A HYBRID PREDICTIVE MODEL FOR CLASSIFICATION OF DENGUE HAEMORRHAGIC FEVER OUTBREAK RISK

NAPA RACHATA

MASTER OF SCIENCE

IN STRATEGIC MANAGEMENT INFORMATION SYSTEM

MAE FAH LUANG UNIVERSITY

2009

A HYBRID PREDICTIVE MODEL FOR CLASSIFICATION OF DENGUE HAEMORRHAGIC FEVER OUTBREAK RISK

NAPA RACHATA

THIS THESIS IS A PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN STRATEGIC MANAGEMENT INFORMATION SYSTEM

MAE FAH LUANG UNIVERSITY

2009

A HYBRID PREDICTIVE MODEL FOR CLASSIFICATION OF DENGUE HAEMORRHAGIC FEVER OUTBREAK RISK

NAPA RACHATA

THIS THESIS HAS BEEN APPROVED
TO BE A PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN STRATEGIC MANAGEMENT INFORMATION SYSTEM
2009

EXAMINING COMMITTEE

Kosin. Chamnongthai ....CHAIRPERSON

(Assoc. Prof. Dr. Kosin Chamnongthai)

C. Lurs ....MEMBER

(Prof. Dr. Chidchanok Lursinsap)

Yooyativong T ....MEMBER

(Gp. Capt. Dr. Thongchai Yooyativong)

P. Temdee ....MEMBER

(Dr. Punnarumol Temdee)

ACKNOWLEDGEMENT

Successful completion of this thesis could not have achieved without kindness and encouragement from many people.

First of all, I would like to express my thankfulness and appreciation to my parents, who always love, understand, and support me along my master degree and my lifetime. I would also like to express my gratefulness to my major supervisor, Gp. Capt. Dr. Thongchai Yooyativong, for his valuable advice, continuous guidance, strong patience, and pushing me the great opportunity to explore this thesis.

Moreover, my sincere thanks to my thesis committee, Prof. Dr. Chidchanok Lursinsap, Assoc. Prof. Dr. Kosin Chamnongthai, and Dr. Punnarumol Temdee, for thoughtful comments, discussion, and wonderful suggestion in making this thesis an increasing accomplishment.

Additionally, I extend to thank Mr. Phasit Charoenkwan and Mr. Rattasak Pengchata for helpful guidance. I wish to thank to all friends in SMIS program and other friends for a friendly, enjoying, and encouragement, including thank to Mae Fah Luang University and to lecturers provided me the high knowledge and experience for studying. Lastly, I would like to thank Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health and Thai Meteorological Department for supporting in the data collection of this thesis.

Finally, my special thanks go to Naresuan University Phayao who provides a scholarship and financial support to me.

Napa Rachata

Thesis Title	A Hybrid Predictive Model for Classification of Dengue Haemorrhagic Fever Outbreak Risk
Author	Napa Rachata
Degree	Master of Science (Strategic Management Information System)
Supervisory Committee	Gp. Capt. Dr. Thongchai Yooyativong Prof. Dr. Chidchanok Lursinsap

ABSTRACT

Outbreak of Dengue Haemorrhagic Fever is still very crucial in international public health problem. Therefore, the prediction is obviously urgent in order to monitor and prevent a widespread of the fever in advance. However, the traditional predictive models of Dengue Haemorrhagic Fever outbreak are complicated because of the large number of input data which are inconvenient and costly. To deal with this problem, this thesis proposes a predictive model for classification of Dengue Haemorrhagic Fever outbreak risk with hybrid method by using weather features as input data. This predictive model uses the entropy measure and the principal component analysis that are combined with the backpropagation neural network. The entropy measure is used for extracting relevant weather data patterns and the principal component analysis is used for extracting principal features. Then, the backpropagation neural network is deployed to predict the possible risk of Dengue Haemorrhagic Fever outbreak. The experiment and its result show that the proposed model produces the optimal predictive result with 94% accuracy using only data patterns of three principal features including maximum temperature, mean temperature, and relative humidity.

Keywords: Backpropagation / Dengue haemorrhagic fever / Information entropy / Neural network / Prediction / Principal component analysis

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

This thesis emphasizes on the study of dengue virus infection particularly for risk prediction of disease outbreak for being an alternative strategy to monitor and prevent the disease outbreak. This Chapter introduces the principle and motivation, including a perspective proposed thesis. Additionally, it indicates the detail of objectives and scopes concerning the study.

1.1 Principle and Motivation

Dengue is one of the dreaded vector-borne infectious diseases (Centers for Disease Control and Prevention [CDC], 2008) that is pandemic infection in the world and it is a serious international public health problem. Moreover, it greatly affects individual, social, economic, and epidemiological factors. Although, many criterions and campaigns are prepared to deal with this dengue incidence, it still spreads all over the worldwide regions, especially in tropical and sub-tropical regions, mostly in urban and semi-urban areas (World Health Organization [WHO], 2009). The dengue is the infection that ensues from one of four dengue viruses which are in RNA virus of the Flavivirus genus, Flaviviridae family. These viruses have four serotypes, which are closely related but antigenically distinct, including DEN-1, DEN-2, DEN-3, and DEN-4. The transmission of the dengue viruses from human to human is caused by the bite of infected female Aedes mosquito vector to human, which are mostly Aedes aegypti mosquito and some Aedes albopictus mosquito (CDC, 2005). This infectious disease has been divided into three types according to clinic syndromes of severe disease including Undifferentiated Fever (UF), Dengue Fever (DF), and Dengue Haemorrhagic Fever (DHF) which is an acute form of the dengue infection and it may lead to Dengue Shock Syndrome (DSS) (Department of Disease

Control [DDC], 2001, pp. 9-10). During 1950s, DHF was first recognized in the Philippines and Thailand. Nowadays, DHF is led to cause of hospitalization and death among children and most prevalent in Asian countries (WHO, 2009).

In recent decades, trend of the dengue has grown dramatically around the world. The global dengue incidence has increased thirtyfold in the last 50 years (WHO & United Nations Children's Fund [UNICEF], 2006, p. 218). The people that are about 2.5 billion people in the affected area over 100 endemic countries and in any regions which dengue viruses can be transmitted are in risk of this disease. WHO (2009) estimates that there may be 50 million of world's population who infected dengue viruses year by year. Moreover, the 52% of 2.5 billion people, 1.3 billion people, who are at risk of dengue infection, live in South-East Asia (SEA) (WHO, 2007). Especially, DHF emerges endemic and frequent severe infection into epidemic in SEA and also in Thailand.

Thailand is one of the countries in the SEA having a large number of epidemics of DHF infection incidence. Especially, in 1987 and 2001, epidemic disease was severely spread, and the number of patients increased to 174,285 and 139,732 respectively (DDC, 2001, p. 2). Nowadays, it still has been spread continuously and has several outbreak patterns such as every other year, every third year, every fourth year. This disease is seasonal infection and is found mostly in rainy season, but in the present, it is found year-round. Moreover, this infection is represented in reported cases of priority by diseases under surveillance and has to inform on every Wednesday at the Bureau of Epidemiology.

Unfortunately, nowadays, there are no vaccine and specific treatment (WHO, 2009), so any people can be infectious risk. Consequently, only one method can prevent the transmission dengue viruses are to deal with Aedes mosquito because Aedes mosquito is the main cause of the outbreak of disease. Several features affect the propagation of this mosquito such as socio-culture, weather, and topography. Among several features, global warming or weather change is the important feature that may encourage changing life cycle of the mosquito, increasing the number of the mosquito, and transmitting cycle between mosquito and human. Moreover, the weather change may impact to a change pattern of the outbreak.

These occurrences indicate that strategy and urgent countermeasure should be required in order to monitor and prevent the spread of disease. Therefore, to determine appropriate and effective strategy and urgent countermeasure, the prediction is proposed for solving problem in time. For example, if the weather change is important feature on DHF outbreak, when there are some weather feature changes, the risk of DHF outbreak may be occurred. Afterwards, public health organization can prepare an appropriate and an effective strategy and urgent countermeasure for monitoring and preventing of DHF outbreak risk. In the recent past, the researchers developed predictive models for this disease both internal and external problems. The internal problem means clinical criteria, while the external problem means environment criteria. However, those predictive models are complex due to the large number of input data.

To arrange with the large number of input data and to obtain a simple predictive model, this thesis proposes a predictive model for classification of DHF outbreak risk with hybrid method by using weather features including rainfall, minimum temperature, maximum temperature, mean temperature, and relative humidity as the input data. This model observes three phases, including information extraction, feature extraction, and prediction. Furthermore, this proposed hybrid method in this thesis takes advantage of using entropy measure for extracting relevant weather data patterns[1], principal component analysis for extracting principal features[2], and backpropagation neural network for predicting the DHF outbreak risk.

[1] weather data pattern is a normalized five weather features per week.

[2] principal feature is weather feature that is identified by principal component analysis.

1.2 Objectives

There are three main objectives for this thesis as follows:

The first objective is to develop simplified predictive model for classification of DHF outbreak risk with hybrid method. The second objective is to propose an optimal result of the proposed predictive model, and the last objective is to investigate which weather features having the most direct influence to the DHF outbreak risk.

1.3 Scopes

The scopes of this thesis are shown as following:

1.3.1 Case Study and Data Collection

Chiang Rai province is selected for case study. Weather features and DHF cases have been collected from January 1999 to December 2008 and have been obtained from Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health and Thai Meteorological Department, respectively.

1.3.2 Preliminary Process

The preliminary process is divided into two phases which are information extraction and feature extraction. The information extraction phase applies entropy measure for obtaining relevant weather data patterns. Then, the feature extraction phase applies principal component analysis for obtaining principal features. After that, the selected data patterns[1] for the principal features will be fed as input data into data modeling process.

---

[1] data pattern is weather data pattern that is identified by entropy measure.

1.3.3 Data Modeling Process

The data modeling process is a prediction phase. This phase applies an artificial neural network called backpropagation for predicting the DHF outbreak risk.

1. Input data : the data patterns of principal features
2. Output data : DHF outbreak risk and non DHF outbreak risk classification

1.3.4 Result Evaluation

The result evaluation of this proposed predictive model is the experiment based on the condition of the weather features and the DHF cases from January 1999 until December 2008 were conducted.

1.4 Definition and Terminology

1.4.1 Data Pattern is a representative of group of data.

1.4.2 Dengue Haemorrhagic Fever (DHF) is a more severe form of the dengue disease. Characteristic of this fever is high fever, hemorrhagic phenomena such as microscopic hematuria, bleeding gums, and epistaxis. It also includes hepatomegaly and signs of circulatory failure. If the patient is untreated, the patient may lead to be shocked with a rapid called Dengue Shock Syndrome (DSS) and may die within 12-24 hours without sufficient treatment (WHO, 1997).

1.4.3 Prediction is a statement of a particular event that will happen in the future.

1.4.4 Weather Feature is a condition of the day-to-day for a particular place such as rainfall, minimum temperature, maximum temperature, mean temperature, and relative humidity.

1.5 Expectation

1.5.1 The predictive model for classification of DHF outbreak risk can be used for determination of an appropriate and an effective strategy and urgent countermeasure as well as monitoring to prevent the spread of disease.

1.5.2 This predictive model can be applied to other problems such as disaster evacuation and other diseases.

CHAPTER 2

LITERATURE REVIEW

This Chapter reviews and provides the related works and the computational theories used through this thesis. This Chapter composes of three sections. Firstly, the related works on predictive model of DHF are introduced. After that, the computational theories that related on the proposed predictive model are described. Those are entropy measure in information theory, principal component analysis, and backpropagation neural network. Finally, the proposed predictive model is followed.

2.1 Related Works

The DHF, as mentioned before, is abundantly severe and influential mosquito-borne infectious disease. It affects interior and exterior factors such as health, public health organization, nation, and internationality. Moreover, there are no available vaccine and specific treatment for protecting DHF infection, so these can be good evidence that support appearance and increase of this disease.

Accordingly, many of the current research works have focused on predictive model of DHF infection to be alternative channel for monitoring, preventing, and curing this infection. The predictive models that have been proposed with various purposes are internal problem such as day of defervescence in patients and external problem such as trend of incidence and trend of outbreak.

Concisely, conventional predictive models generally are categorized into two groups as follows: only data modeling (Husin, Salim & Ahmad, 2008; Nakhapakorn & Tripathi, 2005; Promprou, Jaroensutasinee, M. & Jaroensutasinee, K., 2006; Wongkoon, Pollar, Jaroensutasinee,

M. & Jaroensutasinee, K., 2007) and data preprocessing-based data modeling (Fu et al., 2007; Lbrahim, Taib, Abas, Guan & Sulaiman, 2005; Promprou, Jaroensutasinee, M. & Jaroensutasinee, K., 2004; Wu, Guo, Lung, Lin & Su, 2007). Furthermore, these conventional predictive models can be found that they are presented with two different methods. The first method is statistics, e.g., time series analysis, nonlinear regression, multivariate regression analysis, and principal component analysis. The second method is data mining that includes genetic algorithm, artificial neural network, and support vector machine.

In the only data modeling group, in the investigation by Husin et al. (2008), they designed Neural Network Model (NNM) and Nonlinear Regression Model (NLRM) by using different architectures and parameters to compare result and defined the best architecture for predicting an outbreak of dengue. The research of Nakhapakorn and Tripathi (2005) used multivariate regression analysis to find relationship between climatic independent features and DF and DHF incidences from 1997 to 2001 for predicting a change of the dengue incidence in rainy and non-rainy seasons. In the study of Promprou et al. (2006), they developed Autoregressive Integrated Moving Average (ARIMA) models for forecasting trend in incidence of DHF by using number of DHF cases for 12 years. Wongkoon et al. (2007) developed Seasonal Autoregressive Integrated Moving Average (SARIMA) models by using DHF cases data for four years in order to predict trend of DHF incidence.

For the data preprocessing-based data modeling group, the research of Fu et al. (2007) used the genetic algorithm-based support vector machine model for predicting dengue incidence trend. The genetic algorithm is used to detect influential climate factors and determine time-lags of the climate factors. The study of Lbrahim et al. (2005) used multinomial logistic regression for identifying significant variables from 37 clinical symptoms and signs variables and developed a system for predicting the day of defervescence of DF and DHF in patients with multilayer feed-forward neural networks by using nine significant input variables. In the work by Promprou et al. (2004), principal component analysis is used for investigating relation of socio-cultural variables and identifying significant variables. Besides, multiple linear regression models are used for predicting incidence of DHF with significant socio-cultural variables. The work by Wu et al. (2007) proved an association between variability of weather and dengue fever occurrence with cross-correlation and predicted dengue fever incidence by using Autoregressive Integrated Moving Average (ARIMA) models as well.

However, those conventional predictive models still use the large number of input data which is a number of data and features for developing data modeling. For this reason, the large number of input data consequently effect complex data modeling. To cope with this problem, this thesis is aimed to propose a method for obtaining a fewer number of input data, which is the preliminary process to the data modeling process, in order to solve complexity of the data modeling and yielding a high accuracy result. This method approaches entropy measure to generate a fewer number of data and principal component analysis to generate a fewer number of features. Both methods generate the outstanding weather data patterns and weather features as input data into the data modeling process.

2.2 Computational Theories

In this section, the related computational theories are described as following:

2.2.1 Entropy Measure (Information Theory)

1. Introduction

The entropy is statistical technique. It is applied in many fields such as Physics, Mathematics, Engineering, and Information Theory. In Information Theory, it is introduced by Shannon (1948) named Shannon entropy. The Shannon entropy is defined in terms of probability distribution of a random variable. Moreover, it is represented to be a good measure of uncertainty or randomness.

2. Process of entropy measure

The Shannon entropy can be computed by the following equation (2.1):

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \tag{2.1}$$

where X is a discrete random variable taking a finite number of possible values $\{x_1, x_2, \ldots, x_n\}$ with probabilities $\{p_1, p_2, \ldots, p_n\}$

$H(X)$ is the entropy H of X

This information entropy is used to calculate distribution or randomness level of data. So, if the data have randomness, the data will have a high level of information entropy, whereas, the data which have low level of randomness will have a low level of information entropy either.

As mentioned before, the data modeling generally uses a large number of data. Those data may have data redundancy. In order to reduce a number of data redundancies, the information entropy plays an important role for generating randomness level of data and showing the direction of data in order to select only the necessary data.

2.2.2 Principal Component Analysis

1. Introduction

The principal component analysis, which is abbreviated as PCA, is reported by Smith (2002). The PCA is statistical technique that is one of feature extraction techniques in terms of dimensionality reduction. Nowadays, it is applied in many fields such as recognition, compression, discovery data patterns in high dimension, and classification. Furthermore, it is widely used for producing predictive model and investigating data analysis as well.

In practice, this method is to convert a large number of associated variables into a less number of unassociated variables. They are called that principal components or eigenvectors. Moreover, the PCA calculates eigenvalues from a covariance matrix. These eigenvalues represent a number of the variance in the data. Additionally, the eigenvectors are closely related to eigenvalues. Then, the eigenvector with the higher eigenvalue is more significant than the eigenvector with the lower eigenvalue.

2. Process of principal component analysis

This method process is shown in six main steps as follows:

Step 1 : Gather and transform some data set

This step is to gather and transform some data set into an n by p matrix (n is a row of the matrix corresponding to observation and p is a column of the matrix corresponding to variable or feature).

Step 2 : Subtract the mean

This step is to subtract the mean of the matrix from element of each dimension. This step produces dataset matrix with zero mean as mean-adjusted data in equation (2.2) and (2.3).

