INVESTIGATING THE 2005 SINGAPOREAN DENGUE OUTBREAK

LUKAS TANNER

B.Sc. (Major in Molecular Biology), University of Basel

A THESIS SUBMITTED

FOR THE DEGREE OF A MASTER OF SCIENCE

IN INFECTIOUS DISEASES,

VACCINOLOGY & DRUG DISCOVERY

YONG LOO LIN SCHOOL OF MEDICINE

NATIONAL UNIVERSITY OF SINGAPORE

&

BIOZENTRUM

UNIVERSITY OF BASEL

2007

Acknowledgments

I would like to thank my supervisor Mark Schreiber for his patience and support during my masters project. His guidance made my time at NITD a scientific and educational experience which is invaluable for my future career.

I also thank Liu Wei for taking the time to introduce me to virus work and to the sequencing project. He provided great support during the vicious times of PCR optimization.

I would also like to thank Subhash Vasudevan for his fair comments on the project and for giving me the opportunity to pursue my masters project at NITD.

I am also very grateful to Feng Gu, Katja Fink and Cedric Ng who critically read the manuscript and provided me with the necessary amendments. I also would like to mention Robert Zweighardt who took the time to give suggestions from a non-dengue perspective.

From GIS, I thank Pauline Aw Poh Kim for her part in the sequencing. Without her effort it would not have been possible to sequence the EDEN isolates. In particular, I am grateful to Ong Swee Hoe for his phylogenetic analyses and for his helpful discussions with regard to sequence alignments. He provided me with the necessary background to be able to understand dengue virus phylogeny. I additionally express my gratitude to Martin Hibberd for his advice on dengue immunology.

I express thanks to the whole EDEN team but especially, to Eng Eong Ooi who provided me with the clinical and immunological data. It was a pleasure to collaborate with him and he gave me useful suggestions.

I wish to thank Reinhard Bergmann from NIBR who was helpful in the statistical analyses. He provided me with the necessary knowledge and always had time to answer my many questions.

I would like to express my thanks to all the people in the dengue unit who made my time at NITD not only a scientific but also a personal experience. In particular, I express my gratitude to Cedric, Viral, Sarah, Joanne, Katja, Indira, Anne, Dina, Mee, Cheryl, Celine, Alex, Jasmin and Selina who provided me with the necessary friendship and motivation. It was nice to have Stevie and Tommy as flat mates and friends. The time with all of you was great and hopefully we will meet someday, somewhere again…

I am most grateful to my parents Suzanne and Marcel who gave me this unique opportunity to come to Singapore. Their support in good but especially in difficult times is invaluable and I am happy to have them as parents. A special thanks goes to my two sisters Sabine and Catherine. Talking and spending time with them is great and gives me new energy to go on in my life.

Last but not least, I show gratitude to my friends back home in Switzerland who always had time to talk to me on the phone. I've realized that without their good friendship I would not have the energy to fulfill my goals…Merci vielmol!

Table of Contents

Summary

In Singapore, resurgence of dengue cases peaked with 14,209 laboratory confirmed cases in 2005. Interestingly, dengue epidemiology in Singapore presents itself with striking differences to other South-East Asian countries, where the main burden is considered to be among children. In Singapore, however, young adults are at the highest risk which exacerbates the implementation of the WHO classification scheme which is exclusively based on data from children. Thus, it is evident that these WHO guidelines on dengue diagnosis and case management have to be reassessed to accommodate the adult dengue cases as well. To address this goal the Early Dengue (EDEN) study has been launched in 2005 by a joint effort between Tan Tock Seng Hospital (TTSH), the Genome Institute of Singapore (GIS), the Novartis Institute for Tropical Diseases (NITD), the Environmental Health Institute (EHI), the Singapore Tissue Network (STN), DSO National Laboratories (DSO) and the NHG Polyclinic Group. EDEN aims at finding early host and viral markers that contribute to dengue pathogenesis and the present thesis was part of this effort.

Firstly, this work harnessed patients clinical and immunological data collected during the 2005 Singaporean dengue outbreak in an attempt to distinguish dengue fever from other febrile illnesses. This resulting clinical model showed a sensitivity of 82%. The integration of immunological data improved the sensitivity to 90%. It identified a low white blood cell count (<=4800 cells/mm^3) along with lower number of lymphocytes (<=500cells/mm^3) as clinical indicators for dengue infection. Generally, lower serum levels of IL-2, decreased levels of TNF-α but, increased IP-10 and IL-10 levels were immunological markers that were predictive for dengue infection in an early stage of

disease. Secondly, we aimed at finding an accurate prediction of disease severity in adult infections. Classification of severe patients based on a low platelet count (<=50,000/mm^3) between days five and seven of illness resulted in a model that predicted severity within 72 hours of fever onset. It showed a sensitivity of 81% and integrated viral load along with secondary infection into one predictive model. Furthermore, we identified IP-10, IL-8 and IL-10 as possible severity markers for dengue pathogenesis in an early stage of disease.

Thirdly, 94 full length genomes of dengue viruses (54 DENV-1 & 40 DENV-3) isolated during the 2005 Singaporean dengue outbreak were analyzed. Phylogenetic analysis revealed high levels of similarity between strains within a serotype. Dengue serotype 3 viruses were possibly maintained by silent maintenance since 2003 before the large 2005 Singaporean dengue outbreak. Further investigations of identified clades within the two serotypes did not result in the finding of severity related genomic differences but were able to detect IFN-α, IL-8 and viral load as serotype specific differences that might have a share in disease severity. We concluded that serotype 1 was more likely to cause severe disease than serotype 3.

Taken together, this work contributes to the understanding of the course of disease in relation to genotype patterns. The performed studies suggest new predictive models for distinguishing dengue fever from other febrile illnesses and for determining disease severity.

List of Tables

List of Figures

List of Abbreviations

ADCC	Antibody-Dependent Cell Mediated Cytotoxicity
ADE	Antibody Dependent Enhancement
AUC	Area Under The Curve
BHK	Baby Hamster Kidney (Cells)
CBC	Complete Blood Cell Count
CFR	Case Fatality Rate
CI	Confidence Interval
COX	Cyclooxygenase
CS	Cyclization Sequence
DALY	Disability Adjusted Life Year
DC	Dendritic Cell
DENV	Dengue Virus
DF	Dengue Fever
DHF	Dengue Hemorrhagic Fever
DNA	Deoxyribonucleic Acid
DR	Directed Repeats
DSS	Dengue Shock Syndrome
EDEN	Early Dengue
ID	Identity
ID3	Induction of Decision Trees
IFN	Interferon
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IL	Interleukin
NC	Nucleocapsid
NO	Nitric Oxid
OR	Odds Ratio
PBMC	Peripheral Blood Monocytic Cells
pfu	Plaque Forming Unit
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid

RNAsin	RNAse Inhibitor
ROC	Receiver Operating Characteristic
RR	Relative Risk
RRR	Relative Risk Reduction
RT-PCR	Reverse Transcription PCR
SE	Standard Error
SLA	Stem Loop A
TNF	Tumor Necrosis Factor
ULV	Ultra-Low Volume
WBC	White Blood Cell Count
WNV	West Nile Virus

1 Introduction

1.1 Epidemiology of Dengue

1.1.1 The Global Emergence of Dengue

Dengue is an old and the most important arthropod-borne viral disease affecting humans in terms of morbidity. There are reports in the medical literature about epidemics caused by an illness comparable to dengue that reach back to the late 17th century. The Chinese, however, described similar symptoms even earlier, more precisely during the Chin Dynasty (265 to 420 A.D.) (Gubler, 1998).

Sporadic and widespread outbreaks of dengue caused a major burden on public health between the 17th and the 20th century (Gubler, 2004). Nowadays, over 2.5 billion people live in risk areas (Figure 1.1) and 50 to 100 million people suffer from dengue fever (DF) every year. The World Health Organization (WHO) estimates that 500,000 cases of Dengue Hemorrhagic Fever / Dengue Shock Syndrome (DHF/DSS) and more than 20,000 deaths occur per year (WHO, 2002). In the last 25 years of the 20th century, dengue has emerged as a major public health problem and epidemics have a tremendous impact on social as well as economic structures of society especially in developing countries of the tropics (Gubler, 2002).

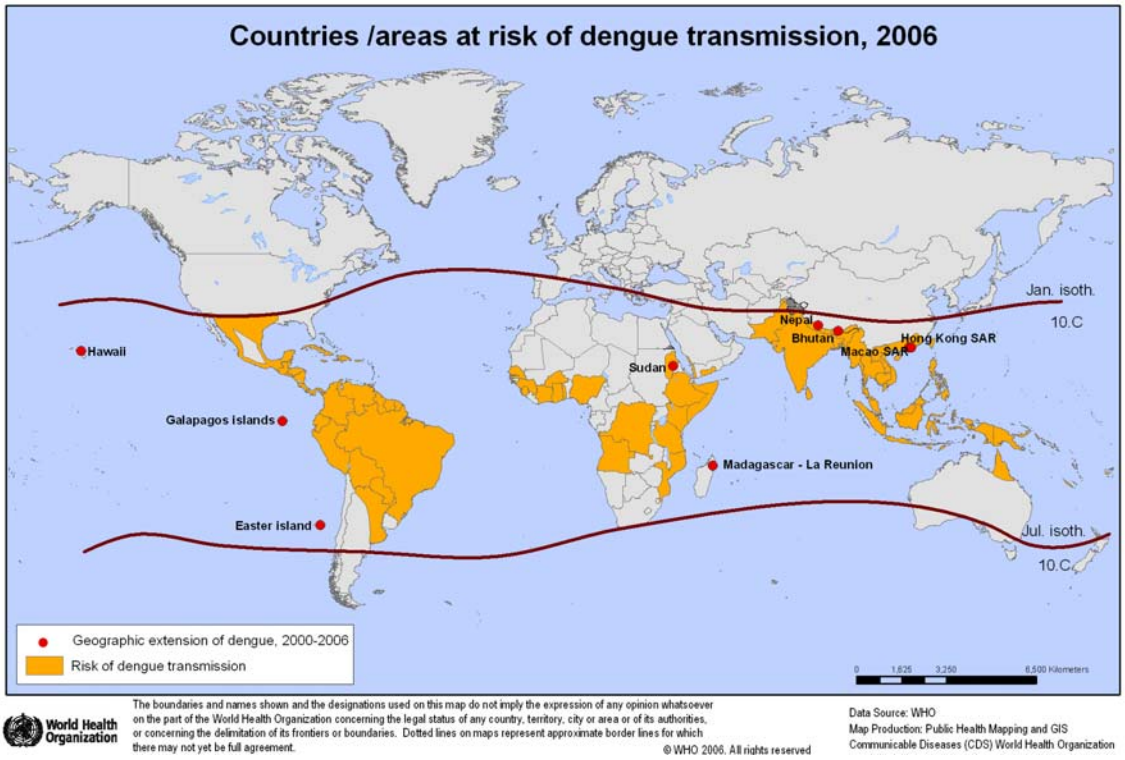


Figure 1.1: Countries and areas at risk of dengue transmission. Yellow color indicates dengue infested areas and countries that experienced new dengue epidemics between 2000-2006 are colored in red (WHO, 2006).

1.1.1.1 Situation in the Americas

In the Americas, major epidemics started to occur periodically in the early 17th century and the USA was confronted with a last major dengue outbreak in 1945 (Ehrenkranz et al., 1971; Gubler, 2004). Epidemics normally had their origin in one country subsequently spreading all over the region and were caused by only one serotype that disappeared after several months. Infections affected thousands of people and were characterised by self-limited classical DF (Gubler, 2004). In the 1900s, implementation of preventive measures and of control programs for subduing *Aedes Aegypti (Ae. Aegypti)* mosquitoes as the vector of yellow fever virus also had a highly

acceptable effect on combating dengue fever. It resulted in the declining or even disappearance of DF throughout the region (Graham et al., 1999; Gubler, 2004).

1.1.1.2 Situation in the Asia/Pacific

However, in the Asia/Pacific region, DF was a common occurrence in the first 50 years of the 20th century. There was an epidemic every 10 to 40 years depending on the introduction of a new virus (Gubler, 2004). Various reports from this time show that dengue virus was endemic during this period but the exact distribution of all four serotypes was unknown. The isolation of all four serotypes in the 1940s (DENV-1 and DENV-2) and 1950's (DENV-3 and DENV-4) finally suggested that dengue virus was already present earlier and was maintained by a monkey-mosquito-monkey cycle (Mackenzie et al., 2004; Weaver and Barrett, 2004). But after and during World War II, the epidemiology of dengue dramatically changed, causing major epidemics and the spread of dengue virus into new geographic areas. The reasons for this change are not completely understood but it is thought that the insertion of hundreds of thousands soldiers, increased population density, troop movement and shipment of war material played a crucial role in causing Asia to become hyperendemic with the co-circulation of multiple dengue virus serotypes (Gubler, 2004). This resulted in an increased transmission of multiple serotypes and in the sequential emergence of dengue hemorrhagic fever (DHF) in the 1950s with a first outbreak in Manila, Philippines. Singapore experienced its first DHF epidemic in the early 1960s (Gubler, 2004).

1.1.1.3 Reasons for the Global Emergence of Dengue

The fledgling stages of the pandemic emergence of dengue originated in the social as well as economic disruptions caused by World War II, that created ideal conditions for mosquito-borne diseases (Gubler, 1998; Gubler, 2002; Gubler, 2004; Mackenzie et al., 2004). Due to the hyperendemic situation in the Asia/Pacific, a newly described disease emerged in the form of DHF in 1960. It spread throughout South-East Asia and by the mid 1970s, DHF had became a major burden among children in this region (Gubler, 2004). On the other hand, the Americas faced dramatic epidemiological changes after discontinuing *Ae. Aegypti* eradication programs in the early 1970s. By the end of the 1990s, *Ae. Aegypti* mosquitoes had nearly regained the geographic distribution leading to major dengue outbreaks in countries that were known to be nonendemic or hypoendemic. From 1981 and 1997, the first DHF cases were reported in the Americas suggesting the same emergence of DHF as it happened 25 years before in the Asia/Pacific (Gubler, 2004).

In summary, we can specify five factors that have been playing a crucial role in the emergence of dengue. Unprecedented population growth combined with unplanned and uncontrolled urbanization led to an increased transmission of arboviral diseases in tropical countries. Additionally, the lack of mosquito control and modern transportation introduced new virus strains and serotypes into new geographic regions that have been regained by the mosquito vector. The last factor of equal importance is represented by the decay of public health infrastructures and the changes in public health policies causing inappropriate outbreak and disease management (Gubler, 2002; Gubler, 2004; Mackenzie et al., 2004).

1.1.2 Public Health, Social & Economic Impacts of Dengue

1.1.2.1 Impact on Public Health

Due to poor dengue surveillance and due to great similarities in disease manifestation to other tropical diseases such us malaria and chikungunya fever (Rigau-Perez et al., 1998), the early stages of dengue outbreaks are mostly not detected and thus, dengue cases are underreported (Gubler, 2002). This results in an unwanted impact on disease management because it is suggested that dengue patients have to be treated as early as possible to prevent the transition from DF to DHF. Furthermore, the case fatality rate (CFR) of DHF varies among countries between <1% to 15% depending on disease surveillance, disposal and condition of health facilities as well as on trained and capable health care worker (Gubler, 2002). Later on in an epidemic, more precisely when transmission peaks, the disease gets recognized and ironically gets over reported. The sequentially implemented emergency mosquito control is often too late and highlights the lack of public health planning and the poor preparedness for future outbreaks.

1.1.2.2 Impact on Society

The social impact of DF/DHF on community must not be underestimated. The complacency that originated between the 1950s and 1960s when the mosquito control programs in the Americas were highly successful resulted in the decrease of public awareness. The global spread of dengue disease was mainly ignored by public health

officials and hence, the people had to tolerate the interruption of their daily life by epidemics occurring every few years (Gubler, 2002). Especially, the insufficient surveillance and case management for DF/DHF in developing countries have a share in the overloading of primary health care centres, resulting in overworked staff and in the sub-standard treatment of life-threatening conditions of DHF and dengue shock syndrome (DSS) as well as in the increasing mortality. These circumstances led to chaos and have forced the governments to attempt preventive measures being mostly insufficient and delivering a false feeling of security (Gubler, 2002).

1.1.2.3 Impact on Economy

In 2002, the WHO estimated a global burden of dengue in the range of 612,000 Disability Adjusted Life Years (DALYs) and 87% of them were represented by children younger than fifteen years of age (WHO, 2002). At first glance, this number seems to be fairly low compared to other infectious diseases such as tuberculosis and malaria. But a study performed a few years ago investigating the burden of DF and DHF in Puerto Rico between 1984-1994, showed that the magnitude of DALYs lies in the range of malaria and tuberculosis (Meltzer et al., 1998). This suggests that, in addition to the epidemic periods, there is a considerable disease burden during inter-epidemic periods. Mainly, lost productivity and time away from work or school are indirect underreported costs that have a huge impact in third world settings where absence of work can easily be the start to the vicious cycle of poverty (Gubler, 2002; Meltzer et al., 1998). Therefore, dengue has to be considered as a major public health burden and more attention has to be paid in terms of research and prevention.

1.1.3 Epidemiological Situation in Singapore

1.1.3.1 Dengue Epidemiology in Singapore

In Singapore, dengue has been successfully prevented through a highly effective vector control program implemented between the 1970s and the 1980s but despite that, dengue cases have resurged since 1990 (Ooi et al., 2006). In 2005, dengue resurgence peaked with 14,209 laboratory confirmed cases (Figure 1.2), an increase of more than 50% compared to 2004. Most cases were reported between June and October, but the month of September reported a weekly incidence of over 700 cases (Ministry of Health, 2006). Studies performed to investigate the seroepidemiology in the Singaporean population showed a completely different picture than observed in other South-East Asian countries where the main burden of dengue usually affects children and female adults (Wilder-Smith et al., 2004). This typical transmission picture is mainly influenced by the highly domesticated lifecycle of the vector *Ae. Aegypti* (Ooi et al., 2006).

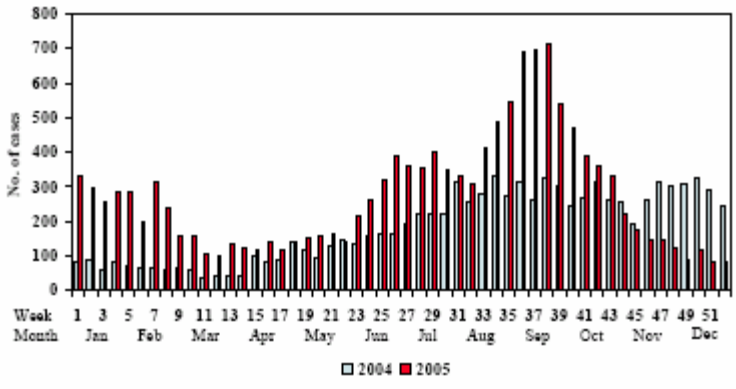


Figure 1.2: Weekly distribution of DF/DHF cases in Singapore between 2004-2005. In 2005, Singapore reported an increase of more than 50% cases compared to 2004 (MOH, 2006).

In Singapore, however, young male adults have the highest risk of being infected with dengue virus (Ooi et al., 2003; Ooi et al., 2006; Wilder-Smith et al., 2004). In 2005, the male to female ratio was 1.4:1 with the highest incidence found in the age group of 15 to 24 year olds (Ministry of Health, 2006). A study performed in 2004 showed that the odds ratio of dengue seroprevalence increases by 4.13 for every 10 years increase in age (Wilder-Smith et al., 2004). The factors involved are not fully understood and several studies revealed lowered herd immunity, increased virus transmission outside the home and shift in the surveillance emphasis of the vector control program as the responsible players (Ooi et al., 2006). In addition, adults suffer more often from symptomatic infections than children and therefore more cases are reported. Furthermore, adults are at lower risk to develop DHF/DSS than are children due to age-dependent differences in vascular permeability. The switch from epidemics causing DHF/DSS to epidemics of classical DF can be nicely observed over the years. Before implementation of the effective mosquito control programs, the disease had a higher prevalence in children and therefore, more cases of DHF/DSS were reported. Nowadays, even though the overall occurrence of cases is much higher, manifestations of classical DF are much more likely to occur than DHF/DSS as a consequence of adult infections (Ooi et al., 2006).

The low incidence rate of the more severe forms of dengue hint towards the unsuitability of the WHO classification scheme (Introduction; Page 37) in Singapore. The scheme is based on clinical data from children and probably does not reflect the present situation in Singapore. It is evident that these guidelines on dengue diagnosis and case management have to be reassessed and improved formulations suitable for adult dengue cases have to be found.

1.1.3.2 The Early DENgue (EDEN) Study

To address this goal, a longitudinal study has been implemented by a joint effort between the Tan Tock Seng Hospital (TTSH), the Genome Institute of Singapore (GIS), the Novartis Institute for Tropical Diseases (NITD), the Environmental Health Institute (EHI), the Singapore Tissue Network (STN), DSO National Laboratories (DSO) and the NHG Polyclinic Group (Low et al., 2006). The six specific goals of the EDEN study are as following:(1) Identify early markers of the disease that are predictive of outcomes such as DF and DHF, (2) Identify pathways that lead to severe disease that may be amenable to therapeutic intervention, (3) Study the epidemiological features of adult infection, (4) Develop robust molecular tools for epidemiological investigation, (5) Correlate virus virulence with their sequences and their replication properties and (6) Refine early dengue clinical and laboratory diagnostic tools (Low et al., 2006). The study has been launched in 2005 and during the 2005 Singaporean dengue outbreak, 455 patients were enrolled and 133 of them were PCR confirmed dengue positive cases. An initial blood and saliva sample is taken on the first visit (1 to 3 fever days) and consequently on the second visit (4 to 7 fever days). Finally, a third sample during convalescence is taken 3 to 4 weeks after the first visit. Afterwards, a full blood count is performed on the samples and the data is analyzed in respect to disease outcome (Materials & Methods; Page 66). The study is expected to be completed by February 2007 (Low et al., 2006).

1.2 Classical Dengue Fever, Dengue Haemorrhagic Fever & Dengue Shock Syndrome

1.2.1 Clinical Manifestations of Dengue

1.2.1.1 Classical Dengue Fever

Dengue is a mosquito-borne viral disease and has an incubation period of 3 to 14 days. The disease can be asymptomatic or lead to a large range of clinical symptoms. Classical DF is usually an acute flu-like febrile illness including fever, frontal headache, myalgias and frequently arthralgias, nausea, vomiting and rash (Rigau-Perez et al., 1998; WHO, 1997). This classical form of DF normally occurs in older children and adults whereas younger children are mostly asymptomatic or minimally symptomatic (Burke et al., 1988). The disease lasts five to seven days and the virus disappears after an average of five days at the starting point of defervescence. Dengue fever may be followed by a convalescence of several weeks presenting a major impact on people's productivity as discussed above (Gubler, 1998; Gubler, 2002; Hayes and Gubler, 1992).

1.2.1.2 Dengue Hemorrhagic Fever and Dengue Shock Syndrome

In rare cases, dengue virus infections can lead to a more severe form called DHF/DSS. This vascular leakage syndrome is observed in all age groups but it is still considered the main burden among children. DHF/DSS starts with a sudden onset of temperature and similar symptoms found in classical DF (Gubler, 1998). The fever period is accompanied by high temperatures and endures for 2 to 7 days with minor or major bleeding. It is believed that an immune mediated mechanism commencing with the infection of cells of the monocytic lineage, which secrete cytokines and other chemical mediators, ultimately leads to the DHF/DSS typical increased vascular permeability or leakage (Fink et al., 2006; Green and Rothman, 2006). If not properly treated, this condition can lead to death. Viral and host factors suggested to be involved in triggering this immunological cascade include previous infection by a heterotypic dengue virus, virus strain, age as well as immunological and genetic background of the patient (Gubler, 1998; Mackenzie et al., 2004; Rigau-Perez et al., 1998).

There are some few reports about other severe manifestations caused by dengue virus infections such as massive haemorrhage, organ failure, cardiomyopathy and neurological diseases in form of encephalitis (Nimmannitya et al., 1987; Rigau-Perez et al., 1998).

1.2.2 Dengue Diagnosis and Discussion of the WHO Classification Scheme

1.2.2.1 Important Laboratory Tests for Dengue Diagnosis

Besides the description of clinical symptoms, there are also clinical laboratory tests that are useful in the diagnosis of dengue. These clinical tests include a complete blood cell count (CBC), especially the white blood cell count (WBC), platelet count and haematocrit levels. The other useful tools for diagnosis are albumin, liver function and urine tests (CDC, 2005).

Confirmation of suspected dengue infections is commonly done with basic serologic and dengue specific tests (Gubler, 1998). An acute-phase blood sample should always be taken as soon as possible after the onset of suspected illness. Optimally, there should be a second blood sample taken after 2 to 3 weeks to affirm the disappearance of dengue virus from the blood. The follow up of non-hospitalized cases is difficult and laborious. Therefore, it is suggested to take a second blood sample of hospitalized cases before hospital discharge. Direct virus isolation is helpful in detecting the serotype of the infecting virus by RT-PCR and is performed by either using mosquito cell cultures or by mosquito inoculation. The IgM ELISA is the basic serologic test and involves the detection of anti-dengue neutralizing IgM antibodies (CDC, 2005; Gubler, 1998).

1.2.2.2 WHO Classification Scheme

The WHO scheme for dengue diagnosis consists of the above mentioned dengue case definition (having dengue or non-having dengue) and dengue case classification (classification of disease manifestations into DF/DHF/DSS and other forms) (Bandyopadhyay et al., 2006; WHO, 1997). Dengue case classification has been widely discussed by researchers and resulted in the demand for reassessment of the WHO guidelines. According to WHO guidelines, DHF must fulfill all of the following four criteria: (1) Fever or history of acute fever lasting 2 to 7 days, (2) hemorrhagic tendencies evidenced by at least one of the following: a positive tourniquet test, petechia, purpura, ecchymoses, bleeding from mucosa, gastrointestinal tract, injection sites or other location, haematemesis, melena, (3) thrombocytopenia (<=100,000 platelets/µl) or (4) =>20% rise of hematocrit value relative to the normal baseline or evidence of plasma leakage (e.g. pleural effusion or ascites) (Bandyopadhyay et al., 2006; WHO, 1997).

DHF is further classified into four grades of severity (I-IV). DHF I is only defined by a positive tourniquet test whereas patients suffering from DHF II show spontaneous bleeding in the skin, through the nose or in the internal organs. DHF III is manifested in the form of hypotension, narrow pulse pressure, restlessness as well as rapid weak pulse and DHF IV is accompanied by profound shock with undetectable blood pressure or pulse. DHF grades III and IV constitute dengue shock syndrome (DSS) (Bandyopadhyay et al., 2006; WHO, 1997).

The description of severity was originally established in 1974 and was based on the data of a clinical study of Thai children performed in 1964 (Cohen and Halstead, 1966). Even though, the resulting recommendations for case management resulted in the decrease of the case fatality rate (CFR), it is stringent that an improved and globally used classification scheme is introduced (Deen et al., 2006; Rigau-Perez, 2006). Due to the global expansion and due to the changing epidemiology of dengue disease, several investigators face difficulties in using the old WHO guidelines. As a result, the terms "dengue fever with signs associated with unusual haemorrhage" and "dengue with signs associated with shock" have been introduced to try to find the correct definitions for atypical severe dengue cases (Deen et al., 2006). Using the WHO guidelines, DHF/DSS is only fulfilled when the four above discussed manifestations are present, but in a lot of severe cases only one manifestation is observed. Comparisons of cases classified by the WHO scheme to classification by expert clinicians, resulted in a sensitivity of 82% (Rigau-Perez, 2006). Additionally, the WHO classification scheme is often misquoted and requires different and repeated clinical tests which might be a serious challenge for countries with limited resources (Rigau-Perez, 2006).

Therefore, it has been suggested that the WHO guidelines have to be reassessed employing the following three criteria (Deen et al., 2006): First, there is much overlap between DF, DHF and DSS and the guidelines do not treat this fact appropriately, but rigorously distinguish the three different disease manifestations. Second, not all of the severe cases fulfil the four criteria for DHF/DSS and finally, the term dengue haemorrhagic fever can be misleading because plasma leakage is the life-threatening condition to look-out for.

1.2.3 Prevention and Treatment of Dengue

1.2.3.1 Vector Control

In the Americas, first attempts to eradicate the vector *Ae. Aegypti* were successful in the beginnings of the 20th century and prevention of mosquito to human transmission was possible (Gubler, 2004). Today, implementations of mosquito control programs face major difficulties due to insufficient knowledge, compliance issues and lack of resources. It is also intriguing, that the use of ultra-low volume (ULV) concentrates of insecticides have become a routine during pre- and post-epidemic periods. They were originally supposed to be an emergency response during epidemics and they have very limited impact on adult as well as immature stages of the mosquito (Gubler, 2002; Newton and Reiter, 1992; Rigau-Perez et al., 1998). Therefore, dengue has to be promoted as a priority among health officials and the general public to achieve a common sense of responsibility for prevention of mosquito to human transmission.

1.2.3.2 Dengue Vaccines

To date, there is no dengue specific drug on the market and ongoing research has mostly been focused on the development of vaccines. A live attenuated vaccine seems to be the most promising at the moment and several candidates are in phase I or phase II clinical trials. Chimeric virus, DNA, inactivated and subunit recombinant vaccines are also of interest but they are still in the preclinical development (Chaturvedi et al., 2005). It is estimated that an approved vaccine is not likely to be available for the next

5 to 7 years because of the complex immune reactions that are involved in dengue pathogenesis (Ooi et al., 2006).

1.2.3.3 Drugs against Dengue

Since 2002, the Novartis Institute for Tropical Diseases (NITD) funded by the Singapore Economic Board and by the Novartis Foundation has been the first institution putting immediate effort into R&D of a dengue specific chemotherapy with a strong vision to launch a widely available drug by the end of 2012. The main targets that are followed up are the E Glycoprotein, the NS3 protease, the NS3 helicase and the NS5 RNA dependent polymerase.

1.2.3.4 Current Treatment of DF and DHF/DSS

Currently, treatment of classical DF requires rest, oral fluids to compensate dehydration via diarrhea or vomiting and analgesics (Gubler, 1998; Rigau-Perez et al., 1998). Antipyretics such as paracetamol are given for pain relief but aspirin and non-steroidal anti-inflammatory drugs have to be avoided due to risk of impairment of platelet function (CDC, 2005; Gubler, 1998; Rigau-Perez et al., 1998; WHO, 1997). In suspicion of severe illness, fluids should be provided intravenously dependent on the patient's blood pressure, haematocrit levels, platelet counts, haemorrhagic manifestations, urinary output and on patient's level of consciousness. Isotonoic solutions and plasma expanders are additionally administered to deal with plasma loss

caused by increased vascular permeability (CDC, 2005; Gubler, 1998; Rigau-Perez et al., 1998; WHO, 1997).

1.3 The Causative Agent: Dengue Virus

1.3.1 Phylogeny

Dengue virus belongs to the genus Flavivirus (family: *Flaviviridae)* which consists of more than 70 viruses (Kuno et al., 1998; Mackenzie et al., 2004). Based on the NS5 gene sequence, the flaviviruses are clustered into three distinct groups which nicely correlate with the mode of transmission (mosquito-borne, tick-borne, unknown vector) (Kuno et al., 1998). Dengue virus, Japanese encephalitis virus (JEV), West Nile virus (WNV) and yellow fever virus (YFV) are the most important members of the genus and cause every year large epidemics in many parts of the world.

It is believed that dengue virus (Figure 1.3; 47) evolved from a common ancestor 1,000 years ago and that human transmission started between 125 and 325 years ago (Mackenzie et al., 2004; Twiddy et al., 2003). It is still not clear whether the virus originated from Africa or from Asia but dengue transmission was first maintained in a sylvatic lifecycle with *Aedes* species (Holmes and Twiddy, 2003). Sporadic outbreaks first occurred in predomestic regions due to transmission by *Aedes albopictus* and showed similar dynamics as are now observed during yellow fever epidemics. As discussed above, adaptation of *Aedes* species, mainly *Aedes aegypti*, to urban and densely inhabited areas created optimal conditions for human transmission resulting in the emergence of dengue epidemics (Gubler, 2002; Gubler, 2004; Mackenzie et al., 2004).

It is intriguing that the four serotypes are phylogenetically different to an extent as observed between different flaviviruses (Figure 1.3; Page 47) which, in turn, highlights their independent zoonotic transfer and their separate evolutionary development (Holmes and Twiddy, 2003). Thus, it is suggested that the phenomenon of antibody dependent enhancement (ADE) is not the driving force of genetic diversity but might be the result of recent contact between the four separately evolved viruses showing antigenic dissimilarities (Holmes and Twiddy, 2003). Evidence showing correlation between secondary infections and disease outcomes as well as the hyperendemic situation with the co-circulation of multiple serotypes additionally support this independent evolution hypothesis. A first study to determine intra-serotype variation was performed by Rico-Hesse using the junction between the E/NS1 genes of DENV-1 and DENV-2 (Rico-Hesse, 1990). Nowadays, it is more common to use the entire sequence of the E-protein (1485nt) and recent phylogenetic studies revealed that three to five human genotypes could be found within a serotype (Holmes and Twiddy, 2003; Rico-Hesse, 2003).

1.3.1.1 Dengue Serotype 1 (Figure 1.3; Page 47)

According to Rico-Hesse, DENV-1 can be subdivided into five genotypes representing different geographic locations (Americas/Africa, South Pacific, Asia Thailand) and different host characteristics (human/sylvatic) (Rico-Hesse, 2003). By contrast, a more recent study using 177 DENV-1 E gene sequences grouped serotype 1 into 3 genotypes but reflection of the geographical distribution was still maintained (A et al., 2004). Genotype I consists of viruses from two different clades (Africa/Asia and Asia/Oceania) connected by short branch lengths and therefore suggesting that viruses

of this genotype are mixed regardless of their location. On the other hand, the picture of genotype II (viruses from Asia and Oceania) is very different representing clusters that are nicely defined by their geographic occurrence. Finally, branch interpretation of genotype III composed of three smaller clades (Asia, Africa, Latin America) is more difficult due to lack of resolution. Nevertheless, it has been postulated that genotype III viruses isolated in Latin America had earlier been introduced from Asia (Holmes and Twiddy, 2003; Rico-Hesse, 2003).

1.3.1.2 Dengue Serotype 2 (Figure 1.3; Page 47)

Phylogeny of DENV-2 is the best studied due to an abundance of sequence data. Earlier studies suggested that five genotypes (Malaysia/Indian Subcontinent, South-East Asia, Americas, West Africa, sylvatic) would be present (Rico-Hesse, 2003). However, calculations using the current DENV-2 sample results in a slightly different picture detecting six possible genotypes (Cosmopolitan, Asian 2, Asian 1, American/Asian, American) (Holmes and Twiddy, 2003). Strikingly, genotypes found in both studies are representing different geographic distributions and state that sylvatic DENV-2 strains will have little epidemiological impact on human populations. Furthermore, both studies recognized the potential of DENV-2 genotypes to spread into various distinct geographic locations which has possibly been the cause of the introduction of the South-East Asian genotype into the Americas (Holmes and Twiddy, 2003; Rico-Hesse, 2003). It was shown that this introduction correlated with the occurrence of DHF/DSS (Rico-Hesse et al., 1997) and later experiments underlined the early replicative advantage of the South-East Asian genotype (Cologna et al., 2005).

1.3.1.3 Dengue Serotype 3 (Figure 1.3; Page 47)

DENV-3 consists of 4 or 5 distinct subtypes depending on the analysis performed (Messer et al., 2003; Rico-Hesse, 2003). Subtype I includes viruses from South-East Asia and the Pacific, whereas subtype II is only represented by viruses from Thailand. Subtype III was isolated in the Indian Subcontinent as well as Africa and subtype IV is comprised of viruses from Puerto Rico and Tahiti (Messer et al., 2003). Subtype III shows interesting patterns in terms of evolutionary and clinical aspects. Further clustering of this subtype into smaller clades (clade A (pre-DHF emergence) and clade B (post-DHF emergence)) revealed correlations with DHF (Messer et al., 2003). In Sri Lanka, regular outbreaks causing DHF only occurred after 1988 and it was proven that neither increased transmission nor introduction of a new serotype were the determining factors (Messer et al., 2002). Messer and collaborators showed that DENV-3 isolated before 1988 clustered into clade A whereas DENV-3 isolates after 1989 belonged to clade B. Possible explanations for this emergence might be (1) the introduction of group B viruses from East Africa or India and/or (2) a selective force favouring group B viruses after 1989 (Messer et al., 2003; Messer et al., 2002). A similar picture was also observed in the Americas and it was suggested that the group B subtype III variants spread from the Indian subcontinent to East Africa and subsequently to Latin America (Messer et al., 2003). This might be an additional reason for the emergence of DHF/DSS in the Americas which occurred between 1989 and 1997 (Messer et al., 2003).

1.3.1.4 Dengue Serotype 4 (Figure 1.3; Page 47)

There is still not much data available for DENV-4 which is the first to separate from the phylogenetic tree (Rico-Hesse, 2003). Considering DENV-4 to be the most divergent of all four serotypes, it was suggested that strain migration of DENV-4 in the Americas was influenced by different patterns in host immunity (Carrington et al., 2005). Upon introduction, DENV-4 introduced from Asia was first isolated in 1981 and showed a tremendous increase in genetic diversity reflecting rapid transmission throughout the region. After that, genetic diversity was retained reflecting the low mutation rate and reports of DENV-4 activity declined which is underlined by decreased transmission and rare occurrence in dengue outbreaks (Carrington et al., 2005). Rico-Hesse suggested that there are currently the Malaysia, South-East Asia and Indonesia genotypes (Rico-Hesse, 2003) but further studies are essential to get a better picture of DENV-4 evolution.

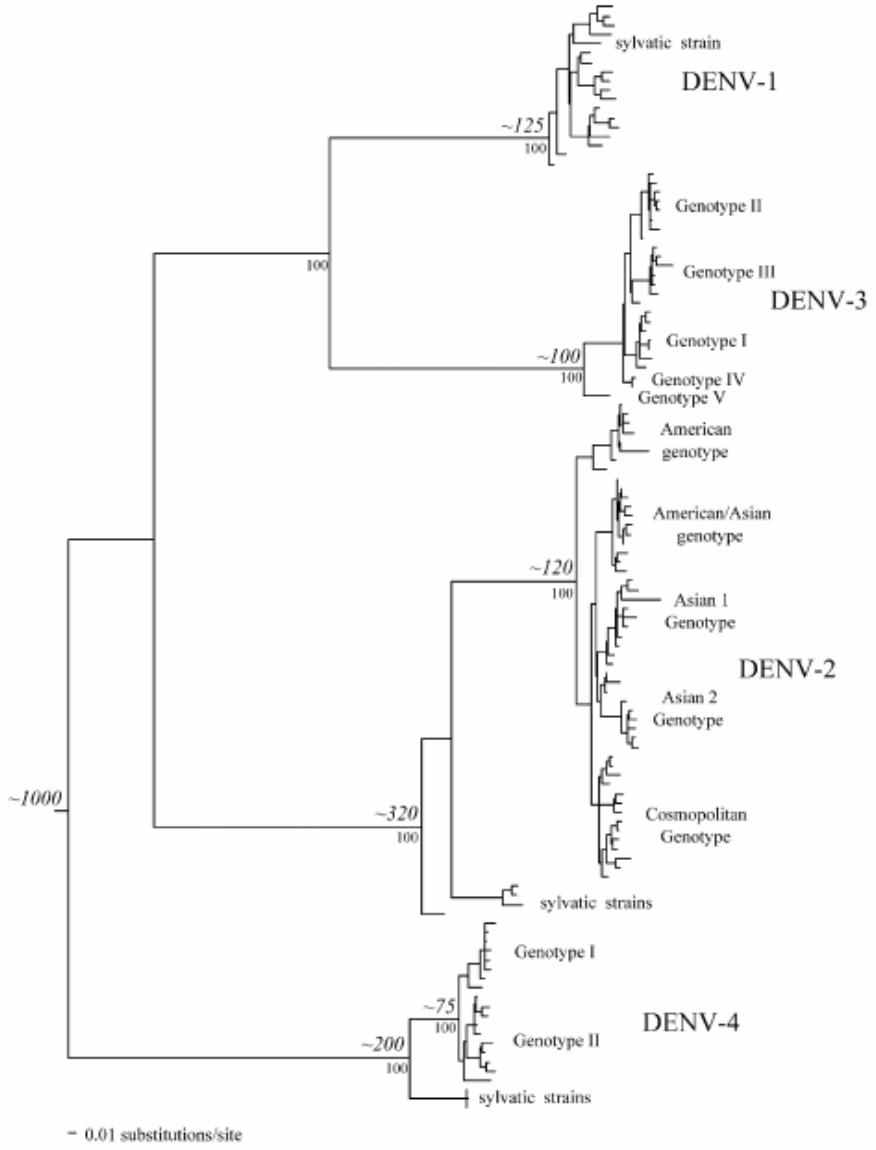


Figure 1.3: Phylogenetic Tree based on 120 E gene sequences (1485 bp) representing the genetic diversity in dengue virus. Within each serotype, the corresponding genotypes are indicated and at each tree node, the approximate divergence times are shown. Claculated bootstrap values are shown below the selected nodes (Holmes and Twiddy, 2003).

1.3.1.5 Serotype Switch & Clade Replacement

Circulation of multiple serotypes in highly endemic regions leads to serotype switch and intraserotypic clade replacement which are commonly observed phenomena that play a crucial role in shaping transmission dynamics. It was observed that

intraserotypic clade replacement increased the overall abundance of a first serotype but simultaneously decreased the population of the second serotype due to mutations affecting viral fitness. In turn, the later decline of the prevalence of the first serotype was associated with the increase in prevalence of the second serotype (Zhang et al., 2005). Finally, it was suggested that still existing clades belonging to the first serotype were antigenically more distinct to the second serotype than the clades that did not survive the period of the second serotype. This clearly hints towards an active role of cross-protection in serotype switch but does not completely explain the phenomenon (Zhang et al., 2005) and in contrast, other studies proposed stochastic explanations for clade replacement as well as serotype switch (Myat Thu et al., 2005). Therefore, further studies are required to determine the complex interactions of selective and stochastic forces that shape DENV-evolution and determine genotypic diversity.

1.3.2 Dengue Virus Lifecycle

1.3.2.1 Structure of Dengue Virions (Figure 1.4; Page 49)

The mature dengue virions have a diameter of approximately 500 Å and consist of a viral genome of around 10.8kb packed by the dimeric capsid proteins. The resulting nucleocapsid is enclosed by a host-derived lipid bilayer surrounded by 180 copies of the E and M protein that form an icosahedral symmetry (T=3) during the fusion process (Kuhn et al., 2002). The viral genome is comprised of a (+)-stranded RNA encoding three structural (C, prM and E-Protein) and seven non-structural proteins (NS1, NS2a, NS2b, NS3, NS4a, NS4b and NS5). The RNA genome is the only viral

mRNA encoding a single open reading frame and is transcribed as a polyprotein. Further processing and cleavage of the polyprotein by the viral serine-protease NS3, host signalases and furin lead to the final ten products that play an important role in viral assembly (C, prM and E proteins), viral maturation (conversion of prM to M) and replication (non-structural proteins) (Mukhopadhyay et al., 2005).

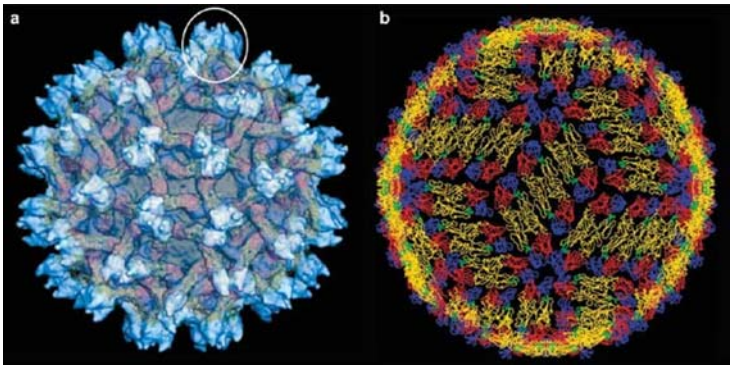


Figure 1.4: Structure of dengue virions. (a) The immature particle is less smooth than the mature form and has 60 ptotein 'spikes' that stick out. (b) The structure of the mature dengue virus is unusually smooth. The membrane is completely enclosed by a protein shell consisting of 30 rafts of three parallel dimers of the E protein (red: *domain I*; yellow: *domain II*; blue*: domain III* and green: *fusion peptide*) (Kuhn et al., 2002).

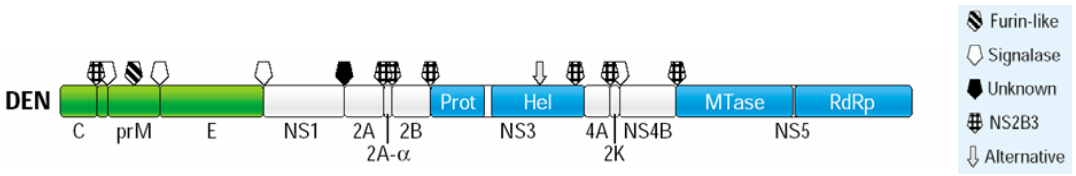


Figure 1.5: Overview of the dengue polyprotein. Further processing and cleavage of the polyprotein by the viral serine-protease NS3, host signalases and furin lead to the final ten products that play an important role during viral lifecycle. C=*capsid protein*; prM=*pre M protein*; E=*envelope protein*; NS1-5=*non-structural proteins 1-5*; Prot=*protease*; Hel=*helicase*; MTase=*methyl transferase*; RdRp=*RNA dependent polymerase* (NITD, 2005).

1.3.2.2 Viral Entry

Dengue virus enters host cells including dendritic cells, monocytes/macrophages, B cells, T cells, endothelial cells, hepatocytes, and neuronal cells by receptor-mediated endocytosis (Clyde et al., 2006) (Figure 1.6; Page 55). Despite some few reports suggesting other entry mechanisms of dengue virus, the receptor-mediated endocytosis is generally accepted as the mode of entry. The primary receptors for this mechanism are DC-SIGN, GRP78/BiP and CD14-associated molecules (Chen et al., 1999; Jindadamrongwech et al., 2004; Navarro-Sanchez et al., 2003; Tassaneetrithep et al., 2003). The interaction between DC-SIGN and the E glycoprotein was recently solved by Pokidysheva and collaborators (Pokidysheva et al., 2006). They have clearly shown that the carbohydrate recognition domain (CRD) of DC-SIGN interacts with the N-glycosylated carbohydrate moieties of the E protein. They propose that this interaction represents a mechanism for avoiding an immune response. In previous studies it was suggested that the carbohydrate moieties on the virus surface might modulate specificity of receptor binding. There is only a ~40% amino acid identity among all the flaviviruses and the number and position of N-glycosylated residues is not conserved among different strains of the same virus. A comparative study of two dengue strains showing differences in infectivity revealed a unique amino acid in the E-protein leading to the loss of a N-linked glycosylation site and therefore to a decreased infectivity (Ishak et al., 2001). Another study with the New York strain of West Nile virus (WNV) suggested that E protein glycosylation is a molecular determinant of neuroinvasiveness (Shirato et al., 2004). Finally, Hanna and collaborators concluded from their experiments that the glycosylation of WNV prM and E proteins can affect

the efficiency of virus release and infection in a cell specific manner (Hanna et al., 2005).

The acidification in the endosome leads to a trimerization of the dimeric E glycoprotein followed by the release of viral RNA due to the class II fusion process (Figure 1.6; Page 55). The E Protein consists of three ectodomains (domains I-III) with a fusion peptide at the tip of domain II. The pH induced conformational change exposes the fusion peptide which then inserts into the host membrane to facilitate the class II fusion steps (Mukhopadhyay et al., 2005). An interesting finding during structural studies of the E protein revealed that the detergent molecule N-octyl-β-D-glucoside can occupy a pocket that forms the hinge region and is positioned at the interface of domains I and II (Modis et al., 2003; Modis et al., 2004). This pocket is essential for membrane fusion and it is thought that the hinge region controls the class II fusion process due to its flexibility. This finding is pertinent for drug discovery since blocking the pocket in the hinge region may prevent membrane fusion and thereby prevent the multiplication of the virus.

1.3.2.3 Viral Replication, Assembly and Exocytosis

Once the genome is released into the host cell, the positive sense RNA is transcribed into the polyprotein at the surface of the endoplasmatic reticulum (ER) (Figure 1.6; Page 55). During the process of translation the structural proteins are translocated and anchored into the endoplasmatic reticulum by membrane anchor sequences. The prM protein, the E protein and the non-structural proteins are found in the lumen of the endoplasmatic reticulum (ER) meanwhile the capsid protein is anchored on the outer

membrane of the ER (Mukhopadhyay et al., 2005). A study showed that translation of the m^7G-capped, nonployadenylated positive-sense RNA genome is controlled by the 3'UTR region in a similar way to cap-dependent and IRES-mediated initiation. Furthermore, the 3'stem loop (3'SL) region contributes significantly to 3'UTR-mediated translation after recognition of the 5'cap structure by cellular proteins (Holden and Harris, 2004). In a later study of dengue virus translation using luciferase reporter mRNA's transfected into Vero cells, it was found that the 3'UTR has translational properties similar to a poly(A) tail (Chiu et al., 2005).

Soon after successful translation of the viral RNA, the virus starts synthesizing viral RNA. Replication of the viral RNA takes place in complexes composed of viral RNA as well as non-structural proteins (Figure 1.6; Page 55). The process is catalyzed by the RNA dependent RNA polymerase (RdRp) represented by the NS5 protein in the viral genome and the replication complexes are associated with intracellular membranes. Furthermore, involvement of interactions of flaviviral specific host proteins and other viral factors with the 3' stem loop of the virus genome have been reported (Blackwell and Brinton, 1995). The 3' non coding region contains adjacent, thermodynamically stable, conserved short and long stem-and-loop structures, formed by the 3'-terminal ~100 nucleotides. It has been shown that the nucleotide sequences in different virus strains are quite variable whereas the secondary structure of the stem-loops stays conserved (Zeng et al., 1998). The 3' stem loop of the viral RNA is preceded by a cyclization sequence 3' CS. This sequence is complementary to the 5' CS leading to a direct RNA-RNA interaction and finally to the cyclization of the viral RNA. This conformation is essential for initiation of replication and plays an additional role in translational control. Studies using atomic force microscopy revealed the RNA

circularization *in vitro* and showed that the presence of the CS regions is not sufficient by itself. An element of 16 nucleotides present at the 3' and 5' UTR is an additional, important determinant for RNA-RNA association (Alvarez et al., 2005). Generally, the proposed mechanism for minus-strand RNA synthesis includes the binding of the viral polymerase to the stem loop A (SLA) at the 5' end of the genome. This results in the translocation of the RdRp to the site of initiation at the 3'end via long-range RNA-RNA interactions (Filomatori et al., 2006). In addition, the SLA serves as a promoter element enhancing replication, meanwhile competence of replication is maintained by bulges and their topological location within the long stem of the 3' stem loop (Yu and Markoff, 2005).

One of the earliest events in virus assembly (Figure 1.6; Page 55) is the formation of a nucleocapsid (NC), which consists of one copy of the genomic RNA and multiple copies of the capsid protein. The assembled nucleocapsid (NC) then buds as a complex into the lumen of the ER which results in the immature virion with a host-derived lipid bilayer including the prM and E proteins. This process is highly coordinated by the membrane-associated capsid protein and the prM-E heterodimer in the ER. Then the immature virion will be transported through the trans-Golgi network where the prM protein is cleaved by the protease furin producing mature, infectious particles that are subsequently released by exocytosis.

1.3.2.4 Extrinsic versus Intrinsic Lifecycle (Figure 1.7; Page 55)

When the mosquito feeds on an infected human being, the virus gets taken up into the mosquito and starts replicating in the mosquito gut cells. This extrinsic period of the

lifecycle lasts for 10 to 14 days. With the next blood meal, the virus gets transmitted to humans in the mosquito saliva and starts infecting its target cells. Viruses can be detected after an incubation period of 3 to 14 days and normally for the first 3 days of fever. This part is called the intrinsic lifecycle.

Dengue viruses are the only known arboviruses that have lost the need for an enzootic lifecycle for maintenance. They are fully adapted to humans and are transmitted by the highly domesticated *Ae. Aegypti* mosquito. It is thought that there is still rare maintenance of the virus through a sylvatic cycle with other *Aedes* species as the vector, but both of the two transmission cycles are separated from each other.

