

# Developing a Dengue Forecasting Model: A Case Study in Iligan City

Ian Lindley G. Olmoguez<sup>1</sup>, Mia Amor C. Catindig<sup>2</sup>, Minchie Fel Lou Amongos<sup>3</sup>, Fatima G. Lazan<sup>4</sup>

Bachelor of Science in Information Technology<sup>1, 3, 4</sup>

College of Computer Studies, IT Department<sup>2</sup>

MSU-Iligan Institute Technology Iligan City, Philippines<sup>1, 2, 3, 4</sup>

**Abstract**—Dengue is a viral mosquito-borne infection that is endemic and has become a major public health concern in the Philippines. Cases of dengue in the country have been recorded to be increasing, however, it is reported that the country lacks predictive system that could aid in the formulation of an effective approach to combat the rise of dengue cases. Various studies have reported that climatic factors can influence the transmission rate of dengue. Thus, this study aimed to predict the probability of dengue incidence in Iligan City per barangay based on the relationship of climatic factors and dengue cases using different predictive models with data from 2008 to 2017. Multiple Linear Regression, Poisson Regression, and Random Forest are integrated in a mini-system to automate the display of the prediction result. Results indicate that Random Forest works better with 73.0% accuracy result and 33.58% error percentage, with time period and mean temperature as predictive variables.

**Keywords**—Dengue; predictive models; Pearson's correlation; multiple linear regression; Poisson regression; random forest

## I. INTRODUCTION

Dengue has been an endemic infection in over 100 countries in the world, in tropical and subtropical regions. One of the four dengue viruses had been classified as dengue serotypes (DENV-1, DENV-2, DENV-3, and DENV-4), carried by the main vectors, the *Aedes Aegypti* and *Aedes Albopictus* [1]. Reports were gathered from the World Health Organization (2018) that the Dengue fever was the most critical and rapidly spreading mosquito-borne viral disease in the world for over the past 50 years with 390 million dengue infections per year in 3.9 billion people in 128 countries at risk of infection.

In the Philippines, Dengue fever has become one of the major health problems among the populace. The Department of Health (DOH) reported a total of 138,444 dengue cases nationwide from January 1 to October 6 2018 representing 21% increase on the number of cases in the same period in 2017 [2]. Just recently, there were 115, 986 dengue cases recorded in the Philippines including 491 deaths reported from January 1 to July 6 2019; 86 % higher than in 2018 [3]. Same report cited Region X as one of the regions with highest incidence rate having 9,354 cases [3]. Incidence of Dengue was caused by several factors, one of which was the climatic conditions referring to temperature, relative rainfall, and relative humidity were reported to be important influential Dengue transmitters [4]. Studies discovered that places with high temperatures and higher rainfall such as in the Philippines

had high dengue transmission rates resulting to more steady water as potential breeding grounds for mosquitoes [3]. Humidity had been a consistent, significant weather factor that provided favorable conditions for Dengue vectors [5].

Despite the effort of the government to look for possible ways to control the increase of dengue cases in the Philippines, there is still no specific solution or response on how to control the dengue outbreaks up to this writing. This circumstance necessitates the implementation of primary safety measures to reduce and prevent dengue infections, to control mosquito populations, and limit the spread of dengue cases nationwide. Since climate conditions influence the dengue transmission cycle [6] [7] [8], a dengue risk-prediction system based on the relationship of dengue incidence and climatic conditions is investigated in this study. The development of a risk prediction system could forecast the locale of possible high incidence rate of dengue thus will have significant contribution in controlling the spread of dengue by reducing the transmission of mosquitoes [9].