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.2}$$

where x_i is the element of each dimension

n is the number of all elements of each dimension

$\overline{X}$ is the mean of each dimension

$$Z = \frac{\sum_{i=1}^{n} x_i - \overline{X}}{n} \tag{2.3}$$

where Z is the mean-adjusted data of each dimension

Step 3 : Calculate the covariance matrix from the dataset matrix whose mean is zero in equation (2.4).

$$Cov(X_j, X_k) = \frac{\sum_{i=1}^{n} (X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k)}{(n-1)} \tag{2.4}$$

where $\overline{X}_j = \frac{\sum_{i=1}^{n} X_{ij}}{n}$, $\overline{X}_k = \frac{\sum_{i=1}^{n} X_{ik}}{n}$, and $j,k = 1,2,..., p$

After that, the covariance matrix has a form as following:

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1p} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2p} \\ . & . & . & . & . \\ c_{p1} & c_{p2} & c_{p3} & \cdots & c_{pp} \end{pmatrix}$$

where C is the covariance matrix

c_{jk} is the covariance of variables X_j and X_k when $j \neq k$

the diagonal element c_{jj} is the variance of variable X_j when $j = k$

Step 4 : Calculate the eigenvalues and the eigenvectors of the covariance matrix in equation (2.5) and (2.6), respectively.

$$\det(C - \lambda I) = 0 \tag{2.5}$$

$$(C - \lambda_i I)v_i = 0 \tag{2.6}$$

where I is the identity matrix

λ is the eigenvalue

v_i is the eigenvector which is closely related to eigenvalue

From equation (2.5), if the covariance matrix is non-singular matrix, the number of all eigenvalues would equal to a number in columns of the covariance matrix or equal to a number of variables.

Step 5 : Choose components and form a feature vector

This step is to eliminate eigenvalues that have low eigenvalues. After that, choose eigenvectors are closely related to retainable eigenvalues and form a matrix with these chosen eigenvectors in the columns of feature vectors.

Step 6 : Derive the new data set

This step is to multiply between the columns of feature vectors and the rows of mean-adjusted data in equation (2.7).

$$\text{New data set} = \text{ColumnFeatureVector} \times \text{RowDataAdjust} \tag{2.7}$$

where ColumnFeatureVector is the matrix with chosen eigenvectors in the columns of feature vectors

RowDataAdjust is the matrix with mean-adjusted data in the rows of feature vectors of which the data items are presented in each row and dimension is presented in each column

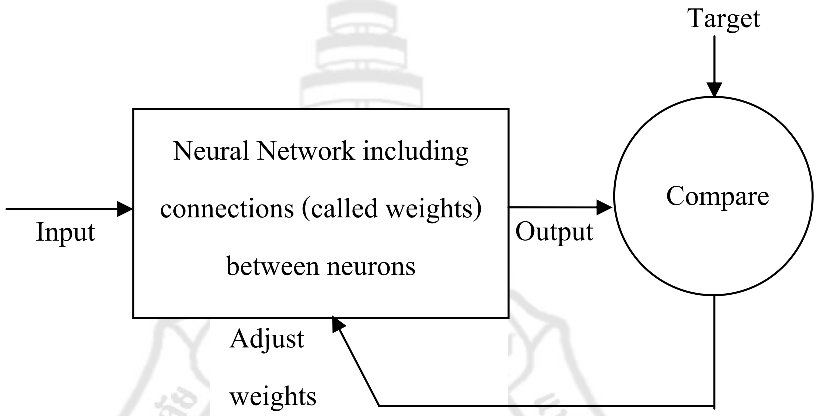
The various numbers of feature cause a problem for developing data modeling because those features may have been related or unrelated to DHF outbreak risk. To deal with this problem, the principal component analysis is used to reduce features or dimensions and generate principal features which related to DHF outbreak risk only.

2.2.3 Backpropagation Artificial Neural Network

1. Introduction

The artificial neural network, commonly called neural network or neural net, is defined as a mathematical model or computational model which is to simulate behaviors based on the biological nervous systems.

In this artificial neural network, the fundamental structure consists of simple elements called neurons which are operated in parallel and the fundamental function is determined by the connections between neurons. Moreover, this neural network is trained by adjusting the values of the connections, called weights, between those neurons. Afterward, neural network will be compared between the result of output and target, until the result of output is matched with the result of target. Hence, a specific input leads to a specific target output. As mentioned above, the neural network is shown in Figure 2.1.



From (Demuth, Beale & Hagan, 2009, p. 2)

Figure 2.1 Common Structure and Function of the Artificial Neural Network

Moreover, the neural network is used widely in non-linear problem. It is used for modeling complex relations of inputs and outputs, finding patterns of input data, predicting, and classifying. Besides, it is used in many fields for a study on Financial, Signal, Image, Energy, and Medicine.

2. Process of artificial neural network

This part is divided into four main fundamental details, including multilayer feedforward neural network, backpropagation algorithm, learning paradigms, and transfer functions.

1) Multilayer feedforward neural network

In a general form, the neural network consists of three network layers which are input layer, hidden layer (not required), and output layer. Each layer consists of one or more neurons.

A feedforward neural network is neural network which is connected by unidirectional form between neurons. It means that data flows in only one forward direction between the input neurons, the hidden neurons (if any), and the output neurons. There is no cycle in the network.

A multilayer feedforward neural network is neural network with one or more hidden layers (at least one hidden layer). That is the input neurons in one layer will be moved in a forward direction only to the neurons in the next layer. There is a connection form on layer by layer. As mentioned above, a multilayer feedforward neural network is shown in Figure 2.2.

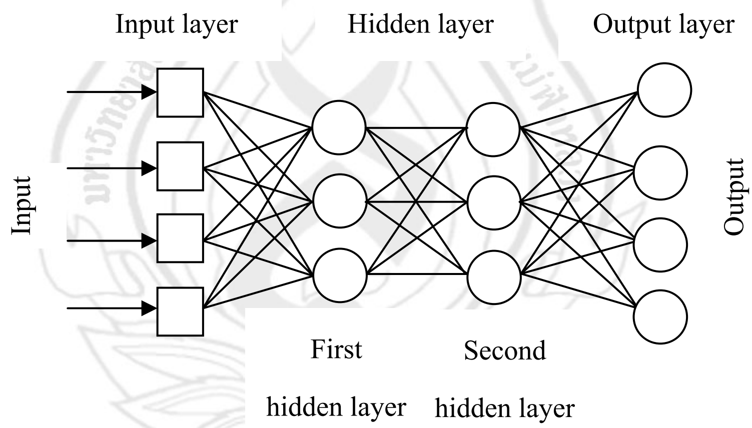


Figure 2.2 Multilayer Feedforward Neural Network with Two Hidden Layers

2) Backpropagation algorithm

A backpropagation algorithm is one of learning algorithms which is the most popular algorithm. In the process of this algorithm, training input pattern is presented to the input layer in the network. After that, the network computes a training input pattern from layer to layer till output pattern is formed with the output layer. If an output pattern differs from an expected output pattern, it means that an error is occurred, and then the error is calculated and propagated backward from the output layer to the input layer. This process of backward propagation is to adjust the weights for reducing that error. Figure 2.3 shows structure of multilayer feedforward backpropagation neural network.

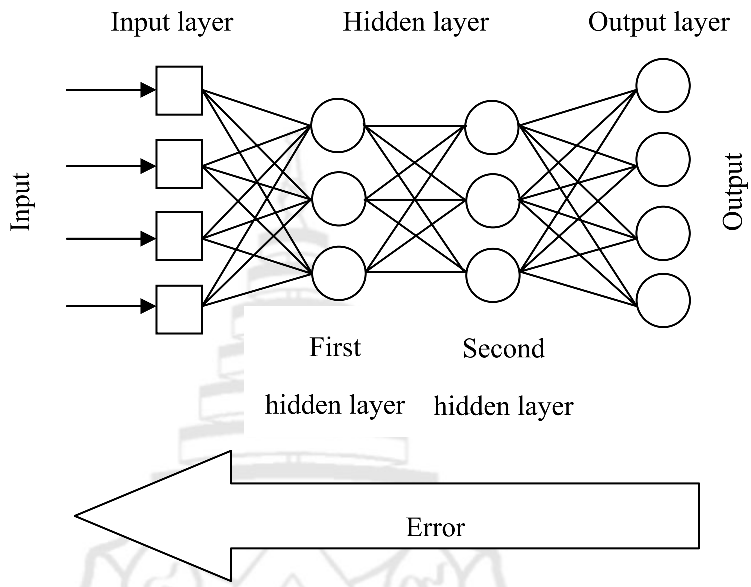
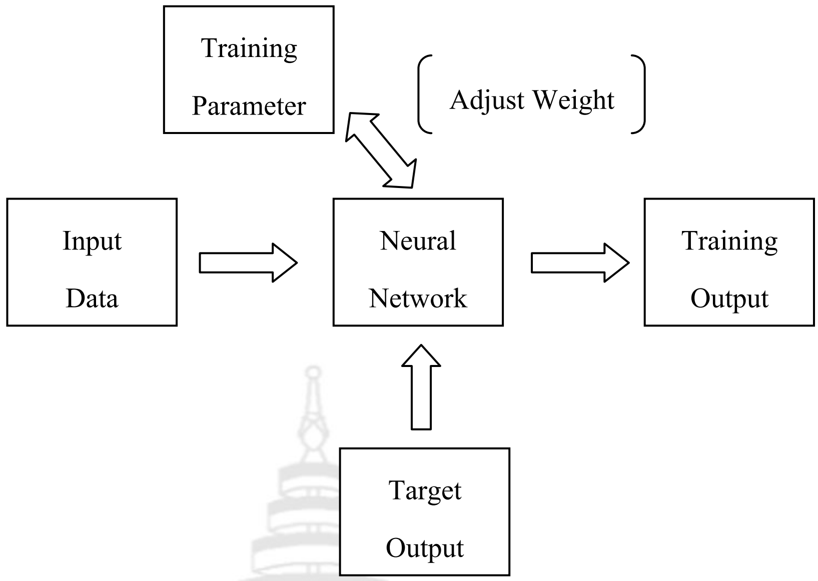


Figure 2.3 Structure of Multilayer Feedforward Backpropagation Neural Network

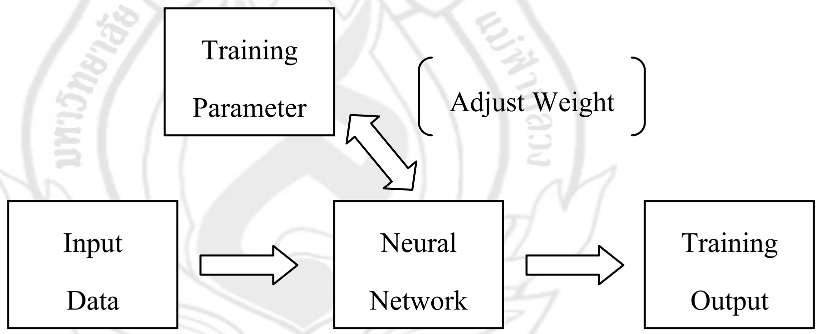
3) Learning paradigms

The learning paradigms are divided into supervised learning and unsupervised learning. The supervised learning is a technique for learning set of examples (called training input patterns). The training input patterns consist of many pairs of input and output (target) training patterns. The learning process is weights and biases of the neural network which are adjusted for moving the network outputs nearer to the targets. In contrast, the process of unsupervised learning is to adjust the weights and biases of the network in network inputs only. There are no target outputs available. Figure 2.4 and 2.5 show the process of supervised learning and unsupervised learning, respectively.



From (Chow & Cho, 2007, p. 17)

Figure 2.4 Process of Supervised Learning



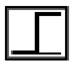
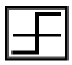
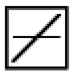
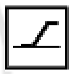
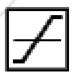
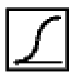
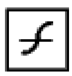
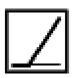
From (Chow & Cho, 2007, p. 17)

Figure 2.5 Process of Unsupervised Learning

4) Transfer functions

The transfer functions, or activation functions, are to transform output of network layer. They calculate a layer's output from its net input. Thus, the outputs from both hidden layer and output layer are computed from these transfer functions. A specific transfer function is selected for satisfying some specification of the problem that will be solved by the neuron. There are varieties of transfer functions as shown in the following Table 2.1.

Table 2.1 Transfer Functions

Name	Input/Output Relation	Icon	Matlab Function
Hard Limit	$a = 0 \quad n < 0$ $a = 1 \quad n \geq 0$		hardlim
Symmetrical Hard Limit	$a = -1 \quad n < 0$ $a = +1 \quad n \geq 0$		hardlims
Linear	$a = n$		purelin
Saturating Linear	$a = 0 \quad n < 0$ $a = n \quad 0 \leq n \leq 1$ $a = 1 \quad n > 1$		satlin
Symmetric Saturating Linear	$a = -1 \quad n < -1$ $a = n \quad -1 \leq n \leq 1$ $a = 1 \quad n > 1$		satlins
Log-Sigmoid	$a = \dfrac{1}{1+e^{-n}}$		logsig
Hyperbolic Tangent Sigmoid	$a = \dfrac{e^n - e^{-n}}{e^n + e^{-n}}$		tansig
Positive Linear	$a = 0 \quad n < 0$ $a = n \quad 0 \leq n$		poslin
Competitive	$a = 1$ neuron with max n $a = 0$ all other neurons	$\boxed{\mathrm{C}}$	compet

From (Hagan, Demuth & Beale, 1996, p. 6)

In terms of prediction application, an artificial neural network is popular for developing predictive model. This neural network is good for implementing a non-linear classification. Moreover, it can give better accuracy that ordinary statistical method for various problems (Munakata, 2008; Smith & Gupta, 2000). For these reasons, the neural network is used as a predictive model for classification of DHF outbreak risk.

2.3 Proposed Predictive Model

The thesis proposes a predictive model for classification of DHF outbreak risk with hybrid method by using weather features as the input data. This hybrid method includes the entropy measure and the principal component analysis that are combined with the backpropagation neural network. The objective of using the entropy measure is to reduce redundant weather data patterns and select relevant weather data patterns for DHF outbreak risk and non DHF outbreak risk classification. The objective of the principal component analysis is to select principal features that are appropriate for dividing DHF outbreak risk class and non DHF outbreak risk class. The objective of using the backpropagation neural network is to predict the DHF outbreak risk. In addition, this predictive model aims to use a few number of data and features for producing useful information and principal features for building strong learning predictive model and building simplified predictive model.

Moreover, the predictive model is designed such that prediction for the next week can be done by using input data of one week ago. That means an input data of the previous week is used for predicting possible risk of DHF outbreak for the next week.

The proposed predictive model is shown in Figure 2.6. There are three phases: information extraction, feature extraction, and prediction.

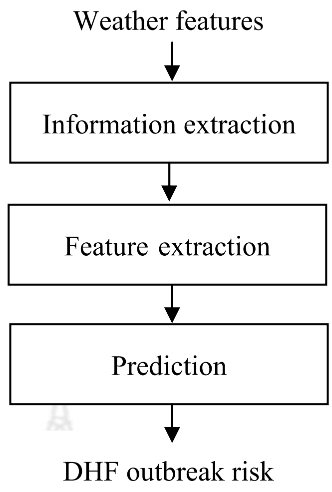


Figure 2.6 Phase of Proposed Predictive Model

CHAPTER 3

METHODOLOGY

This thesis, as mentioned previously, proposes the predictive model for classifying DHF outbreak risk using weather features with hybrid method. Now, in this Chapter, the methodology of the model is described. This Chapter is organized into two sections that are overview and detail of the predictive model.

3.1 Overview of Predictive Model

Prior to explain the detail of the predictive model, the predictive model overview in two main processes that are preliminary process and data modeling process will be mentioned as shown in Figure 3.1.

Firstly, the data collection, of which the detail will be shown in the next Chapter, consists of weather features and DHF cases. The weather features are normalized to adjust its value and become a standard scale because the actual value of each weather feature is used as an input data set during the training process in prediction phase of data modeling process is in different value range. For example, the temperature measured in Celsius may be in the range of [10 , 50] but the amount of rainfall may vary from one millimeter to only few centimeters. Any feature with a large value will dominate any other features with smaller values during the neural computation. This means that the prediction accuracy may not be achieved as expected because the accuracy is dictated by the features with larger values. To resolve this problem, simple value normalization within the same range is applied to all features by using the min-max normalization. Moreover, the percentile of statistical method is used to divide a number of DHF cases into two considered targets, including DHF outbreak risk class and non DHF outbreak risk class. This percentile method is used because it can summarize quantitative terms of a data set and can rank the value of a data set into considered group.

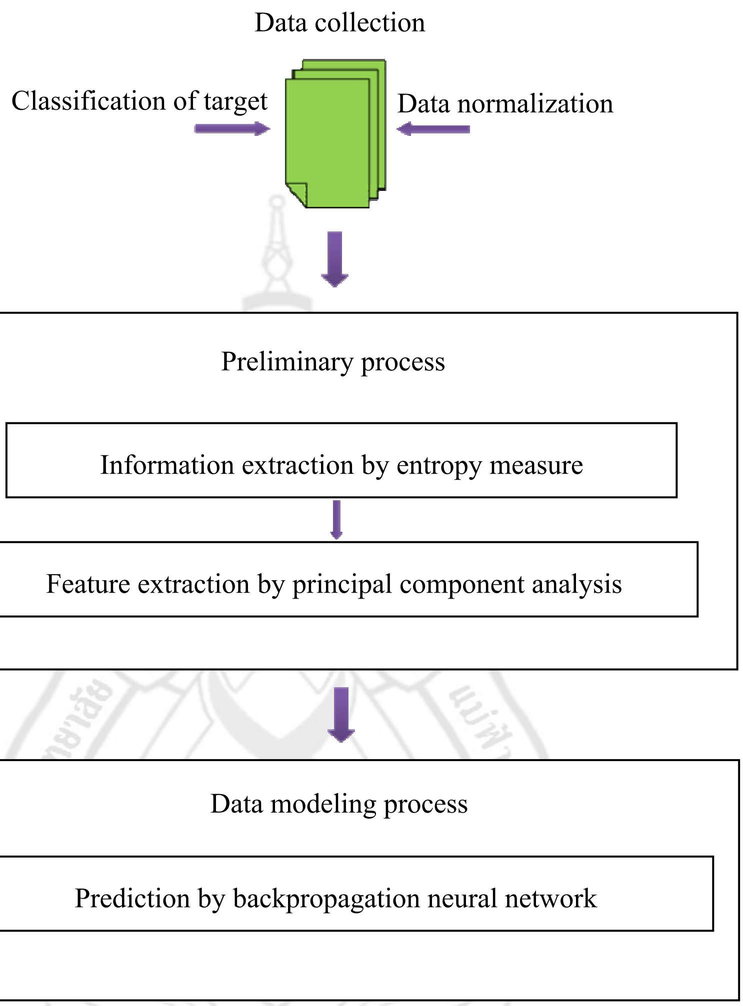


Figure 3.1 Overview of the Predictive Model

Secondly, the weather data patterns that were derived from the normalization of five weather features of every week are used in the preliminary process. This preliminary process is divided into two phases that are information extraction and feature extraction. The information extraction phase is offered to reduce irrelevant weather data patterns and retain only those relevant weather data patterns by using the entropy measure. After that, the feature extraction phase is offered to reduce weather features and select principal features only by using the principal component analysis.