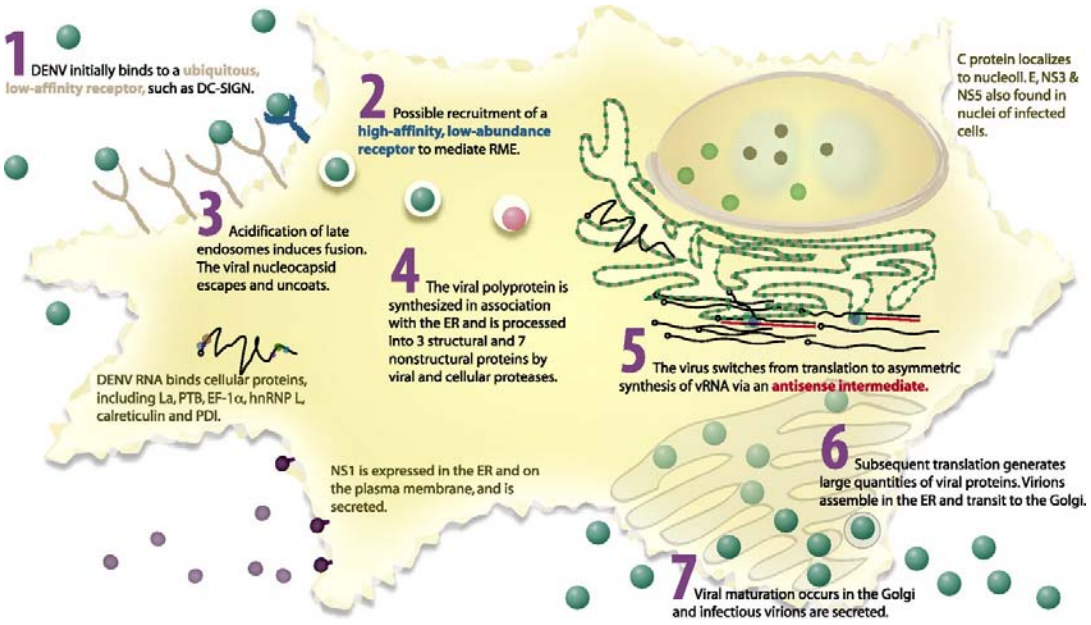


Figure 1.6: Overview of dengue virus lifecycle. (1) Dengue virions bind to target receptors and (2) enter the cell by receptor-mediated endocytosis. (3) Acidification of the late endosome triggers the class II fusion process and the viral nucleocapsid gets released into the cytoplasm. (4) The viral RNA is translated in association with the ER into the polyrotein and further processing results in 3 structural and 7 non-structural proteins. (5) After successful translation, viral replication takes place on intracellular membranes and (6) final assembly of the virions occurs in the ER. (7) The assembled viral particles undergo maturation in the Golgi and are subsequently secreted by exocytosis (Clyde et al., 2006).

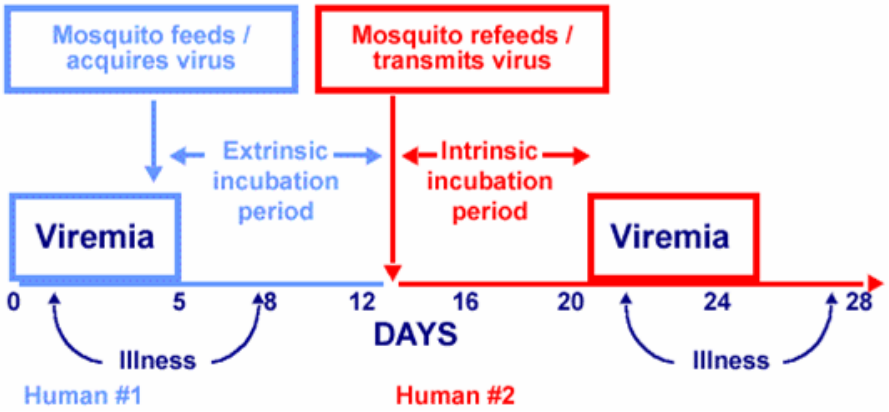


Figure 1.7: Extrinsic versus intrinsic lifecycle of dengue virus. The extrinsic incubation time of the lifecycle lasts for approximately 10 to 14 days. With the next blood meal, the mosquitoes transmit dengue to the human host where the virus starts its replication. The virus can be detected after an intrinsic incubation period of 3 to 14 days (CDC, 2002).

1.4 Immunology of Dengue Virus Infections

1.4.1 Dengue Pathogenesis and Host Immune Response

1.4.1.1 Early Events in the Host after Infection

Inoculation of dengue virus into the skin leads to DC-SIGN mediated infection of dendritic cells (DC) (Tassaneetrithep et al., 2003) and subsequent viral replication within the first 24 hours. It was shown that DCs are much more permissive than either macrophages or monocytic cells which are also considered as a main target of dengue virus infection (Wu et al., 2000). Additionally, the absolute number of lymphocytes is decreased during dengue infection which can be a result of infected or apoptosing cells (Azeredo et al., 2001). The infection of DCs is accompanied by the activation and maturation of the infected as well as neighbouring cells which finally results in the secretion of TNF-α, IFN-α and IL-10 (Green and Rothman, 2006). The antiviral response is mediated by RIG-I that recognizes double-stranded RNA and activates type I interferon signalling/secretion, especially IFN-α (Chang et al., 2006; Green and Rothman, 2006). Dengue virus is able to counteract this immediate antiviral response by downregulation of interferon signalling (I. Umareddy, unpublished data). Furthermore, infected dendritic cells migrate to the lymph nodes but are impaired in their ability to stimulate T-cells. Therefore, only neighbouring DCs undergo appropriate maturation and play a crucial role in the activation of the adaptive immune response (Green and Rothman, 2006).

1.4.1.2 Important Mediators of Innate Immunity during Infection

Natural killer (NK) cells, basophiles, neutrophiles, macrophages and complement are other important mediators of innate immunity during dengue virus infection. NK cells can clear virally infected cells either by direct cytotoxicity or via antibody-dependent cell mediated cytotoxicity (ADCC) (Navarro-Sanchez et al., 2005). They are controlled by cytokines and chemokines released by DCs and macrophages in an early stage of infection. Basophil mast cells and neutrophiles are considered to be involved in dengue virus clearance. Their exact role has yet to be elucidated especially their interaction with endothelial cells, which might explain neutropenia and thrombocytopenia in DHF patients (Navarro-Sanchez et al., 2005). IFN-γ activated macrophages produce nitric oxid (NO) which has recently been shown to inhibit dengue replication *in vitro* (Charnsilpa et al., 2005). In addition, it has been suggested that overproduction of NO may enhance vascular permeability. Furthermore, TNF-α as well as IFN-α increase complement activation and the production of C3a and C5a which may have a significant impact on vascular permeability and therefore on disease outcome (Lei et al., 2001; Navarro-Sanchez et al., 2005).

1.4.1.3 Antibody-Dependent Enhancement (ADE)

Infection of one dengue serotype results in a life-long immunity to that serotype but not to heterologous serotypes. It is suggested that the cross-protective immunity against other serotypes is only active in the first several months after infection. This temporary cross-protection has recently been revealed as a major factor to explain currently observed transmission dynamics of dengue epidemics (Wearing and Rohani,

2006). By contrast, ADE has only a minor role in such transmission dynamics but it is still considered as an important immunological mechanism involved in disease outcome of secondary infections (Halstead, 2003) .

ADE is caused by cross-reactive antibodies that are specific for the serotype of the primary infection, but are unable to neutralize the serotype of the secondary infection. Upon primary infection, neutralizing antibodies are mainly directed against the E glycoprotein (Kaufman et al., 1987) which is the main antigenic determinant but also prM (Kaufman et al., 1989) and NS1 epitopes (Wu et al., 2003) have been identified (Clyde et al., 2006). Following a secondary boost of the immune system, B-cells selected during a primary infection get activated and secrete cross-reactive antibodies. These non-neutralizing antibodies have still the ability to bind to the virus and allow rapid uptake of the virus by Fc-receptor bearing cells (Littaua et al., 1990). It is thought that ADE supports faster viral replication, resulting in greater viremia in an early stage of infection.

1.4.1.4 Differences in Secondary T-Cell Responses

The differences in T-cell responses during secondary infections may have an additional impact on viral clearance and immunopathogenesis. T-cell responses are predominately directed against NS3, NS1, NS2a and E protein. C and NS5 may also boost a T-cell response (Fink et al., 2006). Low affinity memory T-cells from a primary infection need small amounts of interferon to be stimulated and respond much faster than naïve T-cells. Their low-affinity is underlined by higher levels of TNF-α (increased inflammatory response) and lower levels of IFN-γ (impaired activation of

high-affinity T-cells) secreted by CD4 cells during heterologous infection (Green and Rothman, 2006). Furthermore, it was shown that CD8 T-cells boosted by a heterologous serotype are only able to induce lysis of the target cell but not to stimulate cytokine production (Zivny et al., 1999). This subsequently allows the infection to become established and leads to the apoptosis of high-affinity T-cells via the process of activation-induced cell death (AICD) (Fink et al., 2006; Mongkolsapaya et al., 2003). In summary, small numbers of T-cells at an early stage of infection may lead to survival of the virus and therefore to a higher viral load. Finally, the activated low-affinity T-cells have a share in cytokine release and tissue damage. It is suggested that dengue-mediated inhibition of interferon signalling may actively influence T-cell activity (Fink et al., 2006).

ADE and the inadequate immune response caused by low affinity T-cells point towards a crucial role of the 'original antigenic sin' in disease outcome (Mongkolsapaya et al., 2003). One reported study states a clear association of high viremia titers peaking two days before defervescence with progression to DHF (Vaughn et al., 2000). A contradictory study implicates involvement of previous subclinical infection and not viral load in DHF progression (Yeh et al., 2006). However, only 4% of patients experience more severe manifestations during a secondary infection, making clear that additional factors including nutritional status, sex, age, serotype, virus strain and immune status are important for disease outcome.

1.4.1.5 Important Cytokines during Infection

Not only interferon production is induced during dengue virus infection but a broad range of cytokines, mediators and soluble receptors (Fink et al., 2006; Green and Rothman, 2006). By way of example, it was shown that, in a first instance, peripheral blood monocytic cells (PBMC) release IFN-γ, TNF-α, IL-2 and IL-6 followed later on by the secretion of IL-4, IL-10 and IL-5 (Pacsa et al., 2000). This points towards a switch of T_H1 to T_H2 response and leads to an imbalanced T-cell response and may enhance cytophatic effects. In general, a T_H1 response mediates activation of infected monocytes enabling them to destroy intracellular pathogens whereas T_H2 responses stimulate secretion of antibodies and are not sufficient for viral clearance.

1.4.1.6 DHF/DSS: An Immune-Mediated Machanism

A hallmark of DHF/DSS is vascular leakage and its appearance after significant reduction of viremia hints towards an immune-mediated mechanism (Green and Rothman, 2006). Vascular permeability is induced by various cytokines, e.g. TNF-α that may lead to permeability via cyclooxygenase (COX) and the release of prostanoids underlining the risk of using aspirin as a treatment (Fink et al., 2006). Furthermore, plasma leakage is accompanied by a marked thrombocytopenia. The mechanisms involved in platelet reduction are not fully understood but it is suggested that either anti-NS1 IgM antibodies can cross-react with platelets leading to their clearance in the presence of cytokines and complement (Lin et al., 2006) or that haematopoiesis is suppressed in the course of an infection (Murgue et al., 1998).

The discussed immunological response is complex and more studies are required to elucidate epidemiological, virological as well as immunological factors involved in dengue pathogenesis. This, in turn, will have a major contribution towards the ultimate goal of finding either a drug and/or a vaccine against dengue and towards a better understanding of the biology of dengue pathogenesis.

1.4.1.7 Viral Determinants and Disease Outcome

Besides host factors involved in disease pathogenesis, several studies suggest an active role of viral factors determining disease severity (Messer et al., 2002; Pandey and Igarashi, 2000; Rico-Hesse et al., 1997). Both different serotypes and mutational variations in strains are able to interfere with disease outcome, but clear correlations could not be shown (Balmaseda et al., 2006; Clyde et al., 2006; Vaughn et al., 2000). Differences in replication fitness based on sequence variation between serotypes and strains are thought to be a crucial factor leading to more severe disease (Holmes and Twiddy, 2003) which is underlined by the clear association of high viremia titers peaking two days before defervescence with progression to DHF (Vaughn et al., 2000).

A cross-sectional study recently compared DENV-1 to DENV-2 outbreak periods in Nicaragua and pointed out that DENV-1 outbreaks were associated with increased vascular permeability whereas DENV-2 infections resulted in more shock and internal haemorrhage (Balmaseda et al., 2006). Earlier studies indicated associations of certain genotypes of DENV-2 and DENV-3 to DHF versus DF. Firstly, it was shown that the Asian genotype of DENV-2 correlates with high virulence. This is underlined by

mutagenic studies of N390 of the E-protein as well as exchange of 5' and 3' UTRs between Asian and American genotypes and shows an early replicative advantage of the Asian genotype in the host (Cologna et al., 2005; Leitmeyer et al., 1999). Secondly, Messer and collaborators identified a variant of DENV-3 genotype III as a possible factor in the emergence of DHF in Sri Lanka (Messer et al., 2003; Messer et al., 2002)

It is clear that mutations in the 3'UTR have a major impact on viral fitness (Holmes and Twiddy, 2003), and existing directed repeats (DR) could affect virus transmission (Gritsun and Gould, 2006), but their impact on disease outcome is contradictory. Earlier studies mainly performed on DENV-2 suggested that specific mutations either resulting in altered secondary structures of 3'UTRs or in amino acid changes might be directly correlated to disease severity (Clyde et al., 2006; Leitmeyer et al., 1999; Pandey and Igarashi, 2000). However, recent evidence hints towards the activity of multiple gene loci involved in clinical outcome, and different 3'UTRs show only weak associations with viral virulence (Zhou et al., 2006). Therefore, it is unlikely that the 3'UTR by itself represents the ultimate determinant of disease severity. This fact is further highlighted by a recent study investigating the 2004 Indonesian dengue outbreak. The analysis revealed high levels of genetic similarities to severe strains from Sumatra when looking at the whole genome rather than single mutations (M. Schreiber, personal communication).

Therefore, it is evident that only complex host-pathogen interactions dependent on host and viral backgrounds are able to explain disease outcome. This synergy is also emphasized by results showing that some strains have the ability to inhibit interferon signalling (Chang et al., 2006). Recent observations suggest differences in ER stress

mediated response (Yu et al., 2006) and that, in general, flaviviral infections may lead to a lipid-raft mediated interference of the PI3K/AKT pathway, thus inhibiting apoptosis in an early stage of infection (Lee et al., 2005).

1.5 Aims of the Studies

The overall purpose of this study was to investigate clinical parameters in dengue infection that would be useful in distinguishing dengue fever from other febrile illnesses and in predicting the extent of disease severity in an early stage of infection. Furthermore, we aimed at obtaining the full genome sequences of dengue virus isolates from the 2005 Singaporean outbreak to analyze them with regard to their phylogeny and to disease severity.

More specifically we pursued the following four goals:

- Finding of an appropriate model that distinguishes dengue fever from other febrile illnesses in an early stage of disease.

- Identifying early clinical and immunological markers that would predict disease severity in dengue infections.

- Sequencing of dengue viruses isolated during the 2005 Singaporean dengue outbreak and investigate them with regard to their phylogeny.

- Characterizing of identified clades with regard to important markers established by the predictive model for dengue severity.

2 Materials & Methods

2.1 Analysis of Clinical Data

2.1.1 Collection and Preprocessing of Clinical Data

Clinical Data was collected during the 2005 Singaporean outbreak as part of the EDEN study. Briefly, 455 patients were enrolled into the study and a first blood and saliva sample was collected within 72 hours of illness (1st visit) (Figure 2.1). From the 133 dengue positive patients, another sample on day four to seven of illness was collected on the 2nd visit and finally, a third sample was collected three to four weeks after onset of fever on the 3rd visit. The samples were analyzed using standard laboratory methods. For a more detailed description of the collection procedure please refer to the EDEN report (Low et al., 2006).

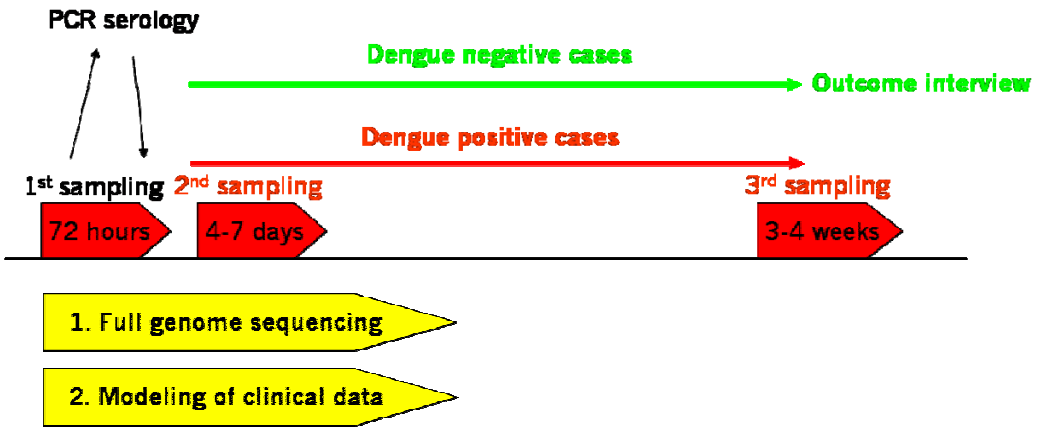


Figure 2.1: Flowchart showing the sequence of events during the EDEN study (Low et al., 2006).

The collected parameters included personal data, symptoms, clinical data as well as many more parameters of all the 455 patients recruited in the EDEN study and were summarized in one datasheet (CRF3) which was inconvenient for our analysis.

Furthermore, for hospitalized dengue patients (72 patients submitted to general hospitals and not private hospitals), more data was collected every day in hospital which was summarized in a separate datasheet (CRF4). The third raw datasheet represented the immunological information of 291 patients (95 dengue patients and 196 cases with other febrile illnesses). Specific information used such as hospitalization status, viral load and additional immunological information was distributed over three other datasheets. For the purpose of analysis, it was necessary to sort, join and clean up the datasheets. Preprocessing of tables and replacement of missing values was performed using the program Inforsense (InforSense Ltd., London, UK). The typical workflow created for data preprocessing is indicated below (Figure 2.2).

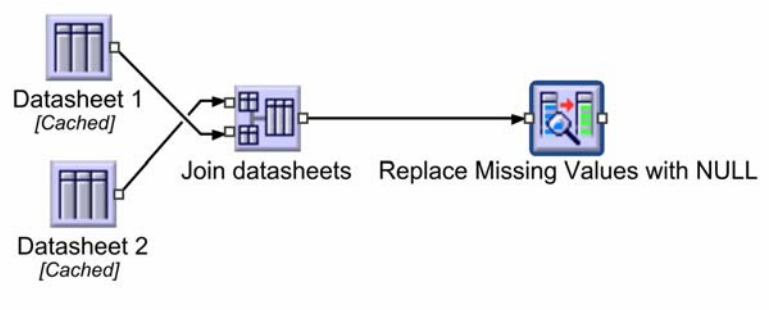


Figure 2.2: Example of Inforsense (InforSense Ltd., London, UK), workflow used for data pre-processing.

2.1.1.1 Clinical Data used for Analyses

For our analyses, we used 25 clinical parameters from the 1st visit, 24 clinical parameters from the 2nd visit along with 19 clinical parameters from the 3rd visit (Table 2.3). In a first step, two main tables were generated including the clinical parameters

along with the identification number, serotype of the causative dengue virus, hospitalization status of the patient as well as our defined severity classification (see below): One table (EDEN_TOTAL) included the 1st visit data of 453 patients (2 patients were excluded due to no follow up or were unlikely to have dengue fever) comprised of 133 dengue positive patients and 320 patients suffering from other febrile illnesses. The second table (DENGUE_TOTAL) included the 1st, 2nd and 3rd visit information of only the 133 dengue positive cases. In both tables, missing values were replaced by null values (Table 2.1).

2.1.1.2 Clinical Data used for Severity Modeling

From the created main table including the 133 dengue positive patients (EDEN_TOTAL), we generated a specific table for severity modeling. For this purpose, we categorized the patients into severe (low) and mild (high) cases based on a low platelet count using a threshold of <=50,000platelets/mm^3 on either day five, six or seven of illness. The categorization was based on the hospitalized cases whereas the non-hospitalized patients were considered as mild cases and were directly classified as high. Seven patients of the hospitalized cases were excluded from the analysis because no samples were collected during hospital stay and one patient from the 72 patients with hospital information was excluded from the analysis due to no data on fever day five, six and seven. This resulted in a dataset (SEVERITY_TOTAL) including 125 patients and was used for severity modeling. Missing values in the table were replaced by null values (Table 2.1).

2.1.1.3 Immunological and Clinical Data used for Analyses

Three tables were generated including immunological (Table 2.2) as well as clinical data. From the 453 patients, immunological data was available for 291 enrolled patients (95 dengue positive cases and 196 patients suffering from other febrile illnesses). 162 patients were excluded from immunological modeling because no immunological data was measured.

The first table (EDEN_IMMUN_TOTAL) joined the immunological information along with the clinical data into one table including all the 291 patients. A second table (DENGUE_IMMUN_TOTAL) included only the 95 dengue positive patients and the third table (SEVERITY_IMMUN_TOTAL) created from the severity table (125 cases) represented only 89 dengue patients (36 patients were excluded because there was no immunological information available) (Table 2.1). In all three tables, missing values were replaced with null values.

It is important to note, that the read out of the standard test for the measurement of immunological parameters was the optical density. The different optical densities were then converted by the software to pg/ml cytokine, based on a series of optical densities from known human cytokine standards which provided a standard curve. A total of 8 different concentrations was used in each run. Samples that were beyond the detection minimum resulted in a test value indicated as <low>. These <low> values were replaced by using the minimum value for each run and measured cytokine, thus not all <low> had the same values. The datasheet with all the <low> values replaced was used for further analysis (Eng Eong Ooi, personal communication).

2.1.1.4 Summary of Clinical and Immunological Data used for Analyses

Table 2.1: Overview of generated tables which were eventually used for analyses. 1=*1st visit*; 2=*2nd visit;* 3=*3rd visit.*

Table Name	Cases	Information included	Description
EDEN_TOTAL	453	Clinical Data [1]	
DENGUE_TOTAL	133	Clinical Data [1, 2, 3]	
SEVERITY_TOTAL	125	Clinical Data [1, 2, 3]	PLT_CLASS (low/high)
EDEN_IMMUN_TOTAL	291	Clinical & Cytokine Data [1]	
DENGUE_IMMUN_TOTAL	95	Clinical & Cytokine Data [1, 2, 3]	
SEVERITY_IMMUN_TOTAL	89	Clinical & Cytokine Data [1, 2, 3]	PLT_CLASS (low/high)

Table 2.2: Overview and explanation of the immunological parameters used for analyses. 1=*1st visit*; 2=*2nd visit*; 3=*3rd visit.*

Abbreviation	Unit	Explanation
IP_10 [1,2,3]	[pg/ml]	Interferon γ inducible protein 10 (IP-10)
I_TAC [1,2,3]	[pg/ml]	Interferon inducible T cell α chemoattarctant (I-TAC)
IFN_ALPHA [1,2,3]	[pg/ml]	Interferon α (IFN-α)
GM_CSF [1,2,3]	[pg/ml]	Granulocyte macrophage colony-stimulating factor (GM-CSF)
IFN_GAMMA [1,2,3]	[pg/ml]	Interferon-γ (IFN- γ)
IL_1 [1,2,3]	[pg/ml]	Interleukin-1 (IL-1)
IL_10 [1,2,3]	[pg/ml]	Interleukin-10 (IL-10)
IL_12 [1,2,3]	[pg/ml]	Interleukin-12 (IL-12)
IL_2 [1,2,3]	[pg/ml]	Interleukin-2 (IL-2)
IL_4 [1,2,3]	[pg/ml]	Interleukin-4 (IL-4)
IL_6 [1,2,3]	[pg/ml]	Interleukin-6 (IL-6)
IL_8 [1,2,3]	[pg/ml]	Interleukin-8 (IL-8)
TNF [1,2,3]	[pg/ml]	Tumor necrosis factor α (TNF- α)

Table 2.3: Overview and explanation of the clinical parameters used for analyses. 1=*1st visit*; 2=*2nd visit*; 3=*3rd visit* (NIH, 2007).

Abbreviation	Unit	Explanation
ID	None	Sample identification number
TEMP [1,2]	°C	Body temperature
PULSERATE [1,2]	Beats per minute	Pulse rate
SYSTOLICBP [1,2]	mmHg	Systolic blood pressure
DIASTOLICBP [1,2]	mmHg	Diastolic blood pressure
WBC [1,2,3]	Cells/microliter [$x10^3$]	White blood cell count. It is almost always part of the CBC (complete blood count); Normal count is between 4,500-10,000 white blood cells/µl.
RBC [1,2,3]	Cells/microliter [$x10^6$]	Red Blood Cell count (varies with altitude): Male: 4.7 to 6.1 million cells/µl; Female: 4.2 to 5.4 million cells/µl
HGB [1,2,3]	Grams/deciliter	Hemoglobin (varies with altitude): Male: 13.8 to 17.2 g/dL; Female: 12.1 to 15.1 g/dL
HCT [1,2,3]	Percentage [%]	The hematocrit is the percent of whole blood that is composed of red blood cells. The hematocrit is a measure of both the number of red blood cells and the size of red blood cells; (varies with altitude): Male: 40.7 to 50.3 %; Female: 36.1 to 44.3 %
MCV [1,2,3]	Femtoliter	MCV is the mean corpuscular volume or the average size of the red blood cells: Normal MCV: 80 to 100 femtoliter.
MCH [1,2,3]	Picograms/cell	MCH is the mean corpuscular hemoglobin or the hemoglobin amount per red blood cell (MCH = HGB/RBC). Normal MCH: MCH: 27 to 31 picograms/cell
MCHC [1,2,3]	Grams/deciliter	MCHC is the mean corpuscular hemoglobin concentration or the hemoglobin concentration (hemoglobin amount relative to the size of the cell) per red blood cell (HGB/HCT). Normal MCHC: 32 to 36 grams/deciliter
PLT [1,2,3]	Number/microliter [$x10^3$]	PLT is the Platelet count or thrombocyte count. Normal PLT: 150,000 to 400,000/µl
LYMPH_PCT [1,2,3]	Percentage [%]	The percentage of the WBC belonging to lymphocytes. Normal LYMPH_PCT: 20% to 40% of the WBC
MXD_PCT [1,2,3]	Percentage [%]	The percentage of the WBC belonging to monocytes, eosinophiles & basophiles. Monocytes: 2%-8%; Eosinophils: 1%-4%; Basophils: 0.5%-1%
NEUT_PCT [1,2,3]	Percentage [%]	The percentage of the WBC belonging to neutrophiles. Normal NEUT_PCT: 40% to 60% of the WBC
LYMPH_NO [1,2,3]	Cells/microliter [$x10^3$]	The absolute number of lymphocytes calculated by multiplying LYMPH_PCT with WBC. (LYMPH_PCT x WBC)

Table 2.3 (continued): Overview and explanation of the clinical parameters used. 1=*1st visit*; 2=*2nd visit*; 3=*3rd visit.*

Abbreviation	Unit	Explanation
MXD_NO [1,2,3]	Cells/microliter [x10^3]	The absolute number of monocytes, eosinophiles & basophiles calculated by multiplying MXD_PCT with WBC. (MXD_PCT x WBC)
NEUT_NO [1,2,3]	Cells/microliter [x10^3]	The absolute number of neutrophiles calculated by multiplying NEUT_PCT with WBC. (NEUT_PCT x WBC)
RDW_CV [1,2,3]	Percentage [%]	Red blood cell distribution width is a measure of the variation of red blood cell volume that is reported as part of a standard complete blood count. Usually red blood cells are a standard size. It's a coeffiecient of variation ((Standard deviation of red cell volume/mean cell volume) × 100); Normal range: 11.5-14.5%
RDW_SD [1,2,3]	Percentage [%]	Standard deviation of RDW_CV.
PDW [1,2,3]	Femtoliter	Platelet distribution width.
MPV [1,2,3]	Femtoliter	Mean platelet volume is a machine calculated measurement of the average size of the platelets.
P_LCR_PCT [1,2,3]	Percentage [%]	Platelet Large Cell Ratio
DV_PCR [1]	None	Detection of dengue Virus by PCR. (positive/negative)
DV_IG_G [1]	None	Detection of dengue virus antibodies: marker for secondary infection. (positive=secondary/negative=primary)
CT_COLLECTION [1]	Number of cycles	The Ct-value is the number of cycles used to reach a specific threshold in the RT-PCR. High Ct-value means a low amount of template and a low Ct-value represents a large amount of template.
HOSPITALIZATION	None	Hospitalization status of dengue positive patients (yes/no).
PLT_CLASS	None	Severity classification of patients defined with regard to the platelet count (low/high); (For more details see Results; Page 123)

2.1.2 Statistical Analyses used for the Description of Clinical and Immunological Data

2.1.2.1 Univariate Analyses

Shapiro-Wilk normality test was used to check for non-normally distributed parameters whereby a p value < 0.05 indicated that the parameter was unlikely to originate from a normal distribution. Non-normally distributed parameters were log-transformed and rechecked for normality. Cases with missing values were excluded from the analysis and thus, the number of cases used for calculations varied between different covariates. All calculations were performed using Systat® for Windows® (SYSTAT Software Inc., 2004).

If the log-transformation still resulted in non-normal distribution, non-parametric *Kruskal-Willis* test was used for continuous variables whereas *Student's* t test was exploited for normally distributed continuous variables. For dichotomous variables, *Chi-square* test was used in case of expected frequencies that were higher than 5, whereas *Fisher's* exact test was performed when the expected table values were smaller than 5. A two-tailed p value < 0.05 was considered as statistically significant. Cases with missing values were excluded from the analysis and thus, the number of cases used for calculations varied between different covariates. All statistical computations were performed using Systat® for Windows® (SYSTAT Software Inc., 2004).

Overall significant differences in clinical manifestations between identified viral clades were assessed by one-way ANOVA (analysis of variance). The *Tukey* post hoc test was used to calculate the pairwise probabilities between the clades and to finally identify the pairs that showed the highest level of significance. Cases with missing values were excluded from the analysis and thus, the number of cases used for calculations varied between different covariates. All statistical computations were performed using Systat® for Windows® (SYSTAT Software Inc., 2004).

2.1.2.2 Multivariate Analyses

Significant differences found by univariate analyses were further entered into multivariate analysis using a logistic regression model to identify genuine risk factors. The complete algorithm was chosen because we aimed at assessing the independent risk for all the entered variables. By contrast, the using of stepwise algorithms results in a model that includes only the most significant covariates. A two-tailed p value < 0.05 was considered as statistical significant. Cases with missing values were excluded from the analysis. All statistical computations were performed using Systat® for Windows® (SYSTAT Software Inc., 2004).

2.1.3 Decision Tree Analyses for Disease Modeling

The thesis exploited decision trees as predictive models to either distinguish dengue from other febrile illnesses or to find an appropriate classification for dengue severity. Decision trees predicting categorical variables are normally referred as classification

trees whereas trees predicting continuous variables are called regression trees (TwoCrowsCorporation, 2005).

Decision trees represent a powerful supervised learning tool commonly used in datamining and are used in medical decision analysis (Peters et al., 2006), in gene expression profiling (Schramm et al., 2005) as well as in molecular biology (Han et al., 2006; Rhee et al., 2006). The top-down induction of decision trees is a simple and potent method of deducing classification rules from a set of labeled data (Quinlan, 1993). At each node a decision rule is implemented that splits the data into two or more partitions. New nodes are created to handle each of the partitions and a node is considered terminal or a leaf node based on a stopping criterium. The advantages of using decision trees in knowledge discovery mainly lie in the simplicity as well as the ability to handle categorical and numerical values at the same time (Kothari and Dong, 2000)

There are several algorithms that are used for tree induction (Holsheimer and Siebes, 1993; Kothari and Dong, 2000). In our case, calculations of classification trees were performed by using the efficient machine learning algorithm C4.5 which is an extension of the ID3 algorithm (Quinlan, 1993). ID3 stands for induction of decision trees and was developed in the first half of the eighties by Quinlan (Holsheimer and Siebes, 1993). It uses an information gain based node splitting where the attribute that minimizes the information needed in the resulting subtrees to classify the elements is selected. In other words, the algorithm takes the parameter as a splitting criterium which has the highest information gain and which reduces the existing data entropy the

most. The algorithm uses a greedy search meaning it chooses the best attribute at the moment and never looks back to reconsider earlier choices.

Another interesting aspect of the C4.5 system is the integration of an extension that is able to handle missing values. Briefly, when coming across a missing attribute which is used for further splitting, the algorithm explores all the possible branches at this point and estimates the probability that this is the correct choice. Finally, the result of the classification is the class with the highest probability (Holsheimer and Siebes, 1993; Quinlan, 1993). The ability to handle missing values represented another advantage of the used algorithm because during pre-processing of clinical and immunological data we encountered a substantial number of missing values. The C4.5 system permitted us to make use of all the cases and thus, we did not have to exclude cases with missing values which is a common problem in multivariate analyses such us logistic regression and discriminant analysis (Acuna and Rodriguez, 2004). Decision tree modeling was performed in Inforsense (InforSense Ltd., London, UK) and the typical workflow used is shown in Figure 2.3.

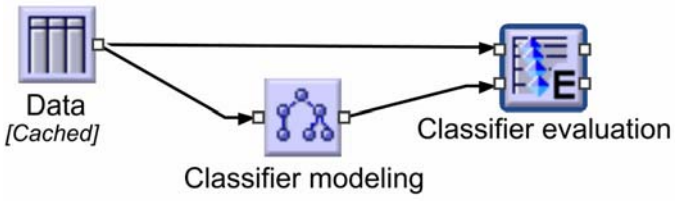


Figure 2.3: Example of Inforsense (InforSense Ltd., London, UK) workflow used for the calculation of decision tree models. Briefly, the data was entered into the model node for decision tree induction. Finally, the calculated tree was validated by k-fold cross validation.

2.1.3.1 Classifier Modeling

Pre-processed data was entered into the model node for decision tree induction. We aimed at finding a decision tree that is simple to understand and that includes only the most important parameters for the addressed prediction task. For this purpose, we had to select both the variables which would serve as predictors and the target class. Furthermore, important tree parameters such as the splitting stopping criterium (minimal cases) and the parameter used for tree simplification (pruning confidence) had to be defined.

In decision tree modeling, pruning of trees is used to remove branches that are not relevant for the generalization of the model and that are more likely to come from data overfitting. In Inforsense (InforSense Ltd., London, UK), the pruning confidence can be adjusted. For all the decision tree calculations the standard pruning confidence of 25% was used meaning that branches were removed when the algorithm was 25% or less confident that the specific branch is not representative for generalization. The higher the pruning confidence is set, the lesser pruning is performed.

The parameter 'minimal cases' represents a stopping criterium for the further partition of data at a specific decision node. Tree growing at a specific decision node is stopped when at least one class includes equal or less cases as the tree parameter 'minimal cases'. This parameter is used to overcome noisy data and a higher value indicates a smaller tree resulting in the inclusion of probably the more important parameters and thus, the risk of data overfitting can be reduced. Choosing an appropriate value for 'missing cases' was done using k-fold cross validation (see below). Briefly, various

decision trees with different 'minimal cases' were calculated and the value resulting in the tree with the best performance was chosen.

2.1.3.2 Classifier Evaluation

The calculated models were validated by using the approach of k-fold cross validation. The fold value was set to the standard k=10 and it was ensured that the order of data instances presented to the cross validation was class and data independent as well as ideally randomized. This was done by picking the standard options stratify and sort in the cross validation node.

K-fold cross validation is considered to be a powerful methodology to overcome data overfitting and a study recommended using stratified ten-fold cross validation for model selection (Kohavi, 1995). Briefly, the original sample is divided into k subsamples. Each subsample is put aside as evaluation data for testing a model, and the remaining k-1 subsamples are used for training the model. The cross-validation process is repeated k times (folds) and each of the k subsamples is used once as the validation data. The k results obtained from the k folds can then be averaged to produce a single estimation of model performance (Moore, 2007)

In Inforsense (InforSense Ltd., London, UK), the output of k-fold cross validation was a report including an overall evaluation of the ten folds as well as the exact details of every fold (Table 2.4). Moreover an averaged receiver operating characteristic curve (ROC curve) was constructed which is a commonly used tool to graphically evaluate classifier performance (Figure 2.4). Briefly, a decision tree is a probabilistic classifier

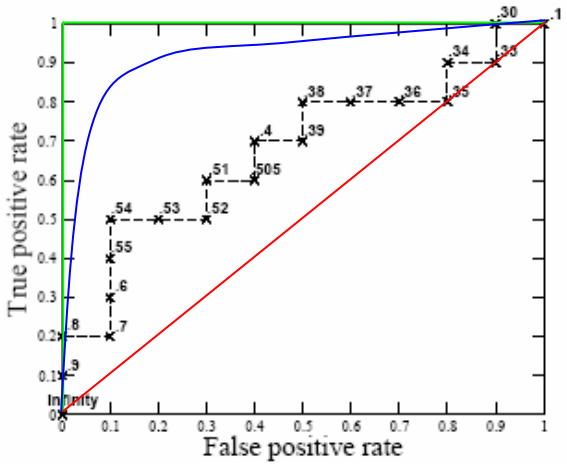
and for each classification made, a confidence value or a probability associated to the decision is obtained. By choosing a specific threshold, the cases are classified as positive in case the confidence value is greater than the threshold and negative if the confidence value is below the threshold. By varying this threshold, it is possible to compute the true positive rate (sensitivity) and the false negative rate for each of these thresholds which finally results in a ROC curve. ROC curves can then be used to judge the discrimination power of a calculated classification model. The area under the curve (AUC) represents a good indicator for the overall performance of a specific classifier (Fawcett, 2003; Hanley and McNeil, 1982). It is worthwhile mentioning, that not only AUCs can be used to choose the best classifier but depending on the situation, other calculated parameters such as profit of the model or the misclassification rate are better to use. However, we chose our classification trees based on the overall performance as well as on the sensitivity. The classifier which had the highest sensitivity along with a high AUC was selected. Furthermore, it was also important to have a classifier with a good discriminative power which was indicated by a low overall error rate and by a higher profit. By way of example, a model was considered inappropriate in case of having a high AUC but, in turn, a low sensitivity. The AUCs of the chosen models were further investigated by calculating their confidence intervals which was based on the method described by Hanley and McNeil (Hanley and McNeil, 1982; McNeil and Hanley, 1984). The overall performance was considered not to be significantly different from the worst classifier when the confidence interval included 0.5. Calculations were performed in Excel® for Windows® (Microsoft, Redmond, WA).

Table 2.4: Explanation of the overall evaluation report obtained after k-fold cross-validation.

Attribute	Explanation
Total misclassifications	Number of misclassified cases after cross validation
Overall error rate	Overall error rate of the model after cross validation
SE of error rate	Estimated standard error of the overall error rate
Average profit	Profit of the model averaged after cross validation
SE of profit	Estimated standard error of the profit
AUC negative	Area under the curve of averaged ROC curve for the classification of negative cases
AUC positive	Area under the curve of averaged ROC curve for the classification of positive cases

Confusion Matrix

		Predicted Class: negative	Predicted Class: positive
Actual Class	negative	specificity	false-positive rate
Actual Class	positive	false-negative rate	sensitivity



Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1

Figure 2.4: A simplified receiver operating characteristic curve obtained by threshold averaging after k-fold cross validation. The table at right shows twenty data and the score assigned to each by a scoring classifier. The black graph at left shows the corresponding ROC curve with each point labeled by the threshold that produces it. The green line indicates the ROC curve of an optimal classifier (100%), the red line represents the ROC curve for the worst classifier (50%) and the blue line shows a typical ROC curve (Fawcett, 2003).

2.1.3.3 Epidemiological Analyses of Parameters included into the Models

Finally, it was important to evaluate the chosen classifier with regard to their epidemiological values. This would give us a more realistic picture of which parameters included in the classification trees might represent genuine risks. For this purpose the odds ratio (OR) along with their confidence intervals and the relative risk (RR) for the chosen parameters with the defined threshold were calculated on the whole dataset as well as on the specific subgroups defined by the obtained decision rules. Statistical significance was assessed by using either *Chi* square test (for expected frequencies > 5) or *Fisher's* exact test (for expected frequencies < 5). Cases with missing values were excluded from the analysis. All statistical computations were performed using Systat® for Windows® (SYSTAT Software Inc., 2004).

The odds ratio (OR) is defined as the ratio of the odds of an event occurring in the exposed group to the odds of it occurring in the non-exposed group. An OR of 1 indicates that the event is evenly likely to occur in both groups. An odds ratio OR > 1 indicates that the event is more likely to occur in the exposed group and an OR < 1 indicates that the event is more likely to occur in the non-exposed group (Rothman and Greenland, 1998).

The relative risk (RR) is the risk of an event relative to exposure. RR is a ratio of the probability of the event occurring in the exposed group versus the probability of the event in the non-exposed group. A relative risk of 1 means that there is no risk difference between the two groups. A RR of < 1 indicates that the event is less likely to occur in the exposed group than in the non-exposed. On the other hand, a RR of > 1

indicates the event is more likely to occur in the exposed group than in the non-exposed group (Rothman and Greenland, 1998).

2.1.3.4 Summary of generated Models included in the Thesis

Table 2.5: Overview of calculated classification tree models. Model names were used throughout the results and discussion sections. Name[a]: DENPRE=*dengue prediction*; HOSP=*hospitalization prediction*; SEVERE=*severity prediction*; SEVHOSP=*severity prediction of hospitalized patients*; EXCYT=*excluding cytokine data*; INCYTA=*including cytokine data*; INCYT=*including cytokine data but excluding IFN-α*; CYTOA=*only including cytokine data*; CYTO=*only including cytokine data but excluding IFN-α*. Data[b]: *See Table 2.1*; Pre[c]=predictor: clin=*clinical data*; cyto=*cytokine data*. Target[d]: denpcr=*dengue PCR positive/negative*; hosp=*hospitalization yes/no*; pltclass= *platelet class low/high*. Min[e]: *tree parameter 'minimum cases'*.1=*1st visit data*; 2=*2nd visit data.*

Name[a] & Number of Cases	Data[b]	Pre[c]	Target[d]	Pruning	Min[e]
DENPRE_TOTAL_453	EDEN_TOTAL	clin[1]	denpcr	25%	19
DENPRE_EXCYT_291	EDEN_IMMUN_TOTAL	clin[1]	denpcr	25%	9
DENPRE_INCYTA_291	EDEN_IMMUN_TOTAL	clin, cyto[1]	denpcr	25%	8
DENPRE_INCYT_291	EDEN_IMMUN_TOTAL	clin, cyto[1]	denpcr	25%	8
DENPRE_CYTOA_291	EDEN_IMMUN_TOTAL	cyto[1]	denpcr	25%	15
DENPRE_CYTO_291	EDEN_IMMUN_TOTAL	cyto[1]	denpcr	25%	14
HOSP_TOTAL_133	DENGUE_TOTAL	clin[1,2]	hosp	25%	12
HOSP_EXCYT_95	DENGUE_IMMUN_TOTAL	clin[1,2]	hosp	25%	5
HOSP_CYTO_95	DENGUE_IMMUN_TOTAL	clin, cyto[1,2]	hosp	25%	10
SEVERE_TOTAL_125	SEVERITY_TOTAL	clin[1]	pltclass	25%	16
SEVERE_EXCYT_89	SEVERITY_IMMUN_TOTAL	clin[1]	pltclass	25%	10
SEVERE_INCYTA_89	SEVERITY_IMMUN_TOTAL	clin, cyto[1]	pltclass	25%	10
SEVERE_CYTOA_89	SEVERITY_IMMUN_TOTAL	cyto[1]	pltclass	25%	23
SEVHOSP_TOTAL_71	SEVERITY_TOTAL	clin[1]	pltclass	25%	4
SEVHOSP_EXCYT_52	SEVERITY_IMMUN_TOTAL	clin[1]	pltclass	25%	4
SEVHOSP_INCYTA_52	SEVERITY_IMMUN_TOTAL	clin, cyto[1]	pltclass	25%	10
SEVHOSP_CYTOA_52	SEVERITY_IMMUN_TOTAL	cyto[1]	pltclass	25%	4

2.2 Full Length Genome Sequencing of Dengue Virus Isolates

2.2.1 Preparative RT-PCR Step

The process of sequencing was divided into several steps undertaken in collaboration with the respective research institutes (Figure 2.5) First, dengue virus isolated from RT-PCR positive samples was quantified by immuno fluorescence (IF) and propagated at the Environmental Health Institute (EHI). Secondly, virus was further passaged up to a titer of 10^5pfu/ml at the Novartis Institute for Tropical Diseases (NITD) followed by RNA extraction and the preparative RT-PCR step. Finally, amplified cDNA was sent to the Genome Institute of Singapore (GIS) for sequencing and concluding sequence analysis was performed at NITD as well as GIS.

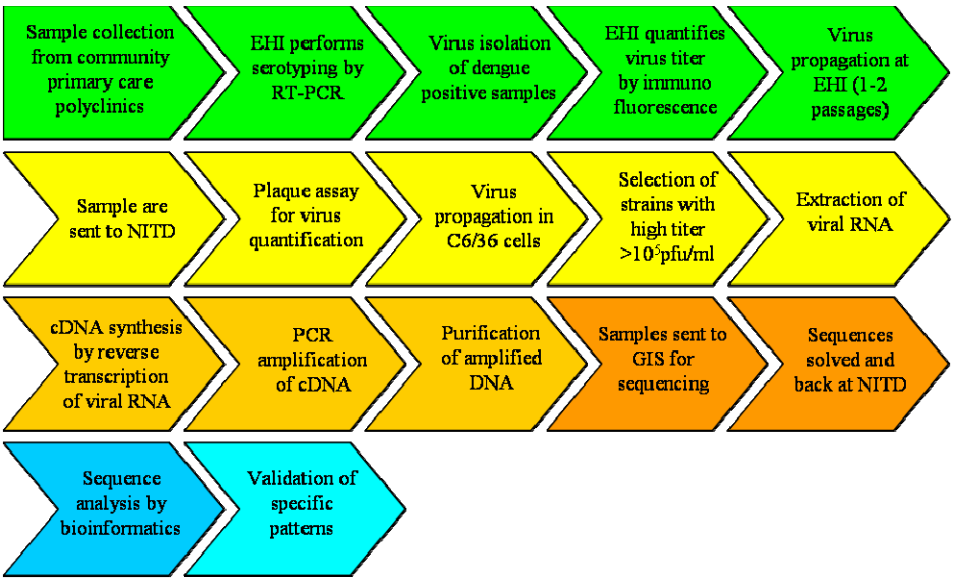


Figure 2.5: Flowchart of the EDEN sequencing project. Green represents the virus work at the Environmental Health Institute (EHI). Yellow indicates the virus work performed at the Novartis Institute for Tropical Diseases (NITD) followed by the RT-PCR step (dark yellow). Orange stands for the sequencing process at the Genome institute of Singapore which is followed by the sequence analysis at NITD and GIS (blue). The turquoise part is the *in vitro* validation of identified patterns.

2.2.1.1 Propagation of Virus

Virus isolates received from EHI were propagated in C6/36 mosquito gut cells. Firstly, virus was inoculated onto C6/36 cells (MOI=0.01) that were split and seeded 24h before in RPMI1640 culture medium (10% fetal bovine serum and 1% penicillin-streptomycin mixture). The cells used were estimated to be $5x10^7$ in number. The virus was incubated with the cells in RPMI1640 medium (5% fetal bovine serum) for an hour at 28°C. The viral infection medium was removed and fresh medium RPMI1640 culture medium (5% fetal bovine serum) was added. After an incubation of five days at 28°C, the supernatant was harvested and 0.2 ml was used for the quantification of the virus titer by plaque assay. The rest of the supernatant was stored at -80°C. Isolates were propagated till a virus titer of $>10^5$pfu/ml and used for viral RNA extraction.

2.2.1.2 Quantification of Virus by Plaque Assay

Two hundred microliters of viral supernatants were inoculated onto $2x10^5$ BHK_{21} cells seeded to each well of a 24 well-plate (RPMI1640 medium with 10% fetal bovine serum and 1% penicillin-streptomycin mixture). Virus and cells were incubated for an hour at 37°C. Afterwards, the inoculums were removed and wells were overlaid with 0.8% methyl-cellulose in RPMI1640 medium containing 2% fetal bovine serum. After an incubation of four to six days at 37°C, 10% formaldehyde was added to each well. The contents of the wells were removed by gentle agitation of the wells and followed by addition of 1ml of 10% formaldehyde for 20 minutes at room temperature. Then, the well contents were removed and stained with 0.2 ml crystal violet for 20 minutes at

room temperature. Subsequently, the wells were washed with running water and dried at 70°C.

2.2.1.3 Extraction of viral RNA and cDNA Synthesis

RNA extraction was performed using the QIAamp Viral RNA Mini Kit (Qiagen Inc., CA, US). Isolated viral RNA was reverse transcribed into five overlapping fragments using serotype specific anti sense primers and Superscript III (Invitrogen Co., CA, US) as the reverse transcriptase. The reactions were performed on the GeneAMP® PCR System 9700 (Applied Biosystems, CA, US). Briefly, 6µl of isolated viral RNA was mixed with 1µl of 10mM dNTP and 1µl of 10µM serotype specific primers. The mixture was done depending on the fragment as well as the serotype and was filled up with RNAse free water to a volume of 13µl. The reaction mixture was incubated at 65 °C for five minutes to denature the RNA secondary structure followed by the immediate shock reaction on ice for at least 1 min. After that, a second mixture of 4µl 5x first strand buffer (Invitrogen), 1µl DTT (Invitrogen), 1µl RNAsin (Promega Co., VA, USA) and 1 µl Superscript III (Invitrogen) was added. The final reaction volume of 20µl was first heated up to 25 °C for five minutes and afterwards to 42°C for an hour allowing cDNA synthesis. Finally, the reaction was inactivated at 70 °C for 15 minutes and kept at 4°C (Table 2.6; Table 2.7).

Table 2.6: Overview of RT-Conditions for cDNA synthesis of DENV-1 & DENV-3.

Serotype	Fragment	RT Conditions					
DENV-1	F1,3,4	65°C 5 min	ice 1 min	25°C 5 min	42°C 1 hour	70°C 15 min	4°C stored
	F2	65°C 5 min	ice 1 min	25°C 5 min	42°C 1 hour	70°C 15 min	4°C stored
	F5	65°C 5 min	ice 1 min	25°C 5 min	42°C 1 hour	70°C 15 min	4°C stored
DENV-3	F1,2,4	65°C 5 min	ice 1 min	25°C 5 min	42°C 1 hour	70°C 15 min	4°C stored
	F3	65°C 5 min	ice 1 min	25°C 5 min	42°C 1 hour	70°C 15 min	4°C stored
	F5	65°C 5 min	ice 1 min	25°C 5 min	42°C 1 hour	70°C 15 min	4°C stored

Table 2.7: Overview of RT-Primers for cDNA synthesis of DENV-1 & DENV-3.

Serotype	Fragment	Primer	Sequence	Nucleotide Position
DENV-1	F1	d1a17	5'-CCAATGGCYGCTGAYAGTCT	2540-2559
	F2	d1a13a	5'-CCTACTTGGGACCTGCCCA	4611-4629
		d1a13b	5'-GGGACCTGCCCAACAGTCCT	4603-4622
		d1a13c	5'-TGCCCAACAGTCCTCTCTGC	4597-4616
	F3	d1a9	5'-CCAGTYARCACAGCTATCAAAGC	6551-6573
	F4	d1a4	5'-CACTCCACTGAGTGAATTCTCTCT	9068-9091
	F5	d1a23	5'-AGAACCTGTTGATTCAACAG	10716-10735
DENV-3	F1	d3a18	5'-GATTCCTATCGCAATGCATG	2361-2380
	F2	d3a14	5'-ACTGTGATCATTAARTTGTGGGA	4334-4356
	F3	d3a10	5'-GCYGCAAARTCCTTGAATTCCT	6339-6360
	F4	d3a6	5'-GCATTRACATGTCGRGTTCC	8342-8361
	F5	d3a23	5'-AGAACCTGTTGATTCAACAG	10688-10707

2.2.1.4 Amplification of cDNA by Polymerase Chain Reaction

Polymerase chain reaction was performed using the GeneAMP® PCR System 9700 (Applied Biosystems). The cDNA of the five fragments was amplified by PCR using *pfu Turbo* polymerase (Stratagene, CA, US) and 10 serotype specific primers. Briefly, 2µl of cDNA was mixed together with 39µl of water, 5µl 10x *pfu* buffer (Stratagene), 1 µl of 10mM dNTP as well as with either 1µl (fragments 1, 2, 3 & 4 for DENV-1 and DENV-3), 1µl (fragment 5 for DENV-3) or 10µl (fragment 5 for DENV-1) serotype specific sense as well as serotype specific anti sense primers. Finally, 1µl of *pfu Turbo* polymerase (Stratagene) was added to amplify the overlapping fragments under fragment specific conditions. The ramp speed [°C/s] was set to the maximum, except the amplification of fragment 5 of DENV-1 used 9600 as the ramp speed (Table 2.8; Table 2.9). The PCR products were analyzed on a 1% agarose gel and the correct bands were cut out, gel purified using the QIAquick Gel Extraction Kit (Quiagen) according to the manufacturers protocol. Purified products were sent to GIS for sequencing.

