Tropical Biomedicine 36(4): 822–832 (2019)

An ensemble model for forecasting infectious diseases in India

Shashvat, K.[1], Basu, R.[1*], Bhondekar, P.A.[2*] and Kaur, A.[2]
[1,2]Department of Computer Science and Engineering, National Institute of Technology, Delhi
[1*]Department of Electronics Communication and Engineering, National Institute of Technology, Delhi
[2*]Department of Agrocincs, Central Scientific Instruments Organization, Chandigarh
[*]Corresponding author e-mail: shashvat.sharma13@gmail.com
Received 27 January 2019; received in revised form 28 May 2019; accepted 5 June 2019

Abstract. Time series modelling and forecasting plays an important role in various domains. The objective of this paper is to construct a simple average ensemble method to forecast the number of cases for infectious diseases like dengue and typhoid and compare it by applying models for forecasting. In this paper we have also evaluated the correlation between the number of typhoid and dengue cases with the ecological variables. The monthly data of dengue and typhoid cases from 2014 to 2017 were taken from integrated diseases surveillance programme, Government of India. This data was analysed by three models namely support vector regression, neural network and linear regression. The proposed simple average ensemble model was constructed by ensemble of three applied regression models i.e. SVR, NN and LR. We combine the regression models based upon the error metrics such as Mean Square Error, Root Mean Square Error and Mean Absolute Error. It was found that proposed ensemble method performed better in terms of forecast measures. The finding demonstrates that the proposed model outperforms as compared to already available applied models on the basis of forecast accuracy.

INTRODUCTION

Time series modelling is a vibrant research area which has captivated attention of many researchers community over the last few decades. The main aim of the time series is to build up a model, which can be used to forecast the future values. There are numerous applications of time series modelling in the practical fields such as the economics, finance and medicine (Anish *et al.*, 2016). A time series is a sequential set of data points, mathematically defined as the set of vectors $y(t), t = 0,1,2 \ldots ..$ where 't' represents the time elapsed, the variable $y(t)$ is treated as the random variable (Bu *et al.*, 2018, Ahmed *et al.*, 2017). In this paper, we analyze two medical data sets by models like support vector regression, linear regression and neural network and developed a simple average ensemble model on dengue and typhoid data sets specifically for Indian population.

Dengue is a vector borne disease of great public health concern in India. The dengue virus is transmitted by mosquito 'Aedes'. Every year thousands of individuals are affected by this deadly disease and which contribute to the burden of health care (Wongkoon *et al.*, 2012). The outbreak of dengue is unremitting from 1950s to present; where the scenario is reported to be getting worse every year. Based upon the data of Integrated Diseases Surveillance Programme, Government of India, the numbers of cases in the year 2017 were 3462. Factors such as population, literacy rate and other environmental factors in the country (Lal *et al.*, 2012; Sitepu *et al.*, 2013) have major contribution in the increase.

Dengue is most outspread viral disease especially in developing countries like India, Pakistan and Bangladesh. Dengue has been endemic in the northern part of India (Xuan *et al.*, 2014; Zhang *et al.*, 2016) where total numbers of cases as well as deaths are at increasing node. In addition to the increased number of the cases and disease severity, there has been also a major shift in the geographical range of the disease. Previously dengue had been restricted to urban but now it's spread can be seen in rural areas as well (Xuan *et al.*, 2014; Zhang *et al.*, 2016; Ramachandran *et al.*, 2016; Mutheneni *et al.*, 2017). The colossal expansion of dengue can be contributed to factors such as unplanned urbanization, changes in environmental factors and increase in population.

Typhoid fever is a noteworthy cause of fatality across the world (Mermin *et al.*, 1999) and especially in developing countries. A current estimate found some 200,000 of consequential in deaths out of 22 million cases worldwide. More than 90% of reported cases are approximated to occur in Asia (Ram *et al.*, 2007). This fever spreads by a bacterium called Salmonella. Salmonella infection in humans can be categorized into two types. First, that grounds by low virulence serotypes of Salmonella enteric, which causes food poisoning. Second, that is caused by Salmonella enteric typhus (*S. typhi*), which is the root of typhoid, and a group of serovars, known as *S. Paratyphi* A, B, and C, which cause Paratyphoid (Kothari *et al.*, 2008). Humans are not only targets of the second type; but are also host and pool causing its spread. In the developing countries like India, Pakistan and Bangladesh (Sur *et al.*, 2007). It is the major cause of distress. In general, typhoid is prevalent in impoverished areas of the world. They all have the common characteristics relating challenges in ensuring safe drinking water supply, sanitations exists are inadequate (Luby *et al.*, 1998) and the pitable quality of life.