The last process is the data modeling process. It is offered to predict between DHF outbreak risk class and non DHF outbreak risk class by using the backpropagation neural network with different data patterns of principal features set as input data.

3.2 Detail of Predictive Model

This section describes the detail of the predictive model which can be categorized into two main processes.

3.2.1 Preliminary Process

This process observes two phases which are information extraction and feature extraction as shown in Figure 3.2.

Number of weather data patterns	Weather features					
	no. of weeks	rainfall	minimum temperature	maximum temperature	mean temperature	relative humidity
	1	0.00	0.40	0.40	0.43	0.64
	2	0.11	0.48	0.43	0.54	0.71
	3	0.01	0.56	0.50	0.50	0.79
	4	0.01	0.42	0.51	0.51	0.65
	⋮	⋮	⋮	⋮	⋮	⋮
	522	0.00	0.47	0.39	0.47	0.58

Feature extraction

Information extraction

Number of weather data patterns	Weather features			
	no. of weeks	maximum temperature	mean temperature	relative humidity
	1	0.40	0.43	0.64
	2	0.43	0.54	0.71
	3	0.50	0.50	0.79
	4	0.51	0.51	0.65
	⋮	⋮	⋮	⋮
	174	0.63	0.68	0.36

Figure 3.2 Diagram of Preliminary Process

From Figure 3.2, the preliminary process shows that the information extraction phase is used to reduce a number of weather data patterns from 522 to 174 data patterns and the feature extraction phase is used to reduce weather features from five to three features. The detail of both phases are described as following:

1. Information extraction phase

1) Phase design

a) This phase uses weather data patterns of DHF outbreak risk class and non DHF outbreak risk class.

b) This phase uses entropy measure to calculate entropy value of each weather data pattern by separating DHF outbreak risk class and non DHF outbreak risk class.

2) Phase detail

The entropy measure is divided into two stages as following: firstly, a histogram is estimated and thereafter the entropy value is calculated.

Stage 1 : Histogram

This stage calculates a number of bars and interval of each the bar in equation (3.1) and (3.2). After that, they are used in creating histogram

$$\text{Number of bars} = \text{Square root of } (n) \tag{3.1}$$

where n is the number of all elements in weather data pattern

$$\text{Interval} = 1/\text{number of bars} \tag{3.2}$$

After that, the each value of weather data pattern is ranged into interval which is closely related to that value by using a histogram.

Stage 2 : Entropy

After that histogram stage, the entropy value is calculated by defining as equation (2.1). Let x_i be the number of all elements in interval i, $p(x_i)$ be the probability of number of all elements in interval i. The probability of each interval is computed from the ratio between the number of all elements in its interval and the number of all elements in weather data pattern. Moreover, box plot is used for selecting number of the entropy values of each class.

2. Feature extraction phase

1) Phase design

a) This phase uses five weather features.

b) This phase uses principal component analysis to present principal features.

2) Phase detail

This phase follows the process of principal component analysis by defining as equations (2.2) - (2.7).

Note that : a number of eliminated eigenvalues depend on accuracy result in the prediction phase of data modeling process, if the accuracy result is dropped, the process of eigenvalues elimination and prediction phase would be terminated.

After that, the result of preliminary process is used as an input data in the data modeling process. The detail of the data modeling process is described as follows:

3.2.2 Data Modeling Process

1. Prediction phase

1) Phase design

a) This phase uses data patterns of principal features of DHF outbreak risk class and non DHF outbreak risk class from preliminary process.

b) This phase uses backpropagation neural network to present accuracy result of prediction.

2) Phase detail

The problem of predicting DHF outbreak risk is transformed to the problem of classifying data patterns of principal features into groups of different outbreak risks or targets. Then, the neural network with a backpropagation algorithm is used for classifying data patterns of principal features. The data patterns of principal features are partitioned into training set and testing set. The ratio between these two sets will be discussed in the Chapter 4, an experiment and result. The proposed neural network consists of three layers as shown in Figure 3.3.

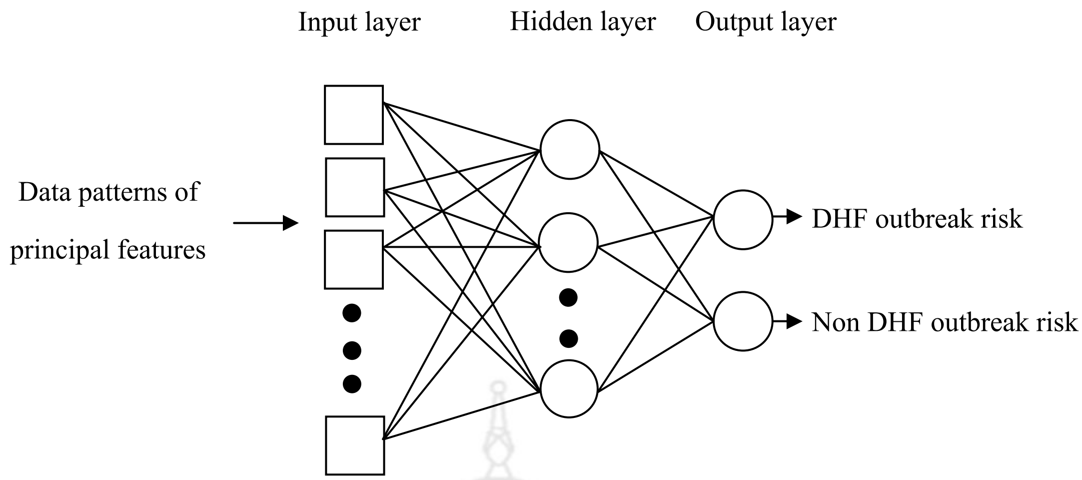


Figure 3.3 Process of Classifier for Prediction in Proposed Neural Network

From Figure 3.3, the structure of each layer is described as follows: the input data uses data patterns of principal features from one week ago that are assigned as the number of neurons in the input layer. The number of neurons in the hidden layer is determined by user's experience and guesswork, and also neurons can be added if the network has barrier learning. The output layer uses two neurons to represent DHF outbreak risk class which is indicated with one and non DHF outbreak risk class which is indicated with zero. Furthermore, the outputs from both hidden layer and output layer are also computed from transfer function (log-sigmoid function). The log-sigmoid function is used because it takes any real values of input to generate output values into the range between 0 and 1 and this transfer function is commonly used in backpropagation algorithm neural network.

From the Figure 3.3, the functional operation of an artificial neural network can be represented as follows:

$$[\text{Non/Outbreak risk}_{t}] = F_{ANN}\,[(\text{data patterns of principal features}_{t-1})]$$

where t is time (in the unit of week) to indicate the week of which the data belong

t-1 is one week before the week t

Non/Outbreak risk $_t$ is the targets at time t

data patterns of principal features $_{t-1}$ is the input data at time t-1

F_{ANN} is function that transfers input data to targets based on the trained artificial neural network

3.3 System Requirements

The minimum requirement of both hardware and software for developing the proposed predictive model is shown in this section.

3.3.1 Hardware

1. CPU : Intel(R) Core(TM) 2 Duo T7200 @ 2.00GHz
2. Ram : 1.00 GB
3. Hard disk : 120 GB
4. Monitor : NVIDIA GeForce Go 7200
5. Peripheral : Keyboard, Mouse USB

3.3.2 Software

1. Operating system : Microsoft Windows XP Professional
2. Model development : MATLAB 6.5

CHAPTER 4

EXPERIMENT AND RESULT

The methodology of the proposed predictive model comprises of mainly information extraction, feature extraction, and prediction phases are firstly explained. The experimental study will be represented in this Chapter. This Chapter is separated into three sections including materials, which can be divided into case study and data collection, data normalization, and an experiment and expected result.

4.1 Materials

4.1.1 Case Study

Chiang Rai province is the northernmost province of Thailand with an approximate of 1,227,317 populations (Department of Provincial Administration [DOPA], 2009). The province is bordered by Myanmar and Lao People's Democratic Republic. It covers an area of 11,678.4 km^2 in the north continental highland with forest, several flat river basins, and a density of 104.9 people / km^2 (DOPA, 2008). Weather of Chiang Rai is influenced by the northeast and southwest monsoon. Also, most of its area is mountainous and forested as well as far away from the ocean.

The reason why Chiang Rai province has been chosen as a case study because this province is one of the provinces which has high number of DHF cases in Thailand. Moreover, in 2007, Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health has reported that it has been the first rate of DHF cases due to its 819 cases. The highest number of cases occurred in only one week is 220 cases that are shown in Figure 4.1.

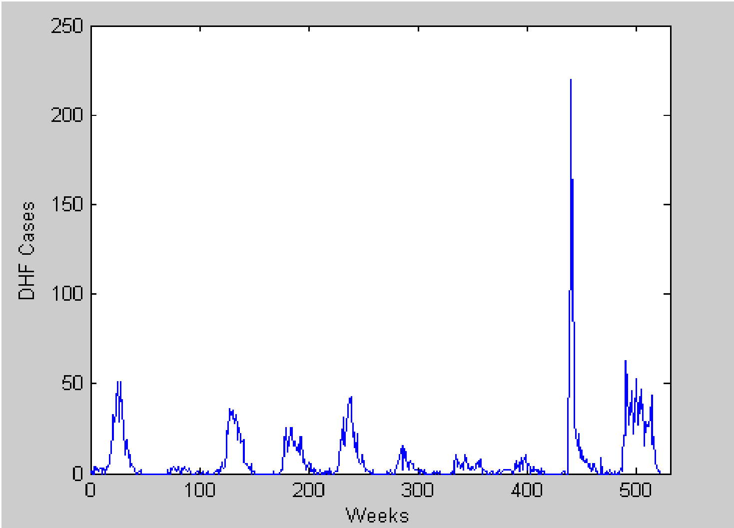


Figure 4.1 Number of DHF Cases per Week in Chiang Rai Province from January 1999 to December 2008

4.1.2 Data Collection

The data collection consists of two data types for experimenting as follows: (1) the daily weather features which include rainfall, minimum temperature, maximum temperature, mean temperature, and relative humidity obtained from Thai Meteorological Department, and (2) the daily DHF cases obtained from Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health. These data types have been provided since January 1999 – December 2008. Then, data collection is totally contained 3,653 data.

Since the proposed predictive model is designed to predict for week unit, so, these data types are converted from the daily data to the weekly weather features and the weekly DHF cases. The size of data collection is reduced to 522 (or 522 weeks). The example of data collection is summarized in Table 4.1.

Table 4.1 Example of Data Collection

no. of weeks	rainfall	minimum temperature	maximum temperature	mean temperature	relative humidity	DHF cases
1	0.00	11.60	29.40	19.59	75.86	1
2	31.90	13.70	30.00	21.74	78.86	0
3	4.30	15.60	31.40	20.91	82.14	4
4	1.70	12.10	31.70	21.11	76.29	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
522	0.00	13.30	29.10	20.29	73.29	0

4.2 Data Normalization

The weather features are normalized by using min-max normalization in equation (4.1) (Priddy & Keller, 2005) for adjusting standard scale within the range of 0 and 1 or [0 , 1] in short. It means that this normalization transforms weather features from its value into a range of 0 and 1.

$$x_i' = \left(\max_{target} - \min_{target}\right) \times \frac{x_i - \min_{value}}{\max_{value} - \min_{value}} + \min_{target} \text{ , } i = 1,2,3,\ldots,n \qquad (4.1)$$

where x_i is the element of each weather feature

$\max_{target}$ is the maximum desired range of values

$\min_{target}$ is the minimum desired range of values

$\max_{value}$ is the maximum data of each weather feature

$\min_{value}$ is the minimum data of each weather feature

An example of the weather data patterns that were derived from the normalization of five weather features of every week is presented in Table 4.2.

Table 4.2 Example of the Weather Data Patterns from Normalization of Five Weather Features of Every Week

	Weather features						
Number of weather data patterns	no. of weeks	rainfall	minimum temperature	maximum temperature	mean temperature	relative humidity	DHF cases
	1	0.00	0.40	0.40	0.43	0.64	1
	2	0.11	0.48	0.43	0.54	0.71	0
	3	0.01	0.56	0.50	0.50	0.79	4
	4	0.01	0.42	0.51	0.51	0.65	0
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	522	0.00	0.47	0.39	0.47	0.58	0

4.3 Experiment

The detail of experiment is described and shown based on methodology of the proposed predictive model by using weather data patterns as an input data as shown in Table 4.2.

Prior to those three phases, the following step is done. The 85^{th} percentile is used to divide the number of DHF cases for setting class of target that is considered i.e. the member between 0 to 17 refers to non DHF outbreak risk class and more than 17 refers to DHF outbreak risk class. Then, the non DHF outbreak risk class has 442 weather data patterns and the DHF outbreak risk class has 80 weather data patterns (442 plus 80 equal to 522 weather data patterns).

4.3.1 Information Extraction Phase

Figure 4.2 shows weather data patterns of DHF outbreak risk class and non DHF outbreak risk class without entropy measure.

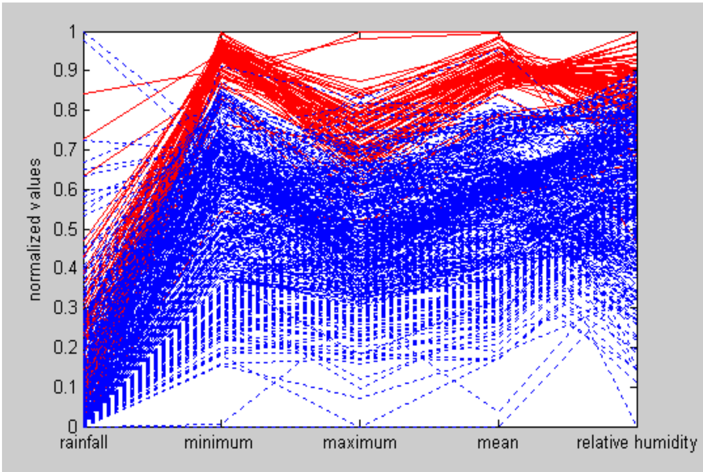


Figure 4.2 Weather Data Patterns without Entropy Measure

From Figure 4.2, the weather data patterns of DHF outbreak risk class is indicated with red solid line and the weather data patterns of non DHF outbreak risk class is indicated with blue dotted line. Obviously, some weather data patterns of DHF outbreak risk class and non DHF outbreak risk class is overlapped.

Therefore, this phase offers relevant weather data patterns of each class using entropy measure. This phase is divided into two stages which are histogram and entropy. In this phase, the first week weather data pattern in Table 4.2 is used as an example in the experimentation.

Firstly, the histogram stage is used for estimating probability frequency of first week weather data pattern.

1. Histogram : this stage is divide into two steps as follows:

Step 1 : Calculate number of bars with equation (3.1)

Number of bars = Square root of (5)

5 = the number of all elements in first week weather data pattern.

= 2.236068 ~ 3 bars

Step 2 : Calculate interval in each the bar with equation (3.2)

Interval = 1/3 = 0.3333

From weather data pattern of the first week, value of rainfall is ranged into 1st interval because it is in the range of 0 - 0.33, while value of minimum temperature, maximum temperature, mean temperature, and relative humidity are ranged into 2nd interval because they are in the range of 0.34 - 0.66. Thus, result of the histogram creation is shown in Figure 4.3.

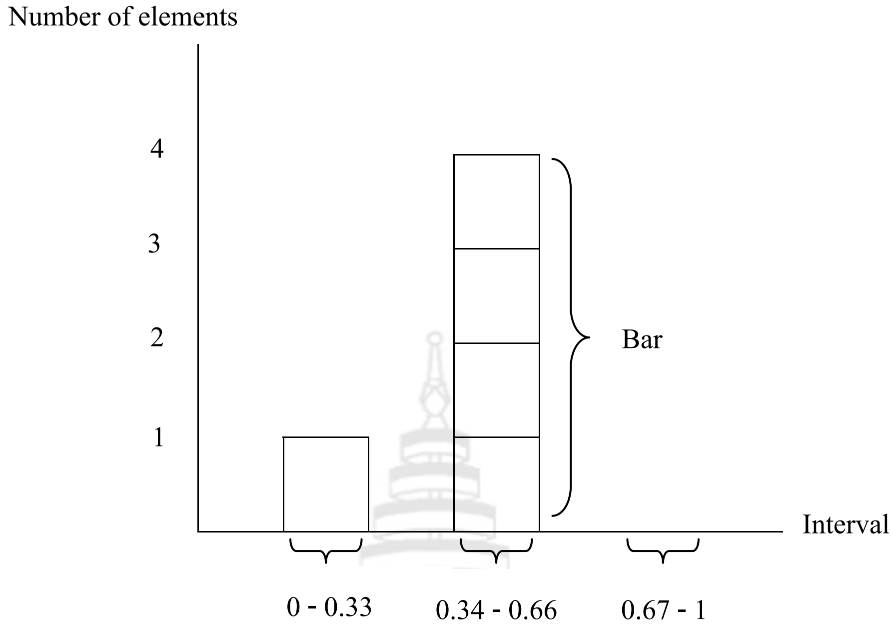


Figure 4.3 The Histogram Creation of First Week Weather Data Pattern

Thereafter, this result of histogram creation is used in entropy stage for calculating entropy value for showing randomness level of first week weather data pattern.