Table 2.8: Overview of PCR conditions for the amplification of DENV-1 & DENV-3; Shading indicates the PCR cycles.

Serotype	Fragment	PCR Conditions						
DENV-1	F1	35 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F2	35 cycles	95°C 2 min	95°C 30 sec	58°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F3	35 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F4	35 cycles	95°C 2 min	95°C 30 sec	58°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F5	40 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 4 min	72°C 2 min	4°C stored
DENV-3	F1	45 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F2	45 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F3	45 cycles	95°C 2 min	95°C 30 sec	63.1°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F4	45 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored
	F5	45 cycles	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3 min 30	72°C 2 min	4°C stored

Table 2.9: Overview of PCR Primers for the amplification of DENV-1 & DENV-3.

Serotype	Fragment	Primer	Sequence	Nucleotide Position
DENV-1	F1	d1a17	5'-CCAATGGCYGCTGAYAGTCT	2540-2559
		d1s22	5'-AGTTGTTAGTCTACGTGGAC	1-20
	F2	d1a13	5'-TTCCACTTCYGGAGGGCT	4544-4561
		d1s6	5'-GGYTCTATAGGAGGRGTGTTCAC	2201-2223
	F3	d1a9	5'-CCAGTYARCACAGCTATCAAAGC	6551-6573
		d1s10	5'-RGCYGGSCCACTAATAGCT	4213-4231
	F4	d1a4	5'-CACTCCACTGAGTGAATTCTCTCT	9068-9091
		d1s14	5'-ATGGRGAAAGGAACAACCAG	6216-6235
	F5	d1a23	5'-AGAACCTGTTGATTCAACAG	10716-10735
		d1s18c	5'-GGTGAGTCCTCTCCAAACCC	8015-8034
DENV-3	F1	d3a18	5'-GATTCCTATCGCAATGCATG	2361-2380
		d3s23	5'-AGTTGTTAGTCTACGTGGAC	1-20
	F2	d3a14	5'-ACTGTGATCATTAARTTGTGGGA	4334-4356
		d3s5	5'-TGAACCTCCTTTTGGGGAA	2035-2053
	F3	d3a10	5'-GCYGCAAARTCCTTGAATTCCT	6339-6360
		d3s9	5'-GAAAACAGATTGGCTCCCAA	4030-4049
	F4	d3a6	5'-GCATTRACATGTCGRGTTCC	8342-8361
		d3s13	5'-CCAGCTCTCTTTGAACCAGAAA	6032-6053
	F5	d3a23	5'-AGAACCTGTTGATTCAACAG	10688-10707
		d3s17	5'-CAACAGTGGAAGAAAGCAGAAC	8025-8046

2.2.2 Phylogenetic Analyses of Dengue Virus Genomes

2.2.2.1 Sequencing Process, Assembly and Quality Control

The sequencing of the amplified cDNA products were performed at GIS. Sequencing of PCR products were carried out using the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). Briefly, for every sequencing reaction, approximately 40 ng of purified PCR product and 3.2 pmol of primers were added into a BigDye Terminator v3.1 master mix. The mixture was then subjected to multiple cycles of PCR in the Eppendorf Mastercycler ep real-time PCR system and cleaned with CleanSEQ (Agencourt Bioscience, MA, US) before being analyzed with an Applied Biosystems 3730xl DNA Analyzer. The sequences were assembled and edited using SeqScape® version 2.5 from Applied Biosystem.

2.2.2.2 Sequence Alignment and Phylogentic Analyses

The nucleotide sequences were aligned with ClustalW-MPI (Li, 2003). Phylogenetic trees were inferred via the Minimum Evolution method and the Tamura-Nei genetic distance (Tamura and Nei, 1993) implemented within MEGA 3.1 (Kumar et al., 2004). Branch topology was verified by generating 1000 bootstraps from the sample sequence. Scores on nodes represent the number of bootstrap replicates supporting each node.

3 Results

3.1 Decision Tree Analyses of Clinical Data

3.1.1 Distinguishing Dengue Fever from other febrile Illnesses

Data were collected as part of the EDEN study during the 2005 Singaporean Dengue Outbreak. In total, 455 patients showing manifestations of febrile illness were recruited and 133 of them were PCR confirmed dengue positive cases (Low et al., 2006). Dengue patients showed significant differences to the control group regarding a lower platelet count, a decreased white blood cell count and lower absolute number of lymphocytes as well as of neutrophiles. Seventy-nine (59.4 %) of the dengue positive patients were hospitalized whereas only four (1.2%) amongst the controls needed hospital care. From the 133 dengue positive cases, there were fifty-seven (42.9%) who experienced a secondary infection which was indicated by a positive result for IgG antibodies at the first visit. There was no significant difference between secondary and primary infections in terms of hospitalization (49.4% versus 50.6%) but logistic regression identified smaller absolute numbers of lymphocytes, decreased platelet count and a higher viral load as genuine risk factors for hospitalization.

3.1.1.1 Dengue Prediction based on Clinical Data

The first results of decision tree modeling were based on the rationale of identifying markers that indicate dengue infection in an early stage of disease. For this purpose, we aimed at predicting a dengue PCR positive result by using clinical data collected on the 1st visit (average length of illness at around 42 hours) which finally resulted in a

dataset including the information of 453 patients. Two patients were excluded because of either a false positive PCR result or because the patient was unlikely to have dengue fever. Missing values were replaced by null, pruning confidence was set to the standard 25% and the tree parameter "minimum cases" was set to 19. The calculated tree (DENPRE_TOTAL_453) (Figure 3.1) used WBC_1<=4.8 (OR: 25.84; 95%CI: 24.14, 27.54), LYMPH_NO_1<=0.8 (OR: 16; 95%CI: 13.30, 18.70), LYMPH_NO_1<=0.5 (OR: 26.88; 95%CI: 24.39, 29.37) and finally TEMP_1>38.5 (OR: 8.23; 95%CI: 4.44, 12.01)1 as splitting criteria to identify dengue positive cases (Table 3.1; Table 3.2). The tree correctly classified 400 (88%) of the patients with a sensitivity of 82% and a specificity of 91% (Table 3.3; Figure 3.2). The overall error rate estimated after k-fold cross validation (k=10) was 12%. The ROC curve had an area under the curve (AUC) of 0.90 for positive (95%CI: 0.86, 0.94) as well as negative (95%CI: 0.87, 0.93) cases. The optimal classifier at a defined threshold for negative and positive cases had an average profit of 0.766.

In addition, we checked for the influence of symptoms on the classifier performance especially to investigate whether one would be able to define dengue disease by only characterizing the patients in terms of symptoms. Using symptoms alone was not at all predictive and resulted in a poor classifier performance (data not shown). We further observed that, when we included symptoms and clinical data for dengue prediction, the tree did not take into account any symptoms and the calculated tree was identical to the tree exclusively based on clinical data (data not shown).

[1] WBC=*white blood cell count*; LYMPH_NO=*absolute number of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval*

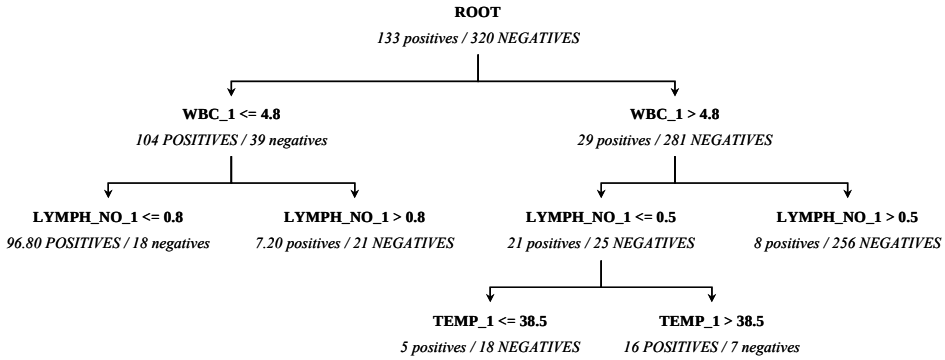


Figure 3.1: DENPRE_TOTAL_453: Decision tree for dengue prediction calculated on 453 patients excluding cytokine data. WBC=*white blood cell count*; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*.

Table 3.1: DENPRE_TOTAL_453: Decision tree for dengue prediction calculated on 453 patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset; WBC=*white blood cell count*; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.8	7.77	25.84	24.14, 27.54	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.8	12.27	28.55	26.65, 30.46	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.5	5.00	14.73	13.08, 16.38	< 0.001
TEMP_1 [°C] *Cut-off* value > 38.5	2.19	3.34	1.78, 4.90	< 0.001

Table 3.2: DENPRE_TOTAL_453: Decision tree for dengue prediction calculated on 453 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. WBC=*white blood cell coun*t; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.8	7.77	25.84	24.14, 27.54	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.8	3.37	16	13.30, 18.70	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.5	15.07	26.88	24.39, 29.37	< 0.001
TEMP_1 [°C] *Cut-off* value > 38.5	3.20	8.23	4.44, 12.01	0.003

Table 3.3: DENPRE_TOTAL_453: Summary of K-fold (k=10) cross validation for dengue prediction based on 453 patients excluding cytokine data.

Overall Evaluation	Value (n=453)	
Total misclassifications	53.0	
Overall error rate	11.691%	
SE of error rate	8.247	
Average profit	0.766	
SE of profit	0.165	
AUC negative	0.9014	95%CI: 0.87, 0.93
AUC positive	0.9013	95%CI: 0.86, 0.94

Confusion Matrix			
		Predicted Class	
		neg	pos
Actual Class	neg	291 (91%)	29 (9%)
	pos	24 (18%)	109 (82%)

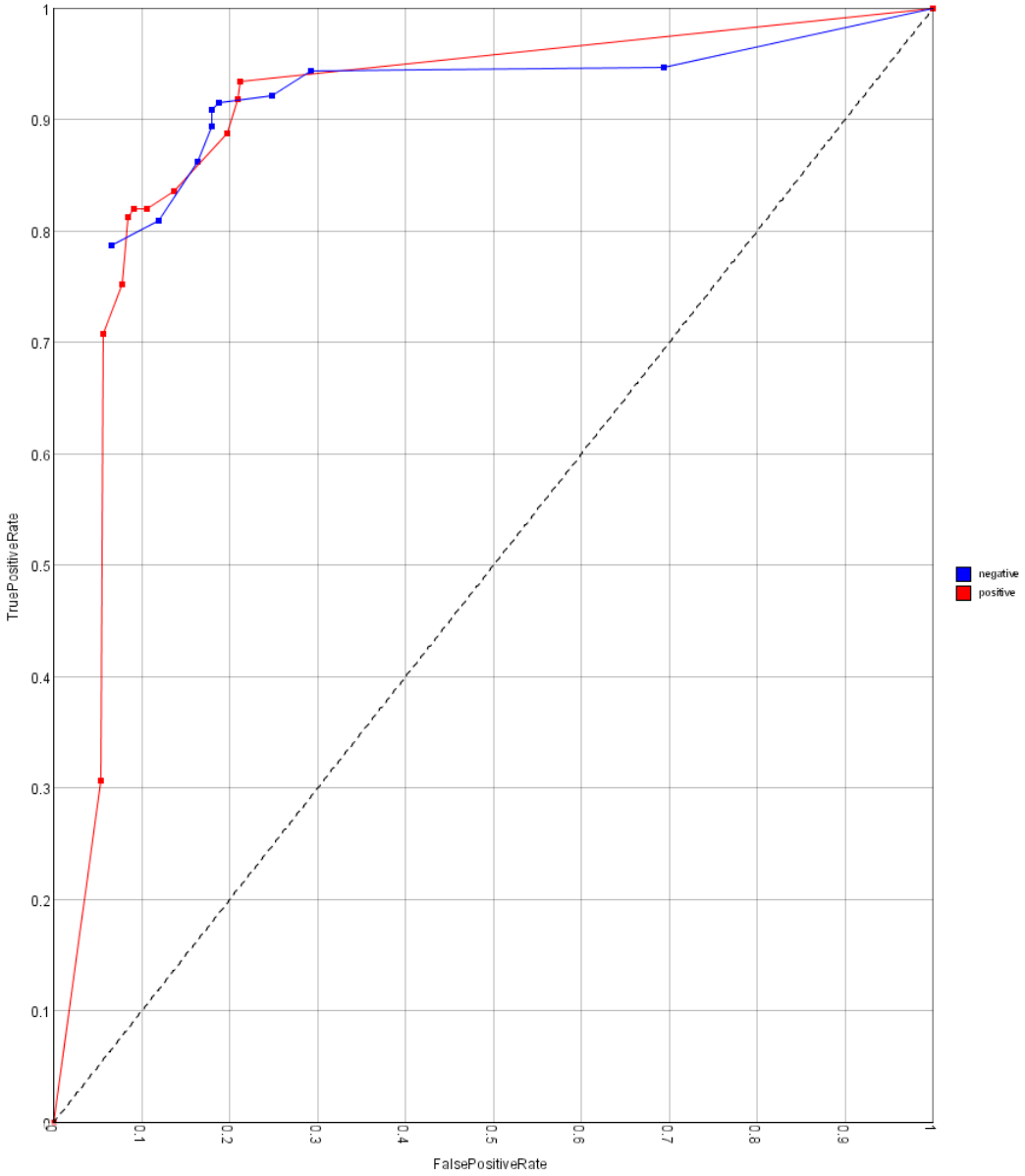


Figure 3.2: DENPRE_TOTAL_453: Receiver operating characteristics (ROC) curve for dengue prediction calculated on 453 patients excluding cytokine data.

3.1.1.2 Dengue Prediction based on Cytokine and Clinical Data

In a second attempt we examined whether we would be able to improve the classification tree by including available cytokine data represented by IP_10_1, IFN_ALPHA_1, I_TAC_1, GM_CSF_1, IL_1_1, IL_10_1, IL_12_1, IL_2_1, IL_4_1, IL_6_1, IL_8_1 and TNF_1[2]. For this analysis, only 291 patients (p(positive)=0.36; p(negative)=0.64) were included due to lack of required information in the remaining 162 patients. As a comparison, a tree excluding the cytokine data was first constructed by using a pruning confidence of 25% and 9 as the "minimum cases". The resulting tree (DENPRE_EXCYT_291) (Figure 3.3) used the same splitting criteria as the first tree constructed on the whole dataset of 453 patients. Besides WBC_1<=4.8 (OR: 26.24; 95%CI: 24.33, 28.16), LYMPH_NO_1<=0.8 (OR: 19.44; 95%CI: 15.27, 23.62), LYMPH_NO_1<=0.5 (OR: 42.12; 95%CI: 38.96, 45.25) and TEMP_1>38.5 (OR: 9.33; 95%CI: 1.51, 17.15), the tree additionally included LYMPH_NO_1<=0.3 (OR: 4.89; 95%CI: -1.06, 10.83) as a splitting criteria for dengue cases (Table 3.4; Table 3.5)[3]. It correctly classified 252 (85%) of the patients with a sensitivity of 79% and a specificity of 90% (Table 3.6; Figure 3.4). The AUC of the ROC curve was 0.91 and was equal regarding the classification accuracy for positive (95%CI: 0.87, 0.95) and negative (95%CI: 0.88, 0.94) cases. K-fold cross validation (k=10) resulted in an average profit of the model of 0.731.

[2] IP_10=*interferon-inducible protein 10*; IFN_ALPHA=*interferon-α*; I_TAC=*interferon-inducible T cell α chemoattractant*; GM_CSF=*granulocyte macrophage colony-stimulating factor*; IL_1=*interleukin*-1; IL_10=*interleukin-10*; IL_12=*interleukin-12*; IL_2=*interleukin-2*; IL_4=*interleukin-4*; IL_6_1=*interleukin-6*; IL_8=*interleukin-8*; TNF=*tumor necrosis factor α*; 1=*1st visit data*

[3] WBC=*white blood cell count*; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval.*

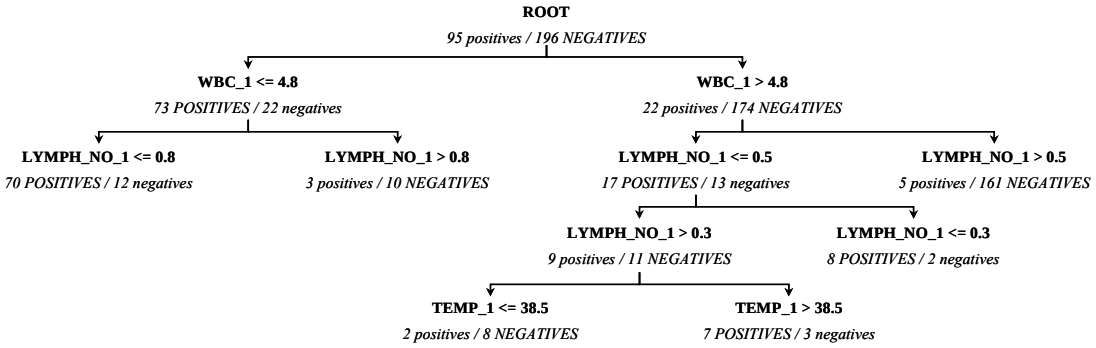


Figure 3.3: DENPRE_EXCYT_291: Decision tree for dengue prediction calculated on 291 patients excluding cytokine data. WBC=*white blood cell count*; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data.*

Table 3.4: DENPRE_EXCYT_291: Decision tree for dengue prediction calculated on 291 patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. WBC=*white blood cell coun*t; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.8	6.85	26.24	24.33, 28.16	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.8	16.00	42.17	39.75, 44.60	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.5	5.61	21.06	19.16, 22.95	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.3	3.72	32.7	29.30, 36.05	< 0.001
TEMP_1 [°C] *Cut-off* value > 38.5	2.26	2.26	2.03, 5.45	< 0.001

Table 3.5: DENPRE_EXCYT_291: Decision tree for dengue prediction calculated on 291 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. WBC=*white blood cell coun*t; LYMPH_NO=*absolute numbers of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.8	6.85	26.24	24.33, 28.16	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.8	3.70	19.44	15.27, 23.62	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.5	18.81	42.12	38.96, 45.25	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.3	1.78	4.89	-1.06, 10.83	0.119
TEMP_1 [°C] *Cut-off* value > 38.5	3.50	9.33	1.51, 17.15	0.07

Table 3.6: DENPRE_EXCYT_291: Summary of K-fold (k=10) cross validation for dengue prediction based on 291 patients excluding cytokine data.

Overall Evaluation	Value (n=291)	
Total misclassifications	39.0	
Overall error rate	13.437%	
SE of error rate	8.054	
Average profit	0.731	
SE of profit	0.161	
AUC negative	0.9099	95%CI: 0.88, 0.94
AUC positive	0.9088	95%CI: 0.87, 0.95

Confusion Matrix			
		Predicted Class	
		neg	pos
Actual Class	neg	177 (90%)	19 (10%)
	pos	20 (21%)	75 (79%)

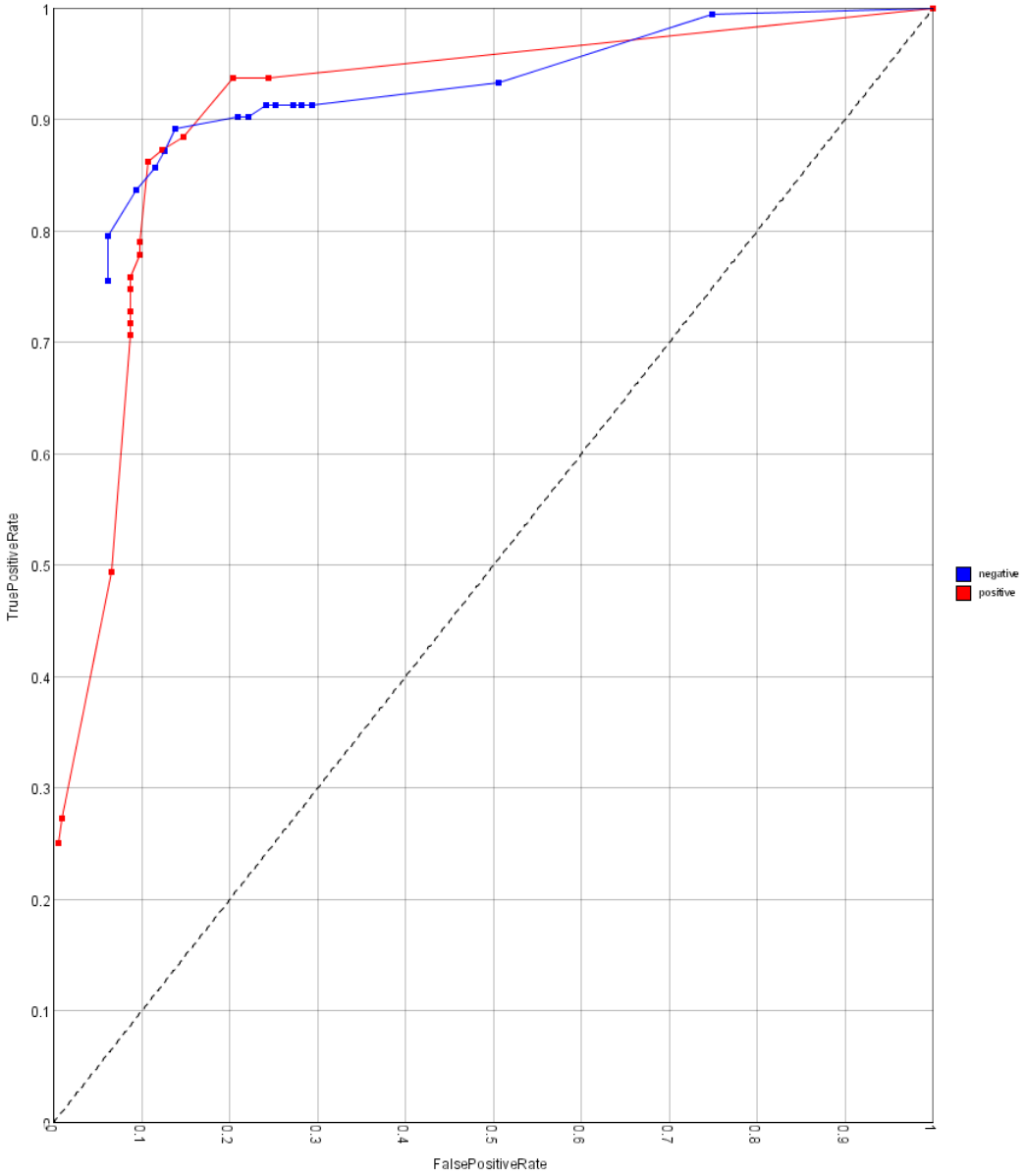


Figure 3.4: DENPRE_EXCYT_291: Receiver operating characteristics (ROC) curve for dengue prediction calculated on 291 patients excluding cytokine data.

Finally, a tree constructed including the cytokine data (DENPRE_INCYTA_291) (Figure 3.5) used IFN_ALPHA_1>389.21 (OR: 79.54; 95%CI: 76.61, 82.48), IL_2_1<=2.24 (OR: 131.59; 95%CI: 123.65, 139.53), WBC<=4.1 (OR: 109.2; 95%CI: 105.23, 113.17) and IL_1_1>2.62 (OR: 45.0; 95%CI: 33.87, 56.13)[4] as splitting criteria (Table 3.7; Table 3.8) and had an overall error rate of 7.9%. Tree parameters such as pruning confidence were left at 25% and "minimum cases" was set to 8. It showed a sensitivity of 84% respectively a specificity of 96% and the AUC was 0.92 for positive (95%CI: 0.89, 0.96) as well as negative (95%CI: 0.90, 0.95) cases, with an average profit of 0.842 (Table 3.9; Figure 3.6).

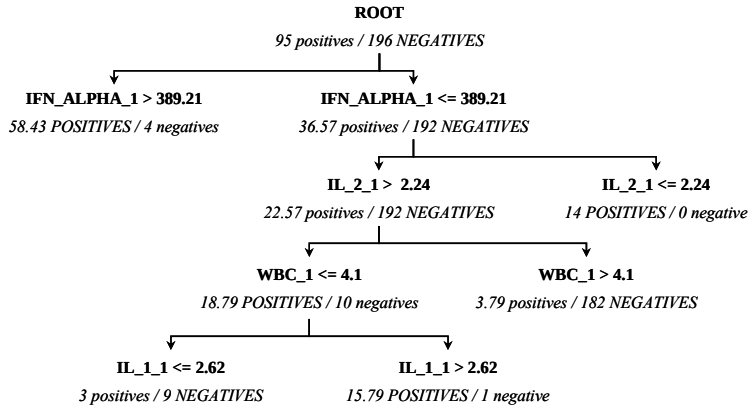


Figure 3.5: DENPRE_INCYTA_291: Decision tree for dengue prediction calculated on 291 patients including cytokine and clinical data. IFN_ALPHA=*interferon-α*; IL_2=*interleukin-2*; WBC=*white blood cell count*; IL_1=*interleukin-1*; 1=*1st visit data.*

[4] IFN_ALPHA=*interferon-α*; IL_2=*interleukin-2*; WBC=*white blood cell count*; IL_1=*interleukin-1*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval.*

Table 3.7: DENPRE_INCYTA_291: Decision tree calculated on 291 patients including cytokine and clinical data. Statistical analysis of splitting criteria performed on the whole dataset. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value [+1]. IFN_ALPHA=*interferon-α*; IL_2=*interleukin-2*; WBC=*white blood cell count*; IL_1=*interleukin-1*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IFN_ALPHA_1 [pg/ml] *Cut-off* value > 389.21	6.07	79.54	76.61, 82.47	< 0.001
IL_2_1[pg/ml] [+1] *Cut-off* value <= 2.24	3.77	84.01	76.53, 91.50	< 0.001
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.1	5.51	28.81	26.75, 30.86	< 0.001
IL_1_1 [pg/ml] *Cut-off* value > 2.62	4.30	7.28	5.36, 9.19	< 0.001

Table 3.8: DENPRE_INCYTA_291: Decision tree calculated on 291 patients including cytokine and clinical data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value[+1]. IFN_ALPHA=*interferon-α*; IL_2=*interleukin-2*; WBC=*white blood cell count*; IL_1=*interleukin-1*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IFN_ALPHA_1 [pg/ml] *Cut-off* value > 389.21	6.07	79.54	76.61, 82.47	< 0.001
IL_2_1 [pg/ml] [+1] *Cut-off* value <= 2.24	9.16	131.59	123.65, 139.53	< 0.001
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.1	39.64	109.20	105.23, 113.17	< 0.001
IL_1_1 [pg/ml] *Cut-off* value > 2.62	3.75	45.00	33.87, 56.13	< 0.001

Table 3.9: DENPRE_INCYTA_291: Summary of K-fold (k=10) cross validation for dengue prediction based on 291 patients including cytokine and clinical data.

Overall Evaluation	Value (n=291)	
Total misclassifications	23.0	
Overall error rate	7.908%	
SE of error rate	4.323	
Average profit	0.842	
SE of profit	0.086	
AUC negative	0.9243	95%CI: 0.90, 0.95
AUC positive	0.9245	95%CI: 0.89, 0.96

Confusion Matrix			
		Predicted Class	
		neg	pos
Actual Class	neg	188 (96%)	8 (4%)
	pos	15 (16%)	80 (84%)

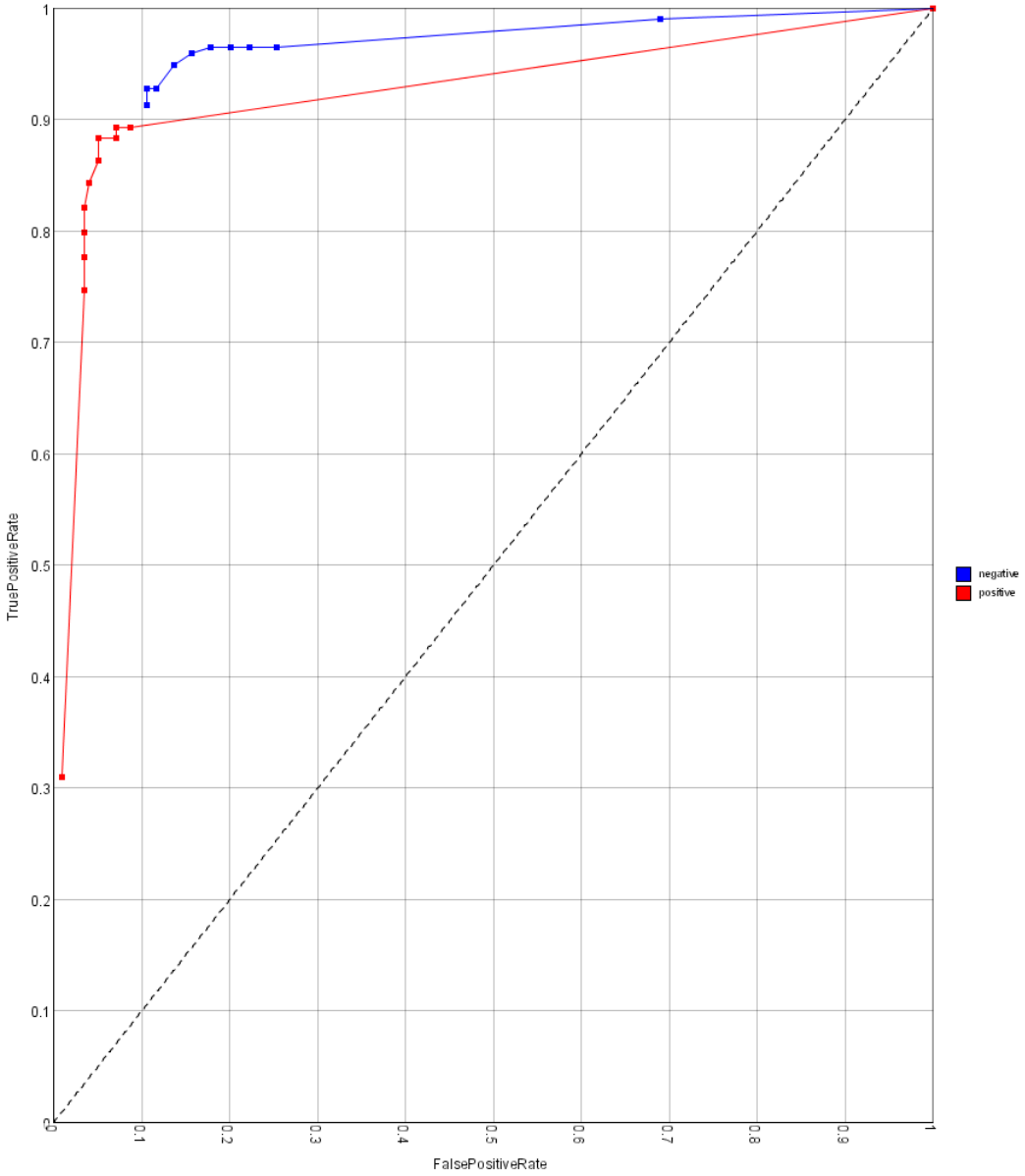


Figure 3.6: DENPRE_INCYTA_291: Receiver operating characteristics (ROC) curve for dengue prediction calculated on 291 patients including cytokine and clinical data.

Considering the fact that interferon-α levels are a general indicator of RNA virus infections we aimed at getting a more detailed picture by excluding this parameter from the decision tree analysis. Interestingly, the resulting classifier (pruning confidence was left at 25% and minimum cases was set to 8) (DENPRE_INCYT_291) (Figure 3.7) was able to correctly classify 270 (92.78%) of the patients with a sensitivity of 90%, a specificity of 94% and an overall error rate of 7.2% (Table 3.12; Figure 3.8). The first split was represented by IL_2_1<=2.54 and patients below this threshold were very likely to be dengue positive (OR: 88.21; 95%CI: 80.73, 95.68). The other arm of the tree which was represented by higher levels of IL_2_1 was further separated by WBC_1<=4.8 (OR: 29.38; 95%CI: 27.28, 31.47) into dengue positive cases which were additionally categorized by using LYMPH_NO_1<=0.8 (OR: 20.83; 95%CI: 15.66, 26.01), IFN_GAMMA_1>9.33 (OR: 26.25; 95%CI: 20.80, 31.70) and WBC_1<=3.2 (OR: 63.0; 95%CI: 44.04, 81.96) as characteristics for positive cases[5] (Table 3.10; Table 3.11). Patients showing WBC>4.8 were further split into dengue positive cases by using IP_10_1>1490 (OR: 68.57; 95%CI: 63.65, 73.49) and TNF_1<=4.08 (OR: 36.0; 95%CI: 25.69, 46.31) as cut off values. The overall classifier performance was 0.94 for positive (95%CI: 0.91, 0.98) as well as negative cases (95%CI: 0.92, 0.97) and the overall profit averaged 0.856.

[5] IL_2=*interleukin-2*; WBC_1=*white blood cell count*; LYMPH_NO=*absolute number of lymphocytes*; IFN_GAMMA=*interferon-γ*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval.*

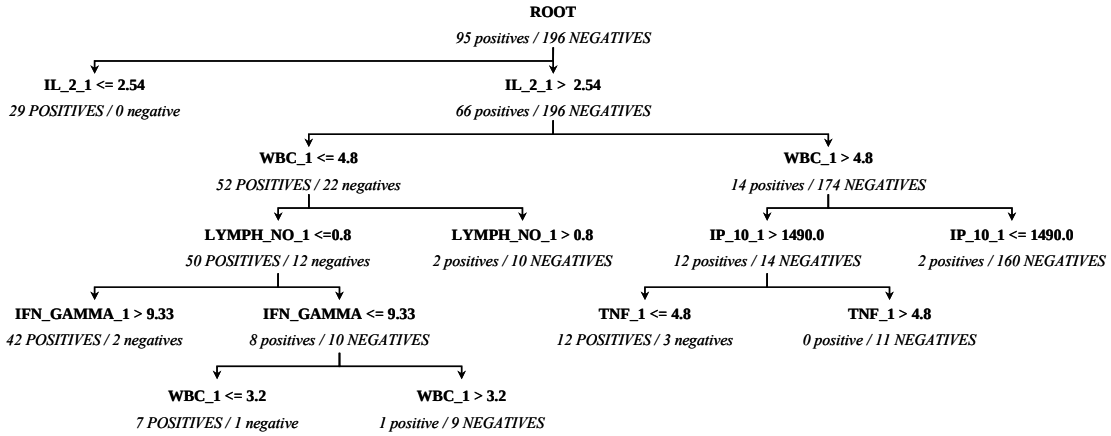


Figure 3.7: DENPRE_INCYT_291: Decision tree for dengue prediction calculated on 291 patients including cytokine (excl. IFN_ALPHA_1) and clinical data. IL_2=*interleukin-2*; WBC_1=*white blood cell count*; LYMPH_NO=*absolute number of lymphocytes*; IFN_GAMMA=*interferon-γ*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data.*

Table 3.10: DENPRE_INCYT_291: Decision tree for dengue prediction calculated on 291 patients including cytokine (excl. IFN_ALPHA_1) and clinical data. Statistical analysis of splitting criteria performed on the whole dataset. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value[+1]. IL_2=*interleukin-2*; WBC_1=*white blood cell count*; LYMPH_NO=*absolute number of lymphocytes*; IFN_GAMMA=*interferon-γ*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IL_2_1[pg/ml] [+1] *Cut-off* value <= 2.54	3.81	88.21	80.73, 95.68	< 0.001
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.8	6.85	26.24	24.33, 28.16	< 0.001
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 3.2	4.60	83.69	79.42, 87.95	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.8	16.00	42.17	39.75, 44.60	< 0.001
IFN_GAMMA_1 [pg/ml] *Cut-off* value > 9.33	1.58	1.94	0.27, 3.61	0.011
IP_10_1 [pg/ml] *Cut-off* value > 1490.0	5.19	20.73	18.05, 22.65	< 0.001
TNF_1 [pg/ml] *Cut-off* value <= 4.8	7.56	15.64	13.59, 17.69	< 0.001

Table 3.11: DENPRE_INCYT_291: Decision tree calculated on 291 patients excluding including cytokine (excl. IFN_ALPHA_1) and clinical data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value[+1]. IL_2=*interleukin-2*; WBC_1=*white blood cell count*; LYMPH_NO=*absolute number of lymphocytes*; IFN_GAMMA=*interferon-γ*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IL_2_1[pg/ml] [+1] *Cut-off* value <= 2.54	3.81	88.21	80.73, 95.68	< 0.001
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 4.8	9.44	29.38	27.28, 31.47	< 0.001
WBC_1 [*1000 cells/mm^3] *Cut-off* value <= 3.2	8.75	63.00	44.04, 81.96	0.003
LYMPH_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.8	4.84	20.83	15.66, 26.01	< 0.001
IFN_GAMMA_1 [pg/ml] *Cut-off* value > 9.33	2.15	26.25	20.80, 31.70	< 0.001
IP_10_1 [pg/ml] *Cut-off* value > 1490.0	37.38	68.57	63.65, 73.49	< 0.001
TNF_1 [pg/ml] [+1] *Cut-off* value <= 4.8	9.75	36.00	25.69, 46.31	< 0.001

Table 3.12: DENPRE_INCYT_291: Summary of K-fold (k=10) cross validation for dengue prediction based on 291 patients including cytokine and clinical data (excl. IFN_ALPHA_1).

Overall Evaluation	Value (n=291)	
Total misclassifications	21.0	
Overall error rate	7.207%	
SE of error rate	3.765	
Average profit	0.856	
SE of profit	0.075	
AUC negative	0.9443	95%CI: 0.92, 0.97
AUC positive	0.9442	95%CI: 0.91, 0.98

Confusion Matrix			
		Predicted Class	
		neg	pos
Actual Class	neg	185 (94%)	11 (6%)
	pos	10 (11%)	85 (90%)

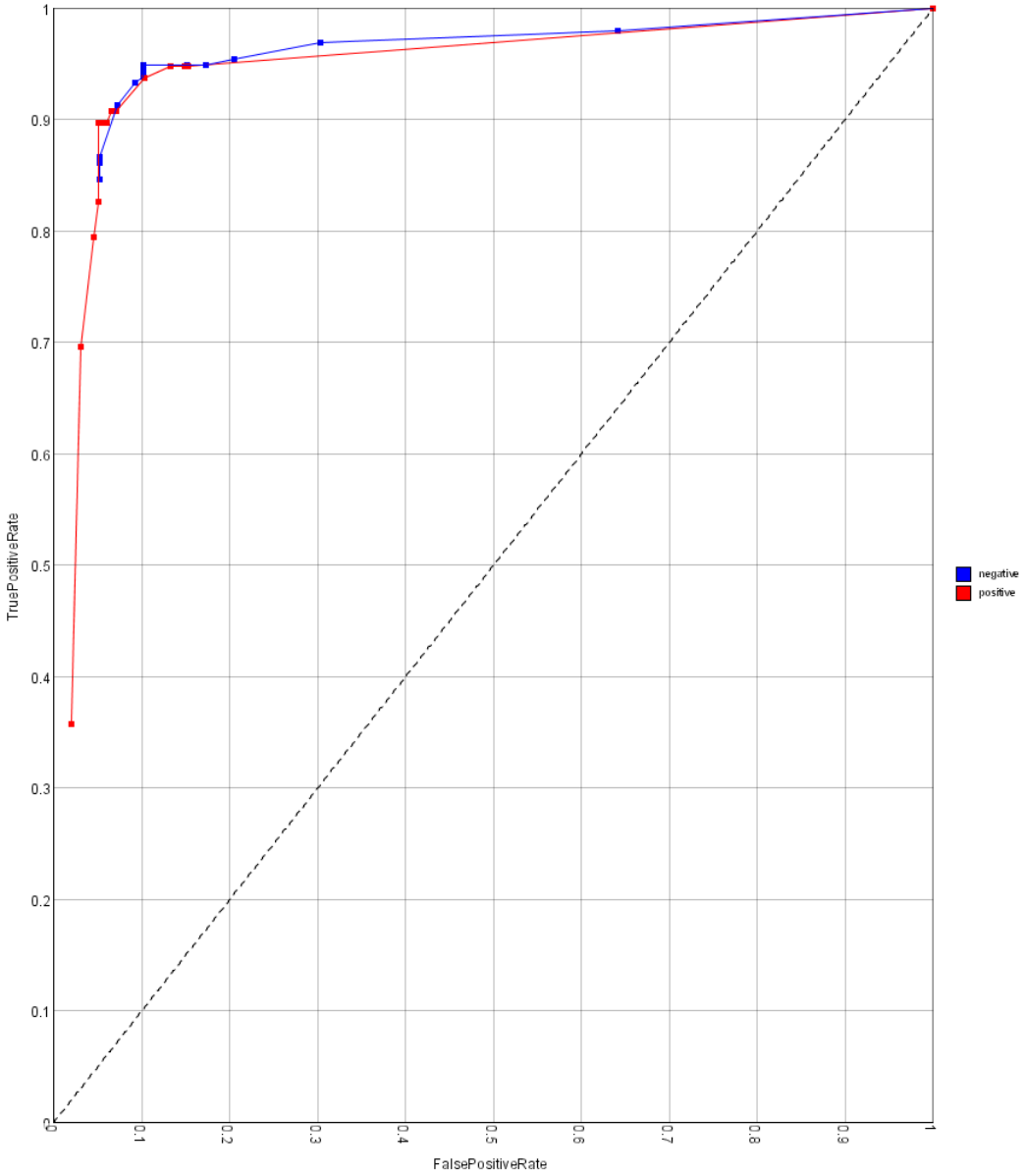


Figure 3.8: DENPRE_INCYT_291: Receiver operating characteristics (ROC) curve for dengue prediction calculated on 291 patients including cytokine (excl. IFN_ALPHA_1) and clinical data.

3.1.1.3 Dengue Prediction based on Cytokine Data

We also addressed the question whether it would be possible to classify dengue patients by only using the cytokine data. The first constructed tree (DENPRE_CYTOA_291) (Figure 3.9) showed a sensitivity of 84% and a specificity of 94% which resulted in an overall error rate of 9.28% (27 cases). The average profit of the model was 0.814 and the positive AUC was 0.92 (95%CI: 0.88, 0.96). The negative AUC had also a value of 0.92 (95%CI: 0.89, 0.95) (Table 3.15; Figure 3.10). The first split of the tree was represented by IFN_ALPHA_1>389.21 (OR: 79.54; 95%CI: 76.61, 82.47) for dengue positive cases. Patients showing lower levels of IFN_ALPHA_1 were further sub divided by IL_2_1<=2.72 (OR: 63.33; 95%CI: 58.63, 68.03) which represented a higher risk for patients to be dengue positive. Classification of cases having IL_2_1>2.72 additionally needed IL_10_1>9.14 (OR: 16.10; 95%CI: 13.34, 18.85) as well as TNF_1<=4.33 (OR: 32.5; 95%CI: 23.07, 41.93) as splitting criteria [6] (Table 3.13; Table 3.14).

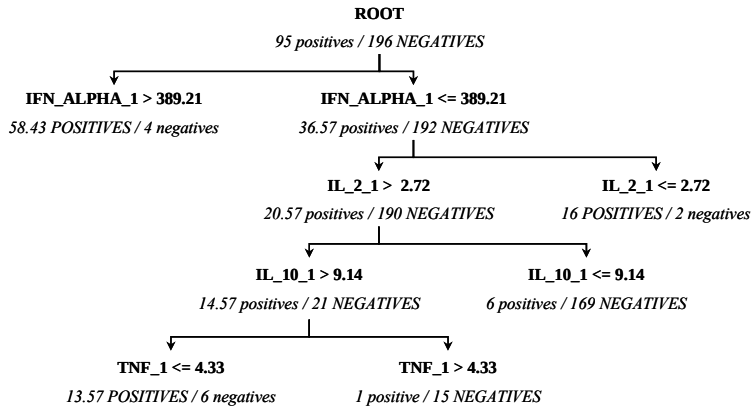


Figure 3.9: DENPRE_CYTOA_291: Decision tree for dengue prediction calculated on 291 patients only using cytokine data. IFN_ALPHA=*interferon-α*; IL_10=*interleukin-10*; IL_2=*interleukin-2*; TNF=*tumor necrosis factor α*; 1=*1st visit data.*

[6] IFN_ALPHA=*interferon-α*; IL_10=*interleukin-10*; IL_2=*interleukin-2*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval.*

Table 3.13: DENPRE_CYTOA_291: Decision tree for dengue prediction calculated on 291 patients only including cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. IFN_ALPHA=*interferon-α*; IL_10=*interleukin-10*; IL_2=*interleukin-2*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IFN_ALPHA_1 [pg/ml] *Cut-off* value > 389.21	6.07	79.54	76.61, 82.47	< 0.001
IL_2_1 [pg/ml] *Cut-off* value <= 2.72	3.69	42.62	38.32, 46.93	< 0.001
IL_10_1 [pg/ml] *Cut-off* value > 9.14	2.80	6.33	4.49, 8.16	< 0.001
TNF_1 [pg/ml] *Cut-off* value <= 4.33	11.49	23.93	21.53, 26.33	< 0.001

Table 3.14: DENPRE_CYTOA_291: Decision tree for dengue prediction calculated on 291 patients only including cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. IFN_ALPHA=*interferon-α*; IL_10=*interleukin-10*; IL_2=*interleukin-2*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IFN_ALPHA_1 [pg/ml] *Cut-off* value > 389.21	6.07	79.54	76.61, 82.47	< 0.001
IL_2_1 [pg/ml] *Cut-off* value <= 2.72	8.79	63.33	58.63, 68.03	< 0.001
IL_10_1 [pg/ml] *Cut-off* value > 9.14	10.06	16.10	13.34, 18.85	< 0.001
TNF_1 *Cut-off* value <= 4.33	10.95	32.50	23.07, 41.93	< 0.001

Table 3.15: DENPRE_CYTOA_291: Summary of K-fold (k=10) cross validation for dengue prediction based on 291 patients only including cytokine data.

Overall Evaluation	Value (n=291)	
Total misclassifications	27.0	
Overall error rate	9.276%	
SE of error rate	5.637	
Average profit	0.814	
SE of profit	0.113	
AUC negative	0.9193	95%CI: 0.89, 0.95
AUC positive	0.9193	95%CI: 0.88, 0.96

Confusion Matrix			
		Predicted Class	
		neg	pos
Actual Class	neg	184 (94%)	12 (6%)
	pos	15 (16%)	80 (84%)

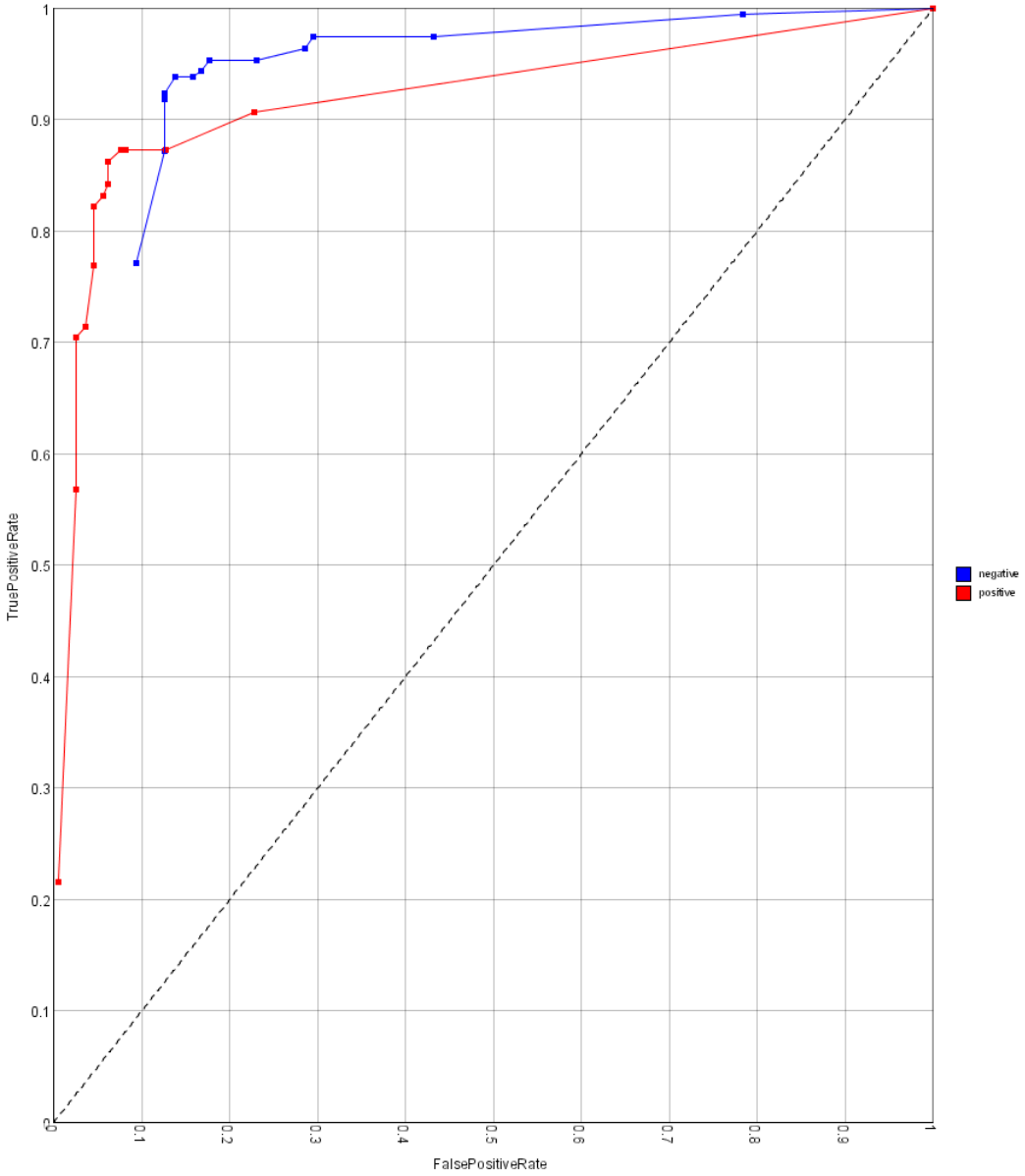


Figure 3.10: DENPRE_CYTOA_291: Receiver operating characteristics (ROC) curve for dengue prediction calculated on 291 patients only including cytokine data.

Finally, we calculated a tree (minimum cases set to 14; pruning confidence set to 25%) based on only the cytokine data and excluding IFN_ALPHA_1 (DENPRE_CYTO_291) (Figure 3.11). This resulted in a correct classification of 264 (90.72%) of the cases with a sensitivity of 84% and a specificity 94% (Table 3.18; Figure 3.12). The overall performance of the tree was slightly worse compared to the inclusion of IFN_ALPHA_1 with an AUC for positive (95%CI: 0.87, 0.95) and negative (95%CI: 0.88, 0.94) cases of 0.91 but had the same average profit of 0.814. IL_2_1<=2.54 (OR: 88.21; 95%CI: 80.73, 95.68) was chosen as the first splitting criteria whereby cases below this threshold were more likely to have dengue. On the other hand, patients having higher levels of IL_2_1 were further classified into positive cases by taking IP_10_1>878.7 (OR: 29.12; 95%CI: 26.77, 31.46) and TNF_1<=4.29 (OR: 25.75; 95%CI: 22.68, 28.82) as characteristics for dengue infections [7] (Table 3.16; Table 3.17).