Predictive analytical approach using a variety of machine learning, modeling, statistics, artificial intelligence, and data mining algorithms could be input data to predict unknown events in the future. Also, the use of statistical methods, correlations between dengue incidence and climatic variables were established to predict potential outbreaks in specific areas. Promprou [6] sampled a predictive model to predict the Dengue Haemorrhagic Fever (DHF) in Thailand using Multiple Linear Regression model to explain the relationship between the household's activities and DHF patients. Results of the study revealed a 26.9% of the variation of DHF patients using a number of water storage containers, *Aedes Aegypti* in drainage of refrigerators, pH and temperature of water in container. In the study by Ong et al. [7] predicting the Dengue incidence with the use of Random Forest approach, predicted the risk rank of dengue transmission in Singapore with dengue cases, population, entomological and environmental data. The evaluation using the latest dengue case data in the study showed a strong predictive ability for the model, compared to the study results of Tilwani, Dave, & Nadurbarkar [8], that adopted a regression approach with Poisson Regression and Negative Binomial to investigate the correlation between dengue incidence and climatic fluctuations including relative humidity, temperature, and pressure. Dengue cases with 70% accuracy showed the impact of climatic fluctuations in dengue transmission.

### III. METHOD

Although studies had been conducted by the foregoing authors, there remained a dearth of the study in Iligan City, Region X of the Mindanao Province, Philippines. It was on this premise that this study was conceptualized and developed to predict the number of dengue incidence based on the correlation between climatic conditions (temperature, relative rainfall, and relative humidity). Three predictive models were used by the study: Poisson Regression, Multiple Linear Regression and Random Forest. The data were limited only to climatic factors including the temperature, relative rainfall, and relative humidity, as provided for by the concerned office. The study will attempt to evaluate and compare the results of these predictive models that best fit in the case of Iligan City and discover and visualize the probability of dengue to arise in a particular area.

#### II. RELATED STUDIES

A study by Ong et al. (2018) [10] used a Random Forest approach that predicts the risk rank of dengue transmission in Singapore with dengue; population, entomological and environmental data. The predicted risk ranks are then categorized and mapped to color-coded risk groups which were evaluated with dengue cases and cluster data. According to its findings, the study demonstrates the potential of Random Forest and its strong predictive ability to stratify Singapore's spatial risk of dengue transmission. It suggests that population density, dengue burden and abundance of *Ae. aegypti* are significant risk factors for dengue transmission. Evaluation using the latest data of dengue cases showed a significant predictive ability for the model. Strong positive correlation between the observed and predicted ranks of risk and an almost perfect agreement between the predicted levels of risk and the density of the case were observed.

In the study of Carvajal, Viacrusis, Hernandez, Ho, Amalin, & Watanabe (2018) [11], various predictive models such as General Additive Modeling, Seasonal Autoregressive Integrated Moving Average with exogenous variables, Random Forest, and Gradient Boosting were used to predict the temporal pattern of dengue incidence in Metropolitan Manila, and to compare their predictive accuracy. Among the statistical modelling techniques, Random Forest showed a better predictive accuracy.

With the above mentioned studies, it has been supported that predictive analytics with the support of statistical methods are useful for predicting the possibilities of dengue incidence in various areas. Moreover, identifying the relationship of climatic factors and dengue transmission is considered to be one of the effective predictor. The presented study intends to predict the probability of dengue as well with dengue incidence rate and climatic factors such as temperature, rainfall, and humidity as the basis. Given the results from the related studies presented, it is perceived that predictive models are promising in predicting dengue incidences as they can provide significant prediction result.

#### A. Study Area

The study area is in Iligan City in Lanao del Norte, Philippines. It is an urbanized city with 44 barangays, with estimated population of 400,000 according to 2015 census.

#### B. Research Design

A quantitative study was utilized using mathematical models and statistics to analyze and provide more objective numerical results [12]. The presented research was quantitative in nature as it relied on the analysis of the secondary numerical data collected to describe and predict the probability of dengue incidence in Iligan City using predictive models.

#### C. Data Collection and Analysis

This study used secondary data obtained from the proper authority. A reported monthly dengue fever incidence data of 44 barangays in Iligan City over the period from 2008 to 2017 were provided by the City Health Office (CHO). The climatic data over the same period comprised monthly average temperature, maximum and minimum temperature, relative humidity and relative rainfall were also collected from the City Health Office (CHO). These data were used to analyze factors affecting the occurrence of dengue cases and to predict the next incidents of dengue in Iligan City.

