THE PREDICTION OF DENGUE HAEMORRAGIC FEVER (DHF) IN CIMAHI USING HYBRID GENETIC ALGORITHM AND FUZZY LOGIC

Fhira Nhita[1], Prof. The Houw Liong[2], Shaufiah[3]

[1]Magister Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Kejadian Demam Berdarah Dengue (DBD) merupakan masalah nasional di bidang kesehatan di Indonesia. Setiap tahun angka kesakitan DBD masih tinggi. Khususnya di Cimahi, salah satu kota di provinsi Jawa Barat dimana angka kesakitan (Incidence Rate) tahun 2005 hingga 2010 di atas standar yang ditentukan oleh Departemen Kesehatan RI.

Banyak faktor yang mempengaruhi kejadian DBD, antara lain iklim dan perilaku hidup bersih dan sehat (PHBS). Oleh karena itu, dibangun Sistem Prediksi Demam Berdarah yang dikaitkan dengan iklim yang diharapkan bisa membantu memberikan informasi bagi Departemen Kesehatan Cimahi tentang prediksi resiko DBD di tahun yang akan datang, sehingga Departemen Kesehatan Cimahi dapat mengambil langkah preventif untuk mengurangi angka kesakitan DBD.

Sistem Prediksi yang dibangun dengan hybrid algorithm yaitu Algoritma Genetika dan Logika Fuzzy mampu menghasilkan akurasi testing 100% dalam memprediksi kondisi DBD di 6 bulan pertama pada tahun 2009 dan 2010 di kecamatan Cimahi Utara dan Cimahi Tengah. Sedangkan pada Cimahi Selatan diperoleh hasil prediksi 6 bulan pertama di tahun 2009 sebesar 100% tetapi pada tahun 2010 terjadi penurunan akurasi sehingga akurasi hanya sebesar 66.67% .

Untuk hubungan antara IR dengan PHBS, karena data PHBS hanya ada per tahun sedangkan data IR dan iklim per bulan sehingga tidak cukup untuk menganalisis hubungan antara IR dan PHBS, dan penurunan atau penaikkan nilai PHBS tidak signifikan sehingga untuk data ini, PHBS tidak mempengaruhi nilai IR.

Kata Kunci : Demam Berdarah, IR, Algoritma Genetika, Logika Fuzzy, PHBS

Abstract

Incidence of Dengue Haemorrhagic Fever (DHF) is a national health problem in Indonesia. Every year dengue morbidity is still high. Particularly in Cimahi, one of the city in West Java province where the morbidity rate (Incidence Rate) in 2005 to 2010 in above the national standard.

Many factors affect the incidence of dengue, such as, climate and living behavior. Therefore, the development of DHF Prediction System which is associated with a climate is expected to help in providing information to Health Department of Cimahi about dengue risk prediction for the coming year. Hence, the Health Departement of Cimahi can take preventive action to reduce morbidity of DHF.

Prediction System that was built with a hybrid algorithm which Genetic Algorithms and Fuzzy Logic is able to obtain 100% testing accuracy in predicting the condition of dengue in the first 6 months in 2009 and 2010 in the district of Northern Cimahi and Central Cimahi. In Southern Cimahi, the prediction results obtained for the first 6 months of 2009 amounted to 100% and with a decrease in accuracy in 2010, the accuracy only 66.67%.

For the relationship between IR and living behavior data, the living behavior data only available yearly while the weather and IR data in monthly. Hence, the data is not adequate to view the relationship between the IR and living behavior and the decreasing or increasing of living behavior is not significantly so that for the data, the living behavior did not influenced IR.

Keywords : Dengue Haemorragic Fever, Incidence Rate, Genetic Algorithm, Fuzzy Logic,Living Behavior

CHAPTER 1
THE PROBLEM

This study is dealing with the prediction of Dengue Haemorragic Fever in Cimahi, West Java, Indonesia. Before it is discussed in great detail, the Rationale, Theoretical Framework, Conceptual Framework/Paradigm, Statement of the Problem, Hypotheses, Assumption, Scope and Delimitation, Importance of the study, and Definition of terms are discussed briefly in chapter 1.

1.1 Rationale

Dengue haemorrhagic fever (DHF) is one of the endemic diseases in various parts of Indonesia. The number of DHF cases tends to increase and spread wider. It is stimulated by many factors such as: agent or vector, living behavior, and immune of human [3,4,7]. The growing up of agent or vector is influenced by weather and living behavioral too [3,7]. Dengue fever is very infectious, its spread cannot be limited by the concept of territory. Up to this moment, no medicine nor vaccine has been found to overcome the disease. The disease is still potential to cause outbreaks and usually occurs with a grace period of 3 to 5 years [7].