2. Entropy : this stage is divided into three steps as follows:

Step 1 : Calculate probability of number of all elements in each interval

The probability in the first interval = 1/5 = 0.2

The probability in the second interval = 4/5 = 0.8

The probability in the third interval = 0/5 = 0 (elimination)

Step 2 : Calculate log base 2 of the probability of number of all elements in each interval

The log base 2 of the probability in the first interval = -2.32193

The log base 2 of the probability in the second interval = -0.32193

Step 3 : Calculate entropy value as shown in Table 4.3

Table 4.3 Calculate Entropy Value of Weather Data Pattern of First Week

	Probability of number of all elements in each interval $p(x_i)$	**Log base 2 of probability of number of all elements in each interval** $\log_2 p(x_i)$	$-\left(p(x_i) * \log_2 p(x_i)\right)$
1st interval	0.2	-2.32193	0.464386
2nd interval	0.8	-0.32193	0.257542
	$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$		**0.721928**

From calculation of this step 3, result of entropy value for weather data pattern of the first week is 0.721928.

After applying entropy measure with weather data patterns of each class, a result for the weather data patterns of DHF outbreak risk class have high randomness level in a comparison with the weather data patterns of non DHF outbreak risk class that have low randomness level. So, high entropy values are selected for DHF outbreak risk class and low entropy values are selected for non DHF outbreak risk class.

Moreover, box plot is used for selecting a number of entropy values for DHF outbreak risk class and non DHF outbreak risk class. The number of entropy values for DHF outbreak risk class is started from maximum entropy value to previous median entropy value. In contrast, the number of entropy values for non DHF outbreak risk class is started from minimum entropy value to previous median entropy value. These are illustrated in Figure 4.4 and 4.5, respectively.

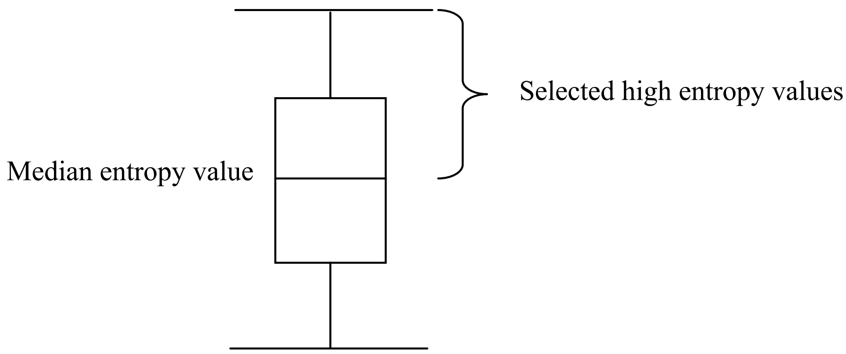


Figure 4.4 Selecting Number of the Entropy Values for DHF Outbreak Risk Class

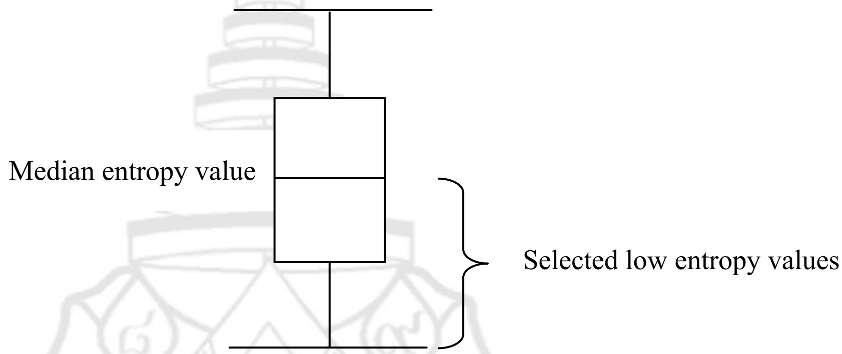


Figure 4.5 Selecting Number of the Entropy Values for Non DHF Outbreak Risk Class

As seen in Figure 4.4 and 4.5, the DHF outbreak risk class has 18 entropy values and the non DHF outbreak risk class has 156 entropy values. After that, weather data patterns which are closely related to selected entropy values of each class are selected as shown in Figure 4.6.

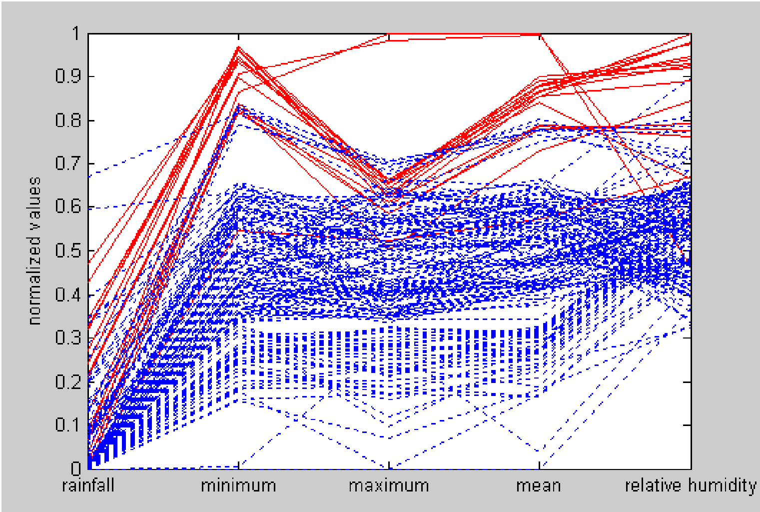


Figure 4.6 Weather Data Patterns with Entropy Measure

From Figure 4.6, the weather data patterns of DHF outbreak risk class are possibly distinguished from the weather data patterns of non DHF outbreak risk class compared with the Figure 4.2.

4.3.2 Feature Extraction Phase

This phase shows principal features set from five weather features using principal component analysis. The process of this method is referred from equation (2.2) to (2.7).

Step 1 : Gather and transform some data set

The weather data patterns of the both classes are gathered and transformed into 522x5 matrixes (five dimensions or features).

Step 2 : Subtract the mean

After that, mean of each the dimension is calculated into zero as shown in Figure 4.7 with equation (2.2) and (2.3).

			Weather features			
Number of weather data patterns	no. of weeks	rainfall	minimum temperature	maximum temperature	mean temperature	relative humidity
	1	0.00	0.40	0.40	0.43	0.64
	2	0.11	0.48	0.43	0.54	0.71
	3	0.01	0.56	0.50	0.50	0.79
	4	0.01	0.42	0.51	0.51	0.65
	⋮	⋮	⋮	⋮	⋮	⋮
	522	0.00	0.47	0.39	0.47	0.58
Mean		**0.12**	**0.61**	**0.51**	**0.60**	**0.64**

			Weather features			
Number of weather data patterns	no. of weeks	rainfall	minimum temperature	maximum temperature	mean temperature	relative humidity
	1	-0.12	-0.21	-0.10	-0.17	0.00
	2	0.00	-0.12	-0.08	-0.07	0.07
	3	-0.10	-0.05	-0.01	-0.11	0.15
	4	-0.11	-0.19	0.01	-0.10	0.01
	⋮	⋮	⋮	⋮	⋮	⋮
	522	-0.12	-0.14	-0.12	-0.14	-0.06
Mean		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

Figure 4.7 Example of Subtract the Mean of Each Dimension

Step 3 : Calculate the covariance matrix as shown in Table 4.4 from weather data patterns matrix whose mean is zero with equation (2.4).

Table 4.4 Calculate the Covariance Matrix

	1st feature	2nd feature	3rd feature	4th feature	5th feature
1st feature	0.02	0.02	0.00	0.01	0.01
2nd feature	0.02	0.05	0.02	0.04	0.03
3rd feature	0.00	0.02	0.03	0.03	0.00
4th feature	0.01	0.04	0.03	0.04	0.02
5th feature	0.01	0.03	0.00	0.02	0.03

Step 4 : Calculate the eigenvalues and the eigenvectors of the covariance matrix as shown in Table 4.5 and 4.6 with equation (2.5) and (2.6).

Table 4.5 Calculate the Eigenvalues

1st eigenvalue	2nd eigenvalue	3rd eigenvalue	4th eigenvalue	5th eigenvalue
0.00054	0.00000	0.00000	0.00000	0.00000
0.00000	0.00354	0.00000	0.00000	0.00000
0.00000	0.00000	0.01311	0.00000	0.00000
0.00000	0.00000	0.00000	0.03396	0.00000
0.00000	0.00000	0.00000	0.00000	0.11406

Table 4.6 Calculate the Eigenvectors

1^{st} eigenvector	2^{nd} eigenvector	3^{rd} eigenvector	4^{th} eigenvector	5^{th} eigenvector
-0.04	0.06	0.85	-0.46	0.26
0.50	-0.58	-0.14	-0.01	0.64
0.36	0.65	0.16	0.55	0.35
-0.79	-0.09	-0.03	0.29	0.53
-0.03	0.49	-0.49	-0.63	0.35

Step 5 : Choose components and form a feature vector

What is done in this step is to reduce dimension by eliminating low eigenvalues. From Table 4.5, for example, the 1^{st} eigenvalue is eliminated because its value is lowest. Since 1^{st} eigenvalue is closely related to 1^{st} eigenvector, therefore the 2^{nd}, 3^{rd}, 4^{th}, and 5^{th} eigenvectors are chosen, respectively and then four eigenvectors are formed in the columns of feature vectors that are the new data set which has only four dimensions including minimum temperature, maximum temperature, mean temperature, and relative humidity.

Step 6 : Derive the new data set

After forming feature vectors previously, what to do in this step is to multiply between the columns of feature vectors and the rows of mean-adjusted data with equation (2.7). The result of the new data set is shown in Table 4.7.

Table 4.7 Example of the New Data Set of Four Dimensions

Number of weather data patterns	Weather features				
	no. of weeks	minimum temperature	maximum temperature	mean temperature	relative humidity
	1	0.06	-0.08	-0.05	-0.29
	2	0.06	-0.03	-0.10	-0.12
	3	0.10	-0.15	-0.08	-0.07
	4	0.12	-0.07	0.02	-0.19
	⋮	⋮	⋮	⋮	⋮
	522	-0.02	-0.07	-0.02	-0.25

4.3.3 Prediction Phase

This phase shows accuracy result of prediction by using backprobagation neural network. In process of this method, it is described about definition of structure and parameters are used for experimenting. After that, experimental result is followed.

1. Definition of structure and parameters

The data patterns of four principal features, which are resulted from information extraction and feature extraction phases, are divided into two sets as follows: training set and testing set. The training set consists of 174 data patterns which are 18 data patterns from DHF outbreak risk class and 156 data patterns from non DHF outbreak risk class. Furthermore, the testing set consists of 348 data patterns which are 62 data patterns from DHF outbreak risk class and 286 data patterns from non DHF outbreak risk class. More details are illustrated in Appendix A for training set of the both classes and Appendix B for testing set of the both classes.

The structure and parameters of prediction phase composes of three layers including input layer, hidden layer, and output layer. The input layer uses data patterns of four principal features including minimum temperature, maximum temperature, mean temperature, and relative humidity from preliminary process are assigned as four neurons in input layer. The hidden layer uses one hidden layer and the number of neurons in hidden layer is between 1 and 15 neurons.

The output layer uses two neurons to present DHF outbreak risk class and non DHF outbreak risk class. Figure 4.8 shows structure and parameters of prediction phase.

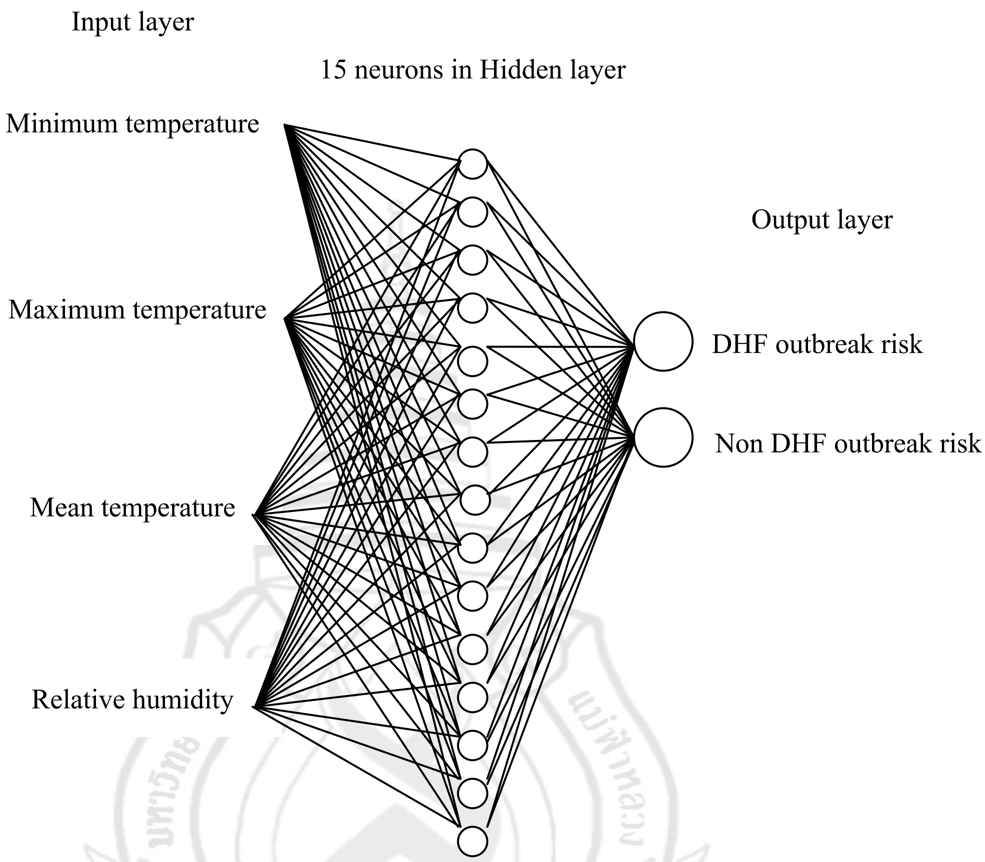


Figure 4.8 Structure and Parameters of Prediction Phase

2. Experimental result

After defining structure and parameters for experimenting, the 348 data patterns of testing set are used for evaluating accuracy result of the proposed predictive model. From Figure 4.8, the experimental result in prediction phase of data patterns of four principal features shows that the accuracy is 89% as illustrated in Table 4.9. Moreover, the experimental results in prediction phase of all cases are shown in Table 4.8 and 4.9 also.

Table 4.8 Accuracy Results Obtained without Using Principal Component Analysis Comparing between Using and without Using Entropy Measure

No. of weather features	**Weather features**	**Prediction accuracy**	
		Without using entropy measure	**Using entropy measure**
5	rainfall, minimum, maximum, mean, and relative humidity	0.70	0.81

Table 4.8 presents accuracy results obtained without using the principal component analysis comparing between using and without using the entropy measure with five weather features, which are rainfall, minimum temperature, maximum temperature, mean temperature, and relative humidity. The results show that using the entropy measure can obtain the highest accuracy around 81%, while without using the entropy measure, the accuracy is only 70%.

Table 4.9 Accuracy Results Obtained from Using Principal Component Analysis Comparing between Using and without Using Entropy Measure

No. of principal features	**Principal features**	**Prediction accuracy**	
		Without using entropy measure	**Using entropy measure**
4	minimum, maximum, mean, and relative humidity	0.77	0.89
3	maximum, mean, and relative humidity	0.80	0.94
2	mean and relative humidity	0.74	0.85
1	relative humidity	0.72	0.83

Table 4.9 presents the accuracy results obtained from using principal component analysis comparing between using and without using the entropy measure with each principal feature. The results show that using the entropy measure acquires more accurate than without using the entropy measure comparing each principal feature. With the entropy measure, the highest accuracy of 94% is occurred in case those data patterns of three principal features including maximum temperature, mean temperature, and relative humidity are used.

Finally, the high accuracy result can be achieved from using the entropy measure combined with the principal component analysis. Meanwhile, without using the entropy measure and without using the principal component analysis give the accuracy results less than the previous one. From the accuracy results of both previous Tables, it can be summarized via illustration in Figure 4.9.

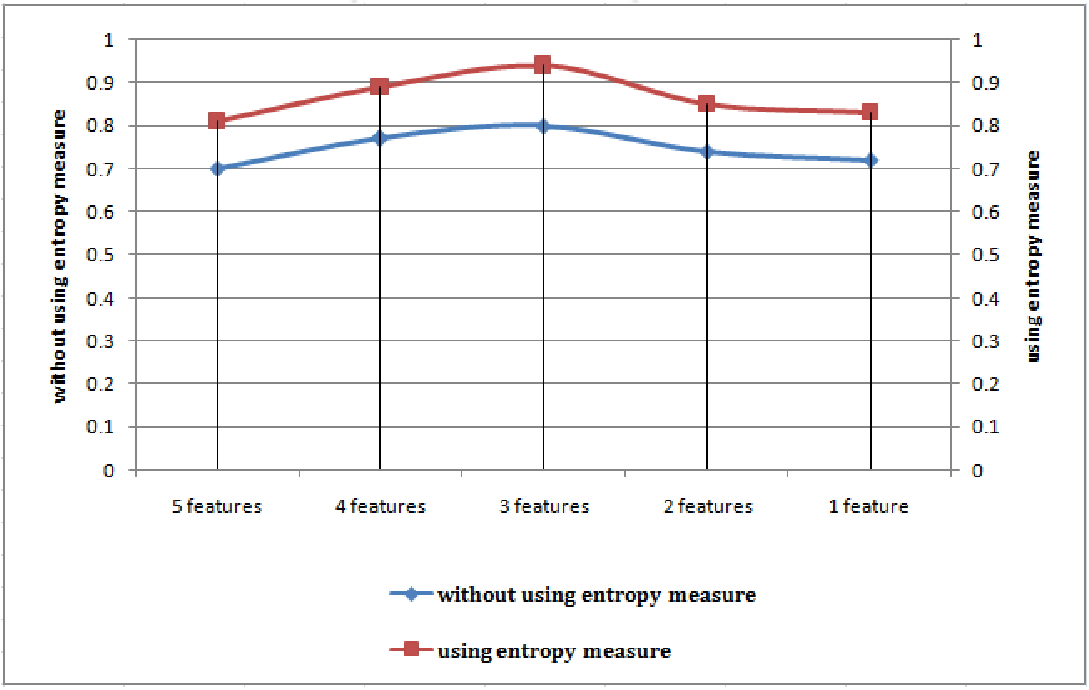


Figure 4.9 Accuracy Results of Using and without Using Principal Component Analysis Comparing between Using and without Using Entropy Measure

Furthermore, the result of accuracy depends on the result of training time in the prediction phase. In other words, if the result of training time decreases, the result of accuracy increases. For this reason, it obviously indicates that using the entropy measure and using the principal component analysis for selecting relevant weather data patterns and principal features effect easily to training input data of neural network. Then, if input data are effective, training time is decreased and accuracy result is increased. More details are illustrated in Appendix C.