Risk factor such as contaminated food (Sur *et al.*, 2007; Luby *et al.*, 1998; Black *et al.*, 1985; Dewan *et al.*, 2013; Ram *et al.*, 2007; Sharma *et al.*, 2009; Shabana *et al.*, 2017) water (Ramachandran *et al.*, 2016; Kumar *et al.*, 2012) have been acknowledged as the foremost concerns for typhoid pervasiveness; other factors such as close contact with typhoid cases (Mermin *et al.*, 1999), lack of educational living near water bodies (Mermin *et al.*, 1999; Kothari *et al.*, 2008), flood (Shabana *et al.*, 2017), neglected personal hygiene, poor standard of living (Sur *et al.*, 2007), and roving to endemic areas (Kumar *et al.*, 2012). In this paper we have proposed a simple average ensemble model and we have constructed ensemble model by ensemble of three regression models such as Support Vector Regression, Linear Regression and Artificial Neural Network. The input variables for construction of an ensemble model is the past values i.e. number of typhoid and number of dengue cases. We collected the data from Integrated Diseases Surveillance Programme from year 2014-2017 collected monthly. We predicted the values from the past values of dengue incidence as well as typhoid incidence and values of ecological variables such as rainfall and humidity for the year 2014-2017.

MATERIAL AND METHODS

The authentic data sets of dengue and typhoid have been collected from integrated diseases surveillance programme, Chandigarh, Government of India. The data obtained was in monthly form (2014-2017). Table 1 shows the scenario of dengue cases in Chandigarh region from 2014-2017.

Table 1. Scenario of dengue cases (2014-2017)

Months	2014	2015	2016	2017
Jan	0	0	1	2
Feb	0	0	1	0
Mar	0	0	0	3
Apr	0	0	0	1
May	0	0	0	11
Jun	0	0	1	3
Jul	0	2	8	13
Aug	0	7	53	298
Sept	3	208	244	928
Oct	3	444	567	1224
Nov	5	368	168	936
Dec	2	16	121	43
Total	13	1045	1164	3462

Table 2. Scenario of Typhoid Cases (2014-2017)

Months	2014	2015	2016	2017
Jan	16	11	15	15
Feb	24	18	20	28
Mar	29	25	25	26
Apr	24	23	21	26
May	23	25	17	42
Jun	42	35	41	42
Jul	42	48	50	37
Aug	51	102	49	37
Sept	62	71	46	40
Oct	26	23	30	21
Nov	16	29	13	43
Dec	35	22	14	32
Total	390	432	341	389

Methodology and Algorithm Description

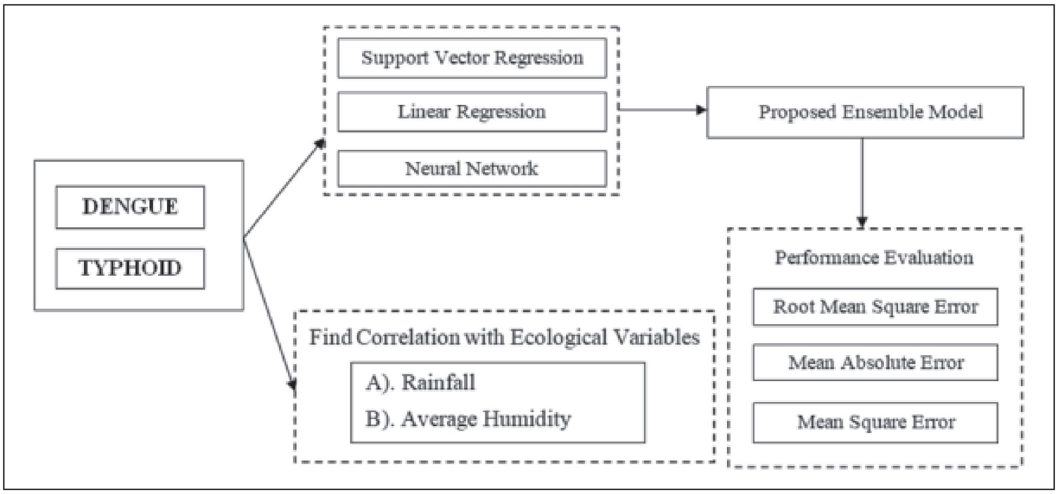

Figure 1. Flow diagram of Proposed Methodology.

Data Collection

The database for the analysis has been collected from integrated diseases surveillance programme (2014-2017) Govt of India.

Finding Correlations

The section 3.2 describes another objective to finding the correlations between the number of typhoid and dengue cases with the ecological variables such as rainfall and average relative humidity. In this type of research the ecological variables plays an important role for spread of the tropical diseases like dengue and typhoid. To accomplish this objective we have applied a statistical P value test to determine the correlation between the number of dengue and typhoid cases. After analysis by P value test we conclude that the rainfall is the prominent ecological variable for the occurrence of dengue incidence in northern India, Chandigarh. In the literature many authors have found the correlation between the infectious diseases and ecological variables by P value test (Potts *et al.*, 2010) (Phakhounthong *et al.*, 2018). The application of P value test concludes that ecological variables like rainfall had a positive correlation with the occurrence of number of cases for all the years 2014-2017. The average humidity also shows the positive correlation for the years 2013-2016 except for the year 2017. The statistical significance was considered $p<0.05$ we reject the null hypothesis. The correlation of dengue cases with the rainfall is significant.

Table 3. Correlation Values of Dengue with Ecological Variables

Year	p value (Humidity)	p value (Rainfall)	Coefficient of Correlation (Rainfall)	Coefficient of Correlation (Humidity)
2014	0.067	0.015	0.79	0.6
2015	0.076	0.015	0.68	0.53
2016	0.084	0.017	0.88	0.52
2017	0.086	0.017	0.88	0.52

Table 4. Correlation Values of Typhoid with Ecological Variables

Year	p value (Humidity)	p value (Rainfall)	Coefficient of Correlation (Rainfall)	Coefficient of Correlation (Humidity)
2014	0.17	0.79	0.79	0.42
2015	0.17	0.055	0.57	0.42
2016	0.23	0.046	0.75	0.38
2017	0.65	0.15	0.45	-0.14

Models applied

(a) Artificial Neural Network

The most widely ANNs used in forecasting problem are multi layer perceptrons (MLPs), which uses the single layer feed forward network (Adhikari *et al.*, 2012). The model is characterized by a network of three layers, viz hidden and output layer. The nodes in various layers are also known as processing elements.

The output of the model is computed using the following mathematical expression

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij}\, y_{t-i}\right) + \varepsilon_t, \forall_t \tag{1}$$

Here $y_{t-i}(i = 1,2 \ldots\ldots p)$ are the p inputs and y_t is the output. The integers p, q are the numbers of input and hidden nodes. $\alpha_j(j = 0,1,2, \ldots. q)$ and $\beta_{ij}(i = 0,1,2 \ldots. p); j = 0,1,2 \ldots. q)$ are the connection weights and ε_t is random shock; α_0 and β_{0j} are the bias terms. Usually the logistic sigmoid function $g(x) = \frac{1}{1+e^{-x}}$ is applied as the non linear activation function.

(b) Linear Regression

In the field of statistics, linear regression is the method for modelling the relationship between the scalar or dependent variables and one or more independent variables (Science *et al.*, 2003; Yang *et al.*, 2005; Barrow *et al.*, 2010; Shanthi *et al.*, 2012; Shaikh *et al.*, 1999). The case of the explanatory variables is called a simple linear regression. In the case of more than one explanatory variable the process is called the multiple linear regressions.

The mathematical formulation of the linear regression is as given below.

$$y = a + bx \tag{2}$$

where x *and* y are two variables of the regression line, b = Slope of the line and a = Y intercept of the line

(c) Support Vector Machines

Support vector machines are a supervised learning method which can be used for the classification as well as the regression problems (Adikari *et al.*, 2012; Shanthi *et al.*, 2012). The main idea of SVM when applied to binary classification problems is to find a canonical hyper plane which maximally separates the two given classes of training samples. In most of the applications the data points are not linearly separable so in such cases, a soft margin hyper plane classifier is constructed as follows

$$X \epsilon R^n : w^T x + b = 0, W\ \ ere\ w \epsilon R^n, b \epsilon R \tag{3}$$

$$\text{Minimize: } K(w, \epsilon) = \frac{1}{2} ||w||^2 + c\left(\sum_{i=1}^{N} \in_i\right) \tag{3.1}$$

$$\text{Subject to } y_i(w^T x_i + b) \geq 1 - \in_i\ \forall_i = 1,2 \ldots N\ \in_i \geq 0 \tag{3.2}$$

Support Vector Machine Regression