Cimahi, one of the city in West Java, had high incidence of DHF. It ranked first from year 2005 to 2007 and second in years 2008 and 2009. The Health Department of Cimahi so far has recorded the incidence number of DHF in each district of Cimahi. But, the data only is only used to obtain the SMP (*Sebelum Masa Penularan*). It is a very limited analysis because no observation of the factors that can cause the disease is ever taken.

In response to those problems, this study analyzes the relationship among the factors influencing the incidence of this disease and generate patterns of the relationship among the factors to predict the DHF conditions in the future. This study is expected to provide optimal patterns or models and accurate predictions so that the Health Department of Cimahi can take into action earlier in the future to prevent DHF spreading in this area.

1.2 Theoretical Framework

This study attempts to predict the DHF conditions in Cimahi for each district in the future. Firstly, the prediction system will search the optimal input for fuzzy logic includes the type and parameter of membership function, number of linguistic value and rules using genetic algorithm and fuzzy grid tables. Whereas, generating the input data for prediction is using the forecast method.

In this study, fuzzy logic is used because fuzzy logic can classify the data with fuzzy or blur, and not crisp so if DHF condition is red, it may not be fully red, it has membership values in range 0-1. Genetic algorithm is used because fuzzy logic system need optimal input to obtain optimal result if there is no information from an expert. Last, fuzzy grid tables are used to generate rules that will be the input to the fuzzy system.

1.3 Conceptual Framework/Paradigm

This sub chapter will discuss research variables and the relationship with the conceptual research. There are at least eight research variables applied in this study namely :

Table 1-1 Research Variables

No	Relationship with Conceptual Research	Variable
1	Class, output, or consequent for fuzzy system	Incidence Rate (IR)
2	The factors that stimulated DHF incidences (Antecendent for fuzzy rule)	Weather data, consist of : temperature, rainfall, solar radiation, relative humidity, and wind velocity
4	The factors that stimulated DHF incidences	Living Behavior, consist of : Using clean water, washing hands, using latrines, fighting larvae, vegetables and fruits consumption, physical activities, and no smoking.
5	Input for GA	Population size (pop size), Probability crossover (pc)
6	Output of GA / Input for fuzzy	Number of linguistic value , Type and parameter of membership function, Rules
7	GA evaluation variable / Optimal input for fuzzy system	Fitness value
8	Comparison of class target and class output	Training, testing or prediction accuracy

1.4 Statement of the Problem

Statement of the problems of this study are :

1. What are the correlations of IR and weather data ?

2. What are the correlations of IR and living behavioral ?
3. What is the optimal input to fuzzy system which influence the accuracy of prediction result ?
4. What is the optimal combination parameter of GA to predict DHF condition in the future based on learning data ?

1.5 Hypothesis

The hypothesis of this study are :

1. There are correlation of IR and weather data, because weather data is one of factors that influenced the growth of agent/vectors.
2. There are correlation of IR and living behavioral data, such as if the human has good living behavioral on any weather conditions then the DHF condition is lower.
3. The optimal input to fuzzy system may result in an accurate system to predict the spread of DHF.
4. There is the optimal combination parameter of GA to predict DHF condition in the future based on learning data.
5. The system can obtain the prediction accuracy more than 80%.

1.6 Assumption

The assumptions of this study are :

1. Living behavioral data is only available for two years, 2008 and 2009, but the data used for this study is the data from 2005 to 2009. So to bridge the gap, regression is used for each living behavioral data of each district because it is assumed that there is a trend of up or down every year.
2. Data for training and testing is labeled based on IR
3. Fitness value and accuracy of testing and prediction by comparing class target and class output
4. Number of linguistic value of DHF condition is 2 based on surveillance information and the rule of the Indonesia Health Departement
5. Minimum number of linguistic values of input variable is 2 and maximum is 3.
6. The membership function type that observe in this study only 3 type namely : Traphesium, Phi, and Triangle.

1.7 Scope and Delimitation

The scopes and delimitations of this study are :

1. This study is conducted based on the data taken from the Health Department of Cimahi
2. This study used weather and living behavioral data from 2005 to 2009 in each district of Cimahi
3. This study used ANN method to forecast the weather data but this study did not discuss in detail about ANN.