CHAPTER 5

DISCUSSION

In this Chapter, there are two aspects that are discussed from the experimental result stated at the end of the Chapter 4. This two aspects include comparison of predictive result and actual result (percentage error), and weather features analysis.

5.1 Comparison of Predictive Result and Actual Result (Percentage Error)

From the experimental result in Table 4.9, it shows that the highest predictive accuracy is 94% on 348 data patterns of testing set (62 data patterns from DHF outbreak risk class and 286 data patterns from non DHF outbreak risk class). Table 5.1 shows the result of predicting DHF outbreak risk class and non DHF outbreak risk class as the following.

Table 5.1 Predictive Result

	Outbreak risk	**Non Outbreak risk**
Test shows “Outbreak risk”	60	2
Test shows “Non outbreak risk”	18	268

From Table 5.1, the accuracy result of predicting DHF outbreak risk is 60, whereas, the inaccuracy result is 2. On the other hand, the accuracy result of predicting non DHF outbreak risk is 268, whereas, the inaccuracy result is 18. Then, the percentage error between predictive result and actual result is 6% ((18+2)/348). The cause of percentage error is the process of selecting number of entropy values for DHF outbreak risk class and non DHF outbreak risk class which may not be good enough because there are a large number of selected weather data patterns of each class as shown in Figure 4.6.

In order to improve accuracy result, the process of selecting a number of entropy values for the both classes should be modified by selecting fewer number of entropy values for the both classes. For example, the number of entropy values for DHF outbreak risk class may select top 10 from the high entropy values and the number of entropy values for non DHF outbreak risk class may select top 80 from the low entropy values. These are illustrated in Figure 5.1 and 5.2, respectively.

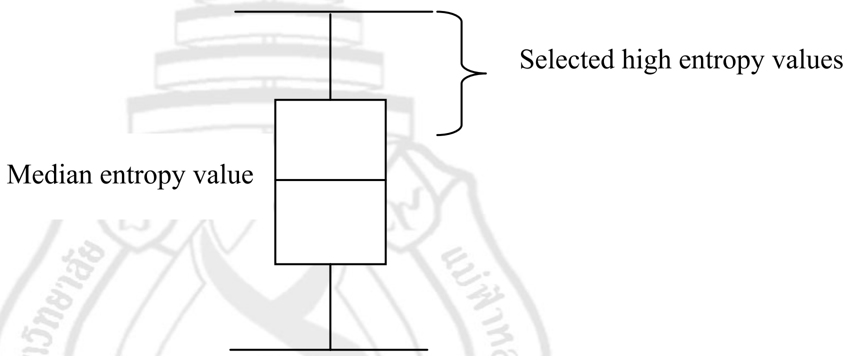


Figure 5.1 Modification of the Number of the Entropy Values Selection for DHF Outbreak Risk Class

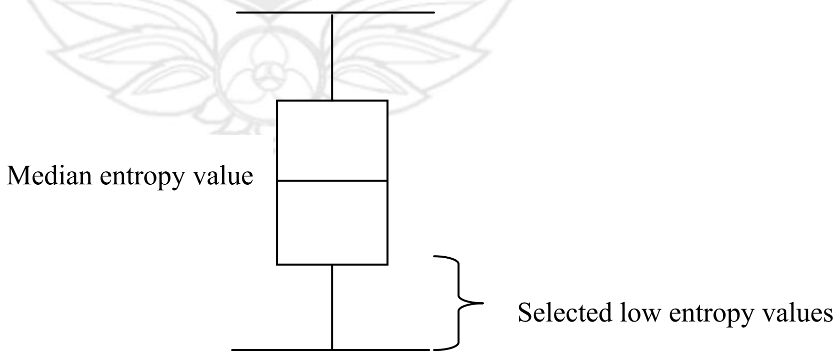


Figure 5.2 Modification of the Number of the Entropy Values Selection for Non DHF Outbreak Risk Class

5.2 Weather Features Analysis

As seen from the experimental result in Table 4.9, the three principal features obtain 94% of highest accuracy result as follows: maximum temperature, mean temperature, and relative humidity.

Therefore, the factor analysis is introduced to explore three principal features whether they significantly affect the risk of DHF outbreak. In this analysis, a One-Way ANOVA, which is statistical analysis, is used to analyze those features. Table 5.2 shows p-value of three principal features with alpha (significance level) which is equal to .05.

Table 5.2 P-Value of Three Principal Features

Three principal features	Maximum temperature	Mean temperature	Relative humidity
p-value	.244	.030	.005

From Table 5.2, it shows that the mean temperature and relative humidity significantly affect the risk of DHF outbreak because the p-value is equal to .030 and .005 (<.05), respectively. On the other hand, the maximum temperature do not significantly affect the risk of DHF outbreak because the p-value is equal to .244 (>.05). The cause of a result from this maximum temperature is affected by slightly different patterns between maximum temperature and minimum temperature. For example, the pattern of the maximum temperature is 24.9 Celsius, whereas, pattern of the minimum temperature is 24 Celsius.

Moreover, the main effect plot of three principal features is shown with mean plot in Figure 5.3, 5.4, and 5.5, respectively also.

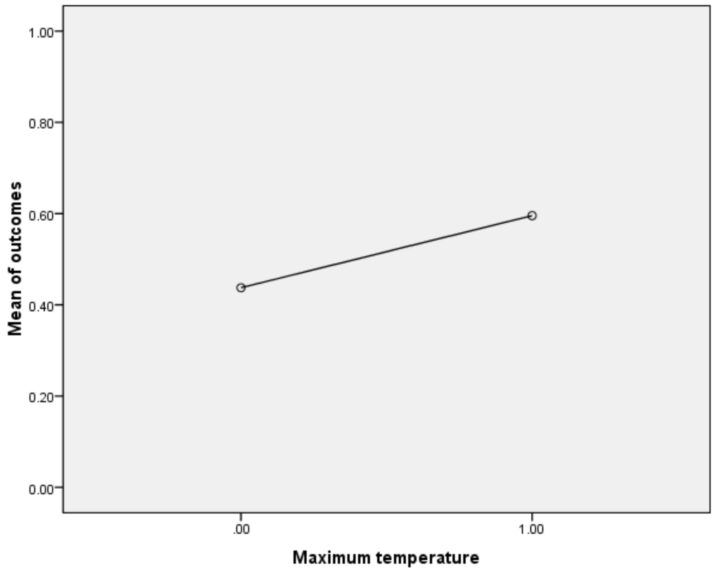


Figure 5.3 Main Effect Plot of Maximum Temperature

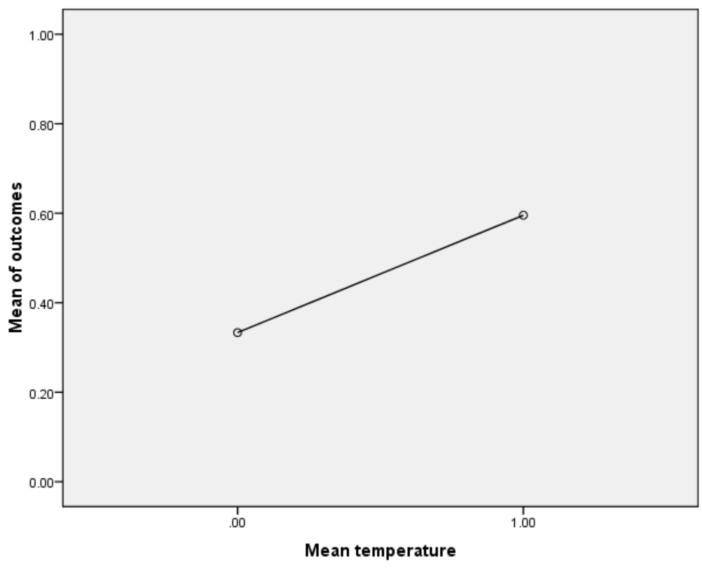


Figure 5.4 Main Effect Plot of Mean Temperature

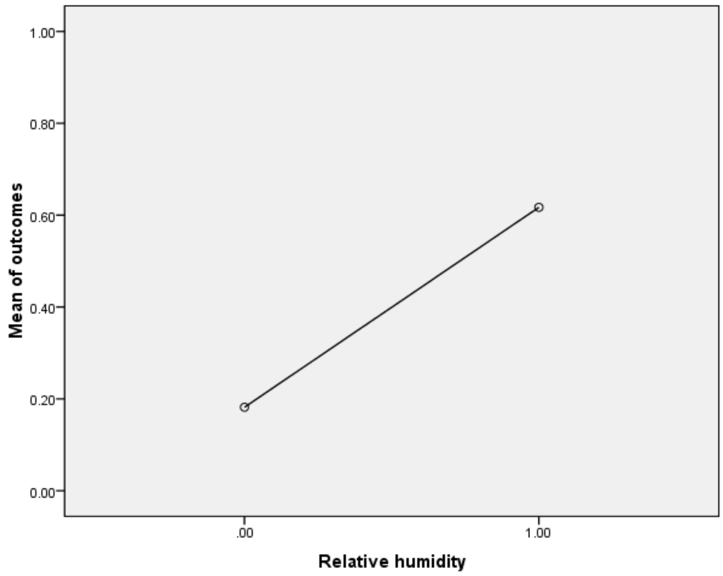


Figure 5.5 Main Effect Plot of Relative Humidity

From the main effect plot of three principal features, the result shows that the graph of relative humidity slopes more than graph of maximum temperature and mean temperature, while, the graph of maximum temperature has the least slope. The slope of graph is ensured from relation between feature and outcome. If the data have inconstant relation, the graph will be steeper than the one which the data that have constant relation. It can be concluded that maximum temperature has an effect on DHF outbreak risk but this effect is not significant, while, mean temperature and relative humidity have an effect on DHF outbreak risk and they are significant.

The p-value and main effect plot results are confirmed from various experimental studies (Parker, 1952; Watts, Burke, Harrison, Whitmire & Nisalak, 1987; Wu et al., 2007). These studies show that the point of low relative humidity may support that Aedes mosquito which can more rapidly seek target hosts and support outbreak of the DHF infection. The high temperature may allow infective female mosquito survive for a long time, extend infection of DHF to vary other regions, and encourage developing virus rate is much faster. Moreover, relative humidity and temperature are crucial features as they affect to the life cycle of mosquito such as mating and oviposition.

Moreover, the both analyses represent that the relative humidity is more significant than the mean temperature because there is fewer p-value and more slope. It can be concluded that the most weather feature significantly affects the risk of DHF outbreak that is relative humidity in Chiang Rai province.

Although the results of analyses show that the relative humidity is the most weather feature significantly affected the risk of DHF outbreak in Chiang Rai province, this result may not be concluded that relative humidity is the most weather feature significantly affected the DHF outbreak risk in all regions. It is so because this feature may be possible in specific region only. Then, other regions should be considered in order to discover and precisely understand about weather features affect to the risk of DHF outbreak in each region.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

DHF has been emerged as a principal international public health problem. Moreover, trend of outbreak and the number of DHF cases increase yet. Therefore, predicting DHF outbreak should be proposed to monitor and prevent an outbreak of this fever in advance.

However, the conventional predictive models use the large number of data and features as input data for developing data modeling. That is the reason why it causes the complex data modeling. Hence, a representative data and features that are preprocessed prior to the data modeling are proposed for solving complexity of data modeling and improving accuracy result of predictive model.

This thesis proposed a predictive model for classifying risk of DHF outbreak using weather features as input data with hybrid method. This hybrid method includes the entropy measure for extracting relevant weather data patterns, the principal component analysis for extracting principal features, and the backpropagation neural network for predicting DHF outbreak risk. This predictive model was tested based on the condition of weather features and DHF cases from January 1999 until December 2008 with 348 data patterns of testing set. The evaluation result of the model achieves 94% optimal accuracy with data patterns of three principal features, including maximum temperature, mean temperature, and relative humidity compared to the actual data and the result of training time which consumed the fewest time for prediction. Besides, the relative humidity is observed to have the most direct influence to the DHF outbreak risk in Chiang Rai province.

Finally, this proposed predictive model indicates that using the entropy measure and the principal component analysis that are combined with the backpropagation neural network give optimal accuracy result and also simplified data modeling.

6.2 Future Work and Suggestion

As the proposed predictive model, there are three main issues that should be a suggestion and development afterwards for the future work in order to improve a model and its accuracy as follows:

First issue, five weather features should be considered with other case study regions for investigating which five weather features significantly affect the risk of DHF outbreak in each region. Additionally, other environment features and weather features should be considered because those other features may affect a risk of DHF outbreak also. Thus, it is led to detect new feature that affects a risk of DHF outbreak in those regions.

Second issue, weather features and environment features are varied according to the geographical region. Therefore, the accordance of data between feature and DHF infection should be considered. It means that data should be in the same level in order to solve directly the point of problem in each region level and improve prediction accuracy as well.

Third issue, characteristics of input data may affect accuracy result because patterns of weather features in Chiang Rai province are slightly different. Then, any region which has various different patterns of weather features should be considered. Moreover, the prediction accuracy can be improved by modifying the number of entropy values selection to have fewer number of entropy values for both classes.

REFERENCES

REFERENCES

Centers for Disease Control and Prevention. (2005, January 13). **Dengue and dengue hemorrhagic fever : questions and answers**. Retrieved May 7, 2008, from http://www.cdc.gov/ncidod/dvbid/dengue/dengue-qa.htm

Centers for Disease Control and Prevention. (2008, November 24). **Introduction : dengue fever**. Retrieved December 17, 2008, from http://www.cdc.gov/ncidod/dvbid/dengue/index.htm

Chow, T. W. S. & Cho, S. Y. (2007). **Neural networks and computing : learning algorithms and applications** (p. 17). London: Imperial College Press.

Demuth, H., Beale, M., & Hagan, M. (2009). Getting started. In **Neural network toolbox™ 6 user's guide** (p. 2). Natick, MA: The MathWorks.

Department of Disease Control. (2001). การติดเชื้อและปัจจัยเสี่ยง. In **โรคไข้เลือดออกฉบับประเกียรณก** (pp. 9-10). N.P.: n.p.

Department of Disease Control. (2001). ระบาดวิทยาของโรคไข้เลือดออก. In **โรคไข้เลือดออกฉบับประเกียรณก** (p. 2). N.P.: n.p.

Department of Disease Control. (2007). **รายงานการเฝ้าระวังทางระบาดวิทยาประจำสัปดาห์**. Retrieved February 19, 2008, from http://203.157.15.4/wesr/

Department of Provincial Administration. (2008, September 1). **บรรยายสรุปจังหวัด 2551**. Retrieved November 21, 2008, from http://www.chiangrai.net/cpoc/2009/Articles/viewArticle.aspx?A=11

Department of Provincial Administration. (2009, January 28). **จำนวนประชากรปี 51**. Retrieved February 7, 2009, from http://www.dopa.go.th/stat/y_stat.html

Fu, X., Liew, C., Soh, H., Lee, G., Hung, T., & Ng, L. C. (2007, September). **Time-series infectious disease data analysis using SVM and genetic algorithm**. Retrieved November 25, 2008, from IEEE database. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4424617

Hagan, M. T., Demuth, H. B., & Beale, M. (1996). Neuron model and network architectures. In **Neural network design** (p. 6). Boston, MA: PWS Publishing.

Husin, N. A., Salim, N., & Ahmad, A. R. (2008, August). **Modeling of dengue outbreak prediction in Malaysia : a comparison of neural network and nonlinear regression model**. Retrieved November 24, 2008, from IEEE database. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4632022

Lbrahim, F., Taib, M. N., Abas, W. A., Guan, C. C., & Sulaiman, S. (2005, April 15). **A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN)**. Retrieved June 1, 2008, from ELSEVIER database. http://www.journals.elsevierhealth.com/periodicals/comm/article/PIIS0169260705000866/abstract

Munakata, T. (2008). **Fundamentals of the new artificial intelligence : neural, evolutionary, fuzzy and more** (2nd ed.). London: Springer-Verlag.

Nakhapakorn, K. & Tripathi, N. K. (2005, June 8). **An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence**. Retrieved June 1, 2008, from BioMed Central database. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1177981

Parker, A. H. (1952). The effect of a difference in temperature and humidity on certain reactions of female *Aedes aegypti*. **Bull. Entomol. Res, 43**, 221-229.

Priddy, K. L. & Keller, P. E. (2005). **Artificial neural networks : an introduction**. Bellingham, WA: SPIE.

Promprou, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2004). Prediction of dengue haemorrhagic fever incidence using principal component analysis (PCA) : a case study in Nakhon Si Thammarat. In **Proceeding of 1st Applied Statistics Conference for Development of the Northeast Thailand** (pp. 370-381). Khonkaen, Thailand.

Promprou, S., Jaroensutasinee, M., & Jaroensutasinee, K. (2006). **Forecasting dengue haemorrhagic fever cases in southern Thailand using ARIMA models**. Retrieved February 19, 2008, from Dengue Bulletin database. http://www.searo.who.int/LinkFiles/Dengue_Bulletins_c12.pdf

Shannon, C. E. (1948). A mathematical theory of communication. **The Bell System Technical Journal, 27**, 379-423.