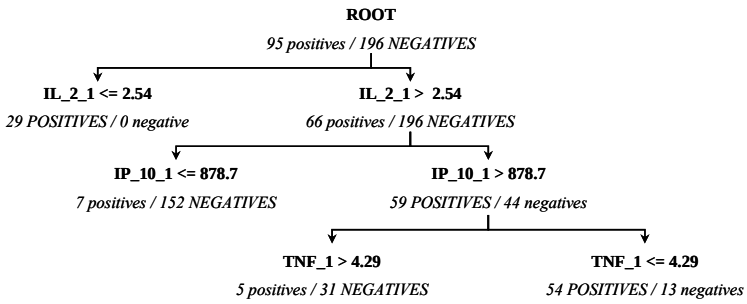


Figure 3.11: DENPRE_CYTO_291: Decision tree for dengue prediction calculated on 291 patients only including cytokine data (excl. IFN_ALPHA_1). IL_2=*interleukin-2*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data.*

[7] IL_2=*interleukin-2*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data.* OR=*odds ratio*; CI=*confidence interval.*

Table 3.16: DENPRE_CYTO_291: Decision tree for dengue prediction calculated on 291 patients only including cytokine data (excl. IFN_ALPHA_1). Statistical analysis of splitting criteria performed on the whole dataset. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value [+1]. IL_2=*interleukin-2*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IL_2_1 [pg/ml] [+1] *Cut-off* value <= 2.54	3.81	88.21	80.73, 95.68	< 0.001
IP_10_1 [pg/ml] *Cut-off* value > 878.7	10.67	29.36	27.28, 31.45	< 0.001
TNF_1 [pg/ml] *Cut-off* value <= 4.29	11.65	24.45	22.05, 26.86	< 0.001

Table 3.17: DENPRE_CYTO_291: Decision tree for dengue prediction calculated on 291 patients only including cytokine data (excl. IFN_ALPHA_1). Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value[+1]. IL_2=*interleukin-2*; IP_10=*interferon-inducible protein 10*; TNF=*tumor necrosis factor α*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IL_2_1 [pg/ml] [+1] *Cut-off* value <= 2.54	3.81	88.21	80.73, 95.68	< 0.001
IP_10_1 [pg/ml] *Cut-off* value > 878.7	13.01	29.12	26.77, 31.46	< 0.001
TNF_1 [pg/ml] *Cut-off* value <= 4.29	5.80	25.75	22.68, 28.82	< 0.001

Table 3.18: DENPRE_CYTO_291: Summary of K-fold (k=10) cross validation for dengue prediction based on 291 patients only including cytokine data (excl. IFN_ALPHA_1).

Overall Evaluation	Value (n=291)	
Total misclassifications	27.0	
Overall error rate	9.287%	
SE of error rate	5.415	
Average profit	0.814	
SE of profit	0.108	
AUC negative	0.9126	95%CI: 0.88, 0.94
AUC positive	0.9132	95%CI: 0.87, 0.95

Confusion Matrix			
		Predicted Class	
		neg	pos
Actual Class	neg	184 (94%)	12 (6%)
	pos	15 (16%)	80 (84%)

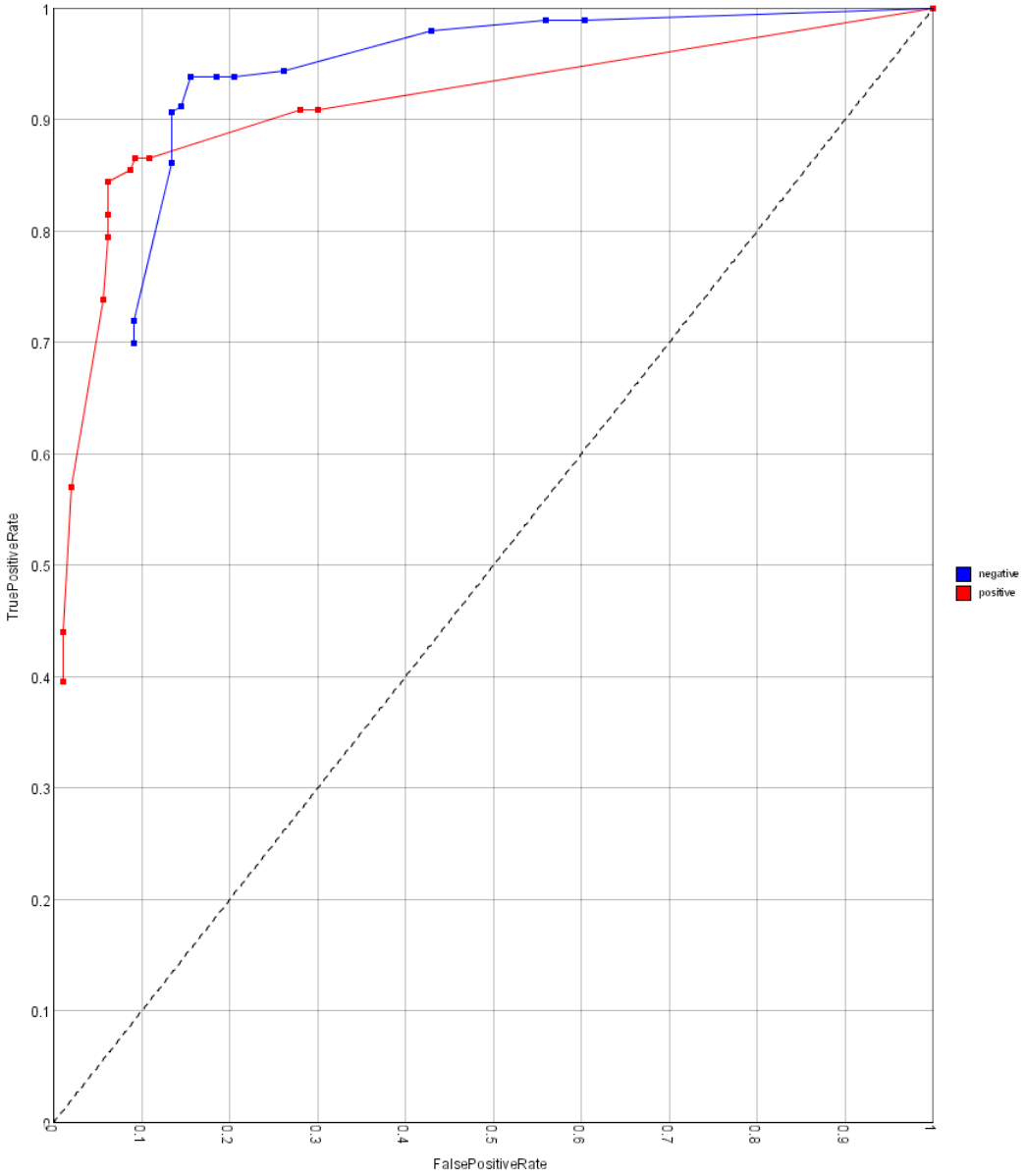


Figure 3.12: DENPRE_CYTO_291: Receiver operating characteristics (ROC) curve for dengue prediction calculated on 291 patients only including cytokine data (excl. IFN_ALPHA_1).

3.1.2 Prediction of Disease Severity in Dengue Patients

3.1.2.1 Prediction of Hospitalization based on Clinical Data

Besides distinguishing between dengue and non-dengue patients, we also elucidated whether we can identify any severity related differences within the 133 dengue positive patients. In a first attempt, we tried to classify dengue patients by using hospitalization (yes/no) as the outcome. A first tree (HOSP_TOTAL_133) was calculated including the total dataset and missing values were replaced by null. The pruning confidence was set to the standard 25% whereas the "minimum cases" was defined to 12. Using PLT_2<=84.0 (OR: 75.87; 95%CI: 68.25, 83.49), LYMPH_NO_1<=0.7 (OR: 15.97; 95%CI: 8.00, 23.94) and TEMP_2<=37.0 (OR: 11.73; 95%CI: 8.44, 15.02)[8] as splitting criteria (Table 3.19; Table 3.20), the tree (Figure 3.13) correctly identified 107 (80.45%) of the cases which accounted for an overall error rate of 19.45% (Table 3.21; Figure 3.14). The tree had a higher sensitivity (85%) but, in turn, a lower specificity (74%) and the profit averaged 0.611. The overall accuracy of the tree was high with an AUC of 0.91 for the classification of yes (95%CI: 0.86, 0.96) and no (95%CI: 0.85, 0.97) cases. Due to our interest in finding early markers of severity of dengue infection we tried to predict hospitalization by only using data from the first visit, but the model did not result in a reliable prediction (data not shown).

[8] PLT=*platelet count*; LYMPH_NO=*absolute number of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; 2=*2nd visit data;* OR=*odds ratio*; CI=*confidence interval*.

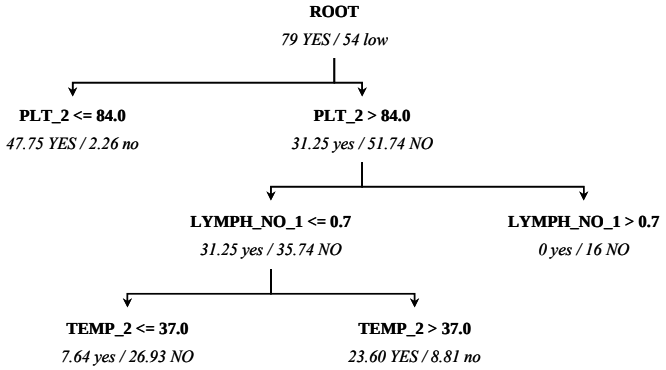


Figure 3.13: HOSP_TOTAL_133: Decision tree for hospitalization calculated on 133 patients excluding cytokine data. PLT=*platelet count*; LYMPH_NO=*absolute number of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; 2=*2nd visit data.*

Table 3.19: HOSP_TOTAL_133: Decision tree for hospitalization calculated on 133 patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. PLT=*platelet count*; LYMPH_NO=*absolute number of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; 2=*2nd visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_2 [*1000/mm^3] *Cut-off* value <= 84	2.53	75.87	68.25, 83.49	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] $^{+1}$ *Cut-off* value <= 0.7	2.76	6.15	3.21, 9.09	0.001
TEMP_2 [°C] *Cut-off* value > 37.0	1.61	3.60	1.44, 5.77	0.001

Table 3.20: HOSP_TOTAL_133: Decision tree for hospitalization calculated on 133 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value $^{+1}$. PLT=*platelet count*; LYMPH_NO=*absolute number of lymphocytes*; TEMP=*body temperature*; 1=*1st visit data*; 2=*2nd visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_2 [*1000/mm^3] $^{+1}$ *Cut-off* value <= 84	2.53	75.87	68.25, 83.49	< 0.001
LYMPH_NO_1 [*1000 cells/mm^3] $^{+1}$ *Cut-off* value <= 0.7	8.72	15.97	8.00, 23.94	< 0.001
TEMP_2 [°C] *Cut-off* value > 37.0	3.50	11.73	8.44, 15.02	< 0.001

Table 3.21: HOSP_TOTAL_133: Summary of K-fold (k=10) cross validation for prediction of hospitalization based on 133 patients excluding cytokine data.

Overall Evaluation	Value (n=133)	
Total misclassifications	26.0	
Overall error rate	19.451%	
SE of error rate	10.962	
Average profit	0.611	
SE of profit	0.219	
AUC no	0.9097	95%CI: 0.85, 0.97
AUC yes	0.9067	95%CI: 0.86, 0.96

Confusion Matrix			
		Predicted Class	
		no	yes
Actual Class	no	40 (74%)	14 (26%)
	yes	12 (15%)	67 (85%)

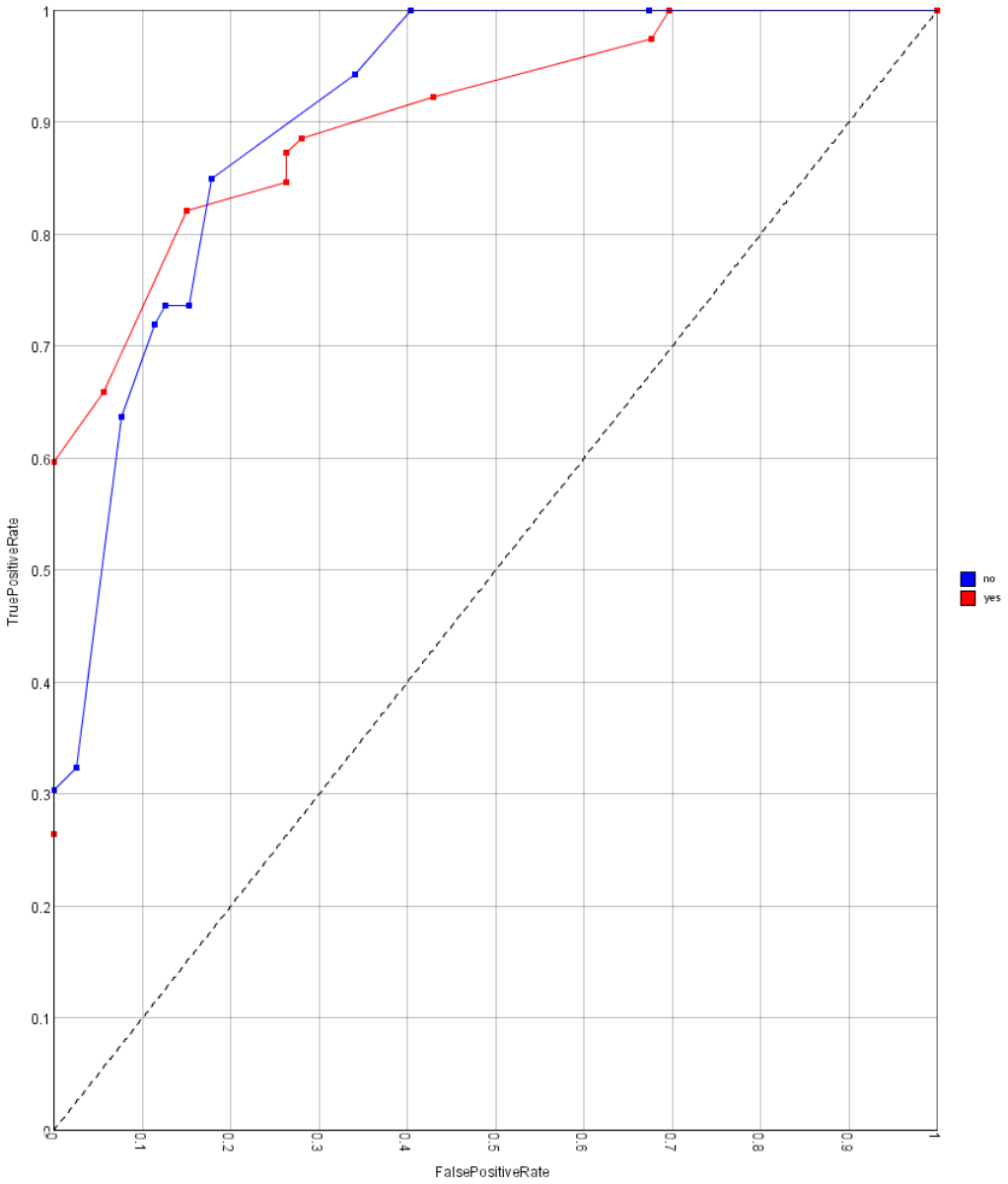


Figure 3.14: HOSP_TOTAL_133: Receiver operating characteristics (ROC) curve for prediction of hospitalization calculated on 133 patients excluding cytokine data.

3.1.2.2 Prediction of Hospitalization based on Cytokine and Clinical Data

In a next step, we investigated the question whether using cytokine data can improve the model for hospitalization. Hence, we first calculated a tree (pruning confidence set to 25% and "minimum cases" set to 5) only using the information of 95 patients and excluding the cytokine data. This tree was the reference tree to investigate the effect of cytokines on the prediction of hospitalization. The remaining 38 patients were excluded from the analysis due to no available cytokine data. The resulting tree (HOSP_EXCYT_95) (Figure 3.15) showed a sensitivity of 81% and a specificity of 70% with an overall error rate of 22.778% which reflected a correct classification of 73 cases (Table 3.24; Figure 3.16). The overall performance was worse than the accuracy of the tree constructed on all the 133 dengue patients and the AUC of the ROC curve was 0.85 (95%CI: 0.76, 0.93) for classification of non-hospitalized cases and 0.85 (95%CI: 0.77, 0.92) for the prediction of hospitalized cases. The probabilistic classifier had an average profit of 0.544. The tree chose different decision nodes and thresholds unlike the tree calculated on the total dataset of 133 patients (Figure 3.13). PLT_2 (OR: 16.88; 95%CI: 14.04, 19.71) was still chosen as the first splitting criteria but with a slightly higher threshold of <=116 for hospitalized cases. Patients having a higher platelet count than the defined threshold were additionally divided by PULSERATE_1 (OR: 26.00; 95%CI: 14.63, 37.37) whereby patients equal or below the threshold 75.00 were more likely to be hospitalized [9] (Table 3.22; Table 3.23).

[9] PLT=*platelet count*; PULSERATE=*pulse rate*; 1=*1st visit data*; 2=*2nd visit data;* OR=*odds ratio*; CI=*confidence interval.*

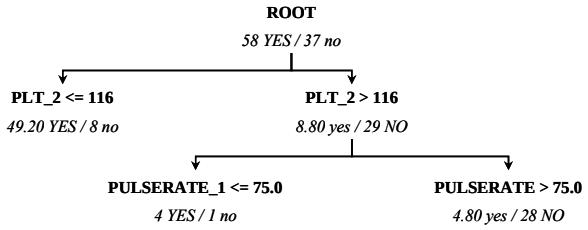


Figure 3.15: HOSP_EXCYT_95: Decision tree for hospitalization calculated on 95 patients excluding cytokine data. PLT=*platelet count*; PULSERATE=*pulse rate*; 1=*1st visit data*; 2=*2nd visit data.*

Table 3.22: HOSP_EXCYT_95: Decision tree for hospitalization calculated on 95 patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. PLT=*platelet count*; PULSERATE=*pulse rate*; 1=*1st visit data*; 2=*2nd visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_2 [*1000/mm^3] *Cut-off* value <= 116	3.65	16.88	14.04, 19.71	< 0.001
PULSERATE_1 [beats/min] *Cut-off* value <= 75.0	1.40	3.21	-1.70, 8.13	0.193

Table 3.23: HOSP_EXCYT_95: Decision tree calculated on 95 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. PLT=*platelet count*; PULSERATE=*pulse rate*; 1=*1st visit data*; 2=*2nd visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_2 [*1000/mm^3] *Cut-off* value <= 116	3.65	16.88	14.04, 19.71	< 0.001
PULSERATE_1 [beats/min] *Cut-off* value <= 75.0	6.00	26.00	14.63, 37.37	0.006

Table 3.24: HOSP_EXCYT_95: Summary of K-fold (k=10) cross validation for prediction of hospitalization based on 95 patients excluding cytokine data.

Overall Evaluation	Value (n=95)	
Total misclassifications	22.0	
Overall error rate	22.778%	
SE of error rate	11.139	
Average profit	0.544	
SE of profit	0.223	
AUC no	0.8481	95%CI: 0.76, 0.93
AUC yes	0.8481	95%CI: 0.77, 0.92

Confusion Matrix			
		Predicted Class	
		no	yes
Actual Class	no	26 (70%)	11 (30%)
	yes	11 (19%)	47 (81%)

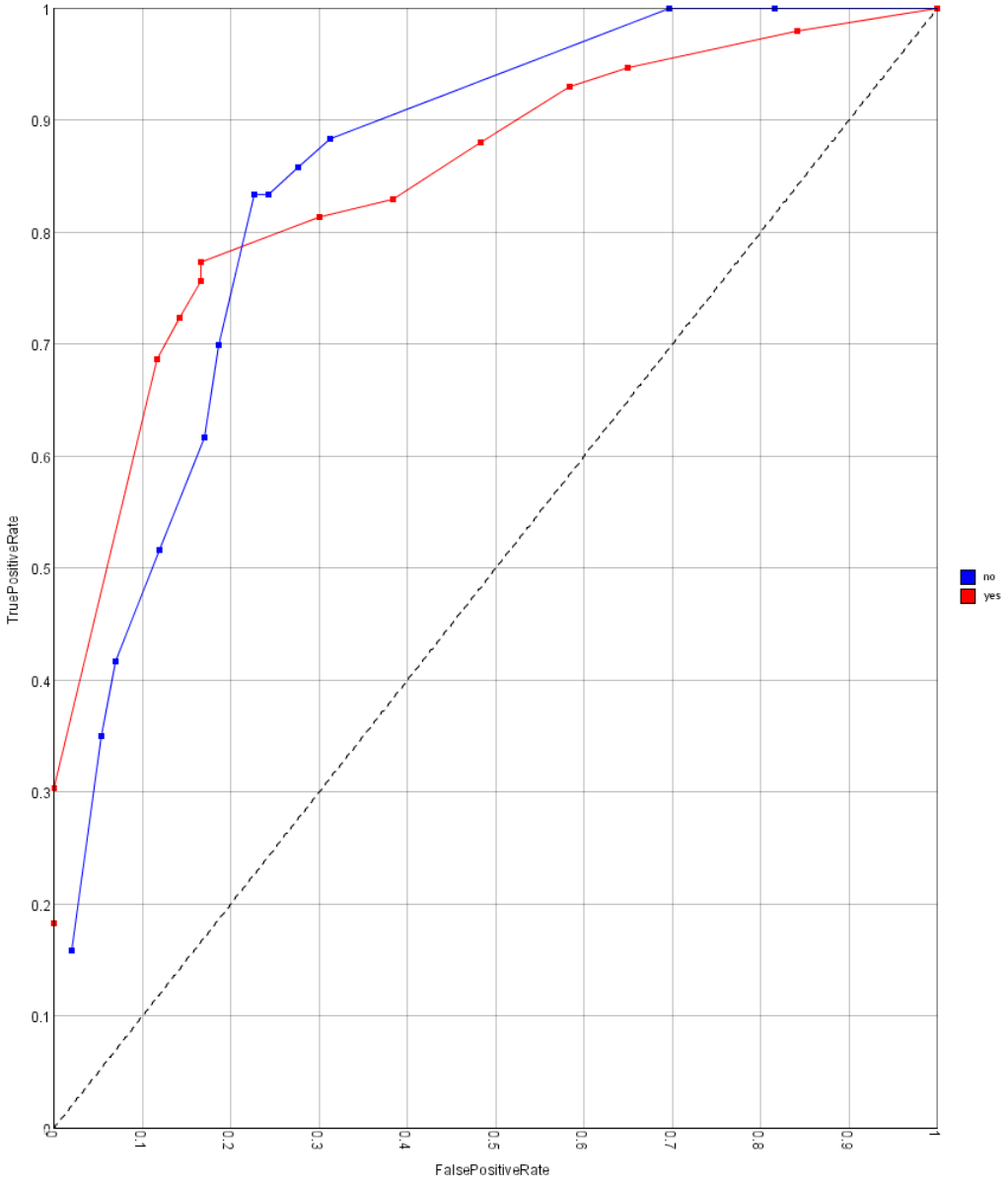


Figure 3.16: HOSP_EXCYT_95: Receiver operating characteristics (ROC) curve for prediction of hospitalization calculated on 95 patients excluding cytokine data.

In a second step, a tree using cytokine as well as clinical data was calculated. A small change in classifier performance (data not shown) could be observed but the splitting criteria remained the same. The tree had an overall performance of 0.86 for hospitalized (95%CI: 0.78, 0.93) as well as non-hospitalized (95%CI: 0.77, 0.94) cases with a higher sensitivity (83%) and a lower specificity (68%). The overall profit of the tree accounted for 0.542.

3.1.2.3 Prediction of Hospitalization based on Cytokine Data

The little improvement with the splitting criteria remaining the same, suggested a minor effect of the cytokine data that was worthwhile investigating. For this purpose, a second classification tree (pruning confidence was set to 25% and minimum cases was defined as 10) which was based only on cytokine data was constructed which turned out to be rather difficult. Nevertheless, a tree (HOSP_CYTO_95) (Figure 3.17) including IP_10_1<=1697.9 (OR: 7.09; 95%CI: 4.56, 9.61), IL_1_1<=7.06 (OR: 8.82; 95%CI: 0.07, 17.58), TNF_1>1.68 (OR: 5.78; 95%CI: -0.08, 11.63) and IL_12_2>15.9 (OR: 7.88; 95%CI: 0.75, 15.00)[10] as splitting criteria (Table 3.25; Table 3.26) had the best overall performance and an AUC of 0.69 for hospitalized (95%CI: 0.58, 0.79) as well as non-hospitalized (95%CI: 0.57, 0.80) cases. It showed an error rate of 26.1% with a sensitivity of 81% plus a specificity of 63% and the model accounted for an average profit of 0.478 (Table 3.27; Figure 3.18).

[10] IP_10=*interferon-inducible protein 10*; IL_1=*interleukin-10*; TNF=*tumor necrosis factor α*; IL_12=*interleukin-12*; 1=*1st visit data*; 2=*2nd visit data;* OR=*odds ratio*; CI=*confidence interval.*

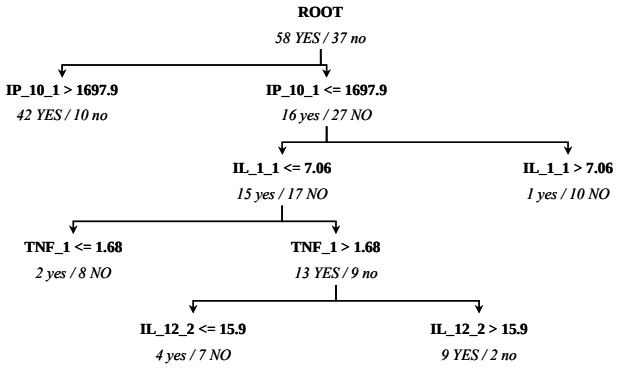


Figure 3.17: HOSP_CYTO_95: Decision tree for hospitalization calculated on 95 patients only including cytokine data. IP_10=*interferon-inducible protein 10*; IL_1=*interleukin-10*; TNF=*tumor necrosis factor α*; IL_12=*interleukin-12*; 1=*1st visit data*; 2=*2nd visit data.*

Table 3.25: HOSP_CYTO_95: Decision tree for hospitalization calculated on 95 patients only including cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. IP_10=*interferon-inducible protein 10*; IL_1=*interleukin-10*; TNF=*tumor necrosis factor α*; IL_12=*interleukin-12*; 1=*1st visit data*; 2=*2nd visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IP_10_1 [pg/ml] *Cut-off* value > 1697.9	2.17	7.09	4.56, 9.61	< 0.001
IL_1_1 [pg/ml] *Cut-off* value <= 7.06	1.45	2.30	-0.42, 5.02	0.098
TNF_1 [pg/ml] *Cut-off* value > 1.68	1.18	1.52	-0.85, 3.88	0.343
IL_12_2 [pg/ml] *Cut-off* value > 15.9	1.01	1.02	-1.37, 3.41	0.968

Table 3.26: HOSP_CYTO_95: Decision tree for hospitalization calculated on 95 patients only including cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. IP_10=*interferon-inducible protein 10*; IL_1=*interleukin-10*; TNF=*tumor necrosis factor α*; IL_12=*interleukin-12*; 1=*1st visit data*; 2=*2nd visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval.*

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IP_10_1 [pg/ml] *Cut-off* value > 1697.9	2.17	7.09	4.56, 9.61	< 0.001
IL_1_1 [pg/ml] *Cut-off* value <= 7.06	5.16	8.82	0.07, 17.58	0.033
TNF_1 [pg/ml] *Cut-off* value > 1.68	2.95	5.78	-0.08, 11.63	0.060
IL_12_2 [pg/ml] *Cut-off* value > 15.9	2.25	7.88	0.75, 15.00	0.080

Table 3.27: HOSP_CYTO_95: Summary of K-fold (k=10) cross validation for prediction of hospitalization based on 95 patients only including cytokine data.

Overall Evaluation	Value (n=95)	
Total misclassifications	25.0	
Overall error rate	26.111%	
SE of error rate	10.776	
Average profit	0.478	
SE of profit	0.216	
AUC no	0.6853	95%CI: 0.57, 0.80
AUC yes	0.6867	95%CI: 0.58, 0.79

Confusion Matrix			
		Predicted Class	
		no	yes
Actual Class	no	23 (63%)	14 (37%)
	yes	11 (19%)	47 (81%)

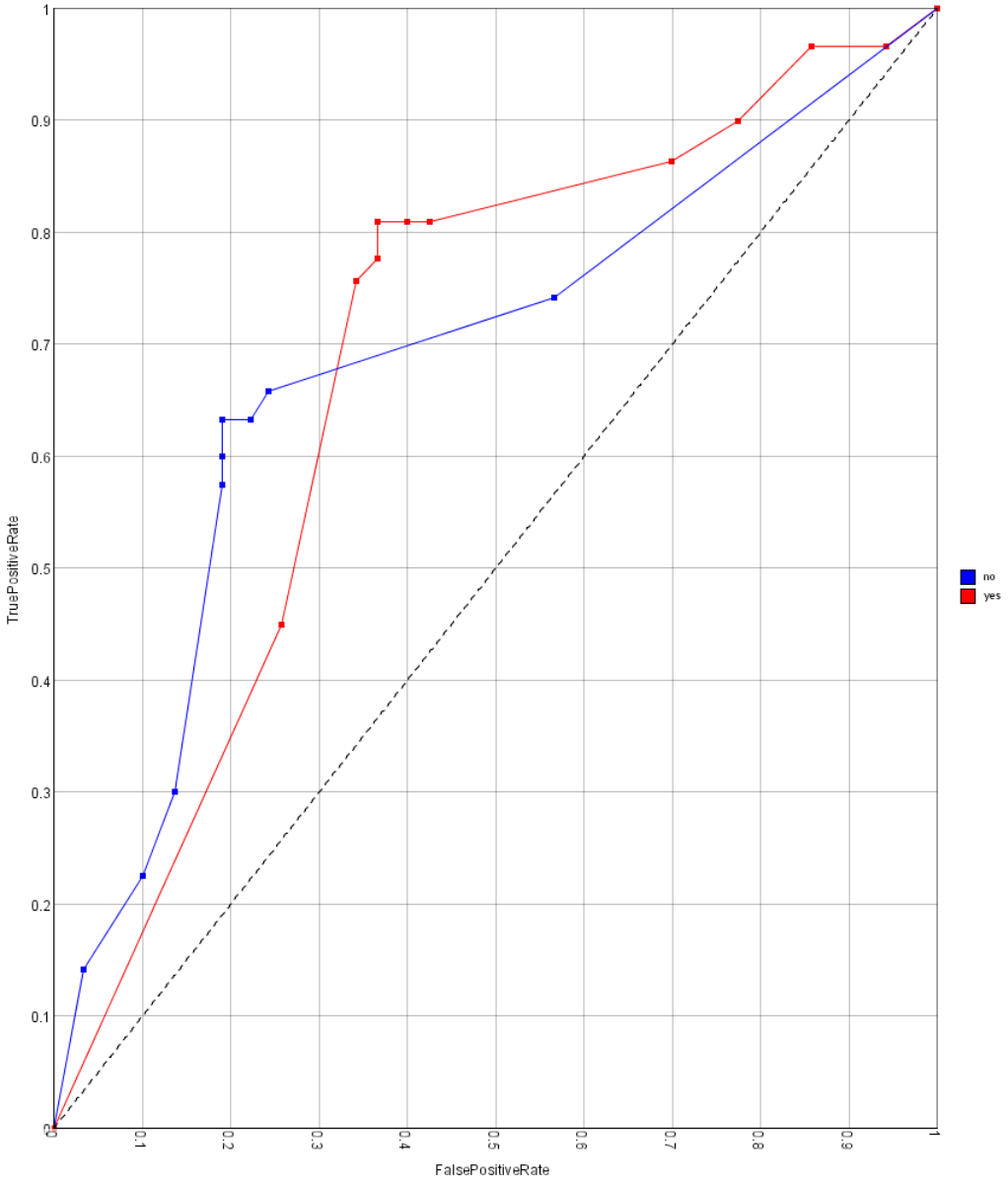


Figure 3.18: HOSP_CYTO_95: Receiver operating characteristics (ROC) curve for prediction of hospitalization calculated on 95 patients only including cytokine data.

3.1.2.4 Using a Platelet Count <=50,000/mm^3 as a Marker of Severity

Although we were able to classify the 133 patients into hospitalized and non-hospitalized groups, we wanted to further investigate disease severity because of several reasons. First, parameters chosen by the calculated classifier included data collected on 2nd visit which corresponded to the day of hospitalization and the calculated trees excluding the cytokine data included parameters such as platelet count and body temperature which were probably used by the treating doctors to decide for hospitalization. Second, we aimed at having early markers of severity that are represented by data from the 1st collection. Both trees excluding the cytokine data only included either the absolute numbers of lymphocytes or the pulse rate as first visit data and strikingly, excluding the 2nd visit platelet count from the analysis was not at all predictive. On the other hand, the classification tree based only on cytokine data included three first visit parameters but did not result in a reliable classification meaning that the two groups were overlapping in terms of cytokines. Third, hospitalization is biased due to its dependency on the treating doctor as well as on the patient and therefore, it is not a good, indicative categorization for severity. This was underlined by the fact that the calculated trees showed different splitting and threshold criteria pointing towards instability of classifier performance. Fourth, we could naturally assume that hospitalized cases felt sicker but on the other hand, the chance to become hospitalized might be higher as soon as the patient did not feel well. Hence, hospitalization might also be considered as a safety measure. Fifth, the ratio of hospitalized to non-hospitalized cases was 1.46 and probably did not represent the real severity situation in the 2005 Singaporean dengue outbreak.

To address the above mentioned concerns arisen from hospitalization and to improve severity prediction, we decided to take the platelet count as a marker of severity. Therefore, we first analyzed the 133 dengue patients regarding their platelet count during the course of infection. From the 71 hospitalized patients that were submitted to general hospitals, blood count as well as other clinical parameters were daily measured and the average platelet count at each day was plotted regarding length of illness (Figure 3.19). The plot indicated a nadir of platelet counts at either day five (60,500/mm^3), day six (54,400/mm^3) or day seven (55,100/mm^3) of illness. For that reason, we defined day five, six and seven as the time points for more severe disease and a platelet count <=50,000/mm^3 was chosen as a severity marker. Using a threshold of <=50,000/mm^3 as a severity marker was additionally supported by evidence of several studies (Balmaseda et al., 2006; Hammond et al., 2005; Malavige et al., 2006) showing clear associations with more severe disease manifestations. Hence, patients having a platelet count <=50,000/mm^3 on either day five, six or seven were classified as more severe cases and named as 'low'. Eight hospitalized patients were excluded from the analysis due to either no clinical information during their hospital stay (seven patients) or due to no data on days five, six and seven (one patient). This resulted in a total of 125 patients that were included into severity modeling using the approach of decision trees. The hospitalized patients showing a platelet count > 50,000/mm^3 as well as the non-hospitalized cases assumed to experience mild disease were categorized into the mild group named as 'high'.

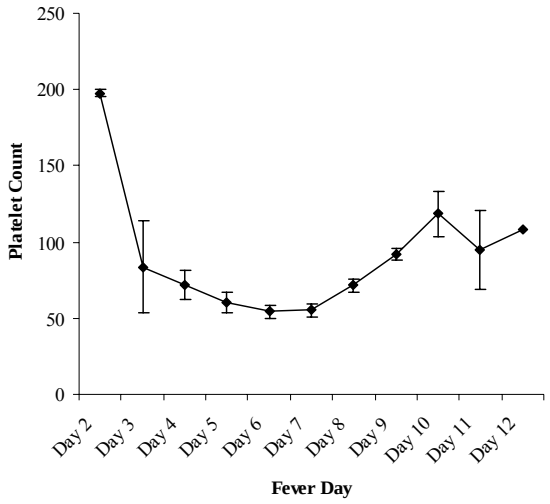


Figure 3.19: Platelet Counts [*1000/mm^3] plotted as a function of days of illness. In total, 71 hospitalized patients were included. Average for each time point was calculated with available information (missing values were excluded); error bars representing +/- standard error.

The two severity groups along with hospitalization as a grouping criterium were compared regarding their clinical parameters that were measured on all of the three visits (Table 3.28; Table 3.29; Table 3.30). There were 42 (33.6%) of the patients in the low group and 83 (66.4%) of the patients belonged to the high group. The overall male-to-female ratio was 1.6 (1.31 in the high group and 2.5 in the low group). Multivariate analysis performed on first visit data revealed higher viral load (OR: 0.86; 95%CI: 0.75, 0.98), increased body temperature (OR: 1.84; 95%CI: 1.00, 3.40), decreased platelet count (OR: 0.98; 95%CI: 0.97, 0.99) and a secondary infection (OR: 4.49; 95%CI: 1.70, 12.40) as genuine risk factors for the development of low platelet counts whereas a higher body temperature (OR: 2.03; 95%CI: 1.05, 3.89) as well as a higher mean corpuscular haemoglobin concentration (MCHC) (OR: 1.74; 95%CI: 1.01, 3.00) represented real risks at the second visit (Table 3.32). On the other hand, using hospitalization as a severity indicator, resulted in a slightly different but still overlapping picture in terms of possible risk factors. Higher viral load (OR: 0.86;

95%CI: 0.77, 0.97), a decreased platelet count (OR: 0.98; 95%CI: 0.97, 0.99) and a smaller absolute number of lymphocytes (OR: 0.06; 95%CI: 0.01, 0.48) represented real risks for hospitalization on the first visit while a higher body temperature (OR: 10.21; 95%CI: 2.15, 48.63) along with low platelets (OR: 0.96; 95%CI: 0.93, 0.98) were identified as real significant differences between hospitalized and non-hospitalized patients on the second visit (Table 3.31). It is intriguing that the cytokine patterns were strikingly different between hospitalization and platelet group classification with regard to univariate analyses (Table 3.33; Table 3.35; Table 3.38). We observed IP-10, I-TAC, IL-10 and IL-8 as significant differences on the first visit between the two platelet group whereas IP-10, IL-10 as well as IL-8 represented the main differences on the second visit. On the other hand, hospitalization as a categorical variable was able to only detect IP-10 on the first and IP-10 along with IL-10 on the second visit. However, logistic regression did only identify IP_10 as a independent risk between the two platelet groups on the second visit (Table 3.34; Table 3.36; Table 3.37).

Table 3.28: Mean for normally [a] and median for non-normally [b] distributed clinical data collected on the first visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters ($p<0.05$) which were first log-transformed [c]. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. *Chi-square* test [d] was used to assess for significance between secondary infections; absolute counts for negatives (neg) / positives (pos) are indicated below [e]. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate (1st visit)	Dengue positive cases (n=133) * Hospitalization					Dengue positive cases (n=125) * Platelet Group				
	yes	SD_{yes}	no	SD_{no}	p value	low	SD_{low}	high	SD_{hi}	p value
CT_1ST_COLLECTION [b]	16.84		20.77		0.001	16.50		19.56		0.002
TEMP_1 [a]	38.48	0.873	38.24	0.800	0.112	38.60	0.91	38.23	0.801	0.029
PULSERATE_1 [a]	92.96	15.04	93.93	16.27	0.730	94.21	15.91	92.17	15.31	0.475
SYSTOLICBP_1 a	114.5	15.92	112.9	13.88	0.548	116.6	15.81	113.0	14.36	0.215
DIASTOLIC1 [a]	70.90	11.72	73.63	9.444	0.141	71.64	11.64	72.34	10.38	0.745
WBC_1 [c]	0.508	0.215	0.582	0.178	0.034	0.509	0.241	0.551	0.186	0.319
RBC_1 [a]	4.993	0.649	4.900	0.746	0.457	5.051	0.678	4.937	0.707	0.382
HGB_1 [a]	14.75	1.879	14.25	2.209	0.175	15.16	1.957	14.35	2.018	0.032
HCT_1 [a]	43.41	5.256	42.38	6.275	0.324	44.38	5.660	42.59	5.677	0.100
MCV_1 [b]	87.80		87.00		0.561	87.70		87.45		0.736
MCH_1 [b]	30.00		29.45		0.206	30.10		29.80		0.163
MCHC_1 [b]	33.90		33.55		0.031	34.05		33.65		0.009
PLT_1 [b]	133.0		188.5		0.001	112.5		168.5		0.001
LYMPH_PCT_1 [b]	13.75		16.35		0.363	13.20		16.20		0.266
MXD_PCT_1 [b]	5.700		5.350		0.865	5.100		5.500		0.820
NEUT_PCT_1 [b]	82.30		77.95		0.326	82.40		78.70		0.476
LYMPH_NO_1 [b]	0.450		0.550		0.001	0.400		0.500		0.005
MXD_NO_1 [b]	0.200		0.200		0.754	0.200		0.200		0.642
NEUT_NO_1 [c]	0.411	0.254	0.477	0.217	0.135	0.406	0.275	0.455	0.225	0.345
RDW_CV_1 [b]	13.30		13.60		0.183	13.25		13.45		0.302
RDW_SD_1 [a]	42.16	2.870	42.82	2.605	0.172	42.51	2.713	42.41	2.732	0.854
PDW_1 [c]	1.086	0.072	1.062		0.041	1.091	0.079	1.070	0.063	0.141
MPV_1 [c]	1.031	0.047	10.47		0.112	1.033	0.053	1.023	0.041	0.282
P_LCR_PCT_1 [c]	1.467	0.144	1.431		0.123	1.474	0.162	1.443	0.121	0.284
DV_IG_G_1 [d, e]	40 [neg]	39 [pos]	36 [neg]	18 [pos]	0.067	15 [neg]	27 [pos]	55 [neg]	28 [pos]	0.001

Table 3.29: Mean for normally [a] and median for non-normally [b] distributed clinical data collected on the second visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters ($p<0.05$) which were fist log-transformed [c]. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate (2nd visit)	Dengue positive cases (n=133) * Hospitalization					Dengue positive cases (n=125) * Platelet Group				
	yes	SD_{yes}	no	SD_{no}	p_{value}	low	SD_{low}	high	SD_{hi}	p_{value}
CT_2ND_COLLECTION [b]	28.16		29.13		0.169	28.47		28.29		0.609
TEMP_2 [b]	37.20		36.70		0.001	37.30		36.90		0.021
PULSERATE_2 [a]	80.84	14.00	76.95	13.36	0.124	78.98	13.77	79.08	13.61	0.969
SYSTOLICBP_2 [a]	108.3	16.56	109.0	15.93	0.800	111.4	17.17	108.5	15.25	0.374
DIASTOLIC2 [a]	70.90	10.73	73.10	10.96	0.376	72.10	10.14	72.57	11.25	0.819
WBC_2 [c]	0.398	0.222	0.523	0.202	0.002	0.423	0.239	0.472	0.214	0.271
RBC_2 [b]	5.130		4.870		0.156	5.125		4.940		0.471
HGB_2 [b]	15.10		14.30		0.058	15.25		14.40		0.064
HCT_2 [b]	44.10		42.10		0.138	44.30		43.05		0.174
MCV_2 [b]	87.20		87.00		0.883	86.65		87.35		0.993
MCH_2 [b]	29.90		29.45		0.039	29.85		29.65		0.087
MCHC_2 [b]	34.30		33.70		0.002	34.45		33.80		0.001
PLT_2 [b]	67.00		149.5		0.001	44.50		126.0		0.001
LYMPH_PCT_2 [a]	30.18	10.43	33.23	10.24	0.113	29.99	10.41	32.27	10.20	0.263
MXD_PCT_2 [b]	7.400		9.100		0.726	6.500		8.900		0.839
NEUT_PCT_2 [b]	65.20		54.25		0.107	65.90		59.20		0.142
LYMPH_NO_2 [c]	-0.13	0.287	0.020	0.271	0.003	-0.12	0.317	-0.04	0.273	0.150
MXD_NO_2 [b]	0.200		0.350		0.054	0.200		0.300		0.107
NEUT_NO_2 [b]	1.300		1.800		0.049	1.350		1.700		0.109
RDW_CV_2 [b]	13.30		13.60		0.308	13.30		13.50		0.434
RDW_SD_2 [a]	41.96	2.780	42.53	2.636	0.245	42.31	2.532	42.12	2.702	0.705
PDW_2 [c]	1.127	0.079	1.085	0.064	0.003	1.130	0.090	1.099	0.069	0.114
MPV_2 [c]	1.055	0.045	1.030	0.041	0.003	1.060	0.052	1.038	0.041	0.048
P_LCR_PCT_2 [c]	1.545	0.114	1.471	0.119	0.002	1.558	0.128	1.495	0.116	0.030

Table 3.30: Mean for normally [a] and median for non-normally [b] distributed clinical data collected on the third visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters ($p<0.05$) which were first log-transformed [c]. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate (3rd visit)	Dengue positive cases (n=133) * Hospitalization					Dengue positive cases (n=125) * Platelet Group				
	yes	SD_{yes}	no	SD_{no}	p_{value}	low	SD_{low}	high	SD_{hi}	p_{value}
WBC_3 [b]	6.100		6.200		0.532	6.200		6.100		0.788
RBC_3 [a]	4.603	0.660	4.567	0.572	0.764	4.523	0.632	4.611	0.569	0.502
HGB_3 [c]	1.131	0.065	1.117	0.063	0.275	1.132	0.071	1.123	0.059	0.523
HCT_3 [b]	40.70		38.90		0.330	40.50		39.70		0.648
MCV_3 [b]	88.80		88.35		0.514	89.10		88.35		0.089
MCH_3 [b]	30.05		29.70		0.080	30.10		29.70		0.017
MCHC_3 [a]	33.64	0.995	33.15	1.030	0.013	33.80	0.918	33.32	1.073	0.020
PLT_3 [b]	318.0		318.5		0.508	314.0		319.5		0.879
LYMPH_PCT_3 [a]	33.64	7.677	34.43	7.184	0.582	35.76	7.684	33.14	7.496	0.110
MXD_PCT_3 [a]	9.320	3.949	8.912	3.751	0.632	9.448	4.592	8.845	3.536	0.553
NEUT_PCT_3 [a]	57.29	9.091	56.24	7.942	0.575	54.70	9.269	58.08	8.341	0.118
LYMPH_NO_3 [c]	0.304	0.142	0.329	0.103	0.289	0.339	0.147	0.311	0.109	0.335
MXD_NO_3 [b]	0.500		0.600		0.914	0.500		0.600		0.743
NEUT_NO_3 [a]	3.725	1.190	3.721	1.015	0.984	3.526	1.179	3.894	1.051	0.177
RDW_CV_3 [b]	13.55		13.80		0.445	13.60		13.65		0.918
RDW_SD_3 [c]	1.635	0.031	1.636	0.025	0.912	1.643	0.034	1.631	0.025	0.080
PDW_3 [b]	10.70		10.70		0.972	11.10		10.60		0.557
MPV_3 [b]	9.800		9.900		0.723	10.10		9.700		0.218
P_LCR_PCT_3 [c]	1.381	0.147	1.372	0.121	0.721	1.398	0.153	1.370	0.134	0.367

Table 3.31: Logistic regression results for the assessment of genuine risk factors of significant group differences which were found by univariate analysis on 1st visit as well as 2nd visit data. Cases with missing values were excluded and complete algorithm was used; table showing the risks for hospitalization (yes/no). Shading indicates a significant *p* value < 0.05.

Covariate (1st visit)	OR	95% CI (OR)	*p* value (n=125)
CT_1ST_COLLECTION	0.86	0.77, 0.97	0.012
MCHC_1	1.12	0.64, 1.97	0.699
PLT_1	0.98	0.97, 0.99	0.001
WBC_1	1.21	0.88, 1.66	0.239
LYMPH_NO_1	0.06	0.01, 0.48	0.008
PDW_1	1.06	0.82, 1.37	0.633

Covariate (2nd visit)	OR	95% CI (OR)	*p* value (n=80)
TEMP_2	10.21	2.15, 48.63	0.004
WBC_2	0.10	0.01, 1.86	0.124
MCH_2	1.02	0.57, 1.81	0.954
MCHC_2	2.36	0.89, 6.24	0.085
PLT_2	0.96	0.93, 0.98	0.001
NEUT_NO_2	4.22	0.24, 74.19	0.325
MPV_2	26.84	0.39, 1867.08	0.129
PDW_2	2.22	0.83, 5.91	0.111
LYMPH_NO_2	35.66	0.48, 2645.61	0.104
P_LCR_PCT_2	0.59	0.32, 1.12	0.105

Table 3.32: Logistic regression results for the assessment of genuine risk factors of significant group differences which were found by univariate analysis on 1st visit as well as 2nd visit data (PLT_2 was excluded). Cases with missing values were excluded and complete algorithm was used; table showing the risks for platelet group (low/high). Shading indicates a significant *p* value < 0.05.

Covariate (1st visit)	OR	95% CI (OR)	*p* value (n=124)
CT_1ST_COLLECTION	0.86	0.75, 0.98	0.027
TEMP_1	1.84	1.00, 3.40	0.050
HGB_1	0.96	0.70, 1.28	0.784
MCHC_1	1.54	0.85, 2.79	0.158
PLT_1	0.98	0.97, 0.99	< 0.001
LYMPH_NO_1	0.37	0.04, 3.78	0.398
DV_IG_G_1	4.49	1.70, 12.40	0.003

Table 3.32 (continued): Logistic regression results for the assessment of genuine risk factors of significant group differences which were found by univariate analysis on 1st visit as well as 2nd visit data (PLT_2 was excluded). Cases with missing values were excluded and complete algorithm was used; table showing the risks for platelet group (low/high). Shading indicates a significant *p* value < 0.05.

Covariate (2nd visit)	OR	95% CI (OR)	*p* value (n=100)
TEMP_2	2.03	1.05, 3.89	0.036
MCHC_2	1.74	1.01, 3.00	0.048
MPV_2	0.70	0.09, 5.38	0.728
P_LCR_PCT_2	1.11	0.87, 1.43	0.399

Table 3.33: Mean for normally [a] and median for non-normally [b] distributed cytokine data collected on the first visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters ($p<0.05$) which were first log-transformed [c]. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate (1st visit)	Dengue positive cases (n=95) * Hospitalization					Dengue positive cases (n=89) * Platelet Group				
	yes	SD_{yes}	no	SD_{no}	p_{value}	low	SD_{low}	high	SD_{hi}	p_{value}
IP_10_1 [b]	2016		1508		0.001	2298		1614		0.001
I_TAC_1 [b]	1226		1096		0.576	1456		1033		0.004
IFN_ALPHA_1 [b]	859.1		798.4		0.329	855.8		798.5		0.149
GM_CSF_1 [b]	4.365		6.32		0.196	4.57		6.18		0.313
IFN_GAMMA_1 [b]	13.30		17.10		0.213	14.00		14.85		0.865
IL_1_1 [b]	5.30		5.30		0.430	5.30		5.23		0.652
IL_10_1 [b]	8.13		5.61		0.427	9.37		5.00		0.039
IL_12_1 [b]	21.00		23.0		0.746	20.00		22.45		0.919
IL_2_1 [b]	10.75		11.50		0.164	10.60		11.30		0.625
IL_4_1 [b]	4.24		5.56		0.792	4.04		5.65		0.304
IL_6_1 [b]	18.15		23.00		0.751	14.90		22.55		0.990
IL_8_1 [b]	3.44		3.26		0.587	4.55		2.61		0.001
TNF_1 [b]	1.80		1.87		0.765	1.87		1.80		0.746

Table 3.34: Logistic regression results for the assessment of genuine risk factors of significant group differences which were found by univariate analysis on 1st visit. Cases with missing values were excluded and complete algorithm was used; table showing the risks for platelet group (low/high). Shading indicates a significant *p* value < 0.05.

Covariate (1st visit)	OR	95% CI (OR)	*p* value (n=56)
IP_10_1	1.001	1.000, 1.002	0.099
I_TAC_1	1.000	1.000, 1.001	0.273
IL_10_1	1.012	0.986, 1.039	0.369
IL_8_1	1.151	0.885, 1.497	0.295

Table 3.35: Mean for normally [a] and median for non-normally [b] distributed cytokine data collected on the second visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters (p<0.05) which were first log-transformed [c]. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate (2nd visit)	Dengue positive cases (n=95) * Hospitalization					Dengue positive cases (n=89) * Platelet Group				
	yes	SD_{yes}	no	SD_{no}	p_{value}	low	SD_{low}	high	SD_{hi}	p_{value}
IP_10_2 [b]	1502		949.4		0.005	1925		998.1		0.001
I_TAC_2 [b]	1070		1184		0.903	1147		996.7		0.713
IFN_ALPHA_2 [b]	499.6		458.2		0.870	295.7		503.0		0.113
GM_CSF_2 [b]	4.97		12.65		0.057	4.15		8.75		0.150
IFN_GAMMA_2 [b]	10.75		12.80		0.932	11.70		11.80		0.604
IL_1_2 [b]	4.64		5.30		0.222	4.64		5.30		0.390
IL_10_2 [b]	25.25		8.51		0.021	33.50		9.07		0.046
IL_12_2 [b]	17.60		18.60		0.400	19.40		17.40		0.882
IL_2_2 [b]	7.43		12.09		0.175	6.98		10.40		0.777
IL_4_2 [b]	4.04		5.26		0.461	4.27		5.23		0.351
IL_6_2 [b]	14.55		20.60		0.862	15.80		20.30		0.608
IL_8_2 [b]	3.55		2.79		0.194	4.00		2.63		0.020
TNF_2 [b]	1.80		1.87		0.398	1.80		1.80		0.969

Table 3.36: Logistic regression results for the assessment of genuine risk factors of significant group differences which were found by univariate analysis on 2nd visit. Cases with missing values were excluded and complete algorithm was used; table showing the risks for hospitalization (yes/no). Shading indicates a significant *p* value < 0.05.

