

CHAPTER 1: THE PROBLEM

This study was dealing with the prediction of malaria incidence in Banggai Regency, Central Sulawesi, Indonesia. This section discusses the rationale, theoretical framework, conceptual framework/ paradigm, statement of the problem, hypothesis, assumption, scope and delimitation, importance of the study, and definition of terms of the study.

1.1 Rationale

Malaria is an endemic disease in most of Indonesian area, especially in rural and remote areas. On a national scale, malaria is one of diseases as a part of global commitment of Millenium Development Goals (MDG's). Its spreading and incidence are targeted to be stop and reduced in 2015 [11]. Banggai, one of regencies in Central Sulawesi province, is one area that has high endemic of malaria. Data in 2010 shows that malaria Annual Parasite Incidence (API) in Banggai Regency reached 7.88%, this number were further above than targeted API in malaria control program that is $API < 1\%$ [6]. Several studies [9][12][17] show that malaria incidence and spreading are influenced by environmental and weather factors, particularly rainfall and temperature. This study was aimed at developing a malaria incidence prediction system based on environmental and weather factors. This was to provide information about prediction of malaria incidence, so that the Indonesian Ministry of Health might take strategic planning in order to control malaria.

Evolving Neural Network (ENN) is a method that integrating Evolutionary Algorithms (EAs) and Artificial Neural Network (ANN). In prediction system, ANN is widely used because of its accuracy and speed. Several studies on malaria prediction [12][17] use back-propagation (BP) algorithm for ANN learning, but there are two shortcomings of this algorithm, namely (1) correction weight which is limited to searching space of steepest descent method and less able to explore searching space which can lead to worse solution or early convergence, (2) sensitive to parameter setup such as number of

hidden layers, number of neuron in hidden layer, learning rate, etc, because the setup was very dependent to the problem encountered [5]. Because of those shortcomings, some studies recommend Genetic Algorithm (GA) as an EAs to replace commonly learning algorithm in ANN, like BP. The advantage of using GA for ANN learning are the definition of ANN parameters can be encoded genetically and evolved in the same time, so that at the same time there are many solutions which can be selected [5][8]. A study [4] shows that the availability of multiple solutions provide learning performance which is more superior instead of only one solution available.

1.2 Theoretical Framework

This study attempts to predict malaria incidence in Banggai, Indonesia based on weather factors. The input for this system is time series data of weather factors, while the output of this system is a number of malaria incidence in a particular month. In learning process, prediction model is constructed using ENN. First, GA is used to get the optimum ANN structures; then GA is reused to train the ANN. In order to prevent overfitting, cross-validation is performed to select the most optimum structure and weights. To measure performance of the prediction model, the model is tested using testing dataset. Finally, prediction model with best performance (most high accuracy) is used to predict malaria incidence.

1.3 Conceptual Framework/Paradigm

There are three variables applied to conduct measurement on this study, namely:

Variable	Variable's Information
Data composition	Composition of data used as training dataset, validation dataset, and testing dataset.
Number of time series	This variable indicates length of time series of past data that is used as input.
GA parameters	Parameter to setup GA, consists of population size, crossover probability, and mutation probability.

1.4 Statement of the Problem

The main problem discussed in this study was applying ENN to develop prediction system of malaria incidence in Banggai Regency based on time series data of weather factors.

1.5 Hypothesis

ANN is widely used for prediction in a number of areas because of its accuracy. The advantages of ANN for prediction are data error tolerance and ability to approximate complex non-linear mappings [20]. However, there are some difficulties in designing ANN. It does not have general method to determine the optimum structure for solving any problem and the correction weight is limited to searching space of steepest descent method which can lead to worse solution or early convergence. Those are impacted to the accuracy. Therefore this study proposed ENN which used GA to design ANN. ENN can generate many structures and weights (solutions) at the same time. The availability of multiple solutions provides learning accuracy which is more superior than only one solution available, because from those solutions can be selected the most optimum solution which produces the highest accuracy. Thus from those premises, using ENN can produce high accuracy in predicting malaria incidence.

1.6 Assumption

The assumptions of this study were:

1. This study did not conduct further analysis about correlation between weather factors and malaria incidence. All weather factors obtained from BPS-Statistical of Banggai Regency were assumed to have good correlation with malaria incidence in Banggai.
2. Weather forecasting was conducted using time series forecasting program from other study and it was assumed that the accuracy was reliable [16].

1.7 Scope and Delimitation

The scopes and delimitations of this study were:

1. The architecture of ANN was fully connected.
2. This study was conducted based on monthly weather data from BPS-Statistical of Banggai Regency and monthly malaria incidence data from Indonesian Ministry of Health from 2004 – 2009.
3. Maximum number of ANN hidden layer and hidden neuron were 2 and 16.
4. Number of time series of each weather factor was from 1 previous month to 4 previous month.

1.8 Importance of the Study

This study might reveal the ENN performance for predicting malaria incidence, especially ENN which applied GA to optimize ANN structure and to train ANN. Besides that, this study might be useful for the government of Indonesia, especially Ministry of Health, because it can be used to predict the malaria incidence in each regency based on environmental and weather factors at the regency, so that they could make a strategic planning to prevent outbreak and reduce the number of malaria incidence in order to control malaria incidence in Indonesia.

1.9 Definition of Terms

Malaria incidence	The number of malaria incidence that were recorded in Indonesian Ministry of Health.
Weather factors	Factors that indicate the weather condition in an area.
Outbreak	A condition characterized by the increasing in incidence of morbidity/ mortality that epidemiologically significant in a region within a certain period and can lead to the epidemic.