The collected data were imported to Python to evaluate inaccuracy or inconsistency in the data such as duplicate columns, not a number (NAN) values, and columns and rows that were not part of the explanatory factors. NAN values were replaced with 0, while gender and age were excluded as independent variables for these variables were presented daily while climatic data are presented by month. Fig. 1 shows a sample of the cleaned data while Fig. 2 is a sample of the climatic values.

A separate the data per year was kept by separating them to tabs and changing the variable names to max temperature, min temperature, mean temperature, relative rainfall, max relative humidity, min relative humidity, and average relative humidity.

Barangay	January	February	March	April	May	June	July	August	September	October	November	December
0	ABUNO	1	0	0	0	0	0	0	0	0	0	0
1	ACMAC	0	1	0	0	0	0	0	0	0	0	0
2	BAGONG SILANG	0	1	0	0	0	1	0	0	0	0	0
3	BONBONON	0	0	0	0	0	0	0	0	0	0	0
4	BUNAWAN	0	0	0	0	0	0	1	0	0	0	0

Fig. 1. Cleaned Sample Data of Dengue Cases.

2008	CLIMATIC DATA											
	January	February	March	April	May	June	July	August	September	October	November	December
Temperature												
max	30.1	29.9	31.1	32.3	32.2	31.5	31.6	31.6	31.9	31.3	30.9	30.6
min	21.1	21.3	21.3	21.8	22.2	21.8	21.9	21.4	21.7	21.7	21.7	22.3
mean	25.6	25.6	26.2	27	27.2	26.6	26.7	26.5	26.5	26.3	26.3	26.4
Relative Rainfall	183.8	90.1	83.4	165.8	170.7	228.2	240.7	193.2	253.2	101.4	101.4	143.4
Relative Humidity												
max	95	95	90	90	93	91	90	88	94	92	92	92
min	66	72	75	72	70	74	79	77	78	79	79	74
average	84.39	84.79	82.74	80.57	74.26	82.57	84.48	82.16	85.55	85.27	85.27	84.61

Fig. 2. Sample Climatic Data in 2008.

#### D. Development of the Predictive Model

Correlation analysis was used to determine the strength of the relationships between the monthly number of dengue cases, dependent variables and climatic factors (minimum temperature (tmin), maximum temperature (tmax), mean temperature (tmean), relative rainfall (rr), minimum relative humidity (rhmin), maximum relative humidity (rhmax), and average relative humidity (rhmean)) and time period (timeperiod) as independent variables, covering 120 months in total from 2008 to 2017.

To see how the data sets were correlated, Pearson's Correlation Coefficient was used. Given that the data are continuous, this method is suitable to perform as it is generally used when variables are continuous in nature such as in ratio or interval scale variables. Pearson's correlation coefficient is indicated by  $r$  and defined by:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\{n\sum x^2 - (\sum x)^2\}\{n\sum y^2 - (\sum y)^2\}}} \quad (1)$$

The value of  $r$  always ranges from -1 to +1. The relationship between the variables is said to be not related when the value of  $r$  comes down to 0. If the value of  $r$  lies to +1, then the variables are said to be positively correlated, while the variables are said to be negatively correlated if the value of  $r$  is -1.

The development of the forecasting model is based on Multiple Linear Regression (MLR), Poisson Regression, and Random Forest.

Multiple Linear Regression analysis is a statistical technique that used several explanatory (independent) variables to predict the outcome of a response (dependent) variable. The model is developed with the following equation.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n \quad (2)$$

Where

$b_i$  = y-intercept

$y$  = dengue cases

$x_i$  = climatic factors and time period

The value of R-squared and adjusted R-squared were calculated in order to test how well the data fit the regression model. Usually, the higher R-squared value between 0 and 100 means the better the regression model fitted the observations. To identify the significance of each of the independent variables, P-value was also computed.