Covariate (2nd visit)	OR	95% CI (OR)	*p* value (n=87)
IP_10_2	1.001	1.000, 1.001	0.084
IL_10_2	1.017	0.998, 1.037	0.085

Table 3.37: Logistic regression results for the assessment of genuine risk factors of significant group differences which were found by univariate analysis on 2nd visit. Cases with missing values were excluded and complete algorithm was used; table showing the risks for platelet group (low/high). Shading indicates a significant *p* value < 0.05.

Covariate (2nd visit)	OR	95% CI (OR)	*p* value (n=81)
IP_10_2	1.001	1.000, 1.001	0.031
IL_10_2	1.010	0.997, 1.023	0.143
IL_8_2	1.155	0.835, 1.597	0.384

Table 3.38: Mean for normally [a] and median for non-normally [b] distributed cytokine data collected on the third visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters (p<0.05) which were first log-transformed [c]. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate (3rd visit)	Dengue positive cases (n=95) * Hospitalization					Dengue positive cases (n=89) * Platelet Group				
	yes	SD_{yes}	no	SD_{no}	p_{value}	low	SD_{low}	high	SD_{hi}	p_{value}
IP_10_3 [b]	153.9		145.3		0.753	198.4		143.6		0.062
I_TAC_3 [b]	NA		NA		NA	NA		NA		NA
IFN_ALPHA_3 [b]	NA		NA		NA	NA		NA		NA
GM_CSF_3 [b]	5.20		8.67		0.360	5.19		7.64		0.336
IFN_GAMMA_3 [b]	8.86		8.78		0.483	9.88		9.15		0.874
IL_1_3 [b]	5.30		5.32		0.834	5.30		5.31		0.621
IL_10_3 [b]	2.22		3.08		0.626	2.07		3.06		0.648
IL_12_3 [b]	17.80		18.47		0.627	17.30		18.49		0.547
IL_2_3 [b]	9.39		12.40		0.181	9.41		12.10		0.458
IL_4_3 [b]	3.56		4.44		0.104	2.98		4.48		0.007
IL_6_3 [b]	13.60		16.70		0.309	11.35		18.45		0.255
IL_8_3 [b]	2.33		2.27		0.724	2.33		2.15		0.210
TNF_3 [b]	1.80		1.80		0.727	1.80		1.80		0.812

3.1.2.5 Severity Prediction based on Clinical Data

A first tree (SEVERE_TOTAL_125) (Figure 3.20) based on first visit data was calculated using inputs of 125 patients and the pruning confidence was set to 25% with "minimal cases" defined as 16. Missing values were treated as null and the tree was evaluated by k-fold cross validation (k=10). Excluding TEMP_1, this resulted in a correct classification of 103 (82.4%) cases with a sensitivity of 81% and a specificity of 83% (Table 3.41; Figure 3.21). The AUC of the low ROC curve was 0.83 (95%: 0.74, 0.91) whereas the AUC of the high classification averaged 0.83 (95%CI: 0.76, 0.90). The threshold for the optimal classifier was set to an average profit of 0.646. The tree took PLT_1 <= 108 (OR: 24.24; 95%CI: 20.57, 27.92) as a first decision node and patients classified under this threshold were mainly cases belonging to the low group. Patients that had higher platelet counts at the first decision node were further classified using CT_1ST_COLLECTION <= 20.9 (OR: 16.29; 95%CI: 8.53, 24.06) and DV_IG_G_1 = 'positive' (OR: 9.87; 95%CI: 6.65, 13.09) as splitting criteria [11] (Table 3.39; Table 3.40).

[11] PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval.*

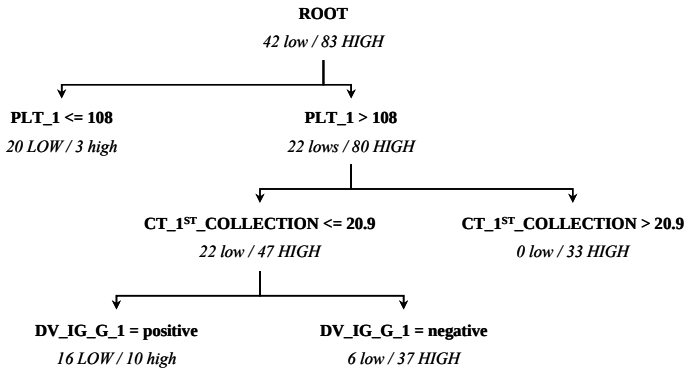


Figure 3.20: SEVERE_TOTAL_125: Decision tree for severity prediction calculated on 125 patients excluding cytokine data. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data.*

Table 3.39: SEVERE_TOTAL_125: Decision tree for severity prediction calculated on 125 patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	4.03	24.24	20.57, 27.92	< 0.001
CT_1ST_COLLECTION *Cut-off* value <= 20.9	3.36	5.13	2.33, 7.94	0.001
DV_IG_G_1 *Cut-off* value = positive	2.29	3.54	1.36, 5.71	0.002

Table 3.40: SEVERE_TOTAL_125: Decision tree for severity prediction calculated on 125 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value $^{+1}$. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	4.03	24.24	20.57, 27.92	< 0.001
CT_1ST_COLLECTION $^{+1}$ *Cut-off* value <= 20.9	11.34	16.29	8.53, 24.06	< 0.001
DV_IG_G_1 *Cut-off* value = positive	4.41	9.87	6.65, 13.09	< 0.001

Table 3.41: SEVERE_TOTAL_125: Summary of K-fold (k=10) cross validation for severity prediction based on 125 patients excluding cytokine data.

Overall Evaluation	Value (n=125)	
Total misclassifications	22.0	
Overall error rate	17.692%	
SE of error rate	9.369	
Average profit	0.646	
SE of profit	0.187	
AUC high	0.8264	95%CI: 0.76, 0.90
AUC low	0.8262	95%CI: 0.74, 0.91

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	69 (83%)	14 (17%)
	low	8 (19%)	34 (81%)

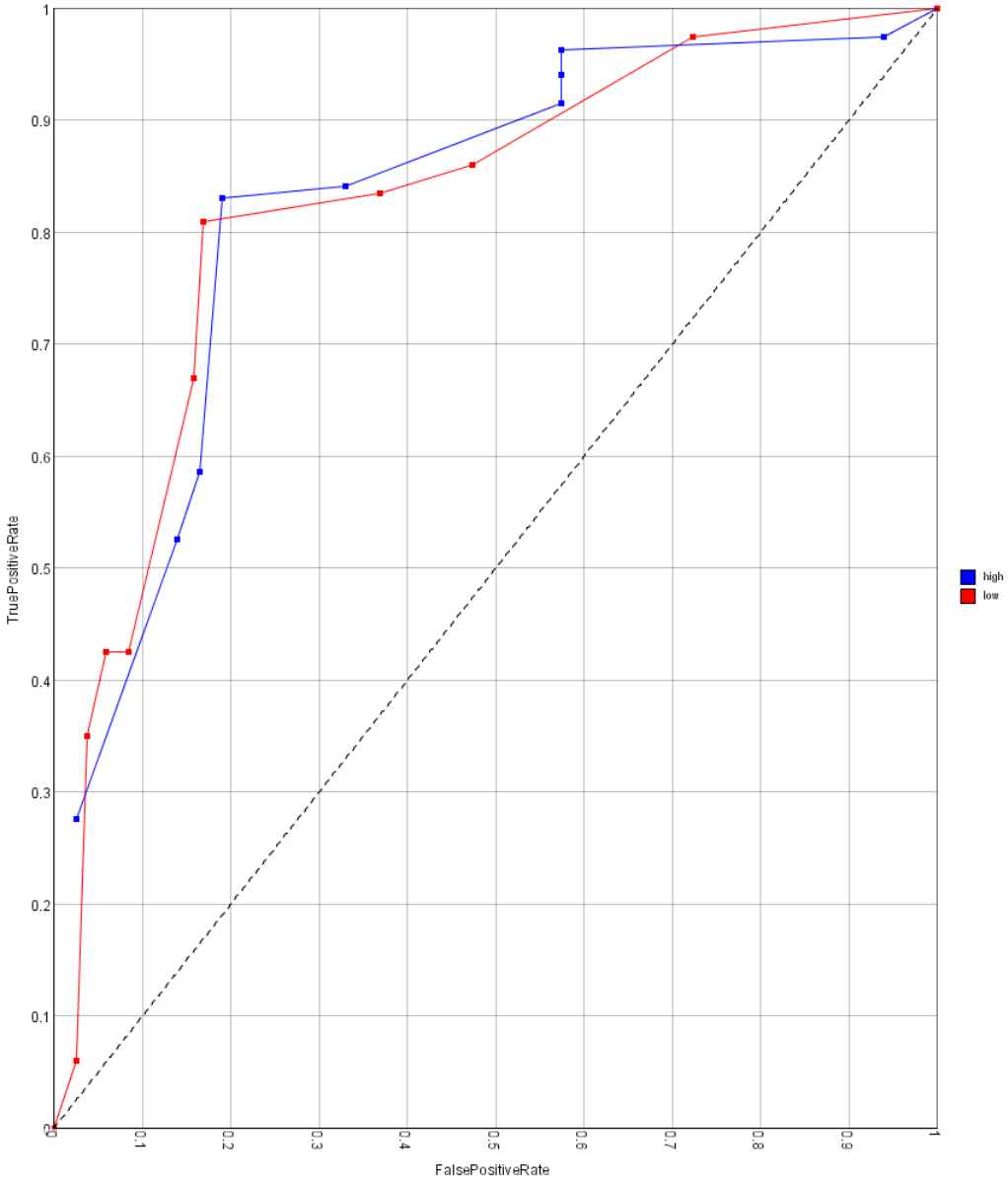


Figure 3.21: SEVERE_TOTAL_125: Receiver operating characteristics (ROC) curve for severity prediction calculated on 125 patients excluding cytokine data.

3.1.2.6 Severity Prediction based on Clinical Data and only using hospitalized Cases

The second tree (SEVHOSP_TOTAL_71) was constructed by only including first visit data from hospitalized cases leading to a dataset of 71 samples. Pruning confidence was set to the standard of 25% and a value of 4 was given to "minimum cases". The overall performance of the decision tree excluding the body temperature (TEMP_1) was slightly worse and resulted in a higher sensitivity (84%) but lower specificity (70%) (Table 3.44; Figure 3.23). The AUC of the ROC curve for more severe (95%CI: 0.69, 0.90) as well as mild cases (95%CI: 0.69, 0.91) was 0.80. The classifier correctly predicted 55 cases with an overall misclassification rate of 22.7%. Interestingly, the tree (Figure 3.22) still used PLT_1 <= 108 (OR: 25.45; 95%CI: 17.41, 33.50) as the main decision criterium to detect the low cases. Patients that were found to have higher platelet counts were further separated in secondary/primary infections. DV_IG__1 = 'positive' (OR: 5.63; 95%CI: 2.21, 9.05) patients were sub-grouped by first using CT_1st_COLLECTION <= 20.94 (OR: 29.75; 95%CI: 19.14, 40.36) as a criteria followed by MCHC_1 <= 34.5 (OR: 32.00; 95%CI: 18.00, 46.00) as a last decision node, both indicating classification into the low group. On the other hand, cases that were categorized as primary infections (DV_IG_G_1 = 'negative') were further grouped by using DIASTOLIC_BP_1 <= 67.0 (OR: 11.38; 95%CI: 1.67, 21.08) as a threshold for low cases. However, patients that had a higher diastolic blood pressure were separated into groups having either lower or higher systolic blood pressure and SYSTOLICBP_1<=119.0 (OR: 14.00; 95%CI: 1.67, 26.33) was an indicator for the more severe group low [12] (Table 3.42; Table 3.43).

[12] PLT=*platelet count*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* MCHC=*mean corpuscular hemoglobin concentration;* DIASTOLIC=*diastolic blood pressure*; SYSTOLICBP=*systolic blood pressure*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval*.

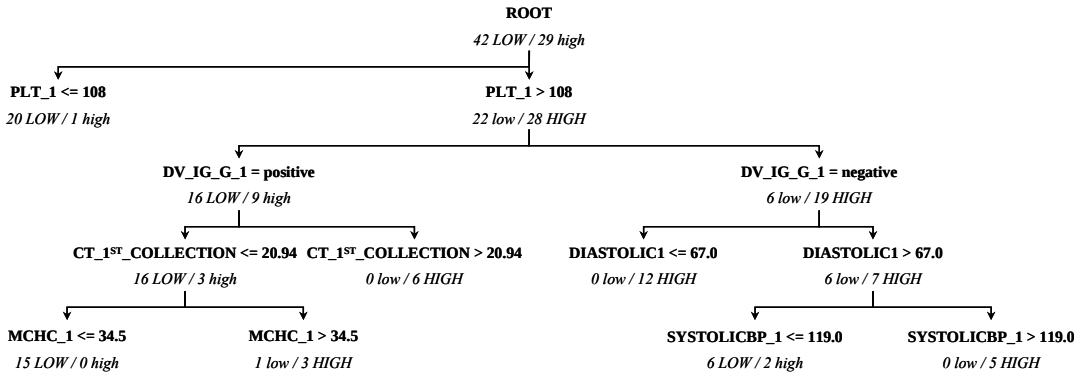


Figure 3.22: SEVHOSP_TOTAL_71: Decision tree for severity prediction calculated on 71 hospitalized patients excluding cytokine data. PLT=*platelet count*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* MCHC=*mean corpuscular hemoglobin concentration;* DIASTOLIC=*diastolic blood pressure*; SYSTOLICBP=*systolic blood pressure*; 1=*1st visit data.*

Table 3.42: SEVHOSP_TOTAL_71: Decision tree for severity prediction calculated on 71 hospitalized patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. PLT=*platelet count*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* MCHC=*mean corpuscular hemoglobin concentration;* DIASTOLIC=*diastolic blood pressure*; SYSTOLICBP=*systolic blood pressure*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	2.16	25.45	17.41, 33.50	< 0.001
DV_IG_G_1 *Cut-off* value = positive	1.65	3.42	0.72, 6.12	0.017
CT_1ST_COLLECTION *Cut-off* value <= 20.94	1.74	3.02	-078, 6.83	0.11
MCHC_1 *Cut-off* value <= 34.5	0.78	0.52	-2.71, 3.75	0.397
DIASTOLIC1 *Cut-off* value > 67.0	1.21	1.57	-1.1084, 4.26	0.451
SYSTOLICBP_1 *Cut-off* value <= 119.0	0.96	0.90	-1.74, 3.54	0.451

Table 3.43: SEVHOSP_TOTAL_71: Decision tree for severity prediction calculated on 71 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value [+1]. PLT=*platelet count*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* MCHC=*mean corpuscular hemoglobin concentration;* DIASTOLIC=*diastolic blood pressure*; SYSTOLICBP=*systolic blood pressure*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	2.16	25.45	17.41, 33.50	< 0.001
DV_IG_G_1 *Cut-off* value = positive	2.67	5.63	2.21, 9.05	0.01
CT_1ST_COLLECTION [+1] *Cut-off* value <= 20.94	6.48	29.75	19.14, 40.36	< 0.001
MCHC_1 [g/dl] [+1] *Cut-off* value <= 34.5	2.82	32.00	18.00, 46.00	< 0.001
DIASTOLIC1 [mmHg] [+1] *Cut-off* value > 67.0	6.53	11.38	1.67, 21.08	0.015
SYSTOLICBP_1 [mmHg] [+1] *Cut-off* value <= 119.0	1.28	14.00	1.67, 26.33	0.015

Table 3.44: SEVHOSP_TOTAL_71: Summary of K-fold (k=10) cross validation for severity prediction based on 71 hospitalized patients excluding cytokine data.

Overall Evaluation	Value (n=71)	
Total misclassifications	16.0	
Overall error rate	22.679%	
SE of error rate	19.380	
Average profit	0.546	
SE of profit	0.388	
AUC high	0.7991	95%CI: 0.69, 0.91
AUC low	0.7962	95%CI: 0.69, 0.90

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	20 (70%)	9 (30%)
	low	7 (17%)	35 (83%)

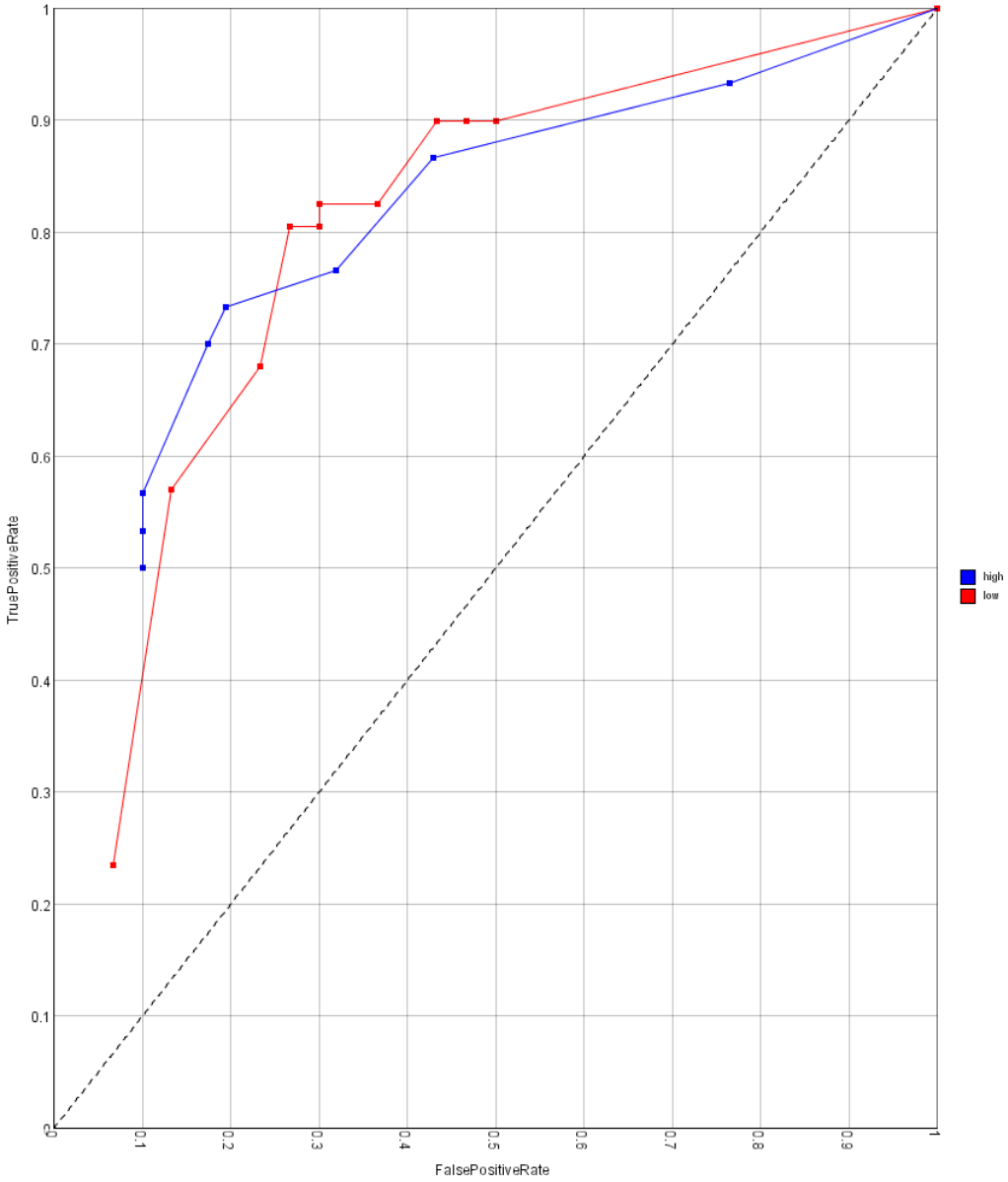


Figure 3.23: SEVHOSP_TOTAL_71: Receiver operating characteristics (ROC) curve for severity prediction calculated on 71 hospitalized patients excluding cytokine data.

3.1.2.7 Severity Prediction based on Cytokine and Clinical Data

Furthermore, we checked for the influence of cytokines on the classifier performance especially to elucidate the possibility of using specific cytokines as a marker for more severe infections. For this kind of analysis, we calculated a decision tree based on 89 dengue positive patients excluding the cytokine data to be able to compare the influence of cytokines on the overall classifier performance. The remaining 36 patients were excluded due to no cytokine data. The resulting tree (SEVERE_EXCYT_89) (Figure 3.24) (excluding TEMP_1), having a pruning confidence of 25% and "minimum cases" set to 10, was identical to the tree (Figure 3.20) calculated on the total of 133 dengue patients using PLT_1 <= 108 (OR: 40.53; 95%CI: 32.40, 48.65), CT_1ST_COLLECTION <= 20.9 (OR: 13.53; 95%CI: 5.55, 21.51) as well as DV_IG_G_1 = 'positive' (OR: 2.28; 95%CI: 6.04, 13.46)[13] as the classification criteria for more severe infections (Table 3.45; Table 3.46). The tree had a misclassification rate of 20.28% with a sensitivity of 77% and a specificity of 83% (Table 3.47; Figure 3.25). The AUC of the high classification was minimally higher (95%CI: 0.70, 0.88) than the AUC for the low classification (95%CI: 0.69, 0.89) indicating similar overall performance of 0.79. The average profit of the chosen probabilistic classifier was 0.594.

[13] TEMP=*body temperature*; PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval*.

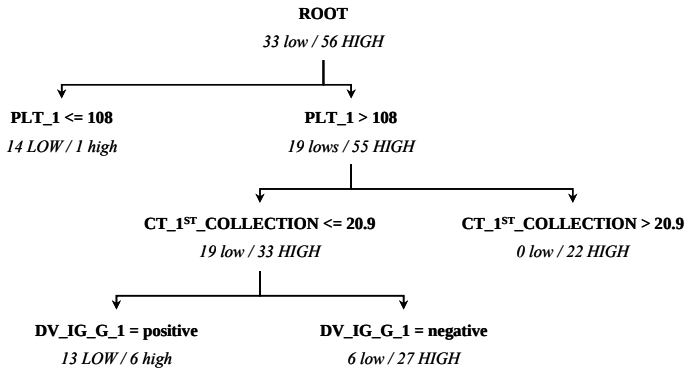


Figure 3.24: SEVERE_EXCYT_89: Decision tree for severity prediction calculated on 89 patients excluding cytokine data. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data.*

Table 3.45: SEVERE_EXCYT_89: Decision tree for severity prediction calculated on 89 patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	3.64	40.53	32.40, 48.65	< 0.001
CT_1ST_COLLECTION *Cut-off* value <= 20.9	6.05	10.80	6.20, 15.40	< 0.001
DV_IG_G_1 *Cut-off* value = positive	1.91	2.90	0.43, 5.30	0.026

Table 3.46: SEVERE_EXCYT_89: Decision tree for severity prediction calculated on 89 patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value [+1]. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	3.64	40.53	32.40, 48.65	< 0.001
CT_1ST_COLLECTION [+1] *Cut-off* value <= 20.9	8.89	13.53	5.55, 21.51	< 0.001
DV_IG_G_1 *Cut-off* value = positive	3.76	2.28	6.04, 13.46	0.001

Table 3.47: SEVERE_EXCYT_89: Summary of K-fold (k=10) cross validation for severity prediction based on 89 patients excluding cytokine data.

Overall Evaluation	Value (n=89)	
Total misclassifications	18.0	
Overall error rate	20.278%	
SE of error rate	8.886	
Average profit	0.594	
SE of profit	0.178	
AUC high	0.7933	95%CI: 0.70, 0.88
AUC low	0.7901	95%CI: 0.69, 0.89

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	46 (83%)	10 (17%)
	low	8 (23%)	25 (77%)

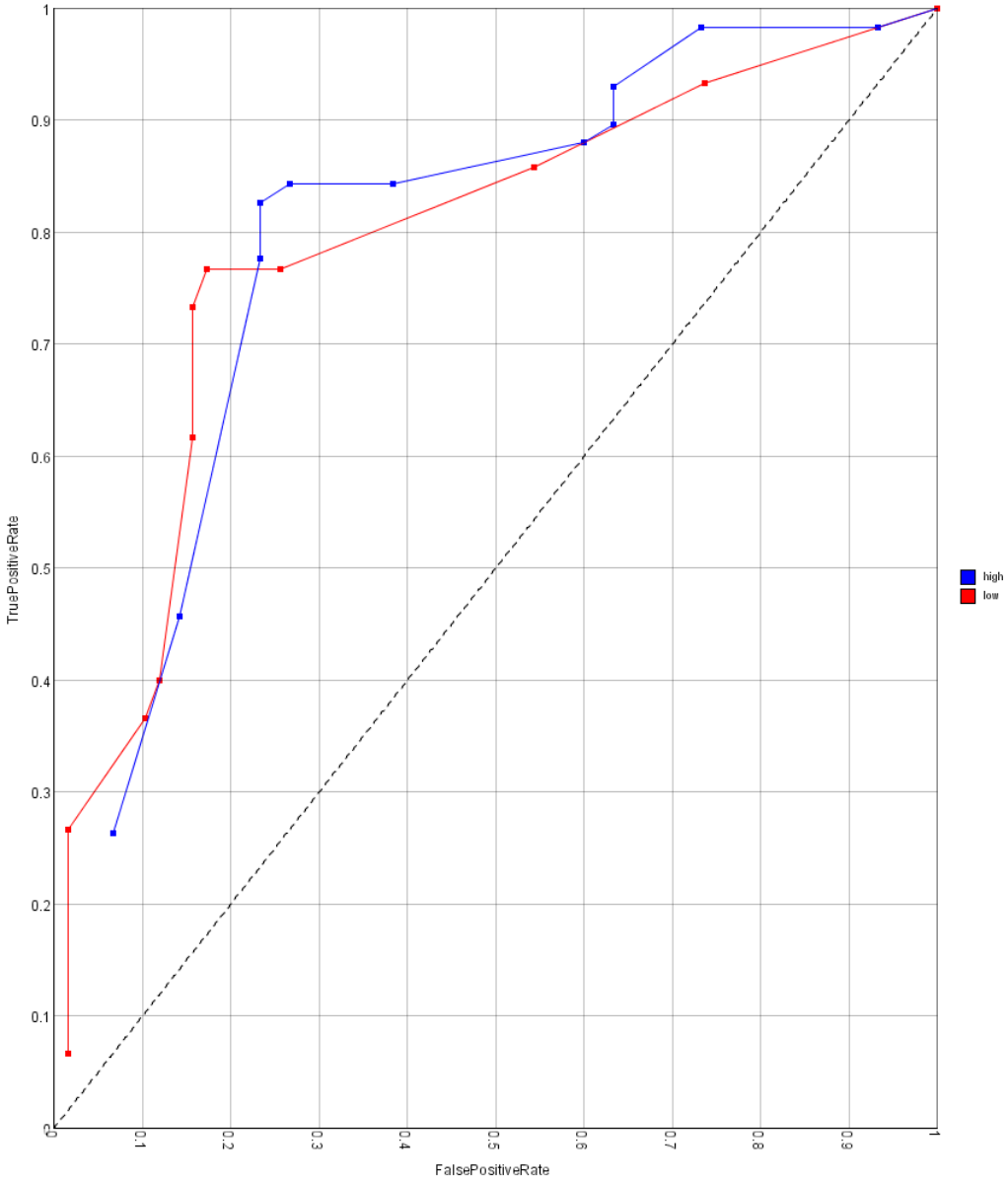


Figure 3.25: SEVERE_EXCYT_89: Receiver operating characteristics (ROC) curve for severity prediction calculated on 89 patients excluding cytokine data.

Including the cytokine data and leaving the technical tree parameters unchanged (minimum cases was set to 10 with a pruning confidence of 25%), no changes in the tree splitting criteria were observed but the overall performance of the tree decreased suggesting noisiness caused by interference of cytokine and clinical data.

Therefore, we constructed a tree that excluded TEMP_1 and CT_1ST_COLLECTION. This resulted in a tree similar to the one calculated without the cytokine data (SEVERE_INCYTA_89) (Figure 3.24), with the split represented by CT_1ST_COLLECTION exchanged by IP_10_1 (OR: 12.75; 95%CI: 7.98, 17.52) [14] (Table 3.48; Table 3.49). The chosen classifier (Figure 3.26) had a higher profit (0.617) with higher specificity (86%) but lower sensitivity (74%) and the resulting overall error rate was 19.17% (Table 3.50; Figure 3.27). The AUC for low and high classification was 0.79 but the two groups showed different confidence intervals (low 95%CI: 0.68, 0.89; high 95%CI: 0.69, 0.88).

[14] TEMP=body temperature; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* PLT=*platelet count*; IP_10=*interferon-inducible protein 10*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data.* OR=*odds ratio*; CI=*confidence interval*.

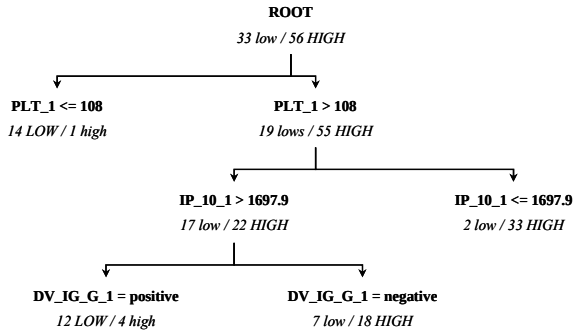


Figure 3.26: SEVERE_INCYTA_89: Decision tree for severity prediction calculated on 89 patients including cytokine data. PLT=*platelet count*; IP_10=*interferon-inducible protein 10*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data.*

Table 3.48: SEVERE_INCYTA_89: Decision tree for severity prediction calculated on 89 patients including cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. PLT=*platelet count*; IP_10=*interferon-inducible protein 10*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	3.64	40.53	32.40, 48.65	< 0.001
IP_10_1 [pg/ml] *Cut-off* value > 1697.9	5.40	11.20	7.97, 14.44	< 0.001
DV_IG_G_1 *Cut-off* value = positive	1.91	2.87	0.43, 5.30	0.026

Table 3.49: SEVERE_INCYTA_89: Decision tree for severity prediction calculated on 89 patients including cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. PLT=*platelet count*; IP_10=*interferon-inducible protein 10*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] *Cut-off* value <= 108	3.64	40.53	32.40, 48.65	< 0.001
IP_10_1 [pg/ml] *Cut-off* value <= 1697.9	7.63	12.75	7.98, 17.52	< 0.001
DV_IG_G_1 *Cut-off* value = positive	3.45	10.80	6.30, 15.30	0.003

Table 3.50: SEVERE_INCYTA_89: Summary of K-fold (k=10) cross validation for severity prediction based on 89 patients including cytokine data.

Overall Evaluation	Value (n=89)	
Total misclassifications	17.0	
Overall error rate	19.167%	
SE of error rate	7.686	
Average profit	0.617	
SE of profit	0.154	
AUC high	0.7865	95%CI: 0.69, 0.88
AUC low	0.7874	95%CI: 0.68, 0.89

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	48 (86%)	8 (14%)
	low	9 (26%)	24 (74%)

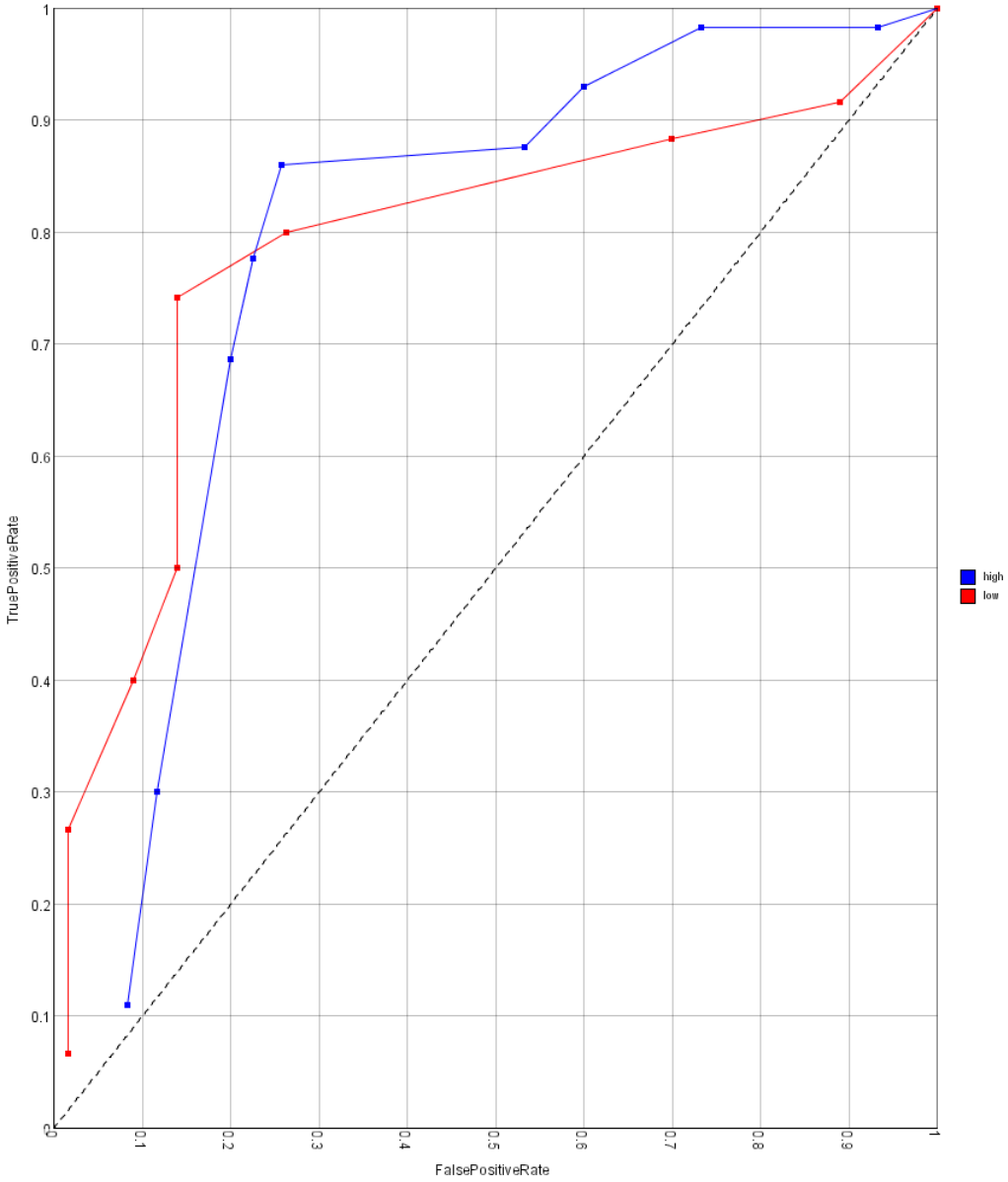


Figure 3.27: SEVERE_INCYTA_89: Receiver operating characteristics (ROC) curve for severity prediction calculated on 89 patients including cytokine data.

3.1.2.8 Severity Prediction based on Cytokine Data

It is known that cytokines are important factors in determining severity of disease and thus, a tree only including the cytokine data was constructed. This would provide us with a better picture of data interference and second, with a more detailed view of which cytokines could be important for disease pathogenesis. The following calculations revealed that classifier performance was strongly dependent on which cytokines were included into the tree. Therefore, we only picked the cytokines IP_10_1, IL_10_1 and IL_8_1[15] which showed the highest accuracy for severity prediction and this resulted in a good classifier performance of 0.75 for the prediction of 'low' (95%CI: 0.64, 0.86) as well as of 'high' (95%CI: 0.65, 0.85) cases (minimum cases was set to 23 with a pruning confidence of 25%). K-fold cross validation (k=10) determined an overall error rate of 29.028% with a sensitivity of 66% and a specificity of 76% (Table 3.53; Figure 3.30). The model (SEVERE_CYTOA_89_IL8) (Figure 3.28) picked IP_10_1>1697.9 (OR: 11.20; 95%CI: 7.97, 14.44) as the first and IL_8_1>3.34 (OR: 5.93; 95%CI: 2.52, 9.33) as the second splitting criterium (Table 3.51; Table 3.52). IL_10_1 was not integrated into the tree, but exclusion of this cytokine from the analysis, resulted in a decreased performance. This suggested that IL_10_1 had a supporting role on either IP_10_1 or IL_8_1 but as a predictor would be inferior to IL_8_1. Even decreasing of the tree parameter "minimum cases" did not result in the inclusion of IL_10_1.

Therefore, we checked for the influence of IL_8_1 in the tree and excluded this cytokine for further calculations. This resulted in a tree (SEVERE_CYTOA_89_IL10)

[15] IP_10=*interferon inducible protein 10*; IL_10=*interleukin-10*; IL_8=*interleukin-8;* OR=*odds ratio*; CI=*confidence interval*.

(Figure 3.29) that had IL_10_1>6.66 (OR: 5.63; 95%CI: 2.27, 8.98) as the second splitting criterium (Table 3.52; Table 3.52) and showed a higher sensitivity (70%) but in turn a lower specificity (72%). The overall accuracy was decreased which was represented by an AUC of 0.74 (95%CI: 0.62, 0.85) for 'low' and a AUC of 0.73 (95%CI: 0.63, 0.84) for 'high' cases. The profit averaged 0.397 with an overall error rate of 30.139% (Table 3.54; Figure 3.31).

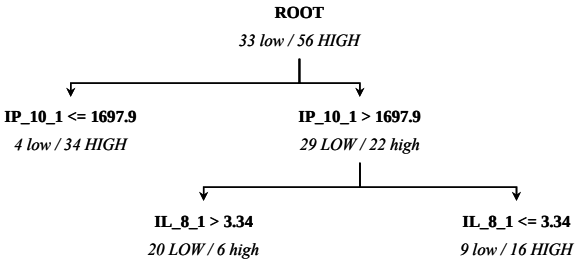


Figure 3.28: SEVERE_CYTOA_89_IL8: Decision tree for severity prediction calculated on 89 patients only including cytokine data and using interleukin-8 as the last splitting criteria. IP_10=*interferon inducible protein 10*; IL_8=*interleukin-8.*

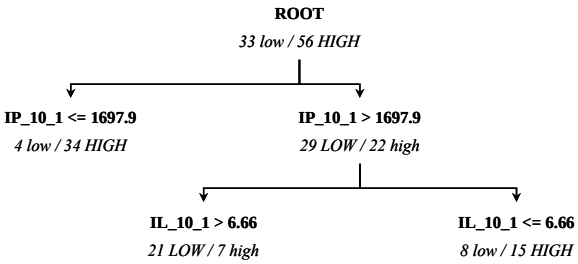


Figure 3.29: SEVERE_CYTOA_89_IL10: Decision tree for severity prediction calculated on 89 patients only including cytokine data and using interleukin-10 as the last splitting criteria. IP_10=*interferon inducible protein 10*; IL_10=*interleukin-10*.

Table 3.51: SEVERE_CYTOA_89: Decision tree for severity prediction calculated on 89 patients only including cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. IP_10=*interferon inducible protein 10*; IL_10=*interleukin-10*; IL_8=*interleukin-8;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IP_10_1 [pg/ml] *Cut-off* value > 1697.9	5.40	11.20	7.97, 14.44	< 0.001
IL_8_1 [pg/ml] *Cut-off* value > 3.34	2.57	4.48	1.95, 7.00	0.001
IL_10_1 [pg/ml] *Cut-off* value > 6.66	1.50	1.86	-0.68, 4.40	0.004

Table 3.52: SEVERE_CYTOA_89: Decision tree for severity prediction calculated on 89 patients only including cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. IP_10=*interferon inducible protein 10*; IL_10=*interleukin-10*; IL_8=*interleukin-8.* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IP_10_1 [pg/ml] *Cut-off* value > 1697.9	5.40	11.20	7.97, 14.44	< 0.001
IL_8_1 [pg/ml] *Cut-off* value > 3.34	2.14	5.93	2.52, 9.33	0.003
IL_10_1 [pg/ml] *Cut-off* value > 6.66	2.16	5.63	2.27, 8.98	0.005

Table 3.53: SEVERE_CYTOA_89_IL8: Summary of K-fold (k=10) cross validation for severity prediction based on 89 patients only including cytokine data and using interleukin-8 as the last splitting criteria.

Overall Evaluation	Value (n=89)	
Total misclassifications	26.0	
Overall error rate	29.028%	
SE of error rate	12.838	
Average profit	0.419	
SE of profit	0.257	
AUC high	0.7506	95%CI: 0.65, 0.85
AUC low	0.7531	95%CI: 0.64, 0.86

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	42 (76%)	14 214%)
	low	12 (34%)	21 (66%)

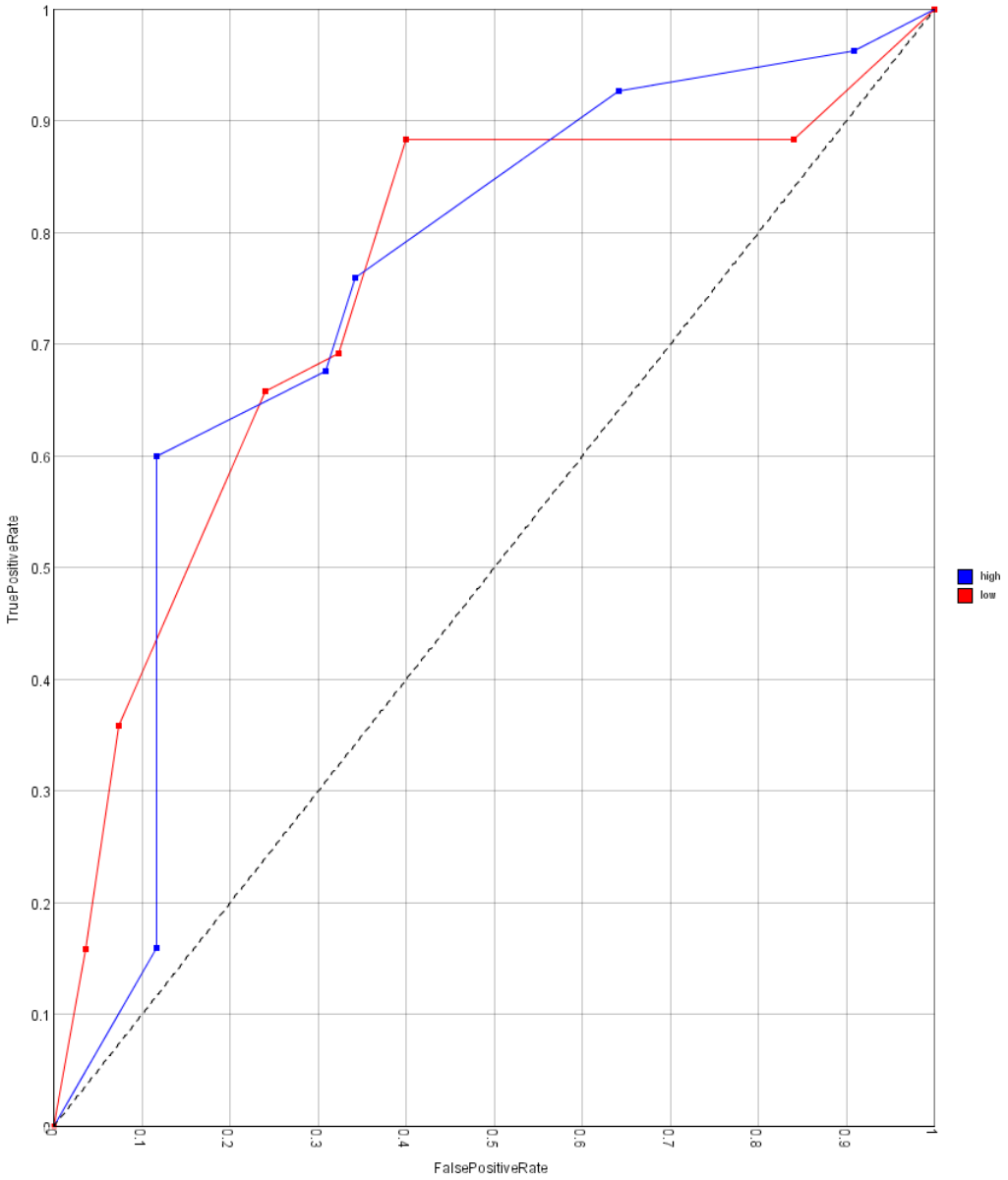


Figure 3.30: SEVERE_CYTOA_89_IL8: Receiver operating characteristics (ROC) curve for severity prediction based on 89 patients only including cytokine data and using IL-8 as the last splitting criteria.

Table 3.54: SEVERE_CYTOA_89_IL10: Summary of K-fold (k=10) cross validation for severity prediction based on 89 patients only including cytokine data and using IL_10_1 as the last splitting criteria.

Overall Evaluation	Value (n=89)	
Total misclassifications	27.0	
Overall error rate	30.139%	
SE of error rate	18.026	
Average profit	0.397	
SE of profit	0.361	
AUC high	0.7338	95%CI: 0.63, 0.84
AUC low	0.7354	95%CI: 0.62, 0.85

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	40 (72%)	16 (28%)
	low	11 (30%)	22 (70%)

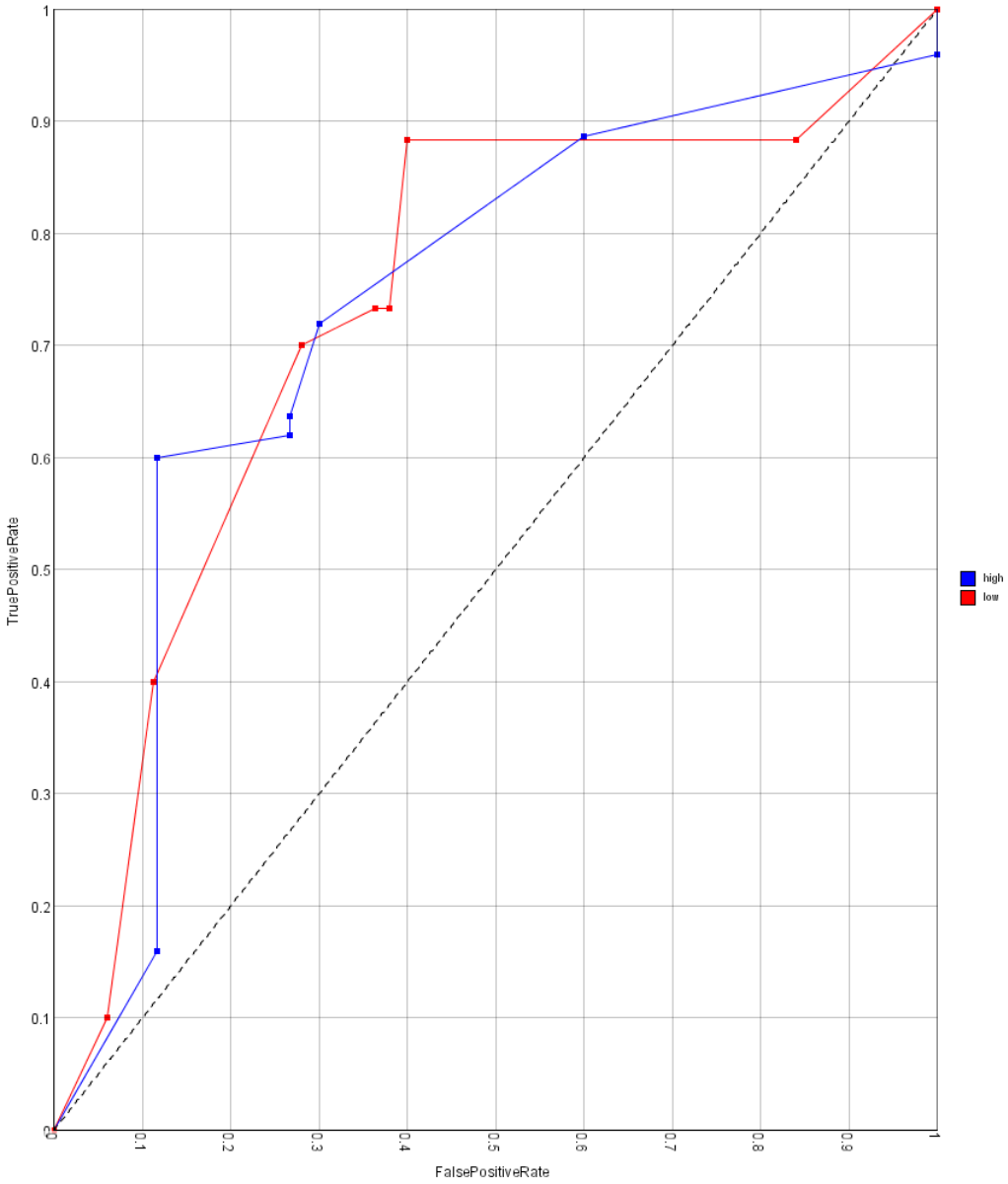


Figure 3.31: SEVERE_CYTOA_89_IL10: Receiver operating characteristics (ROC) curve for severity prediction based on 89 patients only including cytokine data and using IL-10 as the last splitting criteria.

3.1.2.9 Severity Prediction based on Cytokine and Clinical Data but only using hospitalized Cases

For the sake of completion, we separately investigated the influence of the cytokine data on the prediction of severity by excluding the non-hospitalized cases. This kind of analysis might identify cytokines that would be highly specific for severity due to the fact that we would classify two groups within a population (hospitalized patients) that was already more homogenous than using the total of 125 patients. Hence, the analysis was performed on a dataset of 52 cases (36.5% high, 63.5% low) and pruning confidence was set to 25% whereas 'minimum cases' was programmed to 4. The first tree excluding TEMP_1 (SEVHOSP_EXCYT_52) (Figure 3.32) was calculated by leaving the cytokine data on the side. PLT<=108 (OR: 15.00; 95%CI: 6.69, 23.31), DV_IG_G_1='positive' (OR: 4.69; 95%CI: 0.77, 8.62), CT_1ST_COLLECTION<=20.94 (OR: 42.00; 95%CI: 28.75, 55.25) and MXD_NO_1<=0.1 (OR: 15.00; 95%CI: 1.65, 27.35) were chosen as splitting criteria (Table 3.55; Table 3.56)[16] and the decision tree was able to correctly classify 43 cases with an overall error rate of 16.67% (Table 3.57; Figure 3.33). The sensitivity was shown to be 85% and the specificity was 80%. It had a good overall performance of 0.80 for the classification of low (95%CI: 0.69, 0.92) and high (95%CI: 0.67, 0.94). The probabilistic classifier was chosen at an average profit of 0.667.

[16] PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; MXD_NO= *The percentage of the WBC belonging to monocytes, eosinophiles & basophiles*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval*.

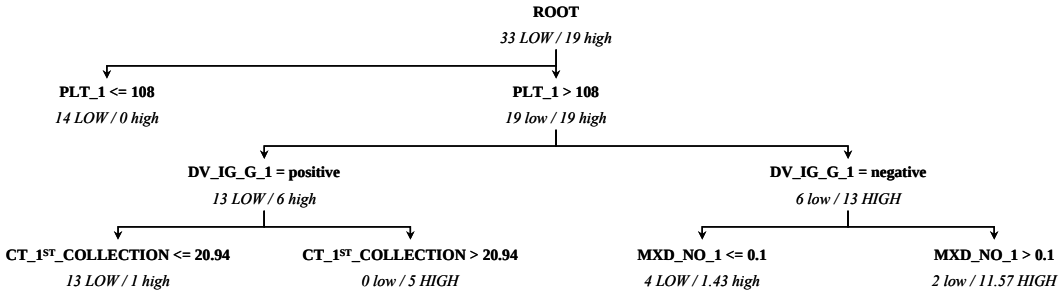


Figure 3.32: SEVHOSP_EXCYT_52: Decision tree for severity prediction calculated on 52 hospitalized patients excluding cytokine data. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; MXD_NO= *The percentage of the WBC belonging to monocytes, eosinophiles & basophiles*; 1=*1st visit data.*

Table 3.55: SEVHOSP_EXCYT_52: Decision tree for severity prediction calculated on 52 hospitalized patients excluding cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value $^{+1}$. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; MXD_NO= *The percentage of the WBC belonging to monocytes, eosinophiles & basophiles*; 1=*1st visit data;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] $^{+1}$ *Cut-off* value <= 108	1.88	15	6.69, 23.31	0.001
DV_IG_G_1 *Cut-off* value = positive	1.47	2.94	-0.34, 6.22	0.071
CT_1ST_COLLECTION *Cut-off* value <= 20.94	5.00	14.77	5.63, 23.91	0.007
MXD_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.1	1.45	3.93	-1.61, 9.46	0.149

Table 3.56: SEVHOSP_EXCYT_52: Decision tree for severity prediction calculated on 52 hospitalized patients excluding cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value $^{+1}$. PLT=*platelet count*; CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load;* DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; MXD_NO= *The percentage of the WBC belonging to monocytes, eosinophiles & basophiles*; 1=*1st visit data;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] $^{+1}$ *Cut-off* value <= 108	1.88	15	6.69, 23.31	0.001
DV_IG_G_1 *Cut-off* value = positive	2.17	4.69	0.77, 8.62	0.05
CT_1ST_COLLECTION $^{+1}$ *Cut-off* value <= 20.94	6.13	42.00	28.75, 55.25	0.001
MXD_NO_1 [*1000 cells/mm^3] *Cut-off* value <= 0.1	3.33	15.00	2.65, 27.35	0.015

Table 3.57: SEVHOSP_EXCYT_52: Summary of K-fold (k=10) cross validation for severity prediction based on 52 hospitalized patients excluding cytokine data.