1.8 Importance of the Study

The importance of this study is the government especially the Health Department will be able to predict DHF conditions in each district earlier so that they can take more comprehensive actions to prevent the outbreak and reduce the number of DHF incidences (IR).

1.9 Definition of Terms

ANN	Artificial Neural Network, one of soft computing algorithm
Fitness value	Maximum accuracy of best individu/solution of GA
SMP (Sebelum Masa Penularan)	The month or time when a district or area has the lowest incidence so that prevention can be carried out and the month is the calculation result of the last 5 years data
IR (Incidence Rate)	The rate of DHF spreading based on standard of Indonesia Health Departement
Population size (pop size)	Number of candidate solution
Probability crossover (pc)	Probability of successfully recombination

CHAPTER V
CONCLUSIONS AND RECOMMENDATIONS

This chapter includes conclusions about the study and recommendations for future work.

5.1 Conclusions

Based on the experiments, the optimal input for fuzzy system is very important if we do not have a priori information from an expert. Genetic algorithm can generate optimal input or best solution for fuzzy input includes number of linguistic values, type and parameter of membership function, and rules so fuzzy system can predict DHF conditions using the model that is generated from hybrid system,

In all district of Cimahi, the maximum testing accuracy was obtained by the different combination paramater of GA. The 1st combination parameter (population size=10 and pc=0.2) for Northern Cimahi and 5th combination parameter (population size=50 and pc=0.5 for Central Cimahi. In Northern and Central Cimahi, the maximum testing accuracy for 6 months in year 2009 and 2010 are 100%, but in Southern Cimahi, the maximum testing accuracy for 6 months in 2009 = 100% but in 2010 for 6 months is 66.67%.

For the relationship between IR and living behavior data, the living behavior data only available yearly while the weather and IR data in monthly. Hence, the data is not adequate to view the relationship between the IR and living behavior and the decreasing or increasing is not significantly so that for the data, the living behavior data did not influenced IR.

5.2 Recommendations

The recommendation for hybrid algorithm are exploring the other forecasting method to obtain the better prediction of DHF condition, modifying the combination parameter of population size and probability crossover, and adding the type of membership function to obtain the better result of testing and prediction.

Whereas, to Health Department, it is recommended to evaluate the data, to check the method of data collection and the data collection can be done by the independent organization. So that, the development of prediction model can obtain more accurate result.

BIBLIOGRAPHY

1. D O Fuller, A. T. (2009). El Nino Southern Oscillation and vegetation dynamics as predictors of dengue fever cases in Costa Rica. *Environmental Research* .

2. Dr. S. Santhosh Baboo, I. S. (2010). An Efficient Weather Forecasting System using Artificial Neural Network. *International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010* , 321-326.

3. Iswari, L. (2008). Pemanfaatan Sistem Inferensi Fuzzy Dalam Pengolahan Peta Tematik (Studi Kasus : Sistem Informasi Geografis Daerah Rawan Penyakit Demam Berdarah). *Seminar Nasional Aplikasi Teknologi Informasi 2008 (SNATI 2008) ISSN: 1907-5022* , E77-E84.

4. Lukito, H. (2007). *Perumusan Pola Penyebaran Demam Berdarah Melalui Data Mining Pada Database Dinas Kesehatan DKI Jakarta.* Bogor: Institut Pertanian Bogor.

5. Mohsen Hayati, a. Z. (2007). Application of Artificial Neural Networks for. *World Academy of Science, Engineering and Technology 28 2007* , 275-279.

6. Mubarok, M. S. (2009). *Sistem Identifikasi Plagiat Menggunakan Kombinasi Algoritma Genetika – System Fuzzy.* Bandung: IT Telkom.

7. RI, D. K. (2005). *Pencegahan dan Pemberantasan Demam Berdarah Dengue di Indonesia.* Jakarta.

8. S. Wongkoon, M. P. (2008). Predicting DHF Incidence in Northern Thailand using Time Series Analysis Technique. *International Journal of Biological and Life Sciences 4:3 2008* , 117-121.

9. *Science Daily*. (2009, June 17). Retrieved January 10, 2011, from http://www.sciencedaily.com

10. Sulistyo, M. D. (2010). *Analisis dan Implementasi Sistem Fuzzy and Evolutionary Programming pada Pengaturan Lampu Lalu Lintas Cerdas.* Bandung: IT Telkom.

11. Suyanto. (2005). *Algoritma Genetika dalam Matlab.* Yogyakarta: Andi Publisher.

12. Suyanto. (2008). *Soft Computing.* Bandung: Informatika.