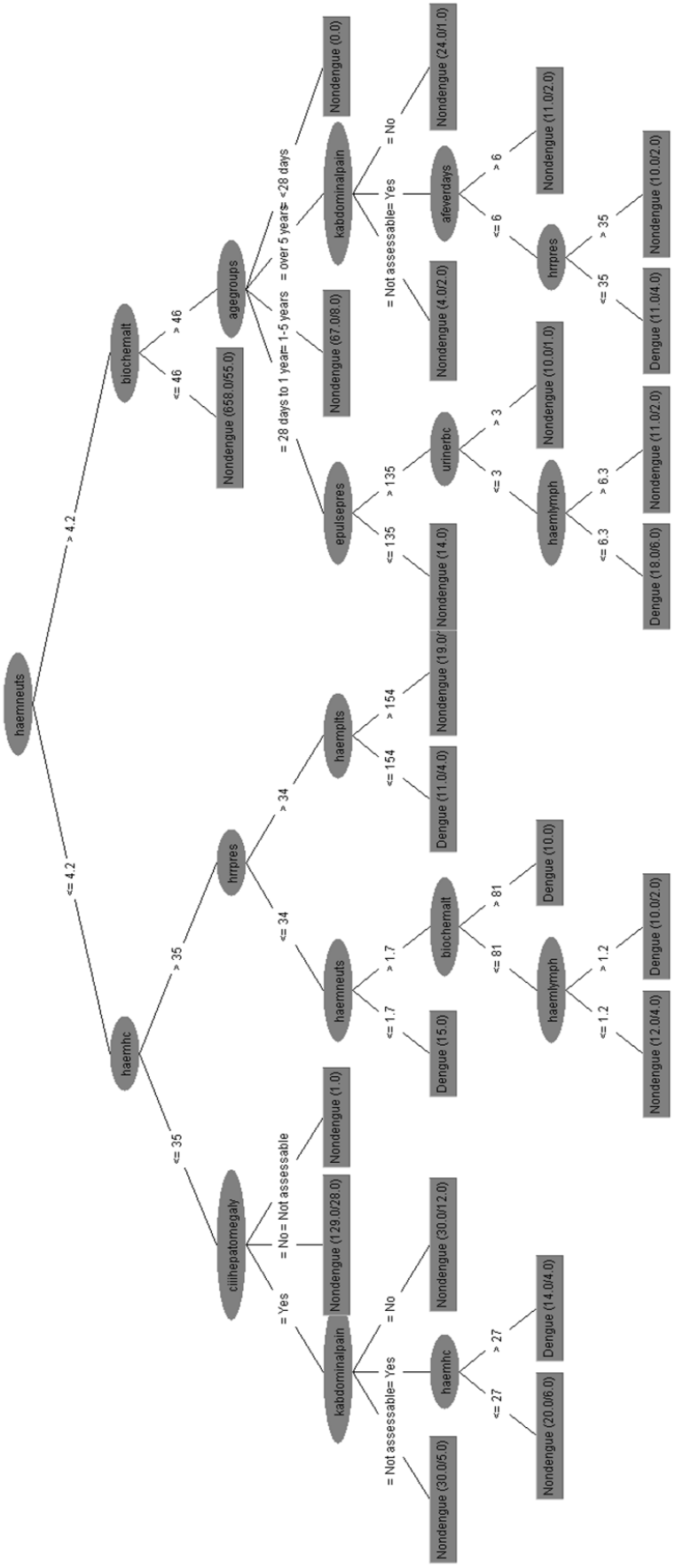


Figure A 1 An example of an inappropriate decision tree (Number of leaves: 24, Size of the tree: 42, prediction nodes: 18)

BIOGRAPHY

NAME	Khansoudaphone Phakhounthong
DATE OF BIRTH	14 November 1989
PLACE OF BIRTH	Vientiane, Lao PDR
EDUCATION	University of Health Sciences, Lao PDR, 2006-2012, Bachelor's degree of medicine.
AWARD	Awarded First Prize in Lao language and literature Competition in Vientiane, 2003**.**
OTHER ACTIVITIES	Worked as an assistant to a French student's hospital director for conducting his thesis in Mahosot hospital, Vientiane, Laos, in November 2012.