Overall Evaluation	Value (n=52)	
Total misclassifications	9.0	
Overall error rate	16.667%	
SE of error rate	12.669	
Average profit	0.667	
SE of profit	0.253	
AUC high	0.8042	95%CI: 0.67, 0.94
AUC low	0.8042	95%CI: 0.69, 0.92

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	15 (80%)	4 (20%)
	low	5 (15%)	28 (85%)

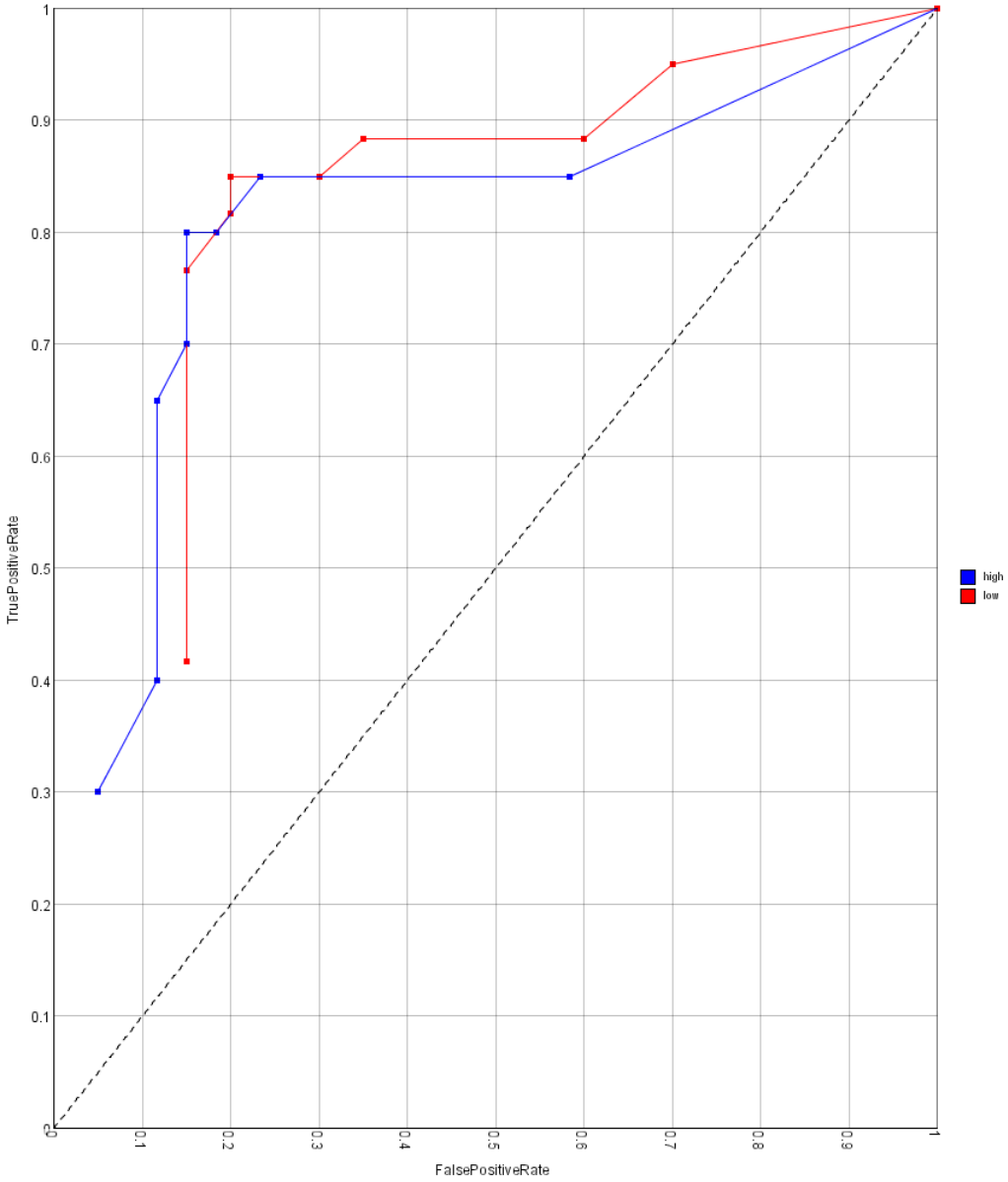


Figure 3.33: SEVHOSP_EXCYT_52: Receiver operating characteristics (ROC) curve for severity prediction calculated on 52 hospitalized patients excluding cytokine data.

Calculating a tree (excluding body temperature) with the same technical parameters and including the cytokine data did not result in an improvement (data not shown). Therefore, we additionally excluded CT_1ST_COLLECTION and found that inclusion of only I_TAC_1 along with IL_10_1 would give the best prediction. Pruning confidence was still set to 25% whereas 'minimum cases' was defined as 10. This, in turn, resulted in a tree (SEVHOSP_INCYTA_52) (Figure 3.34) that only used PLT_1 <=108 (OR: 15.00; 95%CI: 6.69, 23.31) and I_TAC_1<=1255.2 (OR: 26.00; 95%CI: 16.44, 35.56) as decision criteria (Table 3.58; Table 3.59)[17]. The misclassification rate accounted for 20.667% with a sensitivity of 80% and a specificity of 80% (Table 7l; Figure 7h). The AUC of the 'low' ROC curve averaged 0.84 (95%CI: 0.73, 0.94) and classification of 'high' cases showed an AUC of 0.84 (95%CI: 0.72, 0.96). Furthermore, the classifier chosen at a specific threshold had a lower average profit (0.587) (Table 3.60; Figure 3.35).

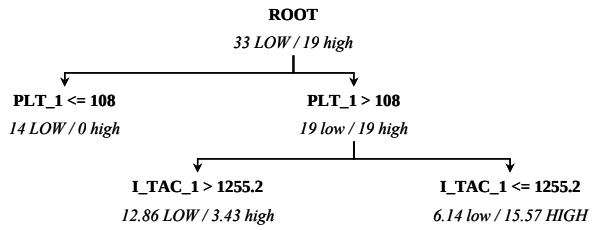


Figure 3.34: SEVHOSP_INCYTA_52: Decision tree for severity prediction calculated on 52 hospitalized patients including cytokine data. PLT=*platelet count*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data*

[17] CT_1st_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load*; PLT=*platelet count*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval*.

Table 3.58: SEVHOSP_INCYTA_52: Decision tree for severity prediction calculated on 52 hospitalized patients including cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value $^{+1}$. PLT=*platelet count*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] $^{+1}$ *Cut-off* value <= 108	1.88	15.00	6.69, 23.31	0.001
I_TAC_1 [pg/ml] $^{+1}$ *Cut-off* value > 1255.2	2.09	22.8	14.00, 31.64	< 0.001

Table 3.59: SEVHOSP_INCYTA_52: Decision tree for severity prediction calculated on 52 hospitalized patients including cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value $^{+1}$. PLT=*platelet count*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
PLT_1 [*1000/mm^3] $^{+1}$ *Cut-off* value <= 108	1.88	15.00	6.69, 23.31	0.001
I_TAC_1 [pg/ml] $^{+1}$ *Cut-off* value > 1255.2	2.79	26.00	16.44, 35.56	< 0.001

Table 3.60: SEVHOSP_INCYTA_52: Summary of K-fold (k=10) cross validation for severity prediction based on 52 hospitalized patients including cytokine data.

Overall Evaluation	Value (n=52)	
Total misclassifications	11.0	
Overall error rate	20.667%	
SE of error rate	15.299	
Average profit	0.587	
SE of profit	0.306	
AUC high	0.8398	95%CI: 0.72, 0.96
AUC low	0.8373	95%CI: 0.73, 0.94

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	15 (80%)	4 (20%)
	low	7 (20%)	26 (80%)

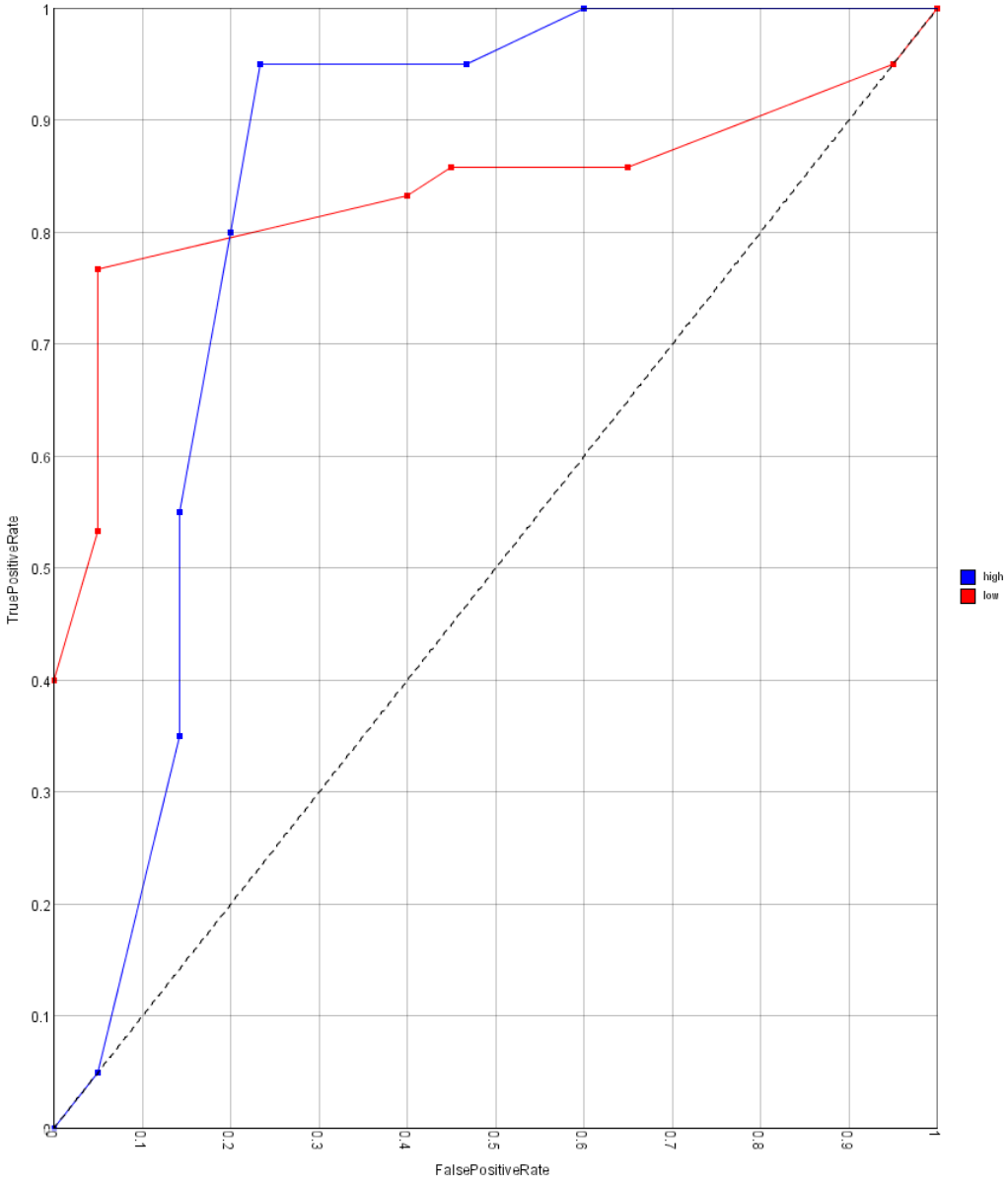


Figure 3.35: SEVHOSP_INCYTA_52: Receiver operating characteristics (ROC) curve for severity prediction calculated on 52 hospitalized patients including cytokine data.

3.1.2.10 Severity Prediction based on Cytokine Data but only using hospitalized Cases

Finally, we calculated a tree on the hospitalized patients that included only the cytokine data. The product was a classifier (SEVHOSP_CYTO_52) (Figure 3.36) (pruning confidence set to 25% and minimum cases defined as 4) that had a high sensitivity of 84% and a specificity of 65% (Table 3.63; Figure 3.37) . The overall performance of the classifier accounted for 0.74 for severe (95%CI: 0.60, 0.87) as well as mild cases (95%CI: 0.59, 0.89) with an error rate of 21.333%. The first branch of the tree was defined by IL_8_1>3.33 (OR: 19.55; 95%CI: 14.38, 24.72) and patients below this threshold were further split by using I_TAC_1>1255.2 (OR: 16.50; 95%CI: 5.40, 27.60)[18]. The tree model having a profit of 0.573 used only these two splitting criteria (Table 3.61; Table 3.62). It was observed that exclusion of IL_10_1 from the decision tree analysis lead to a decrease in classifier performance pointing towards a supportive role of IL_10_1 in the prediction of severity.

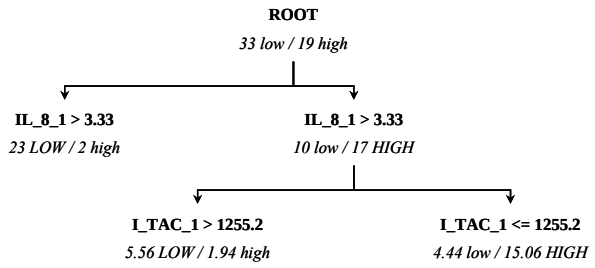


Figure 3.36: SEVHOSP_CYTOA_52: Decision tree for severity prediction calculated on 52 hospitalized patients only including cytokine data. IL_8=*interleukin-8*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data.*

[18] IL_8=*interleukin-8*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data;* OR=*odds ratio*; CI=*confidence interval*.

Table 3.61: SEVHOSP_CYTOA_52: Decision tree for severity prediction calculated on 52 hospitalized patients only including cytokine data. Statistical analysis of splitting criteria performed on the whole dataset. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value [+1]. IL_8=*interleukin-8*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IL_8_1 [pg/ml] *Cut-off* value > 3.33	2.48	19.55	14.38, 24.72	< 0.001
I_TAC_1 [pg/ml] [+1] *Cut-off* value > 1255.2	2.09	22.80	13.96, 31.64	< 0.001

Table 3.62: SEVHOSP_CYTOA_52: Decision tree for severity prediction calculated on 52 hospitalized patients only including cytokine data. Statistical analysis of splitting criteria performed on each subgroup at the decision nodes. In case of 0 values in the original contingency table, OR calculations were adjusted by adding 1 to each table value [+1]. IL_8=*interleukin-8*; I_TAC=*interferon-inducible T cell α chemoattractant*; 1=*1st visit data;* 1=*1st visit data;* RR=*relative risk*; OR=*odds ratio*; CI=*confidence interval*.

Decision Node Feature	RR	OR	95% CI (OR)	*p* value
IL_8_1 [pg/ml] *Cut-off* value > 3.33	2.48	19.55	14.38, 24.72	< 0.001
I_TAC_1[pg/ml] [+1] *Cut-off* value > 1255.2	3.21	16.50	5.40, 27.60	0.007

Table 3.63: SEVHOSP_CYTOA_52: Summary of K-fold (k=10) cross validation for severity prediction based on 52 hospitalized patients only including cytokine data.

Overall Evaluation	Value (n=52)	
Total misclassifications	11.0	
Overall error rate	21.333%	
SE of error rate	20.012	
Average profit	0.573	
SE of profit	0.400	
AUC high	0.7404	95%CI: 0.59, 0.89
AUC low	0.7354	95%CI: 0.60, 0.87

Confusion Matrix			
		Predicted Class	
		high	low
Actual Class	high	13 (65%)	6 (35%)
	low	5 (16%)	28 (84%)

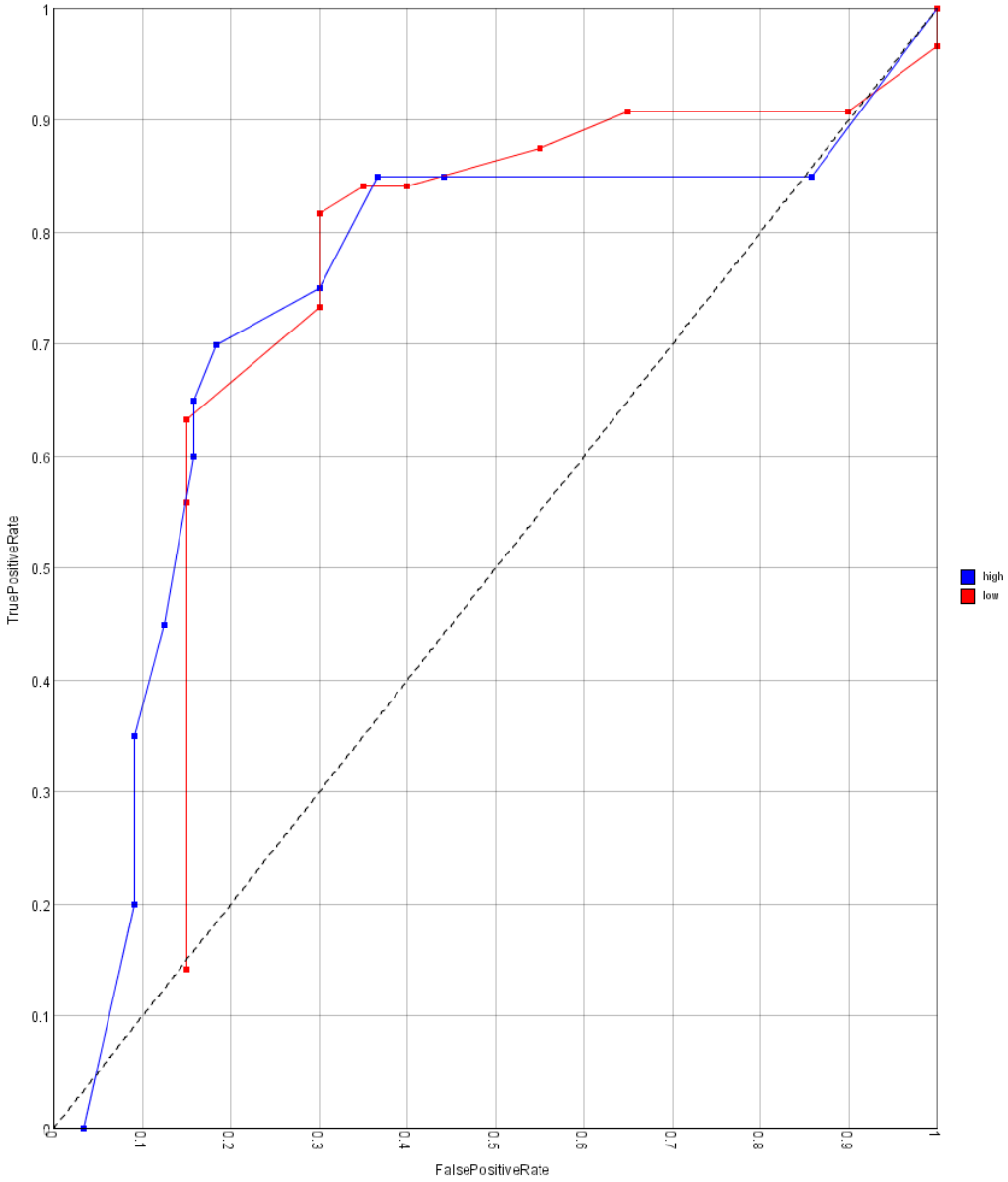


Figure 3.37: SEVHOSP_CYTOA_52: Receiver operating characteristics (ROC) curve for severity prediction calculated on 52 hospitalized patients only including cytokine data.

3.2 Sequence Analyses of Virus Isolates from the 2005 Singaporean Dengue Outbreak

3.2.1 Description of Differences between Serotype 1 and Serotype 3

During the 2005 Singaporean dengue outbreak, there was mainly a co-circulation of DENV-1 and DENV-3. Detection of DENV-2 and especially DENV-4 was very rare. From the 133 dengue positive patients, serotyping detected 66 (48.9%) DENV-1 along with 62 (46.6%) DENV-3 whereas DENV-2 was only the causative agent in 5 patients (3.8%). There was no detection of serotype 4 (Low et al., 2006). Isolation of only 112 (84.2%) viruses was possible. The isolated viruses were mainly represented by DENV-1 (57 isolates; 50.8%) along with DENV-3 (50 isolates; 44.6%). Only five serotype 2 viruses were isolated (Table 3.66).

It is intriguing that 16 (76.19%) of the 21 unisolated viruses (9 DENV-1 and 12 DENV-3) originated from secondary infections. The viral load was significant lower ($p<0.001$) in the non-isolated viruses than in the isolated ones and indicated that there was not enough virus for isolation.

3.2.1.1 Serotype specific Differences with regard to Severity

As a consequence of the severity modeling, we first aimed at investigating whether there were serotype specific differences with regard to disease outcome. For this purpose, we decided to investigate the parameters that were detected to be important in severity modeling and chose as time points the 1st as well as the 2nd visit. Thus, viral

load, platelet count, secondary infection, IL-8, IL-10, IP-10 and I-TAC were included into descriptive statistical analysis. In addition, we included IFN-α to find out whether a virus specific response exists and checked whether there was a difference regarding hospitalization. DENV-2 patients were excluded due to only five samples which would not give any statistically reliable values. This resulted in a datasheet of 128 cases for the analysis of clinical data and 90 cases were used for calculation of cytokine data. 38 samples were excluded from the cytokine analysis due to lack of data. Univariate analysis identified significant differences in viral load and IFN-α on the 1st as well as second visit. On the other hand, IL-8 was significantly different only on the 2nd visit (Table 3.64; Table 3.65).

Table 3.64: Median for non-normally [b] distributed clinical and immunological data collected on the first visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters ($p<0.05$) which were first log-transformed. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. *Chi-square* test [d] was used to assess for significance between secondary infections (positive/negative), platelet class (low/high) and hospitalization (yes/no); for calculations of platelet class, only 86 cases were included [g]; absolute counts are indicated below [e]. Significant differences between cytokine data was calculated by using a dataset of 90 patients. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate	Dengue positive cases (n=128) * grouped by Serotypes 1&3				
	DENV-1		DENV-3		p value
CT_1ST_COLLECTION [b]	17.065		19.755		0.002
PLT_1 [b]	160.500		166.000		0.453
IP_10_1 [b]	1985.950		1722.165		0.205
I_TAC_1 [b]	1219.250		1205.700		0.786
IFN_ALPHA_1 [b]	1270.500		786.865		0.001
IL_8_1 [b]	3.550		2.565		0.120
IL_10_1 [b]	6.965		8.715		0.251
DV_IG_G_1 [d, e]	16 [positive]	26 [negative]	20 [positive]	28 [negative]	0.730
PLT_CLASS [d, e, g]	18 [low]	22 [high]	14 [low]	32 [high]	0.163
HOSPITALIZED [d, e]	31 [yes]	11 [no]	24 [yes]	24 [no]	0.021

Table 3.65: Median for non-normally [b] distributed clinical and immunological data collected on the second visit. *Shapiro-Wilk* normality test was performed to check for non-normally distributed parameters ($p<0.05$) which were fist log-transformed. If the log-transformation still resulted in a non-normal distribution then non-parametric *Kruskal-Willis* test was used whereas *Student's* t test was considered in case of normally distributed data. Significant differences between cytokine data was calculated by using a dataset of 90 patients. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. Shading indicates a significant *p* value < 0.05.

Covariate	Dengue positive cases (n=128) * grouped by Serotypes 1&3		
	DENV-1	DENV-3	p value
CT_2ND_COLLECTION [b]	26.870	29.570	0.001
PLT_2 [b]	92.500	104.000	0.359
IP_10_2 [b]	1491.050	1260.250	0.289
I_TAC_2 [b]	1121.750	1048.650	0.544
IFN_ALPHA_2 [b]	531.840	234.200	0.027
IL_8_2 [b]	3.940	2.580	0.002
IL_10_2 [b]	24.800	9.790	0.144

3.2.2 Phylogenetic Analysis of 94 Virus Strains from the EDEN Study

The viruses isolated from dengue patients from the 2005 Singaporean dengue outbreak were sequenced to identify subgroups that could be related to disease outcome, specifically to compare clades with regard to the clinical parameters that resulted from the severity modelling.

3.2.2.1 Full Length Genome Sequencing of 94 Virus Isolates

In a first attempt, the preparative RT-PCR step had to be optimized to ensure efficient amplification of isolated viral RNA. Amplification of viral RNA was performed by using a strategy of five overlapping fragments which approximately measured 2.5 to 3 kb. Fragment 5 (including the 3'UTR) of dengue serotype 1 represented the main

hurdle for the sequencing process and optimization resulted in the finding that a ramp speed of 9600 °C/s was essential for cDNA synthesis as well as for cDNA amplification. Moreover, a ten fold higher primer concentration with regard to the other fragments and serotypes had to be used (Appendix; Page 234).

On the other hand, we faced another problem with fragment 3 serotype 3. Optimization was performed by running a gradient PCR to identify the appropriate annealing temperature for the two fragment specific primers. This resulted in a temperature increase from the commonly used 55°C to the newly used 63.1°C (Appendix; Page 234).

By using the optimized protocol (Materials & Methods; Page 85), we successfully sequenced a total of 94 dengue virus isolates from the 2005 Singaporean dengue outbreak (Appendix; Page 238). 53 DENV-1 along with 41 DENV-3 full genome sequences were obtained (Table 3.66). Due to the low number of isolated DENV-2, we first aimed at completing serotypes 1 and 3 which were eventually used for further analysis.

Table 3.66: Overview of the 112 virus isolates & their status of sequencing. One DENV-1 strain and four DENV-3 strains were only partially sequenced and thus considered as not sequenced $^{+}$.

Description	DENV-1	DENV-2	DENV-3	Total
Number of isolated virus strains	57	5	50	112
Number of virus isolates that did not grow	2	1	4	7
Number of virus isolates without any PCR product	1	0	1	2
Number of virus isolates that were available for sequencing	54	4	45	103
Total number of sequenced virus isolates	53 $^{+}$	0	41 $^{+}$	94

3.2.2.2 Phylogenetic Analyses of the 94 sequenced Virus Genomes

The obtained full-length genome sequences were aligned and the corresponding phylogenetic trees were inferred. Within both serotypes one outlier was detected and thus, was excluded from further analysis, which finally resulted in 52 DENV-1 and 40 DENV-3 sequences incorporated into the respective trees. The mean nucleotide difference for DENV-1 was 19.8 (0.184%) and for DENV-3 12.6 (0.118%). Moreover, the phylogenetic trees (Figure 3.38; Figure 3.40) revealed one clear early split in both serotypes with a bootstrap value of 100%.

Within serotype 1, we observed that further splitting resulted in three clear splits showing high bootstrap values which was nicely demonstrated by the unrooted tree (Figure 3.39). However, the observed splits were unbalanced and only 11 strains were clearly separated from the remaining 41 strains. The next split showed low reliability indicated by a bootstrap value of 62% suggesting high similarity within this subgroup of EDEN serotype 1. The high variability of bootstrap values made it difficult to identify appropriate clades but for ease of comparison, we could still reliably assign four clades (EDEN 1.1 – 1.4). The bootstrap values were 100% for splitting EDEN 1.1 form the rest, 100% for splitting EDEN 1.2 and 99% between EDEN 1.3 and EDEN 1.4. We decided not to further investigate EDEN DENV-1 clades due to small numbers of viruses in the three first clades which would not give any reliable statistical values in any comparative analysis against EDEN 1.4.

On the other hand, the tree generated from genomes of strains within serotype 3 generally had branches with higher bootstrap values than the serotype 1 tree. In addition, the unrooted tree (Figure 3.41) described higher diversity between the

different strains which was underlined by the longer branch lengths and the star-like branching pattern with regard to DENV-1. Nevertheless, it was noticed that also the EDEN DENV-3 strains were very similar to each other but we still defined four clades (EDEN 3.1 – 3.4) that we would use for further investigations. The bootstrap values were 100% for the first split and afterwards 100% for separating EDEN 3.1 from EDEN 3.2 and 71% for splitting EDEN 3.3 from EDEN 3.4.

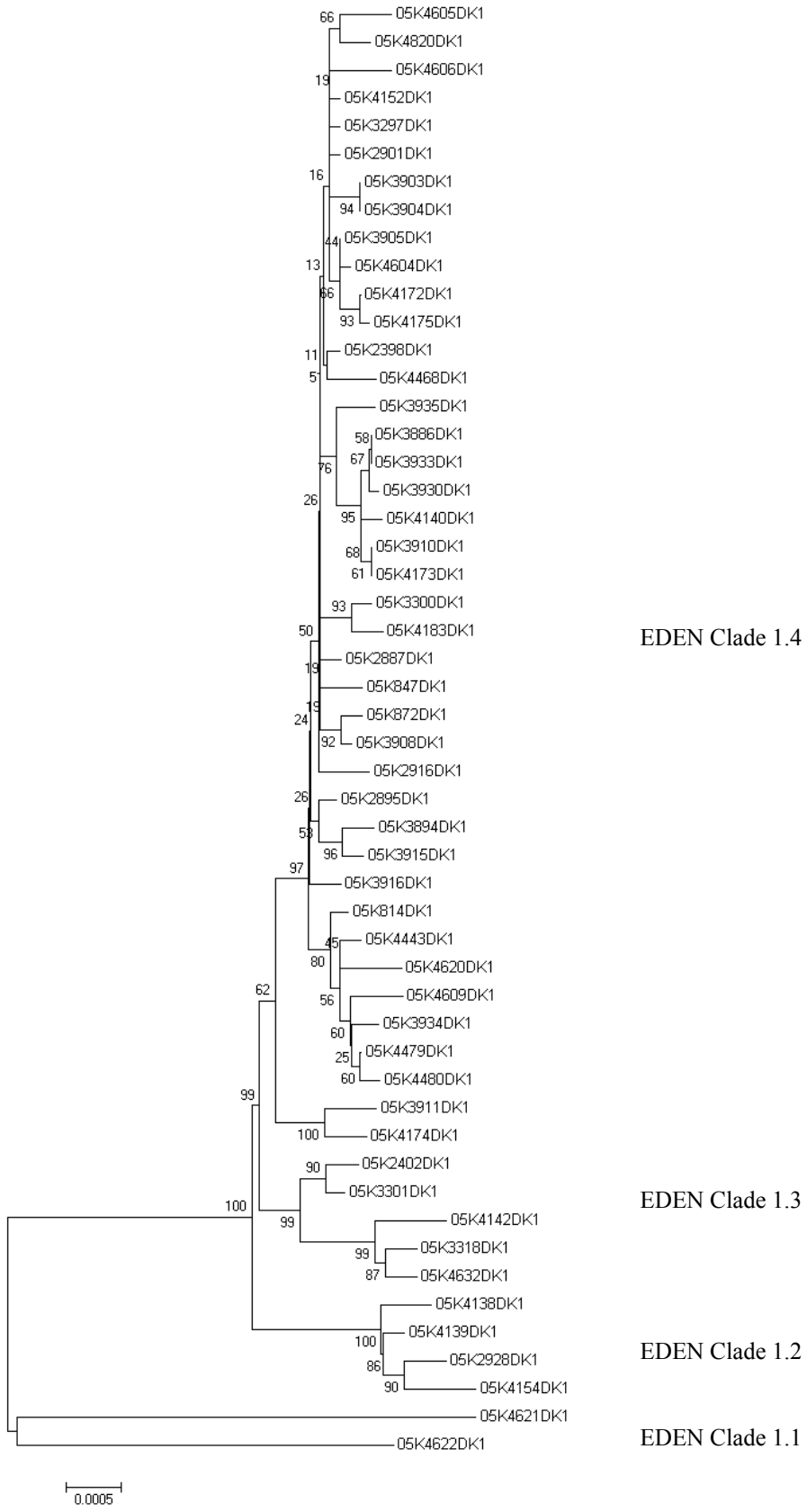


Figure 3.38: Phylogenetic tree based on the whole genome (10,735nt) of 52 EDEN DENV-1 isolates. The mean nucleotide difference within EDEN serotype 1 was 19.8 (0.184%). Strain 05K4107DK1 was an outlier and thus, was removed to better visualize the tree structure. The tree leaves are labeled with the EDEN ID number and the bootstrap value supporting each node is indicated.

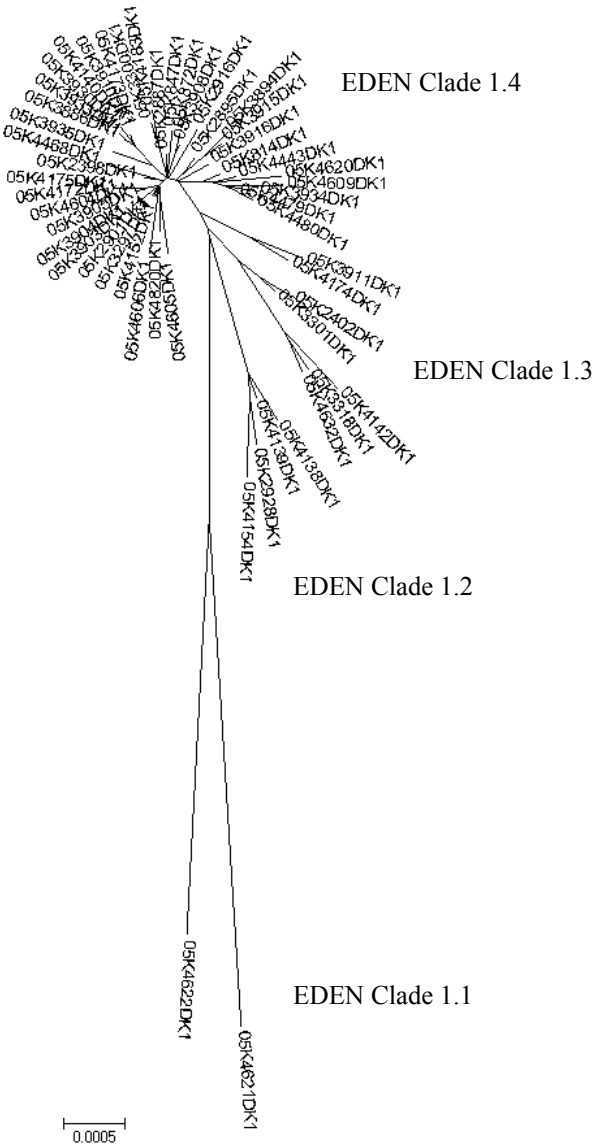


Figure 3.39: The phylogenetic tree from Figure 3.38 drawn in radiation form. The tree leaves are labeled with the EDEN ID number.

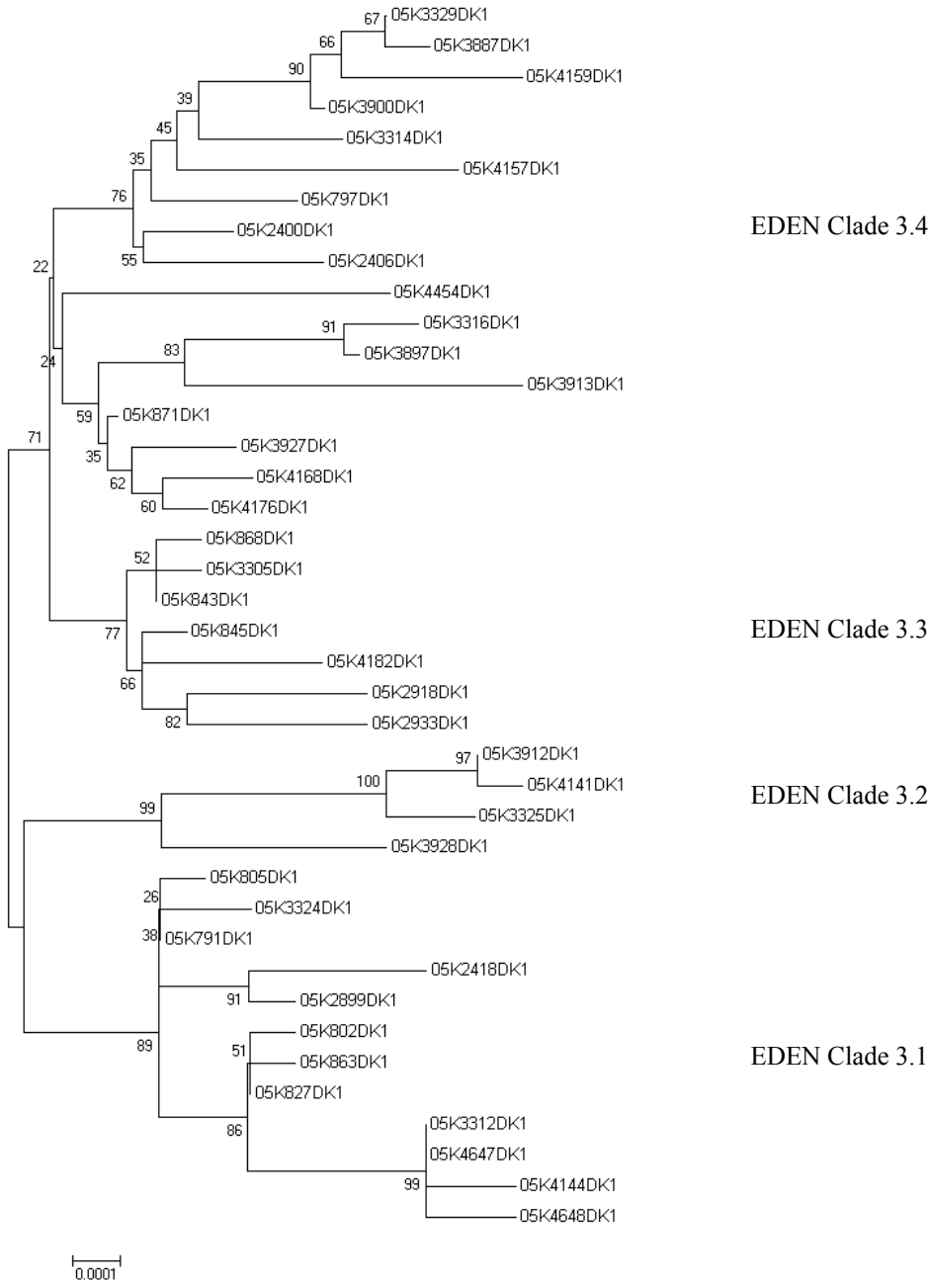


Figure 3.40: Phylogenetic tree based on the whole genome (10,707 nt) of 40 EDEN DENV-3 isolates. The mean nucleotide difference within EDEN serotype 3 was 12.6 (0.118%). Strain 05K4440DK1 was an outlier and thus, was removed to better visualize the tree structure. The tree leaves are labeled with the EDEN ID number and the bootstrap value supporting each node is indicated.

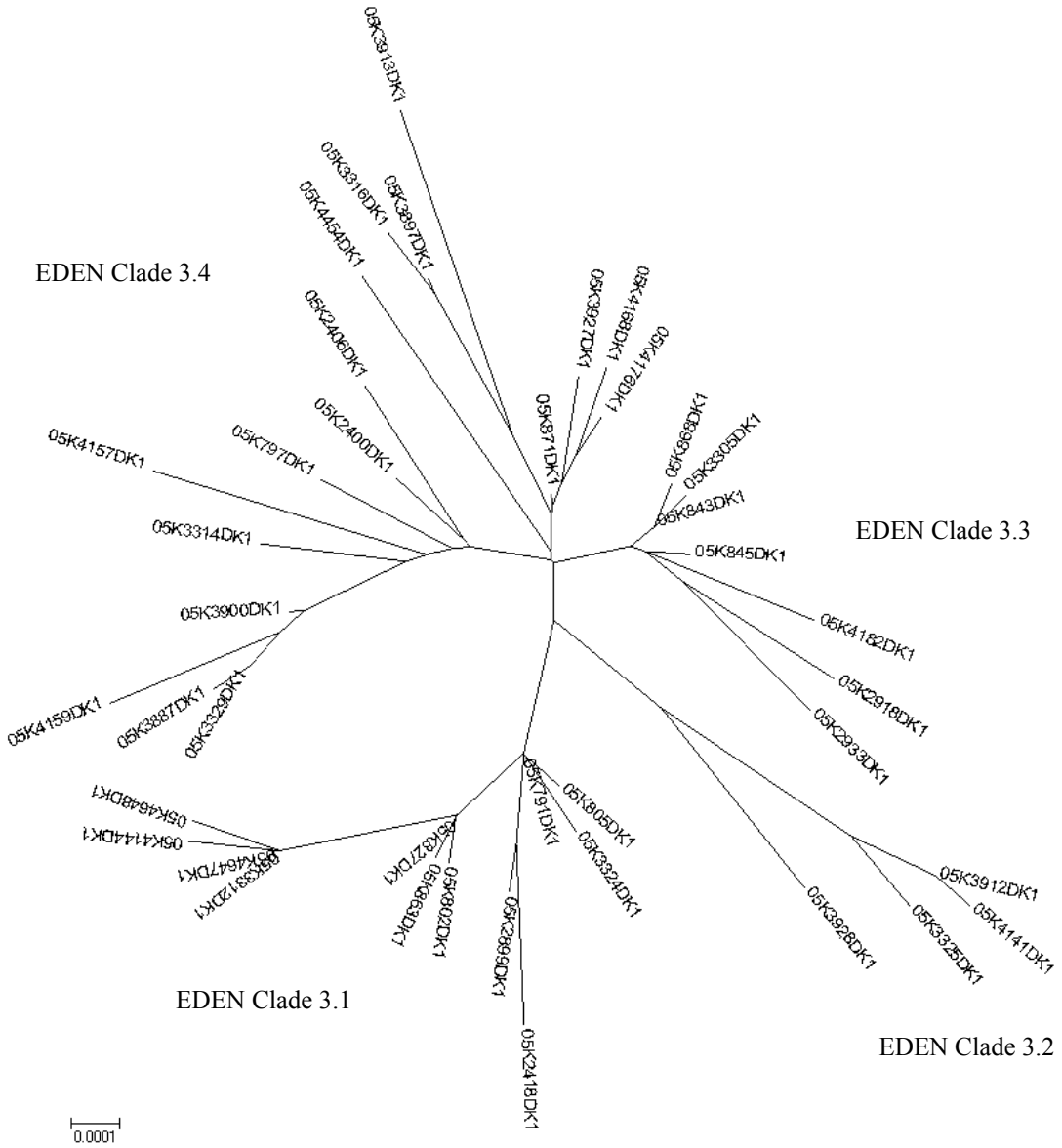


Figure 3.41: The phylogenetic tree from Figure 3.40 drawn in radiation form. The tree leaves are labeled with the EDEN ID number.

3.2.3 Comparison of Clinical Parameters between the four EDEN DENV-3 Clades

After the phylogenetic analysis and after defining significant clades within serotype 1 and 3, we aimed at identifying whether there might be clades correlating with disease severity. For this purpose we compared the clinical parameters identified by the severity modeling. Serotype 1 genomes were excluded from the analysis due to their high similarity (low bootstrap values) which made it difficult to find appropriate clades. Moreover, three of the identified clades (EDEN 1.1, EDEN 1.2, EDEN 1.3) consisted of small numbers of viruses which would not give any statistically reliable values. Thus, the DENV-3 clades were selected for further analyses.

Using ANOVA, we identified viral load as well as platelet count collected on the first visit as the only significant differences between the four DENV-3 clades (Table 3.67; Figure 3.42; Figure 3.43). In addition, post-hoc calculations using the *Tukey* test, resulted in the finding that only clades EDEN 3.1 and EDEN 3.2 showed a real significant difference ($p=0.014$) between each other (Table 3.68). The remaining clades were not significantly different from each other ($p>0.05$). Interestingly, we also revealed a significant difference between clades EDEN 3.1 and EDEN 3.3 ($p=0.041$) as well as between EDEN 3.3 and EDEN 3.4 ($p=0.032$) with regard to the platelet count on the first visit (Table 3.69).

Calculations within DENV-3 using the two clades showing a 100% bootstrap value (EDEN 3.1 combined to EDEN 3.2 and EDEN 3.3 combined to EDEN 3.4) did not result in any significant differences. Moreover, we did not identify any clear

differences between the DENV-3 clades regarding second visit data by either using two or four clades.

Table 3.67: Mean for significant parameters collected on the 1st visit grouped by DENV-3 EDEN clades. Cases with missing values were excluded and therefore, the number of cases for *p* value calculations differed between covariates *. CT_1ST_COLLECTION=*viral load whereby a high Ct-value indicates a low viral load*; PLT=*platelet count*; DV_IG_G=*indicator for primary/secondary infection whereby a positive result indicates a secondary infection*; 1=*1st visit data*; Shading indicates a significant *p* value < 0.05.

Covariate	Serotype DENV-3 (n=40) * grouped by identified DENV-3 EDEN clades				
	EDEN 3.1	EDEN 3..2	EDEN 3.3	EDEN 3.4	p value
CT_1ST_COLLECTION	21.212	14.917	17.291	19.501	0.014
PLT_1	159.583	153.000	241.571	161.294	0.028
DV_IG_G_1	3^{pos} / 9^{neg}	2^{pos} / 2^{neg}	2^{pos} / 5^{neg}	6^{pos} / 11^{neg}	
PLT_CLASS	2^{low} / 10^{high}	2^{low} / 2^{high}	3^{low} / 3^{high}	3^{low} / 13^{high}	
HOSPITALIZED	3^{yes} / 9^{no}	3^{yes} / 1^{no}	5^{yes} / 2^{no}	8^{yes} / 9^{no}	

Table 3.68: Matrix of pairwise comparison probabilities of viral load between the four DENV-3 EDEN clades determined by the *Tukey* post hoc test. CT =*viral load whereby a high Ct-value indicates a low viral load*. Shading indicates a significant *p* value < 0.05.

Post Hoc Test CT	EDEN 3.1	EDEN 3..2	EDEN 3.3	EDEN 3.4
EDEN 3.1	1.000			
EDEN 3.2	0.018	1.000		
EDEN 3.3	0.105	0.704	1.000	
EDEN 3.4	0.572	0.105	0.505	1.000

Table 3.69: Matrix of pairwise comparison probabilities of platelet count between the four DENV-3 EDEN clades determined by the *Tukey* post hoc test. PLT=*platelet count.* Shading indicates a significant *p* value < 0.05.

Post Hoc Test PLT	EDEN 3.1	EDEN 3..2	EDEN 3.3	EDEN 3.4
EDEN 3.1	1.000			
EDEN 3.2	0.998	1.000		
EDEN 3.3	0.041	0.121	1.000	
EDEN 3.4	1.000	0.995	0.032	1.000

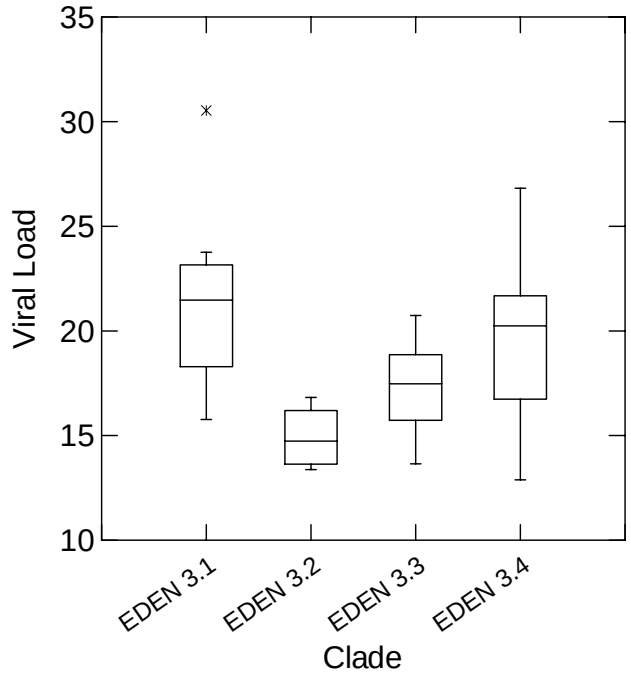


Figure 3.42: Comparison of viral load between the four EDEN DENV-3 clades. Viral load is represented as Ct-value and high Ct-value indicates a low viral load. We observe that EDEN 3.2 has the highest viral load.

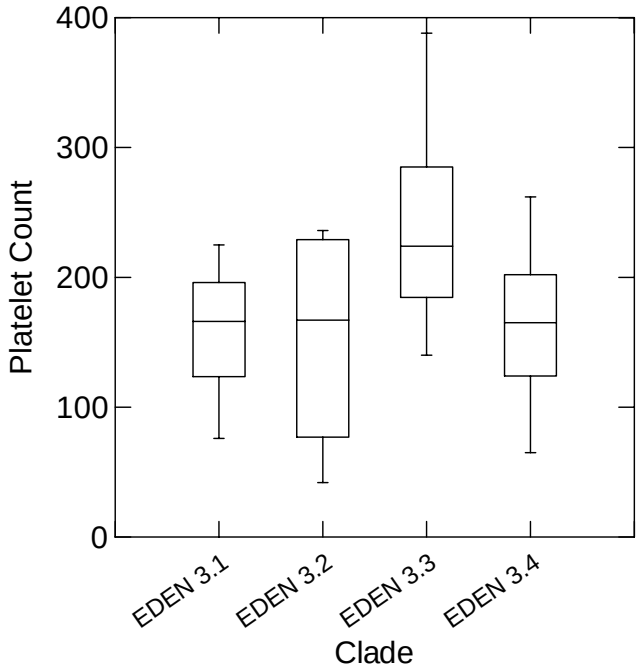


Figure 3.43: Comparison of platelet count between the four EDEN DENV-3 clades. A low platelet count is considered to be a marker of severity. We observe that EDEN 3.3 has the highest platelet count. However, the error bars show a large range and thus, it is unlikely that there is a real difference between the different clades.

4 Discussion

4.1 Decision Tree Analyses

4.1.1 Modeling of Disease Data

The first part of my master thesis investigated the clinical and immunological factors that might play a crucial role in early dengue pathogenesis and might be useful to distinguish dengue virus infection from other febrile illnesses in an early stage of disease. The second part focused on finding a clinical as well as immunological model for severity in dengue disease, especially on finding early markers that might represent potential risk factors for the development of more severe disease manifestations.

4.1.1.1 Decision Tree Models

To address the research questions above, C4.5 classification trees (Quinlan, 1993) were used which have several advantages over other statistical tools. Briefly, classification trees are in principle simple to understand and are able to handle both continous and categorical data. A simple model based on clinical parameters, enabling diagnoses of dengue disease and prediction of disease severity in an early stage of disease is highly desirable and might be of great support in epidemic settings where a lack of infrastructure and under representation of trained personal could result in high morbidity or eventually death. Furthermore, the C4.5 classification tree algorithm used is able to handle missing values which represent a common problem in clinical studies. This permitted us to make use of all the clinical data available and thus, we did not always have to exclude cases with missing values. In contrast, logistic regression as

well as discriminant analyses require much more data preparation and appropriate handling of missing values is crucial for reliable calculations (Acuna and Rodriguez, 2004). Finally, decision trees are easy to interpret, use and validate using common statistical techniques. Our data mining approach was able to combine new findings as well as established disease mechanisms in dengue pathogenesis.

4.1.1.2 Modeling of Dengue Infection

The early time point (average 42h of illness) of the patient blood sample represents a major advantage over other studies that also aimed at dengue diagnosis and/or disease severity which, in contrast, made use of samples collected at later time points. In addition, this study is the first to integrate clinical along with immunological factors such as cytokines and lymphocyte counts into one predictive model for dengue disease which shows a sensitivity of 90%. The use of decision tree modeling is also a novelty in the dengue field.

4.1.1.3 Modeling of Disease Severity

Even though, there were only two reported cases of DHF according to the WHO guidelines, we were able to find an appropriate severity categorization based on platelet counts that might be useful in the assessment of disease severity in adults. The validity of our severity models is supported by the high accuracy achieved as well as by the correct classification of the two DHF cases. Models either based on clinical data, cytokine data or mixed data were always able to correctly classify the two DHF

cases. One aspect that may explain the low DHF incidence in our study is represented by the fact that adults have a decreased risk compared to children for the development of vascular permeability and thus for DHF/DSS. Therefore, in areas such as Singapore where predominantly adults are infected, it is fundamental to get a more detailed picture of dengue pathogenesis in adults and our accurate predictive models make a significant contribution towards understanding of dengue fever in adult patients.

4.1.1.4 Important Considerations of our Approach

Although our model for dengue diagnosis shows high accuracy, we have to keep in mind, that the disease background of the control group is essential. The possibility of an overrepresentation of causative agents that are not of viral origin amongst the control group, may mean that our diagnostic tool might not be specific for dengue infection but might more likely represent a general picture of flaviviral or even viral infections. Therefore, it is evident that our models for dengue prediction have to be validated in different epidemic settings with higher incidence of other flaviviral infections in which infections amongst the controls are clearly identified.