Smith, K. A. & Gupta, J. N. D. (2000, May). **Neural networks in business : techniques and applications for the operations researcher**. Retrieved March 5, 2008, from ScienceDirect database. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6VC5-40797G7-3&_user=10&_coverDate=09%2F30%2F2000&_alid=1051357969&_rdoc=7&_fmt=high&_orig=search&_cdi=5945&_sort=r&_docanchor=&view=c&_ct=71&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=e900d9c82414c80f66f0e5ba93306052

Smith, L. I. (2002, February 26). **A tutorial on principal components analysis**. Retrieved November 10, 2008, from http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Watts, D. M., Burke, D. S., Harrison, B. A., Whitmire, R. E., & Nisalak, A. (1987). Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. **The American Journal of Tropical Medicine and Hygiene, 36**(1), 143-152.

Wongkoon, S., Pollar, M., Jaroensutasinee, M., & Jaroensutasinee, K. (2007, October 15). **Predicting DHF incidence in northern Thailand using time series analysis technique**. Retrieved March 5, 2008, from WASET database. http://www.waset.org/journals/waset/v32/v32-42.pdf

World Health Organization. (1997). **Dengue haemorrhagic fever : diagnosis, treatment, prevention and control** (2nd ed.). Geneva: WHO.

World Health Organization. (2007, July 6). **Dengue/DHF - situation of dengue/dengue haemorrhagic fever in South-East Asia region**. Retrieved March 9, 2008, from http://www.searo.who.int/en/Section10/Section332_1098.htm

World Health Organization. (2009, March). **Dengue and dengue haemorrhagic fever**. Retrieved May 1, 2009, from http://www.who.int/mediacentre/factsheets/fs117/en/

World Health Organization & United Nations Children's Fund. (2006). Part 2. update of the burden of water-related diseases. In **Protecting and Promoting Human Health** (p. 218). Retrieved January 1, 2009, from http://www.unesco.org/water/wwap/wwdr/wwdr2/pdf/wwdr2_ch_6.pdf

Wu, P. C., Guo, H. R., Lung, S. C., Lin, C. Y., & Su, H. J. (2007, May 26). **Weather as an effective predictor for occurrence of dengue fever in Taiwan**. Retrieved February 3, 2008, from ScienceDirect database. http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6T1R-4NTRT5J-1&_user=10&_rdoc=1&_fmt=&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1051452321&_rerunOrigin=google&_acct=C000050221&_version=1&_urlVersion=0&_userid=10&md5=b109b583e5b9be6e39847b03e4f9301d

APPENDIXES

APPENDIX A

TRAINING SET

Table A1 Training Set of DHF Outbreak Risk Class

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
1	-0.02448792	-0.04622004	-0.12265502	0.185279376	0.291856797
2	0.002109874	0.057569538	-0.09172463	0.029958784	-0.067407624
3	-0.01047498	0.03771037	-0.06372009	-0.04516752	0.482866773
4	-0.01120637	0.004668995	-0.07994271	-0.1067399	0.552679918
5	-0.001962	0.019954954	-0.00580347	-0.12262279	0.587227859
6	-0.00536285	0.036058529	0.011865411	-0.11835722	0.559695724
7	-0.01274332	0.01346857	-0.03726352	-0.08443442	0.565168977
8	0.006203585	0.060862133	0.150740858	0.38066947	0.55950269
9	-0.00125527	0.043259344	0.031342459	0.541661209	0.458546427
10	-0.02040328	-0.02243698	-0.21905837	-0.01398033	0.518554043
11	0.004892194	-0.0450829	-0.15625204	-0.08356532	0.272118709
12	-0.0025141	-0.04247241	-0.12549077	0.039602431	0.300303525
13	-0.00714503	-0.01803145	-0.15997582	0.031843091	0.288235132
14	-0.00940971	-0.00621057	-0.07161193	0.016578985	0.319517301
15	-0.01645206	0.041286665	-0.01995895	-0.19797157	0.577608279
16	-0.02280697	0.055690995	0.001073358	-0.15893687	0.582320543
17	-0.01086751	0.053023126	0.10326431	-0.22938798	0.609991344
18	-0.01213967	0.016789861	0.06027758	-0.23697011	0.587036385

Table A2 Training Set of Non DHF Outbreak Risk Class

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
1	-0.032733981	-0.141321766	0.055776895	-0.107822736	-0.80463816
2	0.015577702	-0.204874763	0.047518859	-0.064016637	-0.664249884
3	-0.000584972	-0.08501213	0.310455954	0.085664905	0.031879771
4	-0.001070481	0.061095153	-0.083719015	-0.053516717	-0.285826783
5	-0.009393537	0.120319066	-0.070212372	0.018835898	-0.191518294
6	-0.008692774	0.12823918	-0.051460709	0.090388937	-0.149713082
7	0.005618355	0.126814471	-0.011185349	0.175602219	-0.067541962
8	0.02605661	-0.045071172	0.06572233	0.158736235	-0.040836511
9	-0.005327184	0.00279858	0.021076684	0.093271136	-0.001370737
10	0.024950764	0.042650095	0.084768893	0.006893182	0.054327243
11	0.004449744	0.028264103	-0.079886275	-0.072933283	-0.32540097
12	-0.002283612	0.078461257	-0.07743598	-0.08062246	-0.359245456
13	-0.029779992	-0.01519902	-0.097756226	-0.377055372	-0.478885105
14	-0.010889326	0.067352701	-0.05850421	-0.117366756	-0.486611183
15	0.007718968	-0.063695371	0.051874342	-0.223660229	-1.010973926
16	0.079807442	0.143249261	0.068372151	-0.094512358	-0.868415196
17	-0.007249102	0.092553439	-0.070132794	0.030339539	-0.187863443
18	-0.021066697	0.074244435	0.000322928	0.08306471	-0.323962344
19	-0.005768364	0.043524304	-0.046624011	0.024544996	-0.265029076
20	0.005125643	0.085995702	0.045083529	0.175340027	-0.304946369
21	-0.04660661	0.134442253	0.003338245	0.167558638	-0.132336617
22	-0.001073484	-0.014264825	0.039919004	0.237907277	-0.154473848
23	-0.007227595	0.01962483	0.069958922	0.297829562	-0.154399907
24	-0.0345083	0.066413332	0.04358507	0.282677415	-0.10859913
25	0.022218259	0.020812484	-0.080145577	0.164367496	0.270430147
26	-0.005262235	0.145501993	0.111800118	-0.244926374	0.235258301

Table A2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
27	0.024554508	0.041823187	-0.018702087	0.018517835	0.360133338
28	0.010492593	0.010393004	-0.055970047	0.104292947	0.329399519
29	-0.020994974	0.062709526	-0.091531892	-0.01332628	-0.191028621
30	-0.01342612	0.071730036	-0.065458715	0.005779522	-0.243757557
31	-0.046382364	0.091052591	0.071320733	0.25347511	-0.259792432
32	-0.045278465	0.04417868	0.039948801	0.233014485	-0.18247064
33	-0.024750278	0.042520206	0.083309783	0.299981454	-0.20216896
34	0.03061089	-0.002005924	-0.009782187	-0.154670244	-0.693510722
35	0.027007612	0.051640487	-0.02425738	-0.150896122	-0.651139364
36	-0.023822405	0.041649394	-0.073804666	-0.078375222	-0.356810606
37	0.041678326	-0.036395761	-0.095401475	-0.071629868	-0.252320638
38	0.015140095	-0.025849431	-0.035675015	-0.171872977	-0.635852516
39	-0.008273148	0.032560351	-0.051206324	-0.180757768	-0.616609599
40	-0.012000187	0.014199306	-0.028066196	-0.207399221	-0.589207205
41	0.039668031	0.062666573	-0.057003838	-0.123096429	-0.493589326
42	0.017373896	0.019172079	-0.003395934	0.034325594	-0.374870244
43	0.022295385	0.074656255	0.098997052	0.050459789	-0.336623586
44	-0.00930135	0.058354848	-0.043047649	-0.027514103	-0.291140309
45	0.007807367	0.01470963	0.027529997	0.084185464	-0.394705158
46	0.045602202	0.021615369	0.028736202	0.102675567	-0.338450575
47	0.019187406	-0.002667494	0.051005184	0.202112638	-0.255032452
48	0.002636831	-0.007868777	0.047497801	0.227719801	-0.198682839
49	0.006096882	0.034395056	0.059234116	0.219673395	-0.245699497
50	-0.007464158	-0.057208502	0.074658506	0.326839968	-0.101063275
51	0.016792645	-0.01720615	0.097943547	-0.010761621	0.057634403
52	-0.030993441	-0.02541038	-0.097663617	-0.000918411	-0.130653167

Table A2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
53	0.01413934	-0.00641701	0.124255989	-0.200939744	0.079243706
54	0.010034015	0.040898228	0.301168144	-0.410103445	0.130752732
55	-0.017488506	0.048820503	-0.087865999	-0.053904811	-0.230329627
56	-0.041095546	0.045994418	0.00049768	-0.325175178	-0.502333422
57	-0.002037659	-0.011408592	-0.052142871	-0.142067963	-0.533517328
58	0.035115284	0.047658753	0.003664174	-0.166971419	-0.339117589
59	-0.02299586	0.043212738	-0.055848366	-0.050407851	-0.357976504
60	0.013548209	-0.004564956	0.00701215	0.066076964	-0.354393672
61	0.01931229	-0.016648275	0.076566343	-0.03850722	-0.17972732
62	0.006642079	-0.059449977	0.018995379	0.153970831	-0.225114047
63	0.001931494	-0.06764024	-0.046311907	0.038991774	-0.158950189
64	0.008307581	-0.050293032	-0.004459376	0.173948728	-0.11348568
65	-0.00517297	-0.009279113	0.021657072	0.055224061	-0.043125472
66	0.010437881	-0.068815861	0.008197459	0.06630757	-0.069572406
67	-0.02443418	-0.030960113	-0.100953744	0.033367606	-0.056043862
68	-0.011367755	-0.07210455	-0.095159911	0.069948908	0.001382241
69	-0.004784208	-0.020817311	-0.091382755	0.052982654	-0.0477594
70	-0.027457948	0.018352277	-0.063026008	-0.002217238	-0.256505315
71	0.011005654	-0.018840626	-0.034112728	-0.005626195	-0.344530946
72	0.01138714	-0.013315308	-0.034724847	0.003264435	-0.327691953
73	-0.00818493	-0.000373048	-0.045610495	-0.03394141	-0.364182456
74	0.008292749	0.019222708	-0.048406713	-0.100302553	-0.476059459
75	0.010090972	-0.026237047	-0.018065989	-0.114379273	-0.591412368
76	0.003256416	0.054213064	-0.010151924	-0.054717628	-0.528964646
77	0.033080725	0.019372372	-0.030225825	-0.097634669	-0.527826338
78	-0.000789605	0.076666104	-0.050201881	-0.008435615	-0.31897467

Table A2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
79	-0.032399205	0.091952576	-0.032382259	0.039011005	-0.288269535
80	-0.016131458	0.074998785	0.053945291	0.183245755	-0.320673946
81	-0.008273793	0.102284073	0.067857356	0.254803441	-0.237513544
82	0.047584451	0.014807399	0.004310471	0.1200488	-0.24787793
83	0.028149434	-0.005178554	0.022805955	0.184584103	-0.191497913
84	0.021948486	0.045029541	0.08510558	0.041602009	0.326772965
85	0.039207471	0.051538709	-0.05093608	0.065501267	0.351094172
86	0.032526859	0.001458591	-0.089459177	0.133130086	0.321528877
87	0.00726556	0.0611895	0.375279499	-0.250518067	0.454628407
88	0.025477702	0.051770071	-0.059900931	-0.11131372	-0.462504535
89	0.014928816	0.059805447	-0.095686142	-0.072571153	-0.277930794
90	0.02191088	0.049378101	-0.080288693	-0.174982965	-0.509482169
91	-0.025185035	-0.032144895	-0.014094383	-0.071601222	-0.531465184
92	0.014145873	0.013281456	0.183133844	0.071226854	-0.136698423
93	0.00610768	0.035440966	0.17058513	0.078702417	-0.106839201
94	0.015671566	-0.012958347	0.020701797	0.19900918	-0.158722956
95	0.039399624	0.018645544	-0.00971149	0.047236816	-0.126251781
96	0.057912487	-0.00673477	0.053696586	0.17235019	-0.009017029
97	0.010571271	-0.033498041	-0.025605193	0.072304042	0.022079917
98	-0.032306121	-0.053602608	-0.089311576	-0.028445866	-0.092920645
99	-0.015146991	-0.065902473	-0.114680095	-0.03611538	-0.104759552
100	-0.009896718	-0.057262378	-0.112233328	-0.037609196	-0.124824122
101	-0.01571878	-0.011067802	-0.101688673	-0.191550559	-0.460706511
102	0.003953406	0.051770586	-0.089883095	-0.104784642	-0.354785508
103	-0.036616705	0.027170462	-0.064814645	-0.155453914	-0.529075115
104	-0.043663664	-0.138705273	-0.020826004	-0.231036805	-0.774953926

Table A2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
105	-0.007238216	0.027558092	-0.070021304	-0.207460528	-0.516833849
106	0.008269004	0.061568779	-0.049771547	-0.134146616	-0.541597794
107	-0.018172632	0.038005147	-0.04806287	-0.108130407	-0.499154708
108	0.005210078	-0.003758007	-0.002754479	0.007395883	-0.428892715
109	0.020930237	-0.000107015	0.020094948	0.060208467	-0.40808327
110	0.001729181	-0.028862481	0.020563587	0.158468936	-0.229606291
111	-0.010583602	-0.004937865	0.053893124	0.179306274	-0.282110675
112	0.054662673	0.072334427	0.109946563	0.204254039	-0.159419493
113	0.022083548	-0.01899512	0.063979051	0.261255319	-0.080247831
114	-0.005911689	-0.076570902	-0.026928154	0.117808945	-0.034366115
115	0.039680963	-0.010559644	0.071993183	0.060821031	0.004504772
116	0.003702214	-0.000463319	0.01400114	0.108439606	-0.024898805
117	0.030248467	-0.003333675	-0.025230416	-0.093537434	-0.204649327
118	-0.010810462	-0.089090838	-0.097320677	-0.013402025	-0.137398083
119	-0.019036955	-0.05910595	-0.062290529	-0.019936849	-0.272626942
120	0.003534741	-0.041800257	-0.04585384	0.01849826	-0.258453081
121	0.00585458	-0.066908529	-0.095551747	-0.028378621	-0.172603487
122	-0.013572906	0.012338928	0.004909138	0.049177394	-0.385381028
123	-0.000845423	-0.004648991	-0.007008303	-0.001819808	-0.432062111
124	-0.005186632	-0.013686496	-0.052974137	-0.002137015	-0.252182408
125	0.013629393	-0.073605552	-0.049625524	0.002804576	-0.265779668
126	0.01263343	-0.005303728	0.001032512	-0.149970395	-0.722662772
127	0.006799958	-0.012222407	-0.00477868	-0.064035895	-0.548429931
128	0.025165613	-0.049342254	-0.011744415	-0.066418708	-0.51894581
129	0.044098962	-0.009237515	0.028090419	-0.009873233	-0.555067055
130	0.029471507	-0.08148504	0.138774952	0.101353307	-0.085815851

Table A2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
131	0.000506162	-0.053615938	0.198731827	0.165061173	0.024247219
132	-0.000216555	-0.067038047	0.156886124	-0.029735547	-0.0226874
133	-0.057566568	-0.013502381	0.183376006	-0.133118392	-0.009831796
134	-0.04447429	0.012463708	0.155667203	-0.21090783	0.008932788
135	-0.01792932	-0.099382684	-0.065195282	-0.087486296	-0.092509854
136	-0.01190894	-0.025298904	-0.111304559	-0.080069686	-0.197833862
137	-0.032761904	0.026502629	-0.102586491	-0.082499917	-0.272519789
138	-0.007239391	-0.011481805	-0.058125796	-0.202280179	-0.622627995
139	-0.011088022	0.051699923	-0.046529899	-0.144395948	-0.571337383
140	-0.000685213	0.039532063	-0.055361198	-0.116978273	-0.488756415
141	-0.022680587	0.071127069	-0.036062962	0.03406651	-0.291283282
142	-0.005029562	-0.005442913	-0.036890214	-0.086274488	-0.485438061
143	-0.017111249	0.059073237	-0.040822618	-0.149214339	-0.601326906
144	0.016547208	0.038715942	-0.029633456	-0.072307624	-0.490889717
145	0.025744744	-0.021475388	0.021950155	0.063213424	-0.351774685
146	-0.002668874	-0.029028629	-0.003205367	0.042307644	-0.360301307
147	0.012265703	-0.034438392	0.074674382	0.256126102	-0.230278805
148	0.018138043	-0.03753433	0.077917335	0.294975148	-0.169697065
149	0.008821563	-0.08287737	0.106045654	0.122017939	-0.065602292
150	0.029767469	-0.06642603	-0.03974619	0.125469154	-0.067834537
151	-0.023459725	-0.000238418	0.092071552	0.119675705	-0.010959432
152	-0.001854815	-0.002930315	0.200290911	0.051962974	0.09558206
153	-0.006887917	0.064442056	-0.031489276	-0.078253432	-0.504418839
154	-0.006974523	0.019170795	-0.104049093	-0.093134959	-0.281490022
155	-0.019524197	0.038382115	-0.093515811	-0.101431817	-0.337198041
156	0.001999682	-0.017628687	-0.067574328	-0.015070554	-0.252475433