The Poisson Regression on the other hand is designed to fit a model of regression in which counts were made with the dependent variable  $Y$  (dengue cases). The fitted model  $Y$  to one or more  $X$  predictor variables (climatic factors and time period), which were either quantitative or categorical. A poisson regression model defined as:

$$z = e + b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_kx_k \quad (3)$$

where  $z = \log(y)$ , transformed from  $y$  in generalized linear modelling called link function. This was done so that a linear regression modelling in  $z$  satisfied all required assumptions [11]. The corresponding coefficients ( $b_1, b_2 \dots b_k$ ), which

were originally modelled for  $z$ , have to be transformed back properly to a model for  $y$  for interpretation. To estimate the  $b$  values, a process called maximum likelihood estimation (MLE) or weighted least squares may be used where it fitted a model.

Random forest is an ensemble of simple tree predictors used to determine the final outcome. It used a bootstrap sampling approach to generate  $k$  different training data subsets from an original dataset, and then  $k$  decision trees were constructed by training these subsets. The final value was calculated by taking the average of all the predicted values by all the trees in forest. Since there was no given regression coefficients in building the Random Forest model, variable selection was done through performing the Feature Importance method, which variables had the most effect on the model. Visualizing tree in Random Forest was to evaluate the predictions for each row using all the trees in the model, how each variable contributed to the final prediction. To calculate for the result, proponents have used the RandomForestRegressor class of the sklearn.ensemble library in Python.

However, before performing the prediction process in MLR, selecting highly significant independent variables based on their p-value was done first in order to build a best model, called the Stepwise Regression.

Dengue Prediction System: The proposed system was implemented using Python and PostgreSQL as the backend, while the front end was developed using CSS, Javascript, HTML, Leaflet, and Django. This system will automate the prediction of the model chosen, provided that data are presented in the correct format.

#### IV. RESULTS AND DISCUSSION

As shown in Fig. 3, most of the values were not close to 1 which indicates that there is a weak correlation between independent variables. On the other hand, timeperiod associated with cases has an  $r$  of 0.3318 and has the greatest correlation coefficient among the other independent variables. Other independent variables such as tmin and tmean have positive correlation with dengue cases having  $r$  values equal to 0.1439 and 0.1958, respectively while there is negative correlation with rhmin and rhmean with cases having equal to -0.1298 and -0.1424 consecutively.

Result of the Multiple Linear Regression displays in Fig. 4, shows that  $r^2$  is equal to 0.2304. This indicates that all eight of the independent variables explain 23.04% of the variability on the number of dengue cases. Moreover, it was shown that some independent variables had a p-value greater than 0.05 lowering their significance in predicting dengue incidence. From this model, only the maximum temperature, average relative humidity, and time period have p-values less than 0.05; thus, have significant impact on dengue cases.

Stepwise regression, as seen in Fig. 5, was conducted to eliminate non-significant variables to build a reliable model out of the remaining independent variables. The variables tmin, rhmax, tmean, and rhmin were removed from the model. These variables were omitted in building the Multiple Linear Regression since their respective p-values were higher than 0.05.

	cases	tmax	tmin	tmean	rr	rhmax	rhmin	rhmean	timepe-d
cases	1.0000								
tmax	-0.0094	1.0000							
tmin	0.1439	0.5291	1.0000						
tmean	0.1958	0.6314	0.6718	1.0000					
rr	0.0027	-0.0319	0.0953	-0.0085	1.0000				
rhmax	-0.0440	-0.3085	-0.0971	-0.2017	-0.0870	1.0000			
rhmin	-0.1298	-0.3926	-0.0311	-0.1694	-0.0233	0.3118	1.0000		
rhmean	-0.1424	-0.2965	-0.1539	-0.4117	0.4161	0.2314	0.2763	1.0000	
timeperiod	0.3318	0.2879	0.3452	0.3448	-0.0989	0.1017	-0.1179	0.1473	1.0000

Fig. 3. Correlation Matrix between the Dengue Incidence Rate and Weather Variables with Time Period of 0 to 120 Months from 2008 to 2017 in Iligan City.