Another drawback of this study may lie in the absence of 2nd as well as 3rd visit data for the control group. Having this kind of information might give a better picture of dengue pathogenesis in comparison to other febrile illnesses. With regard to cytokines and their kinetics it would be especially interesting to investigate the differences over the course of an infection and to determine how long cytokine levels remain elevated after dengue infection. The next critical point in the presented study involves using 3rd visit data of dengue patients as healthy control samples which may fail in detecting

elevated cytokine levels because it is not known how long cytokines may remain in patient's serum after dengue infection. Considering the phenomenon of cross-protection against heterologous serotypes, a time window of several months after dengue infection is probably necessary to ensure normal levels of immune mediators that could serve as healthy controls. Hence, it would have been very useful to have data from a 4^{th} visit after several months that would have represented healthy controls.

Finally, the lack of a substantial number of severe cases in our study could be a disadvantage in the generalization of our severity model. Our severity categorization is based on low platelet counts and even though the chosen threshold of <50,000 platelets/mm^3 is underlined by two other studies (Balmaseda et al., 2005; Malavige et al., 2006), it is necessary that our model has to be evaluated in an epidemic setting that shows higher severity than Singapore. Finally, a severity categorization based on more than one parameter will probably be more accurate than our proposed categorization but, in turn, might also result in a more complex picture.

4.1.2 Distinguishing Dengue from other Febrile Illnesses in an early Stage of Disease

4.1.2.1 White Blood Cell Count, absolute Numbers of Lymphocytes and Temperature, but not Symptoms can Diagnose Dengue Infection in an early Stage of Disease

A recent study reported a predictive model based on logistic regression to diagnose dengue fever within the first two days of admission (Chadwick et al., 2006). This study mainly used clinical data combined with symptoms to predict dengue infection and showed a sensitivity of 84%. By contrast, our model (DENPRE_TOTAL_453; Figure 3.1; Page 94) using clinical data within 72 hours of onset of disease also shows a sensitivity of 82%, and integration of cytokine data (DENPRE_INCYTA_291; Figure 3.7; Page 105) resulted in an increased sensitivity of 90%. It is intriguing that our classification tree based on clinical data and Chadwick's model show striking similarity to each other. The first split chosen by our tree is represented by a white blood cell count equal or lower than 4800 cells/mm^3 and this threshold is supported by the logistic regression model which defined a white blood cell count of lower than 5000 cells/mm^3 as a genuine risk factor for dengue infection. The logistic regression model of Chadwick's study further used rash, creatine, bilirubin and prothrombin time as discriminators whereas our presented classification tree included the absolute numbers of lymphocytes and the body temperature as further splitting criteria. We were not able to include the clinical data used by the logistic regression model because the respective variables were not measured in the EDEN study. Furthermore, model calculations only based on symptoms did not result in a reliable classifier suggesting

that detection of dengue specific symptoms is only possible in a later stage of disease. The advantage of our decision tree over the recently published logistic regression model not only lies in the early time point of sample collection but also in the fact that we were predicting a PCR positive result which shows a sensitivity higher than 99%. The IgM assay that Chadwick and collaborators based their logistic regression on, shows a slightly lower sensitivity. PCR is more specific than the IgM assay because IgM can be also found in other flaviviruses although these are not frequent in Singapore.

In addition, we were able to analyze different subgroups of dengue patients which clearly revealed that having a lower white blood cell count represents a main but not the only indicator for a positive dengue PCR result. The decision tree especially emphasized the decreased numbers of lymphocytes as a main characteristic of acute dengue infection. This finding is supported by another study also showing low numbers of lymphocytes in an acute stage of dengue disease (Azeredo et al., 2001).

Our clinical model shows a very high accuracy of over 90% and the stability is underlined by a narrow confidence interval of the calculated AUC of the ROC curve (Table 3.3; Figure 3.2; Page 96). Moreover, the odds ratios calculated either on the whole dataset or at each decision node show high level of significance and the similarity of the two models based either on the total dataset (453 cases) (DENPRE_TOTAL_453; Figure 3.1; Page 94) or on 291 cases (DENPRE_EXCYT_291; Figure 3.3; Page 98) proves the validity of our calculations. It is important to note that the chosen classifier on the ROC curve is skewed towards identifying true negative cases because by definition, the program chooses the model

at a specific threshold with the highest profit for the major class. This resulted in the lower sensitivity of 82% compared to a higher specificity of 91% (Table 3.3; Figure 3.2; Page 96). The model was evaluated by stratified k-fold cross validation (k=10) which is considered to be a powerful method to prevent data overfitting (Kohavi, 1995). Nevertheless, it is necessary that our decision tree model gets evaluated in a new epidemic setting. Assessment of the classification tree on negative patients enrolled in 2006 showed similar classifier performance. However, evaluating the classifier in terms of its ability to identify dengue positive cases is of much higher relevance, especially in a setting where other flaviviruses and symptomatically similar diseases like Chikungunya are frequent.

This presented model includes only basic clinical parameters and is useful in providing assistance in dengue diagnosis in an early stage of disease. To some extent, it might be a useful substitution in epidemic settings where lack of infrastructure and under representation of trained staff exacerbate the implementation of either the commonly used virus isolation, the serologic PCR or the IgM assay. Measuring the white blood cell count and the number of lymphocytes are commonly used basic hematological tests that are not laborious, do not require highly trained staff and are probably more economical than the commonly used dengue diagnostic tools. Hence, the model may allow early dengue diagnosis at low cost. Combination of this diagnostic model with our proposed severity prediction might contribute to better case management and eventually, to a better prevention of DHF/DSS cases.

4.1.2.2 Lower IL-2 and decreased TNF-α, but elevated IL-10 Levels along with an increase of IP-10 may be highly specific for an early Stage of Dengue Disease

It is known that serum concentrations of cytokines and soluble receptors may be significantly increased during dengue virus infection. Different studies showed that, in comparison to healthy controls, levels of IL-10 (Fink et al., 2006; Green and Rothman, 2006; Green et al., 1999a; Libraty et al., 2002; Nguyen et al., 2004), TNF-α (Fink et al., 2006; Green and Rothman, 2006; Hober et al., 1996; Nguyen et al., 2004), IFN-γ (Kurane et al., 1991; Kurane et al., 1986), IFN-α (Chang et al., 2006; Fink et al., 2006; Green and Rothman, 2006) and IL-2 (Kurane et al., 1991; Pacsa et al., 2000) are elevated. To our knowledge, only a few studies were performed that compared serum cytokine levels of dengue patients to controls suffering from other febrile illnesses. Furthermore, these studies were mainly performed on samples collected from children and might not be representative for adult infections. In our study, we compared dengue cases to other febrile illnesses and identified various cytokines as genuine risk factors for dengue infection which were represented by higher levels of IP-10, I-TAC, IFN-α, IL-1 as well as IL-10 but by lower serum concentrations of IL-2 and TNF-α.

A decision tree that only included the available cytokine data but excluded the clinical data, resulted in a classifier (DENPRE_CYTOA_291; Figure 3.9; Page 108) that used IFN-α as the first splitting criterium. Dengue positive patients more often had higher levels of IFN-α possibly reflecting an antiviral mechanism. Generally, interferons have important functions in different kinds of antiviral responses. In particular, the dengue specific immune response is mediated by RIG-1 that recognizes double-stranded RNA and activates type I interferon signaling/secretion, especially IFN-α (Chang et al.,

2006; Green and Rothman, 2006). However, considering the general importance of type I interferons in antiviral responses and the fact that the causes of disease amongst the controls is not known, we must bear in mind that our results naturally support a dengue specific immune mechanism but do not shed light on whether high levels of IFN-α are a useful indicator specifically for dengue prediction. This led us to the assumption that exclusion of IFN-α from our decision tree analysis might be useful in detecting dengue specific cytokines. The resulting tree (DENPRE_CYTO_291; Figure 3.11; Page 111) chose IL-2 as the first decision criterium which, in contrast, was the second splitting criterium in the tree calculated with the inclusion of IFN-α. We observed that very low levels of IL-2 represented a high risk of having a symptomatic dengue infection which was supported by a high and significant odds ratio. Interestingly, we were not able to find any significant changes of IL-2 between the three visits which is in contrast to another reported study that showed elevated levels of IL-2 during course of infection (Kurane et al., 1991; Pacsa et al., 2000). We explain the observed phenomenon with regards to several aspects. IL-2 is an important cytokine mainly produced by activated T-cells which supports the activation of naïve T-cells and their subsequent proliferation into effector T-cells. It is generally produced upon contact with various kinds of antigens and plays a major role in shaping an appropriate immune response. Thus, we can assume that IL-2 levels amongst the controls suffering from other febrile illnesses are also elevated but due to the lack of required data, we were not able to determine differences in cytokine levels between our control group and healthy adults. In addition, comparison of IL-2 levels of dengue patients on the 1st and 2nd visit did not differ significantly from IL-2 levels on the 3rd visit which might be explained by a short time window after infection to detect normal cytokine levels. Nevertheless, our findings might be linked to the observed

phenomenon of impaired T-cell proliferation in acute dengue infection (Mathew et al., 1999). Mathew's study showed that accessory cells obtained from peripheral blood mononuclear cells (PBMCs) during acute dengue infection did not provide an adequate stimulus to enable T-cell proliferation. The decreased responses of T-cells could be partially restored by exogenous IL-2 and IL-4, showing the functionality of T-cells. Our results suggest that patients already having significantly lower levels of IL-2 are at main risk for symptomatic dengue infection. Upon infection, this subgroup of patients is not able to generate an appropriate immune response. In addition, dengue virus aggravates this effect by impairing the beneficial T-cell proliferation. This subsequently leads to an immediate establishment of dengue infection and probably to a higher viral load compared to patients with higher IL-2 levels. Grouping dengue positive patients by low (<=2.54 pg/ml) and high levels of IL-2 revealed a higher viral load in patients with low IL-2 ($p=0.08$). Furthermore, it is worthwhile mentioning that univariate analysis performed on IL-2 grouping of dengue patients detected significant differences with regards to their cytokine levels on all three visits (results not shown) suggesting major differences in their immune responses.

Patients showing higher levels of IL-2 were further split into subgroups by either IL-10 (including IFN-α) or IP-10 (excluding IFN-α). Both trees showed that higher levels of either these cytokines are able to correctly predict dengue infection with a sensitivity of 84%. It is intriguing that several other studies also showed increased higher levels of IL-10 during dengue infection (Fink et al., 2006; Green and Rothman, 2006; Green et al., 1999a; Libraty et al., 2002; Nguyen et al., 2004) which serve as valuable support for our calculated model. IL-10 is an anti-inflammatory cytokine mainly produced by monocytes and regulatory or T_H2 cells and inhibits IL-2 secretion which is a cytokine

specific for T_H1 responses. Its increased secretion in an early stage of an acute dengue infection might be another explanation for the unchanged or minimally decreased levels of IL-2 and for the impaired T-cell proliferation. Interestingly, recent studies showed that IL-10 suppresses the activation of cytotoxic T-cells (Brooks et al., 2006; Ejrnaes et al., 2006). Moreover, we observed that IL-10 levels on the 1st and 2nd visit were significantly higher than compared to the 3rd visit suggesting IL-10 to be highly specific for the early acute phase of dengue infection.

On the other hand, interferon-γ inducible protein 10 (IP-10) is a chemokine specific for the CXCR3 receptors which are selectively expressed on activated T-cells. IP-10 plays a role in the regulation of lymphocyte recruitment as well as in the recruitment of monocytes and NK-cells (Baggiolini et al., 1997). There is still little information about IP-10 function during dengue infection but two *in vivo* studies showed that dengue virus induces expression of IP-10 which, in turn, competitively inhibits viral binding and is required for resistance to primary dengue infections in mice (Chen et al., 2006; Hsieh et al., 2006). Our study is the first to measure IP-10 serum levels in dengue patients and decision tree analysis revealed that high IP-10 levels might be specific for dengue infection which was additionally backed by a high odds ratio and its narrow confidence interval.

Both trees (including or excluding IFN-α) used TNF-α as the last splitting criteria but with slightly different thresholds. We identified a lower level of TNF-α as a significantly higher risk factor for being dengue positive which represents a clear contradiction to other studies which showed that serum concentrations of TNF-α are higher during dengue infection, especially in DHF/DSS (Fink et al., 2006; Green and

Rothman, 2006; Hober et al., 1996; Nguyen et al., 2004). A possible reason for such a contradiction may lie again in our control group which represents patients suffering from other febrile illnesses and not healthy adults. TNF-α is a pro-inflammatory cytokine and is mainly secreted by macrophages, mast cells as well as lymphocytes. TNF-α is the main regulator of pro-inflammatory responses and levels are usually increased during a plethora of infections which may allow the conclusion that TNF levels must also be elevated amongst our control patients. We were not able to determine the levels in respect to healthy controls. Comparison of TNF-α levels of dengue patients on all the three visits, did not detect any significant changes in concentration. These findings demonstrate some similarity with another study showing little to no change in TNF-α levels to healthy control children (Green et al., 1999a). Another possible explanation for our observation of low TNF levels might be the lack of a substantial number of DHF/DSS cases in our study which show stronger associations to the secretion of TNF-α than the classical dengue fever cases.

In summary, we can conclude that prediction of dengue cases only based on cytokines is accurate and that we identified IP-10 as a possible new predictive marker for dengue infection. Furthermore, we find lowered TNF-α levels along with higher serum levels of IL-10 to be associated to dengue infection in an early stage of disease and compared to other febrile illnesses. Our approach is the first to link lower levels of IL-2 derived from clinical samples to low lymphocyte numbers which might represent impaired T-cell proliferation.

4.1.2.3 Lower IL-2 Levels, higher IP-10 Levels as well as increased IFN-γ combined with clinical Data are strong Predictors for Dengue Infection in an early Stage of Disease

Combination of basic clinical data and cytokine data resulted in two different decision trees that showed higher accuracies than trees based only on either clinical or cytokine data. The first constructed tree including IFN-α (DENPRE_INCYTA_291; Figure 3.5; Page 101) did not change its first two splitting criteria compared to the tree which was based only on cytokine data. IFN-α was followed by IL-2 but, in turn, the white blood cell count was used as a third splitting criteria for patients having lower IFN-α along with higher IL-2 levels. This suggests that the white blood cell count is probably a better predictor than IL-10 which is underlined by the better overall performance compared to the tree based only on cytokine data and by the fact that the tree based only on clinical data was the better predictor by itself. Furthermore, the odds ratio of white blood cell count is four times higher than the one of IL-10. The decision node that further separates patients with a lower white blood cell count was represented by IL-1 which is also reported to be secreted during dengue infection (Hober et al., 1993). Generally, IL-1 is mainly secreted by macrophages and dendritic cells and has an important function in the activation of a specific immune response during inflammation. Its effector functions are similar to TNF-α which might be the reason for its position in the tree. It is intriguing that higher levels of IL-1 are a significant characteristic of dengue positive patients whereas high TNF-α levels are more common in other febrile illnesses suggesting that IL-1 is more specific for classical dengue fever and TNF-α might be more important in DHF/DSS as reported in other studies (Kittigul et al., 2000; Nguyen et al., 2004).

Interestingly, our best calculated classification model excluded IFN-α and combined other cytokines as well as clinical data (DENPRE_INCYT_291; Figure 3.7; Page 105). It achieved an overall performance of 94% with a sensitivity of 90%. In the calculated tree, a lower white blood cell count and lower absolute numbers of lymphocytes represented the clinical risks for dengue infection. On the other hand, the immunological factors were represented by IL-2, IP-10, TNF-α as well as IFN-γ. IFN-γ is a pro-inflammatory T_H1 specific cytokine and is mainly produced by macrophages. Our finding that elevated IFN-γ levels could be an indicator for dengue virus infection is consistent with earlier reports demonstrating the induction of IFN-γ secretion upon dengue infection (Kurane et al., 1991; Kurane et al., 1986). Due to the early time point of sample collection, we expected to observe elevated levels of T_H1 cytokines which was shown by Pacsa and collaborators (Pacsa et al., 2000). Hence, it is striking that the observed cytokines included in our trees do not fully support this hypothesis but rather point towards an overlapping of T_H1 and T_H2 responses.

The decision tree explicitly shows the complex picture of cytokines interacting with lymphocytes and is highly predictive in distinguishing dengue patients from patients with other febrile illnesses. We can conclude that the cytokines and clinical features represented in the combined classification model are of significant importance during an early stage of dengue infection. Nevertheless, it is prudent that our classification tree has to be closely evaluated in several aspects. Firstly, more detailed statistical investigation is needed to identify correlations between the chosen parameters at the decision nodes because by only looking at the decision tree, we are unable to detect direct interactions. Secondly, cytokine production has to be investigated on a cellular level and *in vivo* to understand the sources, kinetics and functional impacts of

cytokines specifically produced upon dengue infection. Finally, the model has to be validated in other epidemic settings, especially in regions with a high prevalence of other flaviviruses to determine dengue specific cytokines. Moreover, the combined model is useful in providing a hint of which cytokines might be of greater interest to investigate but it would not be suitable as a diagnostic tool in epidemic settings. The major drawback lies in the procurement of cytokine data which entails more labour and has higher cost than the commonly used hematological methods.

In summary to the first aspect, we integrated immunological and clinical factors into a predictive model for dengue infection, we propose that the combined decision tree may be a biological tool for the better understanding of the interplay between humoral and cellular factors during an early stage of disease and highlight that the tree based only on clinical data may be a useful diagnostic tool in epidemic settings.

4.1.3 Severity Prediction in an early Stage of Disease

4.1.3.1 Platelet Group Classification shows Consistency and Coherence with the two Cases classified as DHF after the WHO Guidelines

Combination of dengue diagnosis with a model that could predict disease severity would be essential for appropriate case management. Hence, we aimed at predicting disease severity in dengue patients. In a first attempt, we defined hospitalization as the severity categorization which resulted in an accurate classification tree (HOSP_TOTAL_133; Figure 3.13; Page 115) including the platelet count at the second visit, the number of lymphocytes during the first visit and finally, the body temperature at the second visit. Considering our goal to predict severity early on in dengue disease by using first visit data only, and considering that hospitalization is biased due to its dependency on various factors (Results; Page 123), we decided to find a new severity categorization that might be more reflective for adult dengue patients. Moreover, a classifier calculated on a smaller dataset (HOSP_EXCYT_95; Figure 3.15; Page 118) resulted in different splitting criteria and used a slightly different threshold than the tree based on the total dataset suggesting instability of classifier performance. Even including cytokine data did not result in an improvement and on the other hand, a decision tree constructed by only using cytokine data (HOSP_CYTO_95; Figure 3.17; Page 121) did not show a reliable overall performance. The observed variation in classifier performance is further underlined by the calculated statistical values on the whole dataset which show that some of the chosen parameters are not real risk factors for hospitalization.

As a consequence, we chose a low platelet count as an indicator for severity. A threshold of <=50,000 platelets/mm^3 on either day five, six or seven of illness defined the cases of a more severe group. A low platelet count as a valid indicator for severity is underlined by several studies showing association of thrombocytopenia with more severe infections (reviewed in (Fink et al., 2006)), especially the threshold of <=50,000/mm^3 is supported by two studies describing significant associations with more severe disease manifestations (Balmaseda et al., 2005; Malavige et al., 2006) and by our study, showing nadir of platelet counts to be in the range of 50,000/mm^3 on either day five, six or seven. Categorization of dengue patients into mild and severe cases was also based on the assumption that non-hospitalized cases might be suffering from mild infections and therefore were classified as non-severe. Naturally, it is deputable that we might have missed some non-hospitalized patients but considering the outline of the EDEN study (Low et al., 2006) and the high hospitalization rate in Singapore it is unlikely that there were severe cases in the group of non-hospitalized cases.

We found that the two platelet groups (low and high) showed significant differences and especially the viral load (higher in more severe patients), the platelet count (lower in more severe patients), the body temperature (higher in more severe patients) as well as the immunological background (more secondary infections in more severe patients) represented the genuine risks on the first visit (Table 3.32; Page 130). Interestingly, categorization by hospitalization resulted in a slightly different picture. Secondary infection as a genuine risk for severity was only found by the platelet categorization. We concluded that significant differences with regard to measured clinical parameters (Table 3.28;Table 3.29; Table 3.30; Table 3.31; Table 3.32; Pages 127 to 130) only

observed by the proposed platelet categorization which are not obvious using hospitalization as a severity indicator are highly interesting early features specifically correlating with a later platelet decay. Moreover, looking at differences between the two categorization methods with regard to cytokine patterns revealed only three differences in hospitalization but, in turn, seven differences in platelet groups which additionally highlights a higher sensitivity of platelet categorization in detecting more severe patients (Table 3.33; Table 3.34; Table, 3.35; Table 3.36; Table 3.37; Table 3.38; Pages 131 to 133). This is backed by the common assumption in the dengue field that cytokines play a crucial role in disease pathogenesis. Finally, it is important to mention that all calculated models predicting severity were able to correctly classify the two DHF cases in our study as more severe cases which confers validity to the used classification scheme.

4.1.3.2 A high viral Load combined with a secondary Infection is a genuine Risk Factor for the Development of Thrombocytopenia

We investigated the possibility of predicting development of a low platelet count on day five, six and seven by including clinical as well as immunological data into one reliable model. In a first attempt, we only exploited clinical data collected on the 1st visit which resulted in a classification tree (SEVERE_TOTAL_125; Figure 3.20; Page 135) with a sensitivity of 81% and a specificity of 83%. The fist decision node made use of the platelet count as a splitting criteria with a threshold set to <=108 platelets/mm^3. It is evident that patients already having a low platelet count on the first visit are at greater risk for the development of an even lower platelet count during

infection. Consequently, we considered the group of patients presenting a higher platelet count at the first visit as more interesting with regard to severity prediction.

Remarkably, we observed that dengue patients having a higher viral load along with a secondary infection are at major risk to become thrombocytopenic during infection. The calculated decision tree explicitly shows that having a high platelet count in combination with a low viremia in an early stage of disease is almost no risk for the progression into severe disease manifestations and shows a relative risk reduction (RRR) of 91%[19]. The RRR value may be an important consideration regarding drug discovery and vaccine development. Our model clearly shows that development of a drug and/or vaccine that may reduce viral load in a very early stage of disease is highly desirable and might prevent the progression to the more severe dengue forms. Furthermore, it highlights secondary infections as a risk factor which is a major challenge in vaccine implementation.

The presented model also supports the finding of another study which clearly showed that high viremia titers 3 days after onset of fever correlated with severe disease 2 days later at the time of defervescence (Vaughn et al., 2000). We observed that there was no correlation regarding viral load and secondary infection on the whole dataset as well as at the specific decision node (results not shown). Hence, we can consider viral load and secondary infections as two independent risk factors that, in combination, may have a significant influence on disease outcome. This is contradictory to the theory of antibody dependent enhancement (ADE) which, strictly applied, defines a direct relationship of viral load with secondary infections (Halstead, 2003). Briefly, it

[19] RRR of people with a high platelet count along with a low viral load is calculated by 1/(Relative Risk of people having a high platelet count along with a high viral load)-1; 1/11.34-1=-0.912 (For values please refer to the results section: Table 3.40; Page 136)

is thought that binding of cross-reactive antibodies from a primary infection enhances uptake of virus via Fc-receptors into target cells and subsequently leads to higher replication causing higher viremia. Our observations would not be against the theory of cross-reactive T-cells which are thought to be triggered upon a secondary infection and which subsequently influence disease outcome (reviewed in (Green and Rothman, 2006)). Looking at the sequence of decision nodes and at the frequency distributions in our proposed model, we hypothesize that a higher viral load might trigger a stronger immune response and a combined secondary infection enhances the risk of an increased inflammatory response which finally causes a more severe disease outcome.

We can assume that the calculated decision tree and its chosen features based on the overall dataset (125 cases) are reliable because constructing a tree on a dataset with smaller sample size (89 cases) results in the exact same tree (SEVERE_EXCYT_89; Figure 3.24; Page 143) with similar performance. The importance of platelet count, viral load and secondary infections in an early stage of infection is further underlined by the models constructed only on the data of hospitalized cases. The two constructed trees (SEVHOSP_TOTAL_71; Figure 3.22; Page 139 and SEVHOSP_EXCYT_52; Figure 3.32; Page 155) (71 cases or 52 cases) both include as a first splitting criteria the platelet count with the same threshold and use as a second decision node secondary infections followed by viral load. The switch in the splitting criteria between secondary infections and viral load might be caused (1) by the smaller datasets used, (2) by the fact that the datasheets represent more similar cases with regard to viral load thus removing statistical significance and creating higher data entropy, (3) by the fact that the frequency distribution of secondary infections still shows the same strength because its dependency on the platelet groups rather than on hospitalization and (4) by the change in class majority meaning that, in this situation, the more severe group

represents the major class. The latter suggests that the further splitting criteria were caused by data overfitting of the tree and do not indicate genuine differences which is finally underlined by comparable weak statistical values on the whole dataset.

By way of example, taking the two correctly classified DHF cases and following them through the classification steps in our decision tree, validates the calculated model. One patient was presented with a platelet count lower than the chosen threshold and was therefore not further classified. The other patient showed a higher platelet count, a higher viral load (meaning the Ct-value was lower than the chosen threshold) and in addition, was positive for IgG antibodies indicating a secondary infection. The simplicity of the model along with the chosen features that are easy to measure as well as the high accuracy of 83% represent advantages with regard to the early assessment of disease severity in dengue patients. Hence, implementation of the model based on clinical data in combination with the severity model may have an important share in tackling severe dengue epidemics.

4.1.3.3 Higher IP-10 Levels show strong Correlation to viral Load and a similar Classifier Performance

In a first round, including cytokine data along with clinical data into the decision tree modeling did not result in an improvement but rather in a decreased overall performance. This suggested data interference of clinical data with immunological data. Thus, we excluded viral load from the analysis and the resulting tree substituted viral load by IP-10 (SEVERE_INCYTA_89; Figure 3.26; 147). Further investigations of the interaction between IP-10 and viral load identified a strongly significant

correlation (R=-0.466; 95%CI: -0.610, -0.292; *p* value < 0.001). This was underlined by the fact that excluding IP-10 from decision tree modeling did result in the same overall performance as calculated by exclusively using clinical data. These findings support that secretion of IP-10 is directly dengue virus induced (Chen et al., 2006; Hsieh et al., 2006). It is not clear from our data whether IP-10 and/or viral load are/is the factor influencing disease outcome. On the other hand, its appearance in all the calculated trees for severity prediction including cytokine data, and its predictive power in distinguishing dengue from other febrile illnesses point towards a possible function as a biomarker for dengue diagnosis as well as severity estimation. However, more epidemiological studies based on patient's serum samples as well as *in vitro* experiments are required to elucidate the exact role of IP-10 in disease pathogenesis.

4.1.3.4 IL-10 as an early Predictor of severe Dengue Infection shows higher Sensitivity than IL-8, but has a narrower predictive Window

In a second round, we calculated a decision tree only based on cytokine data. The goal was to find early immunological predictors of thrombocytopenia in a later stage of disease. This approach resulted in two decision trees which were always using IP-10 as the first splitting criteria but exploited either IL-8 (SEVERE_CYTOA_IL8; Figure 3.28; Page 150) or IL-10 (SEVERE_CYTOA_IL10; Figure 3.29; Page 150) at the second decision node. We observed that the two models still showed a good overall performance but were slightly worse when compared to the clinical trees. Furthermore, the confidence intervals for the AUCs of the ROC curves were comparably large which suggests some instability of the classifiers possibly caused by the smaller dataset. However, the choice of taking IP-10, IL-8 and IL-10 with the defined

threshold into our decision tree modeling is backed by the calculated odds ratios and their confidence intervals. Univariate analysis also showed significant differences of these three variables between the severe and the mild group. It is important to note that analysis of second visit data still detected significant differences of IP-10, IL-8 and IL-10 whereas the discrepancies were reduced between the two groups on the third visit. These findings point towards an important general role of these three cytokines during dengue disease, especially in influencing the development of thrombocytopenia.

We identified IL-8 in combination with IP-10 as a better early predictor for severity which showed an overall accuracy of 75% with a sensitivity of 66%. In addition, it is worthwhile mentioning that IL-8 levels detected on the second visit correlated with viral load on the first visit (R=-0.368; 95%CI: -0.542, -0.162; *p* value = 0.001) which highlights the importance of an early high viral load in disease outcome. Interestingly, we did not detect any significant differences in IL-8 levels between dengue and other febrile illnesses on the 1st visit. This directly supports the finding of another study showing that IL-8 levels of classical dengue fever patients are not elevated within the first 4 days whereas patients suffering from more severe disease already showed elevated IL-8 levels between day one and day four (Raghupathy et al., 1998). Furthermore, a study investigating IL-8 secretion in PBMCs detected delayed peaking of IL-8 levels between days three and four of infection which they explained that a certain threshold of viremia has to be obtained in order to induce chemokine transcription through NS5 (Medin et al., 2005).

IL-8 is secreted by macrophages as well as epithelial cells during an inflammation and is a chemokine that mainly attracts neutrophiles to the site of infection (Baggiolini et

al., 1997). Secretion of IL-8 is linked to pleural effusion (Baggiolini et al., 1995; Pace et al., 1999) and IL-8 levels are elevated in more severe dengue patients (Juffrie et al., 2000; Raghupathy et al., 1998). Induction of IL-8 secretion is controlled by the cytokine network which includes IL-1 and TNF-α. Because of elevated IL-1 levels in dengue patients with regard to our control group but no significant changes in levels TNF-α, we think that IL-1 is the main inducer of IL-8 during dengue infection and that there might be a direct induction by NS5 as proposed by Medin and collaborators (Medin et al., 2005). Expression of IL-8 may be used to counteract the anti-viral effects of innate immunity, allowing further spreading of dengue virus to neighboring cells (Medin et al., 2005). Such a mechanism was also observed to enhance viral replication in other viruses (Lane et al., 2001; Murayama et al., 1994).

Substitution of IL-8 with IL-10 in the final decision tree resulted in a higher sensitivity of 71% but lower overall performance of 73%. This reflects a slightly narrower predictive window but a higher predictive power in detecting severe cases in an early stage of disease. We have to consider the fact that our chosen model is skewed towards the major class which is in our case the mild group. In such class distributions, a better overall performance is achieved by having a higher specificity or lower misclassification rate of mild cases. On the other hand, Increasing the sensitivity and thus decreasing the specificity leads in a higher overall error rate caused by the higher misclassification rate of the major class which decreases the overall performance of the classifier. Thus, we can assume that IL-8 is a better general predictor and that IL-10 is more specific for more severe infections.

There are several reports describing elevation of IL-10 levels during dengue infection (Libraty et al., 2002; Nguyen et al., 2004) and association of IL-10 with thrombocytopenia as well as plasma leakage (Green et al., 1999b) which clearly maintained our calculated model. One explanation for its important role in an early prediction of severity could be its inhibitory effect on IL-2 and therefore on T-cell proliferation. We were able to observe this in the first part where we attempted to find a clinical as well as immunological model that was able to distinguish dengue from other febrile illnesses. This impaired T-cell proliferation may lead to a higher viral load and finally, as observed in our clinical severity model, combined with secondary infections to a more severe disease outcome. Briefly, more IL-10 in an early stage of infection and lower IL-2 may lead to a higher viremia which subsequently affects disease outcome.

In summary, we showed that increased levels of IP-10 in combination with higher serum concentrations of either IL-10 or IL-8 are a good predictive indicator for the development of more severe disease. Due to the fact that adults are at lower risk to develop DHF/DSS it is possible that the observed picture of these three cytokines may be important mostly in disease severity of adult patients which do not show DHF/DSS. This is underlined by the fact that IL-8, IL-10 and IP-10 were still elevated on the second visit and that no differences in TNF-α were observed. Such a picture leads to the assumption that TNF-α is more characteristic for DHF/DSS which, in turn, is very rare in adults and therefore, other immunological factors such as IL-8, IL-10 and IP-10 may have a more important role in disease outcome of adult infections. Another explanation might be the fact that we predicted thrombocytopenia and therefore, only immunological factors specific for a low platelet count were identified.

4.1.3.5 Higher Levels of I-TAC are highly predictive of severe Cases in a more homogenous Population

In a third round, we calculated a cytokine model using only the hospitalized cases (SEVHOSP_INCYTA_52; Figure 3.34; Page 158 and SEVHOSP_CYTOA_52; Figure 3.36; Page 161). Assuming that the hospitalized cases are already more similar in terms of clinical data, we were aiming at identifying markers that might show a higher predictive performance in a more homogenous population. Both trees (either including cytokines along with clinical data or only including cytokine data) showed a good overall performance but had comparable large confidence intervals for the AUC of the ROC curve. A tree based on mixed data showed an accuracy of 83% and used platelet count along with I-TAC as splitting criterium. On the other hand, the tree based only on cytokine data used IL-8 instead of a platelet count as splitting criterium with an accuracy of 73%. This tree supports the important role of IL-8 in disease outcome. Interestingly, both trees showed I-TAC as the last splitting criteria which might be an indicator that I-TAC is more specific for the identification of severity in a more homogenous population. I-TAC is a chemokine that binds the same receptor than IP-10 and also attracts lymphocytes to the site of infection. Currently, the exact role of I-TAC during dengue infection is still not fully understood. Generally, this chemokine is secreted upon virus infections and plays a role in viral clearance. Our analyses did not yield any specific conclusions but were able to observe that I-TAC is induced in dengue virus infections which warrants further investigations.

4.1.4 Concluding Remarks and Outlook

This study is the first to integrate immunological and clinical data into a predictive model for dengue diagnosis as well as to analyze possible risk factors in an early stage of disease by exploiting the approach of classification trees. We showed that decision trees are a powerful tool for data analysis and highlighted the novelty in the dengue field using this kind of approach. The study gave new insights into dengue diagnosis and severity prediction in adult patients. It may have a contribution in the re-assessment of the urgently needed severity definition in adult patients and hopefully has a share in tackling future dengue epidemics.

Using our models for dengue prediction, we were able to distinguish dengue fever from other febrile illnesses with high accuracy. This diagnostic model may be used in epidemic settings where lack of infrastructure and under representation of well trained staffs caused major difficulties in case management. In addition, the combined model of immunological factors and clinical data represents a useful picture of highly important host factors influenced by dengue infection. We were able to provide a link to new mechanisms with already well established findings in dengue pathogenesis and contribute in a better understanding of dengue.

Our models for severity prediction showed interesting aspects and included secondary infection and viral load into one model. The study provides a comprehensible model for the early prediction of dengue severity that may improve the present severity definitions for adults. It could be a useful tool in the assessment of dengue severity and combined to the diagnostic model, may contribute to a better prevention of more

severe disease manifestations in dengue outbreaks. The model using immunological information reveals important cytokines involved in disease severity, especially in thrombocytopenia.

Nevertheless, it is essential that our calculated models have to be further validated in other epidemic settings where other flaviviruses show a higher prevalence than in Singapore. This will hopefully lead to the validation of dengue specific immunological and clinical factors useful for dengue diagnosis along with severity prediction. More *in vitro* and epidemiological studies are required to validate our findings, especially to reveal the complex network of immunological and clinical factors involved in dengue pathogenesis. A better understanding of dengue pathogenesis is urgently needed to finally assist in the development of a drug and/or vaccine.

4.2 The Virus and its Effect on Clinical Outcome

In the third part of my thesis, I pursued the goal to find differences within the viral genome that might correlate with disease severity. For this purpose, we sequenced 94 full length genome sequences and analyzed them with regard to their phylogeny. Finally, clinical parameters found to be important by severity modeling were compared between identified clades within a serotype.

4.2.1 Serotype specific Differences in Clinical Outcome

It is known that distinct serotypes can be associated with different disease outcomes and a recent study compared two periods when either DENV-1 or DENV-2 predominated (Balmaseda et al., 2006). This study revealed serotype specific differences in clinical manifestations which showed that shock and hemorrhage were more likely to occur during the DENV-2 period whereas DENV-1 infection were more strongly associated with vascular permeability. Our study also pointed towards the possibility of serotype specific clinical differences with regard to disease severity. We compared the clinical parameters identified by severity modeling between DENV-1 and DENV-3 (Table 3.64; Page 165 and Table 3.65; Page 166) and found that IFN-α along with viral load collected on the first visit showed higher levels in patients infected by DENV-1. The differences in IFN-α serum levels might be explained by the fact that viruses induce different IFN-α responses and would underline the low levels of phylogenetic similarities observed between the four dengue serotypes (Holmes and Twiddy, 2003). Moreover, IL-8 was additionally increased during DENV-1 infections

on the second visit which, connected to the above presented severity model, might have a major impact on disease outcome.

Our results suggest that serotype 1 had higher potential to cause more severe disease during the 2005 Singaporean dengue outbreak than serotype 3 which is in disagreement with other studies indicating that DENV-3 and DENV-2 may be correlated with higher grades of severity (Nisalak et al., 2003; Vaughn et al., 2000). The observed differences in viral load might point towards a replicative advantage of DENV-1 over DENV-3. A higher virus titer in an early stage of disease as possible risk factors for the progression into more severe disease manifestations was identified in our presented severity model as well as in another study (Vaughn et al., 2000). Moreover, it is thought that mutations affecting viral fitness might be associated with disease outcome which was observed between DENV-2 subtypes (Cologna et al., 2005; Leitmeyer et al., 1999) as well as within serotype 3 variants (Messer et al., 2003; Messer et al., 2002). The finding that DENV-1 still showed higher viral load on the second visit would be in line with another study which showed slower rates of viral clearance in patients with dengue hemorrhagic fever (Wang et al., 2006).

It is intriguing that IL-8 levels were significantly higher in patients suffering from DENV-1 than DENV-3, which underlines the correlation of this chemokine to viral load and supports the assumption that a specific viremia threshold is required for the induction of IL-8 (Medin et al., 2005). Combining this finding to the importance of IL-8 during disease severity, which was clearly shown and extensively discussed in the above calculated severity models, points towards a serotype specific role in disease severity.

Another interesting aspect lies in the idea of a possible direct activation of IL-8 expression by NS5 which was proposed by Medin and collaborators (Medin et al., 2005). Our findings might suggest that DENV-3 NS5 is inferior to DENV-1 NS5 with regard to direct activation of IL-8 expression and it would be highly interesting in comparing the IL-8 in vitro expression induced by DENV-1 NS5 to the expression induced by DENV-3 NS5. Finding differences in such a setting where levels of NS5 are not viral load dependent are essential to exactly elucidate the impact of IL-8 induction during severe dengue infections.

In summary, and with regard to the calculated severity models, we can conclude that serotype specific differences occurred during the 2005 Singaporean dengue outbreak whereby DENV-1 might be considered as the more severe serotype. But we have to keep in mind that disease severity is highly dependent on both host as well as viral factors. Intriguingly, we detected more hospitalized patients suffering from serotype 1 infections than cases which were positive for serotype 3. Furthermore, we must not ignore the possibility that the particular sequence of the infecting serotypes can influence disease severity (Graham et al., 1999). In our study, we did not determine the serotype of the previous exposure of dengue patients suffering from a secondary infection and thus the observed differences in clinical manifestations might be influenced by previous exposures to other serotypes. This might be especially important in Singapore which is a hyperendemic region where co-circulation of multiple serotypes is common. This could be a possible explanation for the contradictory finding, that serotype 1 is more likely to cause severe disease than serotype 3.

4.2.2 Phylogeny of the 2005 Singaporean Dengue Outbreak

4.2.2.1 Isolated Serotype 1 and Serotype 3 Viruses show different phylogenetic Tree Structures suggesting evolutionary Differences

Our phylogenetic analysis revealed a high degree of similarity between strains of the same serotype but identified different tree structures in serotype 1 and serotype 3. DENV-1 isolates did not clearly group into clades of reasonable virus numbers and the tree branches were represented by low evolutionary lengths (Figure 3.38; Page 170 and Figure 3.39; Page 171). On the other hand, DENV-3 isolates showed a clearer tree structure which was defined by longer branch lengths (Figure 3.40; Page 172 and Figure 3.41; Page 151). These two major differences possibly suggest evolutionary differences between DENV-1 and DENV-3 isolates from the 2005 Singaporean outbreak.

In this respect, it is interesting to note that a DENV-3 strain showing a high level of similarity to the EDEN DENV-3 isolates was already detected in 2003 and might represent a common ancestor of the EDEN DENV-3 viruses isolated during the major outbreak in 2005 (Ong Swee Hoe, personal communication). In addition, DENV-1 was predominant in 2004 as well as in 2005 and occurrence of dengue cases was increasing since 2004 and peaking in 2005 with a sharp decrease of dengue cases in 2006 (MOH, 2006). Hence, we can assume that DENV-3 was present in Singapore earlier than DENV-1 and that a newly arisen DENV-1 strain might have triggered the increase of dengue cases. Moreover, the fact that DENV-1 was predominant over DENV-3 might support the idea of a replicative advantage of DENV-1 which was additionally

supported by the observed significant differences in viral load. A possible explanation might lie in host immune recognition. Longer maintenance of DENV-3 in Singapore possibly resulted in a higher proportion of people who had already been exposed to serotype 3, whereas serotype 1 was newer for the host immune system, thus resulting in a higher proportion of susceptible individuals which would finally lead to a replicative advantage of DENV-1 over DENV-3. Such a phenomenon had also been proposed to explain the differences in transmission patterns of DENV-2 and DENV-4 in the Americas (Carrington et al., 2005).

Another study observed that an increase in clade diversity of one serotype leads to a decrease in prevalence of another serotype (Zhang et al., 2005). This might be another possible explanation for the observed differences in the tree structures of DENV-1 and DENV-3. Diversity of DENV-1 strains (0.184%) is higher than diversity of DENV-3 strains (0.118%) and the observation of four clear established clades in DENV-3 stay in contrast to many small subgroups all over distributed in the DENV-1 clade EDEN 1.4. We could suggest that the replicative advantage of DENV-1 led to a higher mutation frequency and thus to an increase in clade diversity whereas the four DENV-3 clades were well established and showed no increase in clade diversity. This was further supported by the longer branch lengths observed in the DENV-3 phylogenetic tree.

Further phylogenetic analysis of EDEN DENV-3 revealed that the strains belonged to DENV-3 subtype 3. It was additionally found that the strains showed a higher level of similarity to a Sri Lankan strain which belonged to clade A of subtype 3 (Ong Swee Hoe, personal communication). This may underline the observation that DENV-3 was

associated to a lower degree of severity because earlier studies showed that subtype 3 clade A (pre-DHF) variants were associated with milder disease manifestations whereas the occurrence of subtype 3 clade B (post-DHF) strains directly correlated with the emergence of DHF in Sri Lanka (Messer et al., 2003; Messer et al., 2002).

The discussed potential reasons for the observed differences in the structures of the phylogenetic trees are hypothetical and more investigation is needed to elucidate the evolutionary mechanisms involved. Furthermore, we have to bear in mind that the mean nucleotide difference was almost negligibly small that an overall comparison of the isolated EDEN strains to geographically and temporally different isolates would naturally result in one EDEN cluster as observed in an overall DENV-3 tree including virus isolates collected throughout the globe (Ong Swee Hoe, personal communication).

4.2.2.2 Serotype 3 Clades show Differences with regard to viral Load and Platelet Count

Comparison of the observed EDEN clades within serotype 3 revealed differences in viral load. More specifically, the significant difference was between EDEN 3.1 (highest viral load) and EDEN 3.2 (lowest viral load) (Figure 3.42; Page 143). This difference is supported by a high bootstrap value and leads to the assumption of a real mutation affecting viral replication and thus disease severity. However, the identified mutation responsible for the clear split between these two clades was a synonymous mutation (Asparagine) in the E protein (EDEN 3.1: 2083C; EDEN 3.2: 2083T) The fact that we do not see any amino acid mutation leads to the assumption that the

observed difference occurred by chance due to only four viruses in clade EDEN 3.2. Furthermore, it is unlikely that a synonymous substitution in the E protein would lead to differences in viral load or host immune responses because the amino acid sequence of the E protein is the main antigenic determinant during dengue virus infection. By way of example, it was reported that a unique amino acid mutation in the E protein of DENV-2 was responsible for the loss of a N-linked glycosylation site which led to a decreased infectivity (Ishak et al., 2001).

4.2.3 Concluding Remarks and Outlook

In summary, we revealed that different serotypes have the ability to influence disease outcome but it turned out to be difficult to find any severity related differences between strains within a serotype. In our study, the lack of a substantial number of DHF/DSS cases combined with the high level of similarity between the dengue isolates were the main reasons that exacerbated the search for viral determinants. In addition, we only had the sequence of one virus that was isolated from a DHF patient. It is worthwhile mentioning that studies determining genome related differences with regard to disease severity mainly used strains isolated during different outbreaks over several years (Leitmeyer et al., 1999; Messer et al., 2003; Rico-Hesse et al., 1997) whereas only a few studies compared more similar viruses isolated from a single outbreak (Pandey and Igarashi, 2000; Rodriguez-Roche et al., 2005). This would support the fact that multi loci mutations would be more likely to be determinants of disease severity than single mutations (Zhou et al., 2006). Thus, it would be interesting to repeat our study during the next big dengue outbreak in Singapore and to compare the sequences to the 2005 EDEN sequences with regard to disease severity. This may

finally lead to a better picture of virus evolution along with viral determinants correlating with disease severity.

Nevertheless, the isolation and sequencing of 94 strains obtained during one outbreak is unique and our study is the first to report such a high number of full length genome sequences. Integration of temporal, geographical and phylogenetic information can be used to determine evolutionary trends along with transmission dynamics that occurred during the 2005 Singaporean dengue outbreak. This would represent a first study of its kind to track virus transmission during a single outbreak.

5 Bibliography

A, A.-N., Berlioz-Arthaud, A., Chow, V., Endy, T., Lowry, K., Mai le, Q., Ninh, T. U., Pyke, A., Reid, M., Reynes, J. M.*, et al.* (2004). Sustained transmission of dengue virus type 1 in the Pacific due to repeated introductions of different Asian strains. Virology *329*, 505-512.

Acuna, E., and Rodriguez, C. (2004). The treatment of missing values and its effect in the classifier accuracy (Berlin-Heidelberg: Springer-Verlag).

Alvarez, D. E., Lodeiro, M. F., Luduena, S. J., Pietrasanta, L. I., and Gamarnik, A. V. (2005). Long-range RNA-RNA interactions circularize the dengue virus genome. J Virol *79*, 6631-6643.

Azeredo, E. L., Zagne, S. M., Santiago, M. A., Gouvea, A. S., Santana, A. A., Neves-Souza, P. C., Nogueira, R. M., Miagostovich, M. P., and Kubelka, C. F. (2001). Characterisation of lymphocyte response and cytokine patterns in patients with dengue fever. Immunobiology *204*, 494-507.

Baggiolini, M., Dewald, B., and Moser, B. (1997). Human chemokines: an update. Annu Rev Immunol *15*, 675-705.

Baggiolini, M., Loetscher, P., and Moser, B. (1995). Interleukin-8 and the chemokine family. Int J Immunopharmacol *17*, 103-108.

Balmaseda, A., Hammond, S. N., Perez, L., Tellez, Y., Saborio, S. I., Mercado, J. C., Cuadra, R., Rocha, J., Perez, M. A., Silva, S.*, et al.* (2006). Serotype-specific differences in clinical manifestations of dengue. Am J Trop Med Hyg *74*, 449-456.

Balmaseda, A., Hammond, S. N., Perez, M. A., Cuadra, R., Solano, S., Rocha, J., Idiaquez, W., and Harris, E. (2005). Short report: assessment of the World Health Organization scheme for classification of dengue severity in Nicaragua. Am J Trop Med Hyg *73*, 1059-1062.

Bandyopadhyay, S., Lum, L. C., and Kroeger, A. (2006). Classifying dengue: a review of the difficulties in using the WHO case classification for dengue haemorrhagic fever. Trop Med Int Health *11*, 1238-1255.

Blackwell, J. L., and Brinton, M. A. (1995). BHK cell proteins that bind to the 3' stem-loop structure of the West Nile virus genome RNA. J Virol *69*, 5650-5658.

Brooks, D. G., Trifilo, M. J., Edelmann, K. H., Teyton, L., McGavern, D. B., and Oldstone, M. B. (2006). Interleukin-10 determines viral clearance or persistence in vivo. Nat Med *12*, 1301-1309.

Burke, D. S., Nisalak, A., Johnson, D. E., and Scott, R. M. (1988). A prospective study of dengue infections in Bangkok. Am J Trop Med Hyg *38*, 172-180.

Carrington, C. V., Foster, J. E., Pybus, O. G., Bennett, S. N., and Holmes, E. C. (2005). Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. J Virol *79*, 14680-14687.

CDC (2005). Dengue Fever.

Chadwick, D., Arch, B., Wilder-Smith, A., and Paton, N. (2006). Distinguishing dengue fever from other infections on the basis of simple clinical and laboratory features: application of logistic regression analysis. J Clin Virol *35*, 147-153.

Chang, T. H., Liao, C. L., and Lin, Y. L. (2006). Flavivirus induces interferon-beta gene expression through a pathway involving RIG-I-dependent IRF-3 and PI3K-dependent NF-kappaB activation. Microbes Infect *8*, 157-171.

Charnsilpa, W., Takhampunya, R., Endy, T. P., Mammen, M. P., Jr., Libraty, D. H., and Ubol, S. (2005). Nitric oxide radical suppresses replication of wild-type dengue 2 viruses in vitro. J Med Virol *77*, 89-95.

Chaturvedi, U. C., Shrivastava, R., and Nagar, R. (2005). Dengue vaccines: problems and prospects. Indian J Med Res *121*, 639-652.

Chen, J. P., Lu, H. L., Lai, S. L., Campanella, G. S., Sung, J. M., Lu, M. Y., Wu-Hsieh, B. A., Lin, Y. L., Lane, T. E., Luster, A. D., and Liao, F. (2006). Dengue virus induces expression of CXC chemokine ligand 10/IFN-gamma-inducible protein 10, which competitively inhibits viral binding to cell surface heparan sulfate. J Immunol *177*, 3185-3192.

Chen, Y. C., Wang, S. Y., and King, C. C. (1999). Bacterial lipopolysaccharide inhibits dengue virus infection of primary human monocytes/macrophages by blockade of virus entry via a CD14-dependent mechanism. J Virol *73*, 2650-2657.

Chiu, W. W., Kinney, R. M., and Dreher, T. W. (2005). Control of translation by the 5'- and 3'-terminal regions of the dengue virus genome. J Virol *79*, 8303-8315.

Clyde, K., Kyle, J. L., and Harris, E. (2006). Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. J Virol *80*, 11418-11431.

Cohen, S. N., and Halstead, S. B. (1966). Shock associated with dengue infection. I. Clinical and physiologic manifestations of dengue hemorrhagic fever in Thailand, 1964. J Pediatr *68*, 448-456.

Cologna, R., Armstrong, P. M., and Rico-Hesse, R. (2005). Selection for virulent dengue viruses occurs in humans and mosquitoes. J Virol *79*, 853-859.

Deen, J. L., Harris, E., Wills, B., Balmaseda, A., Hammond, S. N., Rocha, C., Dung, N. M., Hung, N. T., Hien, T. T., and Farrar, J. J. (2006). The WHO dengue classification and case definitions: time for a reassessment. Lancet *368*, 170-173.

Ehrenkranz, N. J., Ventura, A. K., Cuadrado, R. R., Pond, W. L., and Porter, J. E. (1971). Pandemic dengue in Caribbean countries and the southern United States--past, present and potential problems. N Engl J Med *285*, 1460-1469.

Ejrnaes, M., Filippi, C. M., Martinic, M. M., Ling, E. M., Togher, L. M., Crotty, S., and von Herrath, M. G. (2006). Resolution of a chronic viral infection after interleukin-10 receptor blockade. J Exp Med *203*, 2461-2472.

Fawcett, T. (2003). ROC graphs: Note and practical considerations for researchers (HPL-2003-4) (HP Laboratories).

Filomatori, C. V., Lodeiro, M. F., Alvarez, D. E., Samsa, M. M., Pietrasanta, L., and Gamarnik, A. V. (2006). A 5' RNA element promotes dengue virus RNA synthesis on a circular genome. Genes Dev *20*, 2238-2249.

Fink, J., Gu, F., and Vasudevan, S. G. (2006). Role of T cells, cytokines and antibody in dengue fever and dengue haemorrhagic fever. Rev Med Virol *16*, 263-275.

Graham, R. R., Juffrie, M., Tan, R., Hayes, C. G., Laksono, I., Ma'roef, C., Erlin, Sutaryo, Porter, K. R., and Halstead, S. B. (1999). A prospective seroepidemiologic study on dengue in children four to nine years of age in Yogyakarta, Indonesia I. studies in 1995-1996. Am J Trop Med Hyg *61*, 412-419.