APPENDIX B

TESTING SET

Table B1 Testing Set of DHF Outbreak Risk Class

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
1	-0.001322485	0.124741526	-0.092508873	0.032046801	0.342652208
2	-0.010573516	0.074204564	-0.016623024	0.012513672	0.50853037
3	-0.006084139	0.039341204	-0.060845093	-0.013241522	0.560904758
4	-0.025081778	-0.009533539	-0.193019536	0.091883928	0.485884445
5	-0.000104491	0.004677689	-0.167386866	0.160339346	0.534133048
6	-0.019579952	0.039245512	0.185528789	0.022878684	0.648859549
7	-0.006750367	0.040329475	-0.165396424	0.044052761	0.546585693
8	0.002453364	0.074419301	-0.184689079	0.108301383	0.506084009
9	-0.007681071	0.130119001	0.476138643	-0.306459171	0.683665419
10	0.009950246	0.090453505	0.080505351	-0.028304062	0.612773779
11	-0.023179815	0.082570321	0.335715553	-0.310672607	0.695409424
12	-0.00038872	0.048906327	0.027937511	0.038771092	0.578634774
13	0.001897081	-0.001570536	-0.145986175	0.113399741	0.524918092
14	-0.020842795	-0.003248799	-0.055073522	0.173870464	0.536942482
15	0.002209574	0.01802737	-0.174051011	0.087410804	0.5422595
16	0.013379416	0.035931947	-0.108425781	0.091604675	0.569985582
17	-0.001893512	0.070101071	-0.089516499	-0.074500477	0.554456785
18	0.001163643	0.026363373	-0.14201764	0.160452748	0.576508752
19	-0.000259703	0.045859104	-0.058925083	-0.067854941	0.58643839

Table B1 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
20	0.000417014	0.02113621	-0.183221491	0.203041196	0.594355144
21	-0.027612924	0.051730055	-0.067482634	0.128557253	0.491731882
22	0.003892103	-0.033255625	-0.168750219	0.131073953	0.550697501
23	0.005581758	0.033129502	-0.187617571	0.078940127	0.598979317
24	0.006686297	0.000130297	-0.060271095	0.275252704	0.5245406
25	-0.009404702	-0.034220464	-0.107989749	0.207796675	0.513225411
26	0.014147475	-0.001569786	-0.125069102	0.100566749	0.550105406
27	0.007880686	-0.016578075	-0.101468803	0.060521879	0.539935113
28	0.036881724	0.029133543	-0.138509453	0.134326704	0.533383694
29	-0.019655273	0.034773958	0.123281926	-0.15273506	0.605371413
30	-0.003612062	-0.008296571	-0.094309117	0.312082675	0.536254862
31	0.017847579	0.06239205	0.115177893	-0.025404431	0.649394083
32	0.013928201	0.050791691	-0.002111674	0.057025291	0.587238164
33	-0.033416337	0.090136384	0.018044669	0.029939656	0.554516899
34	0.007468284	-0.005549251	-0.173004713	0.102299022	0.52928946
35	0.00845399	0.008183468	0.015086528	0.139358136	0.520893857
36	-0.034333105	0.091405796	0.292018246	-0.245611279	0.581850965
37	0.026342051	0.044246988	-0.064440487	0.225022523	0.517433636
38	0.010386806	-0.01740787	-0.084219974	0.280041983	0.543204638
39	0.004606152	-0.00084721	-0.118675405	0.138494966	0.523495484
40	-0.001164939	0.009289592	0.068418104	0.089604283	0.595039226
41	0.001012341	0.103446283	0.159634029	0.081035014	0.649644326
42	0.036195068	0.059778796	0.039758375	0.118311066	0.624106456
43	-0.01304888	-0.002077895	-0.028907489	-0.02511229	0.565861144
44	0.021119511	0.024101317	-0.095478986	0.209810281	0.502489588
45	0.003499249	-0.025712299	-0.046835252	0.035080769	0.479632351

Table B1 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
46	-0.013057785	0.069395724	-0.053491414	0.156066994	0.463214389
47	0.009615101	0.018458609	0.056469438	0.043372407	0.604066079
48	0.024069279	0.048205684	-0.090633059	0.035739898	0.533458853
49	-0.012056973	0.016476656	-0.119089518	0.023949021	0.55095041
50	-0.010860799	-0.042671874	-0.134399701	0.114531451	0.524377949
51	-0.020266288	-0.019591168	-0.059005024	0.220474664	0.563856196
52	-0.013894141	0.035886939	-0.093611035	0.100879676	0.521549201
53	-0.003169576	-0.002612676	-0.106708993	0.012979157	0.547462163
54	-0.013904642	0.022731317	-0.176986795	0.082852272	0.503714016
55	-0.011179859	0.0227024	0.000913982	-0.082633852	0.580533772
56	-0.003896752	0.029854607	0.054652204	-0.103139558	0.580488345
57	-0.029533399	0.030051912	-0.095025255	0.115944921	0.497914454
58	-0.007792565	0.008461315	-0.178926794	0.164708034	0.492502875
59	0.007775133	0.042944054	-0.104013525	0.056059678	0.520504896
60	-0.005310495	0.021629237	-0.03641359	-0.043197333	0.518748319
61	-0.019643645	-0.009313428	-0.150239193	0.162894931	0.376232471
62	-0.02063254	0.034812973	0.048458976	-0.056425056	0.518174669

Table B2 Testing Set of Non DHF Outbreak Risk Class

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
1	-0.036774764	0.063565191	-0.032409202	-0.10383183	-0.117563468
2	0.055743255	0.09952713	-0.14801156	-0.081926015	-0.066171926
3	0.012230885	0.09311862	-0.089012353	-0.015559483	-0.073325728
4	-0.002404511	0.172060631	-0.035743303	0.160485278	-0.080342472
5	0.021631471	0.079900403	-0.029457993	0.203320148	-0.00466563
6	0.007163387	0.101289819	0.08834248	0.374715014	-0.092834928
7	-0.028602842	0.024227043	0.138002582	0.445872056	-0.117412068
8	-0.009920451	0.105637931	0.208410402	0.333886286	0.04703725
9	-0.03410501	0.032979904	0.064900462	0.42254103	0.079548223
10	-0.027091701	-0.051880562	0.075155374	0.089421548	0.074807047
11	0.001804631	-0.011120249	-0.029040318	0.181759827	0.025320701
12	-0.015794794	-0.079227773	0.03391587	-0.266531718	0.101579307
13	0.010037629	-0.045444116	-0.0850078	-0.317358302	0.079984369
14	-0.054220378	-0.004971994	0.595737277	-0.625999733	0.260963576
15	-0.000820504	-0.023073418	-0.071727947	-0.184908362	0.082593565
16	0.001301156	0.004624551	-0.076135764	-0.095595103	0.069587542
17	0.005809331	-0.012789563	-0.026366337	-0.149227708	0.102928733
18	0.004375927	-0.018935794	-0.0519536	-0.243011235	0.055529718
19	0.006917688	-0.027649189	-0.16415012	-0.077073533	-0.001009347
20	-0.007738062	-0.023001173	-0.165982629	-0.067749561	-0.022124022
21	-0.004274829	-0.03672721	-0.140022325	-0.046845588	-0.031506256
22	-0.021950129	-0.030763924	-0.109241611	-0.092131335	-0.200720069
23	-0.016737383	-0.002523951	-0.030162519	-0.286711918	-0.073026632
24	0.013168706	-0.028906473	0.01028836	-0.242299214	-0.074710808
25	-0.038953484	0.023282991	-0.128275141	-0.119855864	-0.148829685
26	-0.020519388	0.053632793	-0.060593918	-0.094918684	-0.430966586

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
27	-0.02647725	0.141264925	-0.008872526	0.069163611	-0.333968014
28	0.007647989	0.07479903	-0.009018253	0.036533057	-0.372678235
29	0.002973708	0.06138043	0.055272915	0.27408884	-0.060973287
30	0.117627387	0.187777751	-0.005444693	-0.180812181	-0.053466223
31	-0.008426496	0.037858302	0.127059692	0.024996914	0.158241615
32	-0.026618068	-0.018220889	0.028607209	0.369328577	0.116165282
33	0.009439851	0.102598342	0.251421444	0.029888082	0.296654859
34	0.03082048	0.036060494	-0.085945449	0.108067264	0.21493208
35	-0.026847829	-0.004042749	-0.035689476	0.071237912	0.203531445
36	-0.011480238	0.014894168	0.182769401	-0.05227199	0.284671019
37	-0.025518861	0.025264968	-0.096263704	0.08763435	0.137249631
38	-0.00830139	-0.012658689	-0.040573696	0.015173265	0.331596563
39	-0.011957824	-0.014943196	0.109076426	-0.174667768	0.374642026
40	-0.018579753	-0.013645	-0.062793691	-0.156341831	0.307859565
41	0.012312813	-0.029837183	-0.140951781	0.074170151	0.297245586
42	0.007127316	0.037279361	0.108578286	-0.178790951	0.388764965
43	-0.005984406	-0.03987702	-0.128009674	-0.172843801	0.290461017
44	0.02241065	0.025440909	-0.055569465	0.036460002	0.327058513
45	0.011116014	0.016015977	-0.046287398	-0.119351567	0.321609167
46	-0.002840173	0.03426844	0.166146576	-0.161323593	0.363775641
47	-0.025191514	-0.028979161	-0.132388793	0.017539549	0.267749401
48	-0.008115821	0.011962673	-0.131597556	0.076316861	0.328847791
49	-0.012895082	-0.039699388	-0.080167267	-0.078019421	0.331576732
50	-0.016726519	0.044601554	0.216203318	-0.303355409	0.393249814
51	-0.03431092	0.056471002	0.030582794	-0.128872038	0.25818049
52	-0.029680027	0.054452304	-0.169862933	0.00738716	0.138224819

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
53	0.01358876	0.064796684	-0.170765432	-0.037003015	0.240860653
54	-0.001707425	0.005772086	-0.191083684	-0.029884469	0.213173909
55	-0.028068548	0.009369918	-0.129721147	0.056945491	0.178116435
56	-0.00260608	-0.019280317	-0.101333718	0.120882429	0.13290094
57	-0.015031809	0.053549668	-0.13902724	-0.040397039	0.198961681
58	-0.046019767	0.193977273	-0.003574015	-0.060089203	-0.198109838
59	-0.033319771	0.080244913	-0.093299706	-0.002777089	-0.171837838
60	-0.024622193	0.072769292	-0.100991299	0.023919242	-0.095588703
61	-0.030670054	0.118555135	-0.086935441	-0.00050406	-0.155978202
62	-0.09028469	0.102719334	-0.105477446	0.029550704	-0.036285985
63	-0.059479972	0.099880273	-0.100504813	0.006569914	-0.110720138
64	-0.028063133	0.05427897	-0.12612548	-0.040463378	-0.109666085
65	0.006552525	0.106664352	-0.032764719	-0.027043277	-0.415523021
66	-0.007634559	0.11260982	0.013965307	0.079533341	-0.381846737
67	-0.022873194	0.1094363	-0.092287187	0.001160173	-0.147153631
68	-0.08218851	0.159784543	0.07993777	0.264243082	-0.289081267
69	-0.03757469	0.178175975	0.033711607	0.148054115	-0.342645576
70	0.010169733	0.025467592	0.121053319	0.393195636	-0.15073202
71	0.007988124	0.087438916	0.106545819	0.233509648	-0.006342606
72	0.107064989	0.14465347	-0.005170247	-0.143599049	0.183583645
73	-0.007298227	-0.009298943	-0.046037685	0.25872655	0.164990715
74	-0.020089826	-0.00489417	-0.020376885	0.225374231	0.086852436
75	-0.012494976	-0.086249505	0.012019957	0.269404105	0.008389199
76	0.0033283	-0.007592802	0.124434984	0.222649374	0.018099284
77	-0.01063967	-0.06611761	0.054700903	0.334156138	-0.001439023
78	-0.032909114	-0.142841578	0.147984015	0.512188174	0.00461658

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
79	0.080353962	0.047592808	0.083283068	0.100762825	0.089181501
80	0.004341808	-0.043527586	0.111738048	-0.09247912	0.075964247
81	0.032747743	0.001844107	0.058831682	-0.223535999	0.179553806
82	0.008106732	-0.054556063	0.00821071	-0.163016261	0.155603442
83	-0.031191639	-0.006225365	0.645947609	-0.547652673	0.326384159
84	-0.003812823	-0.03275665	-0.000953971	-0.143234351	0.073503516
85	-0.011804858	-0.00090189	0.208815906	-0.258926777	0.146344786
86	-0.00113895	-0.048906076	-0.014793223	-0.194350403	0.077576451
87	-0.001323396	-0.047179367	0.011740603	-0.162518226	0.067067367
88	0.004985214	0.002061285	-0.049788546	-0.117579845	0.051564798
89	0.000531962	-0.038400467	-0.156367216	-0.052984148	-0.016025896
90	0.041498632	0.001994304	-0.003534691	-0.114942418	0.002545635
91	0.065850623	0.076915866	-0.141381289	-0.124015464	-0.026727854
92	0.008775573	-0.050293404	-0.152735021	-0.05719012	-0.038602464
93	-0.054633965	0.042755315	-0.027619205	-0.056198949	-0.479698241
94	-0.018525161	0.008813375	-0.080747495	-0.11528594	-0.350068902
95	-0.023370245	0.066113923	-0.042142409	-0.089576623	-0.413399652
96	0.016387639	-0.001949382	0.096707589	0.147676239	-0.504393471
97	0.016713708	-0.028185481	0.092667909	0.298511069	-0.211617129
98	0.010090372	-0.001639325	0.253689418	0.286451157	-0.160115874
99	-0.005927014	-0.062853221	0.147697821	0.321167283	-0.049231658
100	0.006410171	-0.078847142	0.08970712	0.299106088	-0.025244394
101	0.022789567	0.004849551	0.249556723	-0.419245662	0.168107509
102	0.021389664	-0.062401127	-0.062446258	0.013643177	0.064005563
103	0.038612502	-0.028605034	-0.10493233	0.060241162	0.109098452
104	0.002084048	-0.100153005	-0.036047037	-0.132691967	0.109738293

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
105	-0.006687235	-0.07303247	-0.018354784	-0.060010296	0.097115949
106	0.00301131	-0.067437363	0.091611235	-0.263820761	0.152557214
107	-0.016659167	-0.027653131	-0.073757198	-0.131183203	0.024045561
108	-0.00112417	-0.109162618	0.013326072	-0.216113914	0.113779868
109	-0.000464715	-0.077506279	0.103442337	-0.315529833	0.15420805
110	0.001515085	-0.036512417	0.057978121	-0.301120498	0.089086112
111	0.025856522	-0.040146081	0.010972397	-0.108029465	0.119515049
112	0.029455882	-0.013683439	0.031589231	-0.166677418	0.108060472
113	0.036361611	-0.020265951	-0.107695539	-0.138180145	0.079292803
114	0.029770461	-0.076913274	-0.124934615	-0.122399632	0.078118027
115	-0.008906055	-0.089109483	-0.128743965	-0.019339217	-0.005864849
116	0.012151762	-0.064954852	-0.150522611	-0.12475623	0.048579717
117	-0.030620865	-0.048015623	-0.043794919	-0.063009251	-0.359435006
118	0.013004234	-0.02309166	-0.062384287	-0.160965986	-0.369784599
119	-0.008849355	-0.031727415	-0.119012543	-0.147040211	-0.190167266
120	0.046623927	0.057148052	-0.078734166	-0.31269527	-0.115473769
121	0.0061385	-0.008742456	-0.148497808	-0.12161279	-0.150577701
122	0.051388646	-0.020351613	-0.155634665	-0.167665753	-0.1680484
123	-0.056849018	0.058041761	-0.079218442	-0.147087483	-0.476097078
124	0.002268255	0.00622591	-0.105540504	-0.306811994	-0.260438993
125	-0.024984382	-0.038688142	-0.082264319	-0.298065009	-0.366056358
126	-0.000887133	-0.04429134	-0.127506296	-0.226795066	-0.429309213
127	-0.006083502	0.011502725	-0.054475235	-0.092669363	-0.442668741
128	0.009985346	0.0083673	0.065543868	0.126728465	-0.44221026
129	0.007212747	-0.023710534	0.106973312	0.227016436	-0.39236695
130	-0.007163789	-0.039172751	0.054383574	0.191078934	0.0187541

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
131	-0.025271392	-0.106947637	0.030364882	0.246426082	-0.063611867
132	-0.008379945	0.001843617	-0.077073529	0.303033531	0.473746011
133	0.01811708	-0.070762583	-0.029553281	-0.129409099	0.113171364
134	0.003608145	-0.057589818	-0.101610652	-0.143135367	0.09046797
135	0.000407242	-0.054535581	0.052894249	-0.179384064	0.140641622
136	0.012045577	0.00709631	0.335081444	-0.422045424	0.221156766
137	0.007042533	-0.00562562	0.292962553	-0.358562255	0.211223672
138	0.006001993	-0.124318273	-0.125598382	0.044426329	0.079546171
139	0.007367998	-0.10543805	-0.111126714	0.034867812	0.037825278
140	0.034733676	-0.002623355	0.022853399	-0.134708158	0.009085302
141	0.004691785	-0.033389939	-0.150160162	-0.088215165	-0.086716945
142	0.023103169	-0.034370212	-0.0397517	-0.039601322	-0.381793833
143	-0.000558481	0.075694227	0.006902007	0.072714562	-0.361670948
144	0.003312906	0.000616619	0.10497074	0.308523165	-0.245224794
145	-0.001173819	-0.010418565	0.133792227	0.384508617	-0.202214438
146	-0.026199554	0.081017624	0.110565035	0.402552121	-0.116621579
147	0.01892465	-0.012336201	0.078482867	0.360578149	0.0904781
148	-0.002705896	-0.045904993	0.188179768	0.582438415	-0.018886799
149	0.016654191	0.026295676	0.131558419	0.489499438	0.05492246
150	0.021339545	0.020642826	0.093424857	0.115509378	0.143507356
151	0.002518012	-0.022239784	0.047293098	0.390956431	0.160881064
152	-0.002199604	-0.026544897	0.079661034	0.316569875	0.178979967
153	0.003145594	0.054171854	0.238580598	0.229201706	0.305561898
154	-0.002318386	-0.038959409	-0.111036369	0.201909365	0.320284899
155	-0.002570832	0.042223266	0.042992439	-0.107296477	0.308470855
156	0.005952844	0.041901711	-0.042752783	-0.016689441	0.308405916