Source	SS	df	MS	Number of obs	=	120
Model	74488.9524	8	9311.11905	F(8, 111)	=	4.15
Residual	248792.973	111	2241.37813	Prob > F	=	0.0002
Total	323281.925	119	2716.65483	R-squared	=	0.2304
				Adj R-squared	=	0.1749
				Root MSE	=	47.343

cases	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
tmax	-14.24543	4.797811	-2.97	0.004	-23.75261 -4.738248
tmin	6.578569	13.35265	0.49	0.623	-19.88059 33.03773
tmean	5.490475	8.760066	0.63	0.532	-11.86818 22.84913
rr	.0713063	.0525447	1.36	0.178	-.0328145 .1754272
rhmax	-1.274295	1.76028	-0.72	0.471	-4.762408 2.213818
rhmin	-1.144215	1.127534	-1.01	0.312	-3.3785 1.090069
rhmean	-4.759513	2.0437	-2.33	0.022	-8.809241 -.7097844
timeperiod	.6520949	.1554286	4.20	0.000	.3441028 .960087
_cons	748.8141	365.6827	2.05	0.043	24.1895 1473.439

Fig. 4. Performance Analysis of Multiple Linear.

p = 0.6232 >= 0.0500 removing tmin  
 p = 0.4801 >= 0.0500 removing rhmax  
 p = 0.3408 >= 0.0500 removing

Fig. 5. Variable Selection in Multiple Linear Regression using Stepwise Regression.

Fig. 6 shows the progress of the result of the Multiple Linear Regression model after eliminating insignificant independent variables.

OLS Regression Results						
Dep. Variable:	Dengue_cases	R-squared:	0.214			
Model:	OLS	Adj. R-squared:	0.186			
Method:	Least Squares	F-statistic:	7.811			
Date:	Fri, 10 May 2019	Prob (F-statistic):	1.32e-05			
Time:	10:35:09	Log-Likelihood:	-629.78			
No. Observations:	120	AIC:	1270.			
Df Residuals:	115	BIC:	1283.			
Df Model:	4					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	815.9833	211.635	3.856	0.000	396.775	1235.192
Time_period	0.7151	0.138	5.200	0.000	0.443	0.987
Temp_max	-9.8282	3.702	-2.655	0.009	-17.161	-2.496
Rel_Rain	0.1004	0.048	2.075	0.040	0.005	0.196
Avg_RH	-6.2039	1.707	-3.633	0.000	-9.586	-2.822
Omnibus:	63.418	Durbin-Watson:	0.670			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	232.563			
Skew:	1.924	Prob(JB):	3.16e-51			
Kurtosis:	8.631	Cond. No.	1.04e+04			

Fig. 6. Performance Analysis of Multiple Linear Regression after Removing Non-Significant Independent Variables.

Based on the result, time period, maximum temperature, relative rainfall, and average humidity, all indicates significant impact on predicting the number of dengue cases with p-value of less than 0.05. Comparing its result from the first model, the value of adjusted R-squared increases, adjusted R2 = 0.186 or 18.6% which conveyed that omitting non-significant independent variables could improve the accuracy result of the model.

Using the coefficient values of the identified significant factors would generate the predictive model in Eq. (4).

$$\text{cases} = 0.7151(\text{Time\_period}) - 9.8282(\text{Temp\_max}) + 0.1004(\text{Rel\_Rain}) - 6.2039(\text{Avg\_RH}) + 815.9833 \quad (4)$$

Fig. 7 shows the relation between the number of cases and time period for multiple linear regressions. Time period represents the total number of months involved in the study from January 2008 to December 2017. It is evident in the results that there are a number of gaps between the actual and predicted values.

### A. Poisson Regression

One of the major assumptions in Modeling Poisson regression model is the equality of the mean and variance. As shown on Table I, the data that was used for Poisson Regression model did not meet the assumption wherein the mean and variance should be equal. The result showed that the variable Time\_period had 60.5 mean and 1210 variance which is an overdispersion whereas Temp\_max has 32.09333 mean and 1.753401 variance which results to underdispersion.