Green, S., and Rothman, A. (2006). Immunopathological mechanisms in dengue and dengue hemorrhagic fever. Curr Opin Infect Dis *19*, 429-436.

Green, S., Vaughn, D. W., Kalayanarooj, S., Nimmannitya, S., Suntayakorn, S., Nisalak, A., Lew, R., Innis, B. L., Kurane, I., Rothman, A. L., and Ennis, F. A. (1999a). Early immune activation in acute dengue illness is related to development of plasma leakage and disease severity. J Infect Dis *179*, 755-762.

Green, S., Vaughn, D. W., Kalayanarooj, S., Nimmannitya, S., Suntayakorn, S., Nisalak, A., Rothman, A. L., and Ennis, F. A. (1999b). Elevated plasma interleukin-10 levels in acute dengue correlate with disease severity. J Med Virol *59*, 329-334.

Gritsun, T. S., and Gould, E. A. (2006). Direct repeats in the 3' untranslated regions of mosquito-borne flaviviruses: possible implications for virus transmission. J Gen Virol *87*, 3297-3305.

Gubler, D. J. (1998). Dengue and dengue hemorrhagic fever. Clin Microbiol Rev *11*, 480-496.

Gubler, D. J. (2002). Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends Microbiol *10*, 100-103.

Gubler, D. J. (2004). The changing epidemiology of yellow fever and dengue, 1900 to 2003: full circle? Comp Immunol Microbiol Infect Dis *27*, 319-330.

Halstead, S. B. (2003). Neutralization and antibody-dependent enhancement of dengue viruses. Adv Virus Res *60*, 421-467.

Hammond, S. N., Balmaseda, A., Perez, L., Tellez, Y., Saborio, S. I., Mercado, J. C., Videa, E., Rodriguez, Y., Perez, M. A., Cuadra, R.*, et al.* (2005). Differences in dengue severity in infants, children, and adults in a 3-year hospital-based study in Nicaragua. Am J Trop Med Hyg *73*, 1063-1070.

Han, P., Zhang, X., Norton, R. S., and Feng, Z. P. (2006). Predicting disordered regions in proteins based on decision trees of reduced amino acid composition. J Comput Biol *13*, 1579-1590.

Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology *143*, 29-36.

Hanna, S. L., Pierson, T. C., Sanchez, M. D., Ahmed, A. A., Murtadha, M. M., and Doms, R. W. (2005). N-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. J Virol *79*, 13262-13274.

Hayes, E. B., and Gubler, D. J. (1992). Dengue and dengue hemorrhagic fever. Pediatr Infect Dis J *11*, 311-317.

Hober, D., Poli, L., Roblin, B., Gestas, P., Chungue, E., Granic, G., Imbert, P., Pecarere, J. L., Vergez-Pascal, R., Wattre, P., and et al. (1993). Serum levels of tumor necrosis factor-alpha (TNF-alpha), interleukin-6 (IL-6), and interleukin-1 beta (IL-1 beta) in dengue-infected patients. Am J Trop Med Hyg *48*, 324-331.

Hober, D., Shen, L., Benyoucef, S., De Groote, D., Deubel, V., and Wattre, P. (1996). Enhanced TNF alpha production by monocytic-like cells exposed to dengue virus antigens. Immunol Lett *53*, 115-120.

Holden, K. L., and Harris, E. (2004). Enhancement of dengue virus translation: role of the 3' untranslated region and the terminal 3' stem-loop domain. Virology *329*, 119-133.

Holmes, E. C., and Twiddy, S. S. (2003). The origin, emergence and evolutionary genetics of dengue virus. Infect Genet Evol *3*, 19-28.

Holsheimer, M., and Siebes, A. (1993). The search for knowledge in databases (Amsterdam, Netherlands: CWI).

Hsieh, M. F., Lai, S. L., Chen, J. P., Sung, J. M., Lin, Y. L., Wu-Hsieh, B. A., Gerard, C., Luster, A., and Liao, F. (2006). Both CXCR3 and CXCL10/IFN-inducible protein 10 are required for resistance to primary infection by dengue virus. J Immunol *177*, 1855-1863.

Ishak, H., Takegami, T., Kamimura, K., and Funada, H. (2001). Comparative sequences of two type 1 dengue virus strains possessing different growth characteristics in vitro. Microbiol Immunol *45*, 327-331.

Jindadamrongwech, S., Thepparit, C., and Smith, D. R. (2004). Identification of GRP 78 (BiP) as a liver cell expressed receptor element for dengue virus serotype 2. Arch Virol *149*, 915-927.

Juffrie, M., van Der Meer, G. M., Hack, C. E., Haasnoot, K., Sutaryo, Veerman, A. J., and Thijs, L. G. (2000). Inflammatory mediators in dengue virus infection in children: interleukin-8 and its relationship to neutrophil degranulation. Infect Immun *68*, 702-707.

Kaufman, B. M., Summers, P. L., Dubois, D. R., Cohen, W. H., Gentry, M. K., Timchak, R. L., Burke, D. S., and Eckels, K. H. (1989). Monoclonal antibodies for dengue virus prM glycoprotein protect mice against lethal dengue infection. Am J Trop Med Hyg *41*, 576-580.

Kaufman, B. M., Summers, P. L., Dubois, D. R., and Eckels, K. H. (1987). Monoclonal antibodies against dengue 2 virus E-glycoprotein protect mice against lethal dengue infection. Am J Trop Med Hyg *36*, 427-434.

Kittigul, L., Temprom, W., Sujirarat, D., and Kittigul, C. (2000). Determination of tumor necrosis factor-alpha levels in dengue virus infected patients by sensitive biotin-streptavidin enzyme-linked immunosorbent assay. J Virol Methods *90*, 51-57.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, In International Joint Conference on Artificial Intelligence (IJCAI), 1995.

Kothari, R., and Dong, M. (2000). Decision Trees for Classification: A Review and Some New Results. World Scientific.

Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G.*, et al.* (2002). Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. Cell *108*, 717-725.

Kumar, S., Tamura, K., and Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform *5*, 150-163.

Kuno, G., Chang, G. J., Tsuchiya, K. R., Karabatsos, N., and Cropp, C. B. (1998). Phylogeny of the genus Flavivirus. J Virol *72*, 73-83.

Kurane, I., Innis, B. L., Nimmannitya, S., Nisalak, A., Meager, A., Janus, J., and Ennis, F. A. (1991). Activation of T lymphocytes in dengue virus infections. High levels of soluble interleukin 2 receptor, soluble CD4, soluble CD8, interleukin 2, and interferon-gamma in sera of children with dengue. J Clin Invest *88*, 1473-1480.

Kurane, I., Meager, A., and Ennis, F. A. (1986). Induction of interferon alpha and gamma from human lymphocytes by dengue virus-infected cells. J Gen Virol *67 (Pt 8)*, 1653-1661.

Lane, B. R., Lore, K., Bock, P. J., Andersson, J., Coffey, M. J., Strieter, R. M., and Markovitz, D. M. (2001). Interleukin-8 stimulates human immunodeficiency virus type 1 replication and is a potential new target for antiretroviral therapy. J Virol *75*, 8195-8202.

Lee, C. J., Liao, C. L., and Lin, Y. L. (2005). Flavivirus activates phosphatidylinositol 3-kinase signaling to block caspase-dependent apoptotic cell death at the early stage of virus infection. J Virol *79*, 8388-8399.

Lei, H. Y., Yeh, T. M., Liu, H. S., Lin, Y. S., Chen, S. H., and Liu, C. C. (2001). Immunopathogenesis of dengue virus infection. J Biomed Sci *8*, 377-388.

Leitmeyer, K. C., Vaughn, D. W., Watts, D. M., Salas, R., Villalobos, I., de, C., Ramos, C., and Rico-Hesse, R. (1999). Dengue virus structural differences that correlate with pathogenesis. J Virol *73*, 4738-4747.

Li, K. B. (2003). ClustalW-MPI: ClustalW analysis using distributed and parallel computing. Bioinformatics *19*, 1585-1586.

Libraty, D. H., Endy, T. P., Houng, H. S., Green, S., Kalayanarooj, S., Suntayakorn, S., Chansiriwongs, W., Vaughn, D. W., Nisalak, A., Ennis, F. A., and Rothman, A. L. (2002). Differing influences of virus burden and immune activation on disease severity in secondary dengue-3 virus infections. J Infect Dis *185*, 1213-1221.

Lin, C. F., Wan, S. W., Cheng, H. J., Lei, H. Y., and Lin, Y. S. (2006). Autoimmune pathogenesis in dengue virus infection. Viral Immunol *19*, 127-132.

Littaua, R., Kurane, I., and Ennis, F. A. (1990). Human IgG Fc receptor II mediates antibody-dependent enhancement of dengue virus infection. J Immunol *144*, 3183-3186.

Low, G. H., Ooi, E. E., Tolfvenstam, T., Leo, Y. S., Hibberd, M. L., Ng, L. C., Lai, Y. L., Yap, S. L., Li, S. C., Vasudevan, S. G., and Ong, A. (2006). Early Dengue Infection and Outcome Study (EDEN) - Study Design and Preliminary Findings (Singapore), pp. 25.

Mackenzie, J. S., Gubler, D. J., and Petersen, L. R. (2004). Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nat Med *10*, S98-109.

Malavige, G. N., Velathanthiri, V. G., Wijewickrama, E. S., Fernando, S., Jayaratne, S. D., Aaskov, J., and Seneviratne, S. L. (2006). Patterns of disease among adults hospitalized with dengue infections. Qjm *99*, 299-305.

Mathew, A., Kurane, I., Green, S., Vaughn, D. W., Kalayanarooj, S., Suntayakorn, S., Ennis, F. A., and Rothman, A. L. (1999). Impaired T cell proliferation in acute dengue infection. J Immunol *162*, 5609-5615.

McNeil, B. J., and Hanley, J. A. (1984). Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making *4*, 137-150.

Medin, C. L., Fitzgerald, K. A., and Rothman, A. L. (2005). Dengue virus nonstructural protein NS5 induces interleukin-8 transcription and secretion. J Virol *79*, 11053-11061.

Meltzer, M. I., Rigau-Perez, J. G., Clark, G. G., Reiter, P., and Gubler, D. J. (1998). Using disability-adjusted life years to assess the economic impact of dengue in Puerto Rico: 1984-1994. Am J Trop Med Hyg *59*, 265-271.

Messer, W. B., Gubler, D. J., Harris, E., Sivananthan, K., and de Silva, A. M. (2003). Emergence and global spread of a dengue serotype 3, subtype III virus. Emerg Infect Dis *9*, 800-809.

Messer, W. B., Vitarana, U. T., Sivananthan, K., Elvtigala, J., Preethimala, L. D., Ramesh, R., Withana, N., Gubler, D. J., and De Silva, A. M. (2002). Epidemiology of dengue in Sri Lanka before and after the emergence of epidemic dengue hemorrhagic fever. Am J Trop Med Hyg *66*, 765-773.

Ministry of Health, S. (2006). Dengue Surveillance in Singapore, 2005. Epidemiol News Bull, 9-14.

Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2003). A ligand-binding pocket in the dengue virus envelope glycoprotein. Proc Natl Acad Sci U S A *100*, 6986-6991.

Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2004). Structure of the dengue virus envelope protein after membrane fusion. Nature *427*, 313-319.

Mongkolsapaya, J., Dejnirattisai, W., Xu, X. N., Vasanawathana, S., Tangthawornchaikul, N., Chairunsri, A., Sawasdivorn, S., Duangchinda, T., Dong, T., Rowland-Jones, S.*, et al.* (2003). Original antigenic sin and apoptosis in the pathogenesis of dengue hemorrhagic fever. Nat Med *9*, 921-927.

Moore, A. W. (2007). Statistical Datamining Tutorials.

Mukhopadhyay, S., Kuhn, R. J., and Rossmann, M. G. (2005). A structural perspective of the flavivirus life cycle. Nat Rev Microbiol *3*, 13-22.

Murayama, T., Kuno, K., Jisaki, F., Obuchi, M., Sakamuro, D., Furukawa, T., Mukaida, N., and Matsushima, K. (1994). Enhancement human cytomegalovirus replication in a human lung fibroblast cell line by interleukin-8. J Virol *68*, 7582-7585.

Murgue, B., Cassar, O., Deparis, X., Guigon, M., and Chungue, E. (1998). Implication of macrophage inflammatory protein-1alpha in the inhibition of human haematopoietic progenitor growth by dengue virus. J Gen Virol *79 (Pt 8)*, 1889-1893.

Myat Thu, H., Lowry, K., Jiang, L., Hlaing, T., Holmes, E. C., and Aaskov, J. (2005). Lineage extinction and replacement in dengue type 1 virus populations are due to stochastic events rather than to natural selection. Virology *336*, 163-172.

Navarro-Sanchez, E., Altmeyer, R., Amara, A., Schwartz, O., Fieschi, F., Virelizier, J. L., Arenzana-Seisdedos, F., and Despres, P. (2003). Dendritic-cell-specific ICAM3-

grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. EMBO Rep *4*, 723-728.

Navarro-Sanchez, E., Despres, P., and Cedillo-Barron, L. (2005). Innate immune responses to dengue virus. Arch Med Res *36*, 425-435.

Newton, E. A., and Reiter, P. (1992). A model of the transmission of dengue fever with an evaluation of the impact of ultra-low volume (ULV) insecticide applications on dengue epidemics. Am J Trop Med Hyg *47*, 709-720.

Nguyen, T. H., Lei, H. Y., Nguyen, T. L., Lin, Y. S., Huang, K. J., Le, B. L., Lin, C. F., Yeh, T. M., Do, Q. H., Vu, T. Q.*, et al.* (2004). Dengue hemorrhagic fever in infants: a study of clinical and cytokine profiles. J Infect Dis *189*, 221-232.

NIH (2007). MedlinePlus.

Nimmannitya, S., Thisyakorn, U., and Hemsrichart, V. (1987). Dengue haemorrhagic fever with unusual manifestations. Southeast Asian J Trop Med Public Health *18*, 398-406.

Nisalak, A., Endy, T. P., Nimmannitya, S., Kalayanarooj, S., Thisayakorn, U., Scott, R. M., Burke, D. S., Hoke, C. H., Innis, B. L., and Vaughn, D. W. (2003). Serotype-specific dengue virus circulation and dengue disease in Bangkok, Thailand from 1973 to 1999. Am J Trop Med Hyg *68*, 191-202.

NITD (2005). Dengue Digest Volume 2.

Ooi, E. E., Goh, K. T., and Chee Wang, D. N. (2003). Effect of increasing age on the trend of dengue and dengue hemorrhagic fever in Singapore. Int J Infect Dis *7*, 231-232.

Ooi, E. E., Goh, K. T., and Gubler, D. J. (2006). Dengue prevention and 35 years of vector control in Singapore. Emerg Infect Dis *12*, 887-893.

Pace, E., Gjomarkaj, M., Melis, M., Profita, M., Spatafora, M., Vignola, A. M., Bonsignore, G., and Mody, C. H. (1999). Interleukin-8 induces lymphocyte chemotaxis into the pleural space. Role of pleural macrophages. Am J Respir Crit Care Med *159*, 1592-1599.

Pacsa, A. S., Agarwal, R., Elbishbishi, E. A., Chaturvedi, U. C., Nagar, R., and Mustafa, A. S. (2000). Role of interleukin-12 in patients with dengue hemorrhagic fever. FEMS Immunol Med Microbiol *28*, 151-155.

Pandey, B. D., and Igarashi, A. (2000). Severity-related molecular differences among nineteen strains of dengue type 2 viruses. Microbiol Immunol *44*, 179-188.

Peters, R. P., Twisk, J. W., van Agtmael, M. A., and Groeneveld, A. B. (2006). The role of procalcitonin in a decision tree for prediction of bloodstream infection in febrile patients. Clin Microbiol Infect *12*, 1207-1213.

Pokidysheva, E., Zhang, Y., Battisti, A. J., Bator-Kelly, C. M., Chipman, P. R., Xiao, C., Gregorio, G. G., Hendrickson, W. A., Kuhn, R. J., and Rossmann, M. G. (2006). Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN. Cell *124*, 485-493.

Quinlan, J. R. (1993). C4.5 : programs for machine learning (San Mateo, Calif.: Morgan Kaufmann Publishers).

Raghupathy, R., Chaturvedi, U. C., Al-Sayer, H., Elbishbishi, E. A., Agarwal, R., Nagar, R., Kapoor, S., Misra, A., Mathur, A., Nusrat, H.*, et al.* (1998). Elevated levels of IL-8 in dengue hemorrhagic fever. J Med Virol *56*, 280-285.

Rhee, S. Y., Taylor, J., Wadhera, G., Ben-Hur, A., Brutlag, D. L., and Shafer, R. W. (2006). Genotypic predictors of human immunodeficiency virus type 1 drug resistance. Proc Natl Acad Sci U S A *103*, 17355-17360.

Rico-Hesse, R. (1990). Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. Virology *174*, 479-493.

Rico-Hesse, R. (2003). Microevolution and virulence of dengue viruses. Adv Virus Res *59*, 315-341.

Rico-Hesse, R., Harrison, L. M., Salas, R. A., Tovar, D., Nisalak, A., Ramos, C., Boshell, J., de Mesa, M. T., Nogueira, R. M., and da Rosa, A. T. (1997). Origins of dengue type 2 viruses associated with increased pathogenicity in the Americas. Virology *230*, 244-251.

Rigau-Perez, J. G. (2006). Severe dengue: the need for new case definitions. Lancet Infect Dis *6*, 297-302.

Rigau-Perez, J. G., Clark, G. G., Gubler, D. J., Reiter, P., Sanders, E. J., and Vorndam, A. V. (1998). Dengue and dengue haemorrhagic fever. Lancet *352*, 971-977.

Rodriguez-Roche, R., Alvarez, M., Gritsun, T., Halstead, S., Kouri, G., Gould, E. A., and Guzman, M. G. (2005). Virus evolution during a severe dengue epidemic in Cuba, 1997. Virology *334*, 154-159.

Rothman, K. J., and Greenland, S. (1998). Modern epidemiology, 2nd edn (Philadelphia (Pa.): Lippincott-Raven).

Schramm, A., Schulte, J. H., Klein-Hitpass, L., Havers, W., Sieverts, H., Berwanger, B., Christiansen, H., Warnat, P., Brors, B., Eils, J.*, et al.* (2005). Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling. Oncogene *24*, 7902-7912.

Shirato, K., Miyoshi, H., Goto, A., Ako, Y., Ueki, T., Kariwa, H., and Takashima, I. (2004). Viral envelope protein glycosylation is a molecular determinant of the neuroinvasiveness of the New York strain of West Nile virus. J Gen Virol *85*, 3637-3645.

Tamura, K., and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol *10*, 512-526.

Tassaneetrithep, B., Burgess, T. H., Granelli-Piperno, A., Trumpfheller, C., Finke, J., Sun, W., Eller, M. A., Pattanapanyasat, K., Sarasombath, S., Birx, D. L.*, et al.* (2003). DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. J Exp Med *197*, 823-829.

Twiddy, S. S., Holmes, E. C., and Rambaut, A. (2003). Inferring the rate and time-scale of dengue virus evolution. Mol Biol Evol *20*, 122-129.

TwoCrowsCorporation (2005). Introduction to Data Mining & Knowledge Discovery, Third Edition edn).

Vaughn, D. W., Green, S., Kalayanarooj, S., Innis, B. L., Nimmannitya, S., Suntayakorn, S., Endy, T. P., Raengsakulrach, B., Rothman, A. L., Ennis, F. A., and Nisalak, A. (2000). Dengue viremia titer, antibody response pattern, and virus serotype correlate with disease severity. J Infect Dis *181*, 2-9.

Wang, W. K., Chen, H. L., Yang, C. F., Hsieh, S. C., Juan, C. C., Chang, S. M., Yu, C. C., Lin, L. H., Huang, J. H., and King, C. C. (2006). Slower rates of clearance of viral load and virus-containing immune complexes in patients with dengue hemorrhagic fever. Clin Infect Dis *43*, 1023-1030.

Wearing, H. J., and Rohani, P. (2006). Ecological and immunological determinants of dengue epidemics. Proc Natl Acad Sci U S A *103*, 11802-11807.

Weaver, S. C., and Barrett, A. D. (2004). Transmission cycles, host range, evolution and emergence of arboviral disease. Nat Rev Microbiol *2*, 789-801.

WHO (1997). Dengue haemorrhagic fever: diagnosis, treatment, prevention and control, 2nd edn (Geneva).

WHO (2002). The World Health Report 2002 - Reducing Risks, Promoting Healthy Life (Geneva, Switzerland: World Health Organization).

Wilder-Smith, A., Foo, W., Earnest, A., Sremulanathan, S., and Paton, N. I. (2004). Seroepidemiology of dengue in the adult population of Singapore. Trop Med Int Health *9*, 305-308.

Wu, S. F., Liao, C. L., Lin, Y. L., Yeh, C. T., Chen, L. K., Huang, Y. F., Chou, H. Y., Huang, J. L., Shaio, M. F., and Sytwu, H. K. (2003). Evaluation of protective efficacy and immune mechanisms of using a non-structural protein NS1 in DNA vaccine against dengue 2 virus in mice. Vaccine *21*, 3919-3929.

Wu, S. J., Grouard-Vogel, G., Sun, W., Mascola, J. R., Brachtel, E., Putvatana, R., Louder, M. K., Filgueira, L., Marovich, M. A., Wong, H. K.*, et al.* (2000). Human skin Langerhans cells are targets of dengue virus infection. Nat Med *6*, 816-820.

Yeh, W. T., Chen, R. F., Wang, L., Liu, J. W., Shaio, M. F., and Yang, K. D. (2006). Implications of previous subclinical dengue infection but not virus load in dengue hemorrhagic fever. FEMS Immunol Med Microbiol *48*, 84-90.

Yu, C. Y., Hsu, Y. W., Liao, C. L., and Lin, Y. L. (2006). Flavivirus Infection Activates the XBP1 Pathway of the Unfolded Protein Response To Cope with Endoplasmic Reticulum Stress. J Virol *80*, 11868-11880.

Yu, L., and Markoff, L. (2005). The topology of bulges in the long stem of the flavivirus 3' stem-loop is a major determinant of RNA replication competence. J Virol *79*, 2309-2324.

Zeng, L., Falgout, B., and Markoff, L. (1998). Identification of specific nucleotide sequences within the conserved 3'-SL in the dengue type 2 virus genome required for replication. J Virol *72*, 7510-7522.

Zhang, C., Mammen, M. P., Jr., Chinnawirotpisan, P., Klungthong, C., Rodpradit, P., Monkongdee, P., Nimmannitya, S., Kalayanarooj, S., and Holmes, E. C. (2005). Clade replacements in dengue virus serotypes 1 and 3 are associated with changing serotype prevalence. J Virol *79*, 15123-15130.

Zhou, Y., Mammen, M. P., Jr., Klungthong, C., Chinnawirotpisan, P., Vaughn, D. W., Nimmannitya, S., Kalayanarooj, S., Holmes, E. C., and Zhang, C. (2006). Comparative analysis reveals no consistent association between the secondary structure of the 3'-

untranslated region of dengue viruses and disease syndrome. J Gen Virol *87*, 2595-2603.

Zivny, J., DeFronzo, M., Jarry, W., Jameson, J., Cruz, J., Ennis, F. A., and Rothman, A. L. (1999). Partial agonist effect influences the CTL response to a heterologous dengue virus serotype. J Immunol *163*, 2754-2760.

6 Appendix

6.1 Sequencing Report and Optimization

6.1.1 Virus Propagation

Viruses where propagated by passages in C6/36 mosquito cell culture. Viral titer was measured using the plaque assay technique described in Liu et al 2004 Report. Viral titers of at least $1x10^6$ were found to be required for optimal success in subsequent RT-PCR steps.

All procedures were performed in a class II biosafety cabinet (BSC) in a certified biosafety level 2 laboratory designated for Dengue virus work at NITD, in accordance with NITD BSL2 policy 5.1 and NITD Work Instructions on working with Dengue virus.

6.1.2 RNA Extraction

Viral RNA was extracted from the culture supernatant using the QiaAmp extraction protocol:

1. Pipet 560 µl of prepared Buffer AVL containing carrier RNA(stored at 4°C, before usage let crystals dissolve by putting in 37°C water bath) into a 1.5 ml microcentrifuge tube. Aliquoting buffer reduces negative effects of repeated thaw-freezing.
2. Add 140 µl virus-cell-culture to the Buffer AVL carrier RNA in the microcentrifuge tube. Mix by pulse-vortexing for 15 s.
3. Incubate at room temperature for 10 min.
4. Briefly centrifuge the tube to remove drops from the inside of the lid.
5. Add 560 µl of ethanol (96-100%) to the sample, and mix by pulse-vortexing for 15 s. After mixing, briefly centrifuge the tube to remove drops from inside the lid.
6. Carefully apply 630 µl of the solution from step 5 to the QIAamp Mini spin column (in a 2 ml collection tube) without wetting the rim. Close the cap, and centrifuge at 8000 rpm for 1 min. Place the QIAamp spin column into a clean 2 ml collection tube, and discard the tube containing the filtrate.
7. Carefully open the QIAamp Mini spin column, and repeat step 6. Repeat this step until all of the lysate has been loaded onto the spin column.
8. Carefully open the QIAamp Mini spin column, and add 500 µl of Buffer AW1. Close the cap, and centrifuge at 8000 rpm for 1 min. Place the QIAamp Mini spin column in a clean 2 ml collection tube (provided), and discard the tube containing the filtrate.
9. Carefully open the QIAamp Mini spin column, and add 500 µl of Buffer AW2. Close the cap and centrifuge at full speed 13,000 rpm for 3 min.
10. Place the QIAamp Mini spin column in a new 1.5 ml collection tube (not provided), and discard the old collection tube with the filtrate. Centrifuge at full speed for 1 min.
11. Place the QIAamp Mini spin column in a clean 1.5 ml microcentrifuge tube (not provided). Discard the old collection tube containing the filtrate. Carefully open the QIAamp spin column and add 60 µl of Buffer AVE equilibrated to room

temperature. Close the cap, and incubate at room temperature for 1 min. Centrifuge at 8000 rpm for 1 min. Store at -80°C

6.1.3 RT-PCR of DENV RNA

The extracted RNA was copied to cDNA using an RT reaction followed by PCR amplification. The virus was amplified as five overlapping fragments. Unless specified below the same conditions were used for all five fragments.

6.1.3.1 RT of DENV RNA

RT of fragments 1, 3 and 4 was performed in a single tube. RT of fragments 2 and 5 was performed in separate tubes. Samples were kept on ice and pipetting was carried out using RNAase free filter tips.

The reaction volume was 20µL and consisted of 8µL of master mix 1, 6 µL of RNA and 6µL of master mix 2. Master mix 1 and the RNA were pre-incubated at 65°C for 5 minutes and then on ice for 3 minutes before the addition of master mix 2.

Table 6.1: RT Master Mix 1

Reagent	Volume
DTT	1µL
dNTPs	1µL
Antisense primer	1µL
RNAase free H_2O	5µL

Table 6.2: RT Master Mix 2

Reagent	Volume
5xFS buffer	4µL
RNAse out	1µL
Super Script III	1µL

The following primers were used in the RT reaction.

Table 6.3: Overview of RT primers

Serotype	Fragment	Primer	Sequence	Postion
DENV-1	F1	d1a17	CCAATGGCYGCTGAYAGTCT	2540-2559
	F2	d1a13a	CCTACTTGGGACCTGCCCA	4611-4629
		d1a13b	GGGACCTGCCCAACAGTCCT	4603-4622
		d1a13c	TGCCCAACAGTCCTCTCTGC	4597-4616
	F3	d1a9	CCAGTYARCACAGCTATCAAAGC	6551-6573
	F4	d1a4	CACTCCACTGAGTGAATTCTCTCT	9068-9091
	F5	d1a23	AGAACCTGTTGATTCAACAG	10716-10735
DENV-3	F1	d3a18	GATTCCTATCGCAATGCATG	2361-2380
	F2	d3a14	ACTGTGATCATTAARTTGTGGGA	4334-4356
	F3	d3a10	GCYGCAAARTCCTTGAATTCCT	6339-6360
	F4	d3a6	GCATTRACATGTCGRGTTCC	8342-8361
	F5	d3a23	AGAACCTGTTGATTCAACAG	10688-10707

Table 6.4: RT reaction conditions

Step	Temperature (°C)	Time
1	65	5 min
2	0 (ice)	1 min
3	25	5 min
4	42	1 hr
5	70	15 min
6	4	Hold

6.1.3.2 PCR of DENV cDNA

The five RT fragments were subsequently amplified by PCR. All components were kept on ice. Each reaction was performed in a volume of 50µL consisting of the following reagents:

Table 6.5: Overview of PCR Reagents

Reagent	Volume
Pfu buffer	5 µL
dNTP's	1 µL
H_2O	39 µL
Pfu Turbopolymerase	1 µL
Primer (antisense)	1 µL
Primer (sense)	1 µL
cDNA	2 µL

The following PCR primers were used to amplify the RT fragments.

Table 6.6: Overview of PCR primers

Serotype	Fragment	Primer	Sequence	Position
DENV-1	F1	d1a17	CCAATGGCYGCTGAYAGTCT	2540-2559
		d1s22	AGTTGTTAGTCTACGTGGAC	1-20
	F2	d1a13	TTCCACTTCYGGAGGGCT	4544-4561
		d1s6	GGYTCTATAGGAGGRGTGTTCAC	2201-2223
	F3	d1a9	CCAGTYARCACAGCTATCAAAGC	6551-6573
		d1s10	RGCYGGSCCACTAATAGCT	4213-4231
	F4	d1a4	CACTCCACTGAGTGAATTCTCTCT	9068-9091
		d1s14	ATGGRGAAAGGAACAACCAG	6216-6235
	F5	d1a23	AGAACCTGTTGATTCAACAG	10716-10735
		d1s18c	GGTGAGTCCTCTCCAAACCC	8015-8034
DENV-3	F1	d3a18	GATTCCTATCGCAATGCATG	2361-2380
		d3s23	AGTTGTTAGTCTACGTGGAC	1-20
	F2	d3a14	ACTGTGATCATTAARTTGTGGGA	4334-4356
		d3s5	TGAACCTCCTTTTGGGGAA	2035-2053
	F3	d3a10	GCYGCAAARTCCTTGAATTCCT	6339-6360
		d3s9	GAAAACAGATTGGCTCCCAA	4030-4049
	F4	d3a6	GCATTRACATGTCGRGTTCC	8342-8361
		d3s13	CCAGCTCTCTTTGAACCAGAAA	6032-6053
	F5	d3a23	AGAACCTGTTGATTCAACAG	10688-10707
		d3s17	CAACAGTGGAAGAAAGCAGAAC	8025-8046

PCR of the RT fragments was carried out in a thermal cycler with the following program. For DENV-1 a ramp speed of 9600 was used for fragment 5. All other fragments from DENV-1 and DENV-3 were cycled at maximum ramp speed.

Table 6.7: PCR program

Serotype	Fragment	Cycles	Initial	Cycle			Extend	Hold
DENV-1	F1	35	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F2	35	95°C 2 min	95°C 30 sec	58°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F3	35	95°C 2 min	95°C 30 sec	63.1°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F4	35	95°C 2 min	95°C 30 sec	58°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F5	40	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 4 min	72°C 2 min	4°C
DENV-2	F1	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F2	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F3	45	95°C 2 min	95°C 30 sec	63.1°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F4	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C
	F5	45	95°C 2 min	95°C 30 sec	55°C 1 min	72°C 3.5 min	72°C 2 min	4°C

6.1.3.3 Gel Electrophoresis

Gel electrophoresis was used for visualization of products as well as gel extraction and purification of products. The products were separated in 1% agarose TBE gels after visualization of ethidium bromide stained bands under UV light.

6.1.3.4 Gel Extraction of DNA

DNA was extracted from bands excised from agarose gels. A Qiagen QiaQuick extraction kit was used with the following protocol:

1. Weigh the gel slice. Add 3 volumes of Buffer QG to 1 volume of gel (100 mg ~ 100 µl).
2. Incubate at 50°C for 10 min (or until the gel slice has completely dissolved). To help dissolve gel, mix by vortexing the tube every 2-3 min during the incubation. IMPORTANT: Solubilize agarose completely.

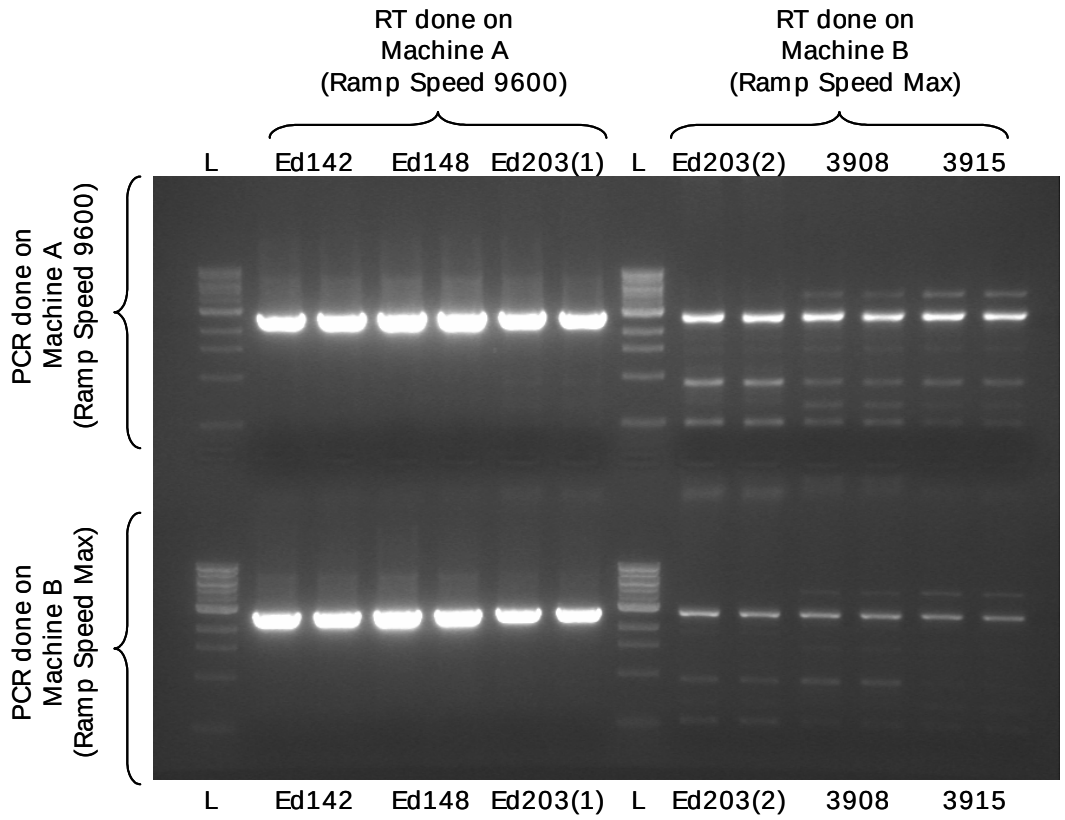
3. After the gel slice has dissolved completely, check that the color of the mixture is yellow .
4. Place a QIAquick spin column in a provided 2 ml collection tube.
5. To bind DNA, apply the sample to the QIAquick column, and centrifuge for 1 min. For sample volumes of more than 750 µl, simply load and spin again.
6. Discard flow-through and place QIAquick column back in the same collection tube. Collection tubes are re-used to reduce plastic waste.
7. Add 0.5 ml of Buffer QG to QIAquick column and centrifuge for 1 min.
8. To wash, add 0.7 ml of Buffer PE to QIAquick column. Let the column stand 5 min after addition of Buffer PE, before centrifuging for 1 min.
9. Discard the flow-through and centrifuge the QIAquick column for an additional 1 min at 13,000 rpm
10. Place QIAquick column into a new clean 1.5 ml microcentrifuge tube.
11. To elute DNA, add 50 µl of Buffer EB (10 mM Tris•Cl, pH 8.5) or H2O to the center of the QIAquick membrane, wait 1 min and centrifuge the column for 1 min. Ensure that the elution buffer is dispensed directly onto the QIAquick membrane for complete elution of bound DNA.

6.1.4 Results (or Method validation)

6.1.4.1 Optimization of Ramp Speed.

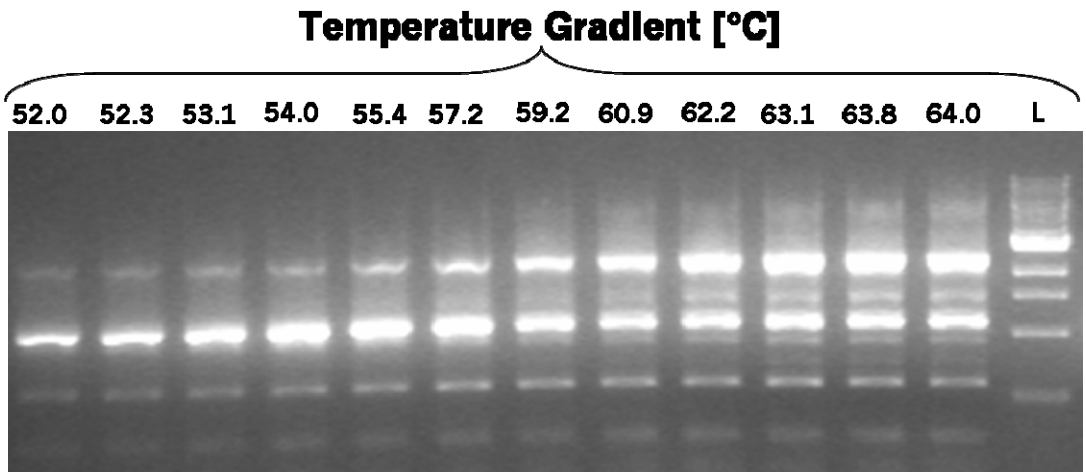
For DENV-1 fragment 5 we found that the specificity of the PCR step of RT-PCR was sensitive to the "ramp speed" (the time it takes to heat and cool the thermal block) of the machine. When the ramp speed was set to the maximum for the machine non-specific products formed and the desired band was weak. The ideal setting was found to be 9600. This result is shown in figure 6.1. Notably this was not found to be the case on all machines even those of the same make. As can be seen from the results of Machine B in figure 6.1 the use of maximum ramp speed was best in this machine. We strongly recommend that the ideal ramp speed is explored for each machine used.

Figure 6.1: Optimization of ramp speed.



L indicates a 1KB ladder. Ed142, Ed148, Ed203, 3908 and 3915 indicate independent isolates.

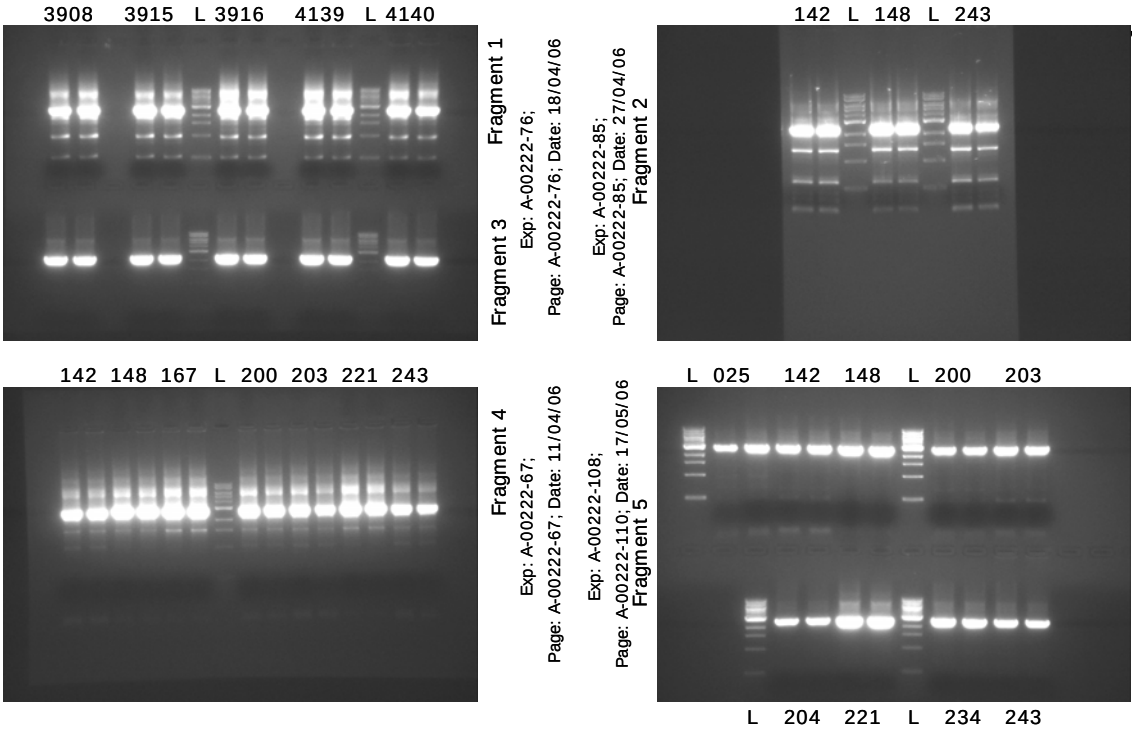
Figure 6.2: Temperature Gradient of Fragment 3 Serotype 3



6.1.4.2 Typical RT-PCR results

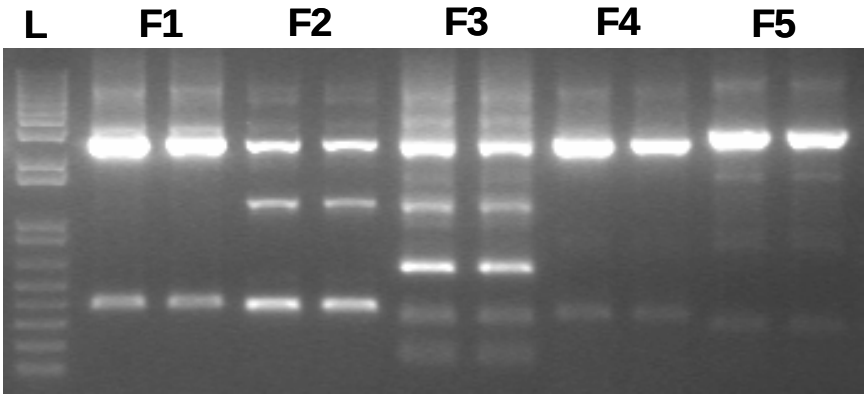
Figure 6.2 shows a typical RT-PCR result for DENV-1 prior to gel purification. Figure 6.3 shows a typical result for DENV-3

Figure 6.3: DENV-1 RT-PCR result



L indicates a lane with a 1KB ladder. Other lanes have viral isolates indicated by their number.

Figure 6.4: DENV-3 RT-PCR result



L indicates a 1KB ladder, F1-F5 indicate fragments 1 through 5.

6.1.4.3 Typical gel purification results

Because of the presence of non-specific products bands gel purification was used to clean up the fragments. The result of gel purification of DENV-1 fragments is shown in figure 6.4 while DENV-3 results are shown in figure 6.5

Figure 6.5: Gel purification of DENV-1 RT-PCR products

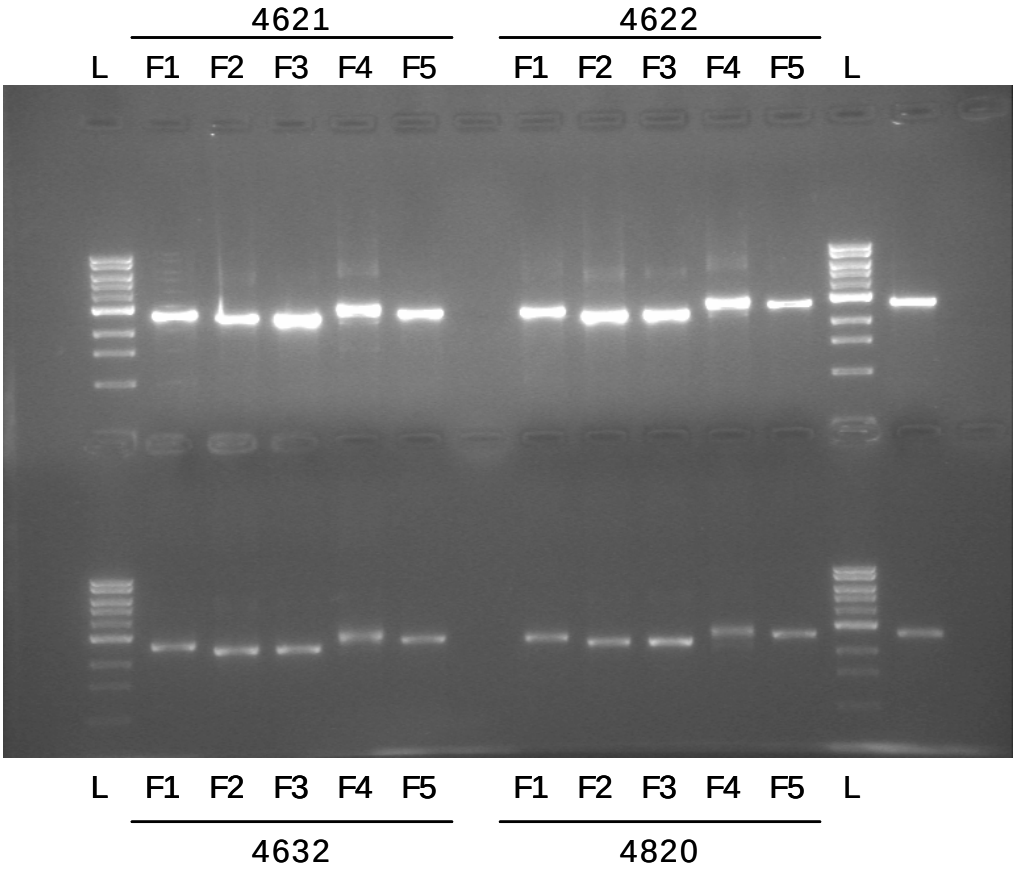
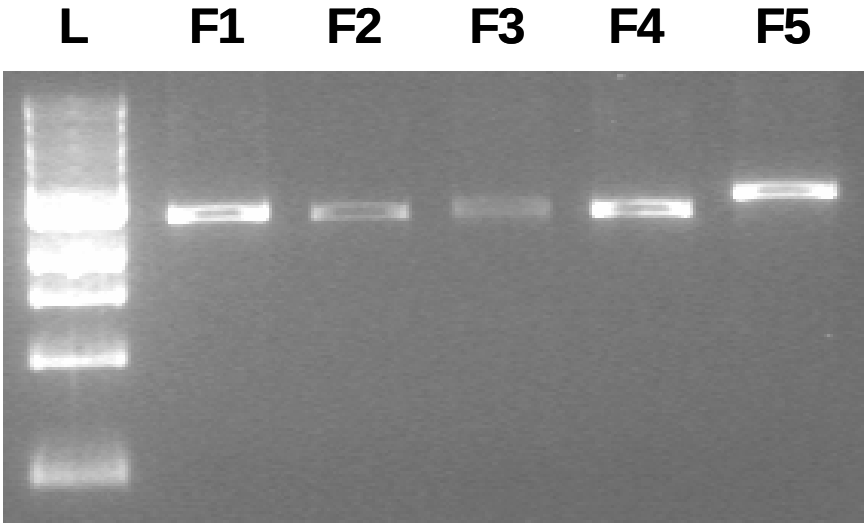


Figure 6.6: Gel purification of DENV-3 RT-PCR products



6.2 Identities of sequenced Virus Isolates

D105k2398DK1	D1SG05ED111	D1	Singapore	2005
D105k2402DK1	D1SG05ED114	D1	Singapore	2005
D105k2887DK1	D1SG05ED134	D1	Singapore	2005
D105k2895DK1	D1SG05ED142	D1	Singapore	2005
D105k2901DK1	D1SG05ED148	D1	Singapore	2005
D105k2916DK1	D1SG05ED167	D1	Singapore	2005
D105k2928DK1	D1SG05ED179	D1	Singapore	2005
D105k3297DK1	D1SG05ED200	D1	Singapore	2005
D105k3300DK1	D1SG05ED203	D1	Singapore	2005
D105k3301DK1	D1SG05ED204	D1	Singapore	2005
D105k3318DK1	D1SG05ED221	D1	Singapore	2005
D105k3886DK1	D1SG05ED234	D1	Singapore	2005
D105k3894DK1	D1SG05ED243	D1	Singapore	2005
D105k3903DK1		D1	Singapore	2005
D105k3904DK1		D1	Singapore	2005
D105k3905DK1		D1	Singapore	2005
D105k3908DK1		D1	Singapore	2005
D105k3910DK1		D1	Singapore	2005
D105k3911DK1		D1	Singapore	2005
D105k3915DK1		D1	Singapore	2005
D105k3916DK1		D1	Singapore	2005
D105k3930DK1		D1	Singapore	2005
D105k3933DK1		D1	Singapore	2005
D105k3934DK1		D1	Singapore	2005
D105k3935DK1		D1	Singapore	2005
D105k4138DK1		D1	Singapore	2005
D105k4139DK1		D1	Singapore	2005
D105k4140DK1		D1	Singapore	2005
D105k4142DK1		D1	Singapore	2005
D105k4147DK1		D1	Singapore	2005
D105k4152DK1		D1	Singapore	2005
D105k4172DK1		D1	Singapore	2005
D105k4173DK1		D1	Singapore	2005
D105k4174DK1		D1	Singapore	2005
D105k4175DK1		D1	Singapore	2005
D105k4183DK1		D1	Singapore	2005
D105k4443DK1		D1	Singapore	2005
D105k4468DK1		D1	Singapore	2005
D105k4479DK1		D1	Singapore	2005
D105k4480DK1		D1	Singapore	2005
D105k4604DK1		D1	Singapore	2005
D105k4605DK1		D1	Singapore	2005
D105k4606DK1		D1	Singapore	2005
D105k4609DK1		D1	Singapore	2005
D105k4620DK1		D1	Singapore	2005
D105k4621DK1		D1	Singapore	2005
D105k4622DK1		D1	Singapore	2005
D105k4632DK1		D1	Singapore	2005
D105k4820DK1		D1	Singapore	2005
D105k814DK1	D1SG05ED025	D1	Singapore	2005
D105k847DK1	D1SG05ED059	D1	Singapore	2005

D105k872DK1	D1SG05ED084	D1	Singapore	2005
D105k4154DK1		D1	Singapore	2005
D305k2400DK1	D3SG05ED112	D3	Singapore	2005
D305k2406DK1	D3SG05ED118	D3	Singapore	2005
D305k2418DK1	D3SG05ED130	D3	Singapore	2005
D305k2899DK1	D3SG05ED146	D3	Singapore	2005
D305k2918DK1		D3	Singapore	2005
D305k2933DK1	D3SG05ED184	D3	Singapore	2005
D305k3305DK1	D3SG05ED208	D3	Singapore	2005
D305k3312DK1	D3SG05ED215	D3	Singapore	2005
D305k3314DK1	D3SG05ED217	D3	Singapore	2005
D305k3316DK1	D3SG05ED219	D3	Singapore	2005
D305k3324DK1	D3SG05ED227	D3	Singapore	2005
D305k3325DK1	D3SG05ED228	D3	Singapore	2005
D305k3329DK1	D3SG05ED232	D3	Singapore	2005
D305k3887DK1	D3SG05ED235	D3	Singapore	2005
D305k3897DK1	D3SG05ED246	D3	Singapore	2005
D305k3900DK1		D3	Singapore	2005
D305k3912DK1		D3	Singapore	2005
D305k3913DK1		D3	Singapore	2005
D305k3927DK1		D3	Singapore	2005
D305k3928DK1		D3	Singapore	2005
D305k4141DK1		D3	Singapore	2005
D305k4144DK1		D3	Singapore	2005
D305k4157DK1		D3	Singapore	2005
D305k4159DK1		D3	Singapore	2005
D305k4168DK1		D3	Singapore	2005
D305k4176DK1		D3	Singapore	2005
D305k4182DK1		D3	Singapore	2005
D305k4440DK1		D3	Singapore	2005
D305k4454DK1		D3	Singapore	2005
D305k4647DK1		D3	Singapore	2005
D305k4648DK1		D3	Singapore	2005
D305k791DK1	D3SG05ED002	D3	Singapore	2005
D305k797DK1	D3SG05ED008	D3	Singapore	2005
D305k802DK1	D3SG05ED013	D3	Singapore	2005
D305k805DK1	D3SG05ED016	D3	Singapore	2005
D305k827DK1	D3SG05ED038	D3	Singapore	2005
D305k843DK1	D3SG05ED055	D3	Singapore	2005
D305k845DK1	D3SG05ED057	D3	Singapore	2005
D305k863DK1	D3SG05ED075	D3	Singapore	2005
D305k868DK1	D3SG05ED080	D3	Singapore	2005
D305k871DK1	D3SG05ED083	D3	Singapore	2005