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
157	-0.001294837	-0.024882353	-0.111460575	-0.021561168	0.287293674
158	-0.013683278	-0.133449411	-0.035967471	0.077445955	0.282491943
159	0.016248614	0.013253568	0.015760711	-0.098789748	0.383326988
160	0.011641868	0.011021162	-0.002512596	-0.052463792	0.314583475
161	0.029092134	-0.012766899	-0.044160676	-0.0388177	0.325298308
162	0.003014496	-0.041625093	-0.099912091	-0.020153495	0.319858385
163	-0.019529887	-0.028952327	0.117506768	-0.284022024	0.356345723
164	0.031614159	0.032342358	0.064127893	-0.130339711	0.386542826
165	0.003287051	-0.029883701	-0.105554833	-0.084621915	0.346664494
166	0.006699596	-0.014626398	0.04041543	-0.047756408	0.407764341
167	-0.027829175	-0.037191691	0.001426149	-0.129150081	0.35638981
168	0.001889268	0.052978006	0.154516409	-0.295889004	0.34885841
169	0.002524762	-0.028521331	0.071580636	-0.312500665	0.317221188
170	-0.005077939	-0.026627299	-0.068027522	-0.112121271	0.278792362
171	0.003821862	-0.004223126	-0.133541857	0.070886942	0.219245585
172	-0.017125268	-0.067930198	-0.033565192	-0.041497246	0.240582259
173	0.001555488	-0.034123269	-0.192350321	-0.074958556	0.204067267
174	0.019663736	0.014476391	-0.168988129	-0.010548171	0.203251515
175	0.010199015	0.031536008	-0.114695067	0.024386923	-0.035460634
176	0.0140611	0.020445816	-0.113443556	0.041691347	-0.005537002
177	0.010197677	0.023514565	-0.12265791	0.078536853	0.089868728
178	0.029039961	0.084293802	-0.096996266	0.027866838	0.025876805
179	-0.059963876	0.006025727	-0.105473022	-0.132190033	-0.210213595
180	-0.034908341	0.036925159	-0.073168774	-0.152768767	-0.261886513
181	-0.00992035	0.0323051	-0.067171693	-0.116199567	-0.448073795
182	0.015562424	0.11025142	-0.055492811	-0.046991883	-0.37574088

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
183	-0.003898514	0.042891769	-0.024950027	-0.022630459	-0.42098077
184	0.00353355	-0.002434046	0.046988702	0.06966779	-0.481949216
185	-0.01664241	0.038850567	0.060855813	0.133472259	-0.425084754
186	0.020959433	0.024929394	0.087569482	0.178303432	-0.424403647
187	0.037273608	-0.013088575	0.202916996	0.390198249	-0.414325384
188	0.015411877	0.007398778	0.136516989	0.286994613	-0.387887615
189	0.06175272	-0.026659328	0.168269635	0.314318617	-0.429650156
190	0.014834761	0.002604375	0.161012945	0.342752112	-0.367719926
191	-0.009808137	-0.051106616	0.031865508	0.280054517	-0.030396416
192	-0.039948115	-0.087898696	0.076511013	0.338152547	-0.08481437
193	-0.029172852	-0.143009342	0.121275472	0.419263272	-0.059861044
194	-0.011395296	-0.078379269	0.04746165	0.331810137	0.029001325
195	-0.004405025	-0.031501707	0.035084906	0.263846251	0.02018416
196	0.035042673	0.010230766	0.039931841	0.035194435	0.135375922
197	0.022830569	-0.049100526	0.009803278	-0.103979725	0.174336917
198	0.008456931	-0.117824347	-0.112973395	0.006946031	0.104081268
199	-0.002530104	-0.07815824	-0.102300044	-0.102100098	0.07078566
200	0.01368939	-0.103675806	0.005963285	-0.011321651	0.113827729
201	0.004260106	-0.130017427	-0.030812826	0.093616266	0.120880426
202	0.016372727	-0.039133805	0.153903515	-0.106746892	0.14024967
203	0.023537292	0.02526764	0.355121467	-0.332913398	0.246064728
204	0.027575913	-0.063099164	-0.128946037	-0.061454776	0.098225315
205	0.012180414	-0.04831403	0.067667943	-0.15400781	0.152689901
206	-0.018336128	-0.124847415	-0.04264208	-0.262248991	0.094419545
207	0.012254391	-0.085807227	-0.128172709	-0.254439891	0.087729867
208	0.00327791	-0.053970646	0.089183491	-0.272837225	0.170040296

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
209	-0.011083348	-0.027292613	0.024183295	-0.159741529	0.1258346
210	0.004727812	-0.017136233	0.085702891	-0.228762008	0.129103076
211	-0.009607068	-0.027363165	0.106020151	-0.225301136	0.09284125
212	0.032966742	-0.042511041	-0.030623216	-0.125501397	0.108799643
213	0.023186441	0.036325979	0.081433237	-0.186974099	0.126269939
214	0.009793694	-0.054937573	-0.095054568	0.026131515	0.030414558
215	-0.015633032	-0.050790832	-0.008958922	-0.150868178	-0.05703216
216	0.014511705	0.040271098	-0.027239057	-0.248665428	-0.094024858
217	0.037924551	-0.027992664	-0.13649563	-0.064107526	0.017016515
218	0.040739084	0.015621077	-0.05616022	-0.263495943	-0.161468728
219	0.007605649	0.02121822	-0.079591225	-0.098856112	-0.370945638
220	0.023268611	0.042589262	0.018878016	-0.015199732	-0.548504851
221	0.009427535	-0.028981241	-0.045306512	-0.117980916	-0.307434355
222	-0.005821774	-0.060692789	0.079803614	0.25903131	-0.238864898
223	0.013862644	0.018551635	0.113484185	0.348175974	-0.204541919
224	0.007634785	-0.031411237	0.122803976	0.341652985	-0.237028838
225	-0.018438833	-0.068308569	0.203955423	0.300386035	0.034302992
226	0.005432714	-0.040026792	0.089689817	0.075917369	0.075314873
227	0.008551371	-0.050262941	0.073791098	-0.145949964	0.076112927
228	-0.015735714	-0.044294634	0.033476284	-0.140113707	0.085828434
229	-0.006198118	-0.104941632	0.037837258	0.010893291	0.086627891
230	0.017049264	-0.085937122	-0.096653313	0.064495915	0.07570129
231	-0.022124321	-0.052909982	0.15266947	-0.204840568	0.12611067
232	-0.006296513	-0.145490918	-0.083838721	-0.024667003	0.096397662
233	0.027652337	-0.042241804	0.088799926	-0.226301443	0.174222721
234	-0.008698117	-0.096114306	0.136617734	-0.146533161	0.176928909

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
235	-0.002068231	-0.118701508	-0.146037694	-0.157113777	0.060149661
236	-0.002146983	-0.048456101	0.111753315	-0.151472682	0.142822425
237	0.004922077	-0.029810848	0.286030341	-0.288806057	0.165842711
238	-0.041276137	-0.074113879	0.163046298	-0.264062919	0.135579046
239	0.014407193	-0.035698255	0.085931325	-0.12832748	0.160506838
240	0.001838496	-0.033557457	0.168475448	-0.328266449	0.16963804
241	0.031314153	-0.025786851	0.026117992	-0.176244106	0.118668787
242	-0.003292725	-0.061096218	-0.065543425	-0.026130322	0.095539196
243	0.015193473	-0.111112281	-0.102296274	-0.050961791	0.033646082
244	-0.009328684	-0.036579109	0.38756938	-0.501447113	0.190377928
245	-0.013390939	-0.028606087	0.032457128	-0.10274957	0.037613111
246	0.012077493	-0.066218322	0.015572865	-0.111639689	0.089310875
247	0.004865859	-0.032048941	-0.036564487	-0.264486471	0.009839785
248	0.004580953	-0.050555295	-0.142693253	-0.045030735	-0.002888255
249	-0.006376466	-5.80E-05	0.08305969	0.086090326	-0.573589222
250	-0.048291854	0.010568666	0.053489678	0.045998033	-0.556217855
251	0.01135039	-0.055773276	0.140646141	0.236959112	-0.478180528
252	0.014152195	-0.058786243	0.189389944	0.27859106	-0.563417744
253	0.00991956	-0.010690989	0.124072197	0.250498943	-0.409149641
254	0.019279688	-0.056739277	0.209701926	0.358108547	-0.487424317
255	0.020918309	0.052339284	0.141778882	0.292220025	-0.404922823
256	-0.014455149	-0.087413511	0.183981584	0.237652558	-0.264673239
257	-0.015964226	-0.131516986	0.172824254	0.460780717	-0.16985194
258	-0.008939518	-0.141492274	0.255450854	0.61026593	-0.171205514
259	-0.000755337	-0.12651912	-0.066448613	-0.105797409	0.106100278
260	-0.011890413	-0.108871157	-0.004563171	0.051500069	0.097076255

Table B2 (continued)

No.	Rainfall	Minimum Temperature	Maximum Temperature	Mean Temperature	Relative Humidity
261	0.025319138	-0.088227653	-0.08236036	-0.120438165	0.025749896
262	0.01890561	-0.053493507	-0.004965607	-0.190669837	0.096388677
263	0.00186091	-0.100603445	-0.07903288	0.0345219	0.102027111
264	-0.024518376	-0.078395398	0.033640972	-0.174855481	0.116118182
265	0.011767346	-0.025656966	0.206137114	-0.250803127	0.184018058
266	-0.005717878	-0.090883544	0.049870356	-0.1087818	0.118590237
267	0.016846587	-0.016913245	0.058056166	-0.156181649	0.138992794
268	-0.028447821	-0.068361146	0.146547641	-0.210924812	0.146042863
269	0.010832997	0.017018155	0.144817475	-0.217537282	0.137733665
270	0.005083861	-0.043544182	-0.075001894	-0.246077178	0.03374402
271	0.027674694	0.038117371	-0.06426464	-0.188269445	-0.180441279
272	-0.02303733	-0.032375493	-0.066607387	-0.112455822	-0.417357492
273	-0.00387384	-0.026520779	-0.021568753	-0.099731419	-0.283459916
274	0.019938163	0.035497186	-0.028424212	-0.046473242	-0.447293009
275	-0.004605436	0.023106591	-0.033799349	-0.048777452	-0.434479584
276	0.071633657	0.084287533	-0.034319517	-0.229593193	-0.245887516
277	-0.003027926	0.012702378	0.011097535	0.022099827	-0.453034785
278	-0.059687066	-0.051091995	0.029806314	-0.132591304	-0.427840997
279	0.009383192	-0.106714144	-0.000504383	0.033248108	-0.366792809
280	-0.014318807	-0.129142802	0.054769716	0.33093567	-0.013244073
281	-0.000109695	-0.006861919	0.125410782	-0.183104631	0.066076341
282	-0.001126686	-0.086885255	-0.105552742	-0.032533787	0.073952792
283	-0.006618481	-0.06537646	0.072600685	-0.210413296	0.042909799
284	-0.045955063	0.004513497	-0.07852186	-0.079574115	-0.343872255
285	-0.048926767	0.054124367	-0.002973207	-0.030651807	-0.516921376
286	-0.048593559	0.095824558	-0.08231779	-0.091392486	-0.338890593

APPENDIX C

TRAINING TIME

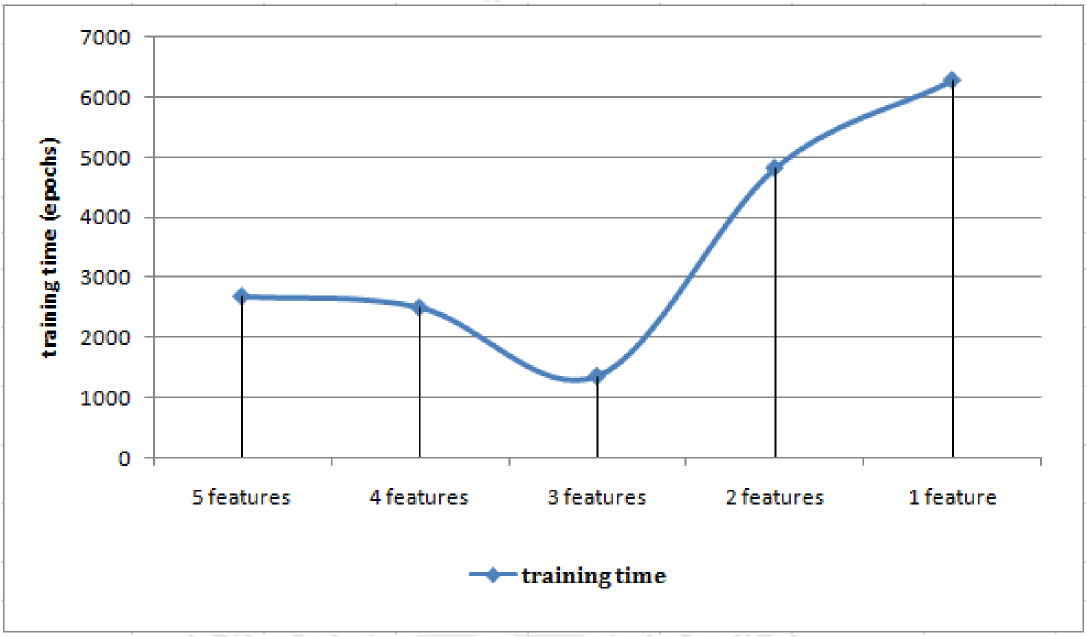


Figure C1 Training Time between Data Patterns of Five Weather Features and Data Patterns of Each Principal Feature

From Figure C1, it represents the result of training time with data patterns of five weather features and data patterns of each principal feature as follows: (1) 2,688 epochs with data patterns of five weather features (rainfall, minimum temperature, maximum temperature, mean temperature, and relative humidity), (2) 2,503 epochs with data patterns of four principal features (minimum temperature, maximum temperature, mean temperature, and relative humidity), (3) 1,358 epochs with data patterns of three principal features (maximum temperature, mean temperature, and relative humidity), (4) 4,825 epochs with data patterns of two principal features (mean temperature and relative humidity), and (5) 6,287 epochs with data patterns of one principal feature (relative humidity).

These results show that the minimum training time is 1,358 epochs with data patterns of three principal features, whereas, the maximum training time is 6,287 epochs with data patterns of one principal feature. It can be concluded that the number of appropriate data patterns and principal features effect the amount of few training time.

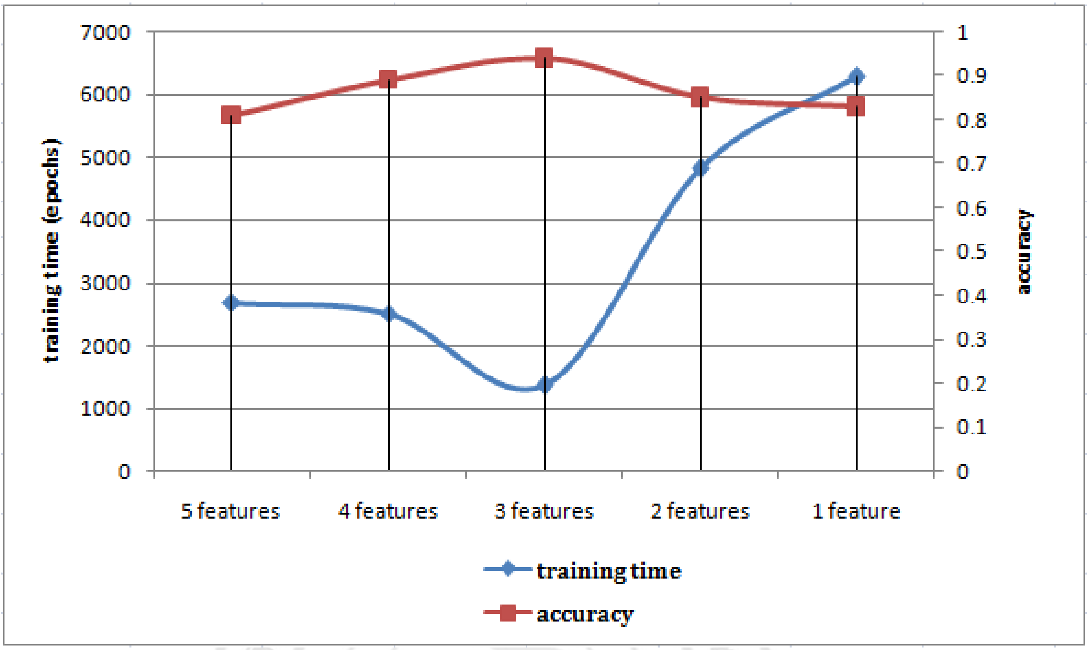


Figure C2 Comparison of Training Time and Accuracy

From Figure C2, it represents a comparison of the result of training time and the result of accuracy with data patterns of five weather features and data patterns of each principal feature as follows: (1) 2,688 epochs and 0.81 accuracy with data patterns of five weather features (rainfall, minimum temperature, maximum temperature, mean temperature, and relative humidity), (2) 2,503 epochs and 0.89 accuracy with data patterns of four principal features (minimum temperature, maximum temperature, mean temperature, and relative humidity), (3) 1,358 epochs and 0.94 accuracy with data patterns of three principal features (maximum temperature, mean temperature, and relative humidity), (4) 4,825 epochs and 0.85 accuracy with data patterns of two principal features (mean temperature and relative humidity), and (5) 6,287 epochs and 0.83 accuracy with data patterns of one principal feature (relative humidity).

This comparison indicates that if the result of training time decreases, the result of accuracy increases, whereas, if the result of training time increases, the result of accuracy decreases. It can be concluded that the effective data patterns of principal features, which is data patterns of three principal features, provide the fewest training time and the most accurate result.

CURRICULUM VITAE

NAME	Miss Napa Rachata
DATE OF BIRTH	12 February 1984
ADDRESS	553 Moo 24, Tambon Robwieng, Amphoe Muang, Chiang Rai 57000

EDUCATIONAL BACKGROUND

2005	Bachelor of Business Administration Major in Business Computer Naresuan University Phayao

WORK EXPERIENCE

2006 – 2007	Teacher Assistance at Naresuan University Phayao

PUBLICATION

2008	Rachata, N., Charoenkwan, P., Yooyativong, T., Chamnongthai, K., Lursinsap, C., & Higuchi, K. (2008). Automatic prediction system of dengue haemorrhagic-fever outbreak risk by using entropy and artificial neural network. In **International Symposium on Communications and Information Technologies** (pp. 210-214). Vientiane, Lao PDR.