### B. Random Forest

By performing the *Feature Importance* method in Python, the significant independent variables with the highest significance out of the eight are time period with a 0.48 and mean temperature with a value of 0.12. Results are shown in Fig. 8. These variables were used to train the random forest algorithm. Fig. 9 illustrates the adjusted R-Squared of using Random Forest having a value of 73.0%. This means that timeperiod and tmean in this model explain 73.0% of the variability of the dependent variable. It is evident in Fig. 10 that the actual and predicted values of random forest from the time period of January 2008 to December 2017 are closer from the actual number of dengue cases. The Mean Absolute Percentage Error (MAPE) of the model using Random Forest is 33.58%.

### C. Evaluation of the Models

Based on the summary of results of the three models shown in Table II, it indicates that Random Forest has a greater accuracy with 73.0% compared with Multiple Linear Regression with an adjusted R-Squared value of 18%. Further, Random Forest has lower MAPE score result with 33.58% while Multiple Linear Regression has 67.14% MAPE value. Comparing the results of the two models, Random Forest having two significant independent variables specifically the time period and the mean temperature clearly performs relatively better having higher accuracy score and lowest MAPE score than Multiple Linear Regression. This output explains the flexibility and the potential of Random Forest in

predicting the number of dengue incidence based on climatic factors.

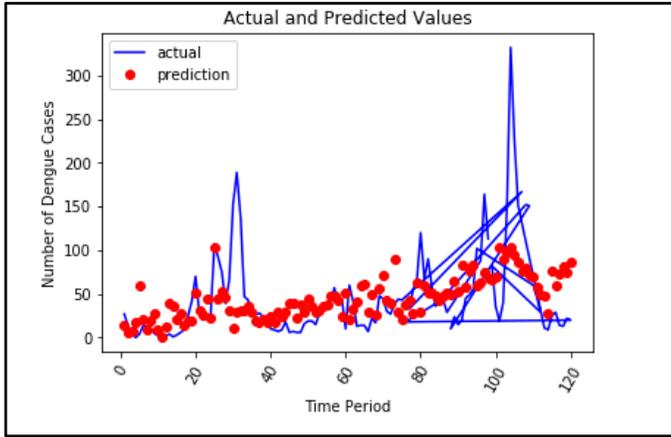


Fig. 7. The Actual Values and Predicted Number of Dengue Incidence from Time Period of January 2008 to December 2017 for Multiple Linear Regression.

TABLE. I. Descriptive Analysis of the Data for Poisson Regression

Variable	Mean	Variance
Time_period	60.5	1210
Temp_max	32.09333	1.753401
Temp_min	22.52833	.2198627
Avg_Temp	27.32917	.708806
Rel_Rain	159.6975	10250.92
RH_max	91.88333	7.566106
RH_min	73.00833	21.40329
Avg_RH	82.913	9.554999

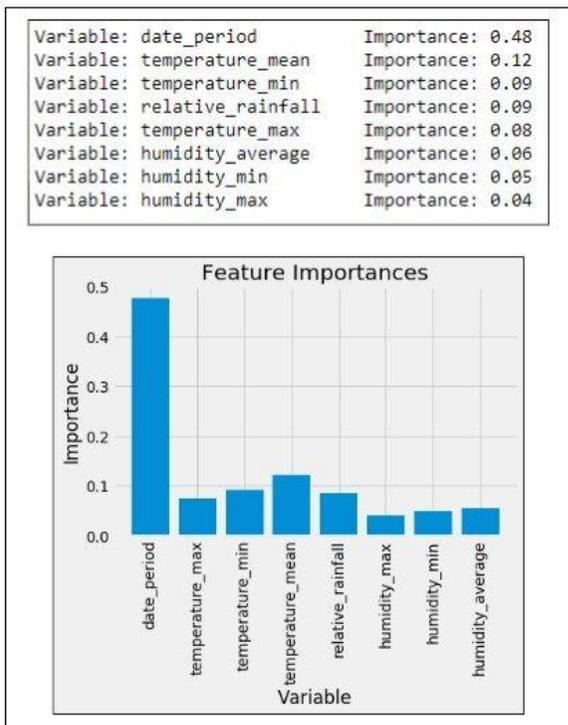


Fig. 8. Feature Importance Performed for Random Forest.

```
prediction = regressor.predict (X_test)
from sklearn.metric import r2_score
print('R-Squared score is: {0}%'.format(round(r2_score(y_test,prediction) * 100, 2)))

from sklearn import metrics
adj = 1 - float (len(y_test)-1)/(len(y_test) - len(coef)-1) * (1 - r2_score(y_test,
prediction))
print('Adjusted R-Squared score is: {0}%'.format(round (adj*100, 2)))

R-Squared score is: 74.54%
Adjusted R-Squared score is: 73.0%
```

Fig. 9. The Adjusted R-Squared using Random Forest.

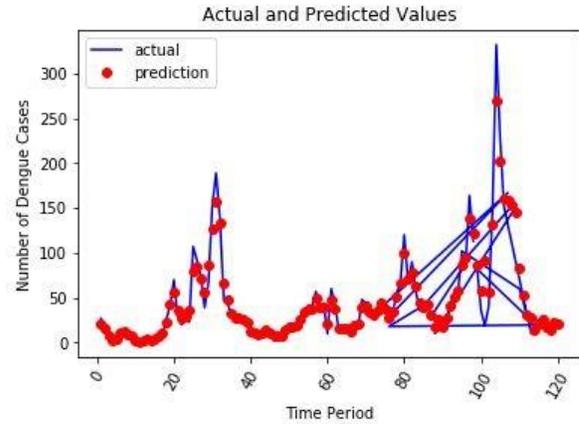


Fig. 10. The Actual Values and Predicted Number of Dengue Incidence of Random Forest from the Time Period of January 2008 to December 2017 using Random Forest.

TABLE. II. Comparison of the different Predictive Models

Predictive Models	Accuracy Result (Adjusted R-Squared)	Mean Absolute Percentage Error (MAPE)
Random Forest	73.0%	33.58%
Multiple Linear Regression	18%	67.14%

The proposed system generates prediction of dengue based on historical data and reported dengue cases in each barangay that can assist the society and local health offices in Iligan City. Fig. 11 is a screenshot of the system as output from the results of this study.

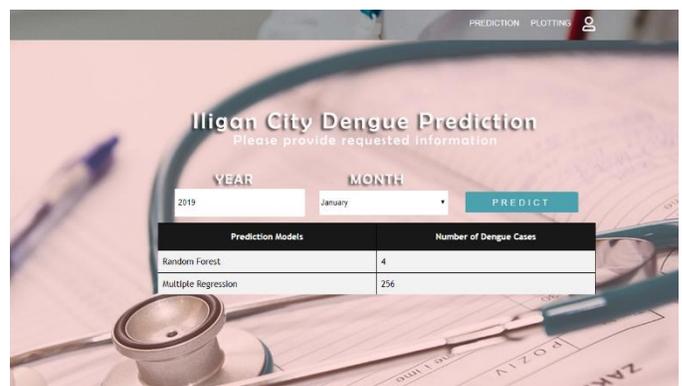


Fig. 11. A Screenshot of the Developed Dengue Prediction System.

## V. CONCLUSION

The correlation analysis between each climatic factor has different level of significance in relation with dengue cases. This means that some of the independent variables can significantly affect the transmission of dengue while some factors are not. Time period is one of the highly correlated variable and even with most of the climatic factors. This only shows that time period has valuable significance as it helps to determine dengue in a given period of time.

Predictive models, specifically Multiple Linear Regression and Random Forest, can potentially predict dengue cases in Iligan City based on the results compared to Poisson Regression which was excluded due to the violation of the assumptions of the model. In building the Multiple Linear Regression model, coefficient of the significant independent variables were considered and formulated as:

$$\text{dengue cases} = 0.1003793 (\text{rr}) - 9.828208 (\text{tmax}) - 6.203866 (\text{rhmean}) + 0.7150807 (\text{time period}) + 815.9834$$

On the other hand, the result of the evaluation with Random Forest performed better in terms of its goodness-of-fit and error difference between its actual and predicted values, having an accuracy percentage of 73% and 33.58% error result, than Multiple Linear Regression with only 18% accuracy percentage and error result of 67.14%.

Considering the results, Random Forest has the highest accuracy output and smallest error measure compared with multiple linear regressions. This indicate that this model has a positive impact in providing a prediction model for dengue cases in Iligan City. However, the accuracy percentage of the predictive model is lower than the expected percentage of the researchers. One factor is possibly due to the limited data of climatic factors used in the study. Considering other factors in dengue prediction may increase accuracy results of the predictive model.

## VI. RECOMMENDATION

It is recommended that additional possible independent variables alongside climatic data must be considered such as the topographic profile of the area, ecological, biological and sociological aspects. Gathering of climatic data is recommended if possible to be daily or weekly rather than monthly and must be specific to every barangay to further investigate the impact of climatic factors in predicting dengue cases specific for every barangay. Further, temperature, rainfall, and humidity alone can contribute in predicting number of dengue cases. But considering other climatic factors

including wind speed, precipitation, air pressure, and other weather data can provide better prediction result. Analysis and the use of other predictive models are highly recommended to test further the accuracy of forecasting dengue cases in Iligan City.

## ACKNOWLEDGMENT

The study would like to acknowledge the Center of Health in Iligan by providing the data concerned in the development of the project.

## REFERENCES

- [1] T. W. Chuang, K. C. Ng, T. Nguyen and L. Chaves, "Epidemiological characteristics and space-time analysis of the 2015 dengue outbreak in the metropolitan region of Tainan City," Taiwan. International journal of environmental research and public health, 2018.
- [2] A. De Vera, "Dengue Cases Up," Manila Bulletin, 24 October 2018. [Online]. Available: <https://news.mb.com.ph/2018/10/24/dengue-cases-up>. [Accessed 22 December 2018].
- [3] Philippines Department of Health Surveillance Reports, "Dengue Outbreak," WHO, Philippines, 2019.
- [4] V. G. Ramachandran and e. al, "Empirical model for estimating dengue incidence using temperature, rainfall, and relative humidity: a 19-year retrospective analysis in East Delhi," Epidemiology and health, vol. 38 e2016052, 2016.
- [5] University Corporation for Atmospheric Research, "Climate Change and Vector-Borne Disease," [Online]. Available: <https://scied.ucar.edu/longcontent/climate-change-and-vector-borne-disease>. [Accessed 20 February 2019].
- [6] S. Promprou, "Multiple Linear Regression Model to Predict Dengue Haemorrhagic Fever (DHF) Patients in Kreang Sub-District, Cha-Uat District, Nakhon Si Thammarat, Thailand," JOURNAL OF APPLIED SCIENCES RESEARCH, pp. 6193-6197, 2014.
- [7] J. Ong, X. Liu, J. Rajarethinam, S. Y. Kok, S. Liang, C. S. Tang and G. Yap, "Mapping dengue risk in Singapore using Random Forest," PLoS Neglected Tropical Diseases, vol. 12:e0006587, 2018.
- [8] T. K. D. G and a. N. V, "Impact of Climatic Fluctuation on Dengue Virus Etiology," Journal of Molecular and Genetic Medicine, vol. 12(1): 331, 2018.
- [9] C. Hettiarachige, S. von Cavallar, T. Lynar, R. I. Hickson and M. Gambhir, "Risk prediction system for dengue transmission based on high resolution weather data," Plos One, vol. 13(12)e0208203, no. e0208203, 2018.
- [10] J. Ong, X. Liu, J. Rajarethinam, S. Yheng Kok, S. Liang, C. S. Tang, A. Cook, L. C. Ng and G. Yap, "Mapping dengue risk in Singapore using," PLoS Negl Trop Dis 12(6):e0006587, 2018.
- [11] T. M. Carvajal, K. M. Viacrusis, L. F. Hernandez, H. T. Ho, D. M. Amalin and K. Watanabe, "Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in Metropolitan Manila, Philippines," BMC Infectious Disease, Vols. vol. 18, no. 1, 2018, 2018.
- [12] F. Moore, "Qualitative vs. Quantitative Design," DOI: 10.13140/RG.2.2.34861.49128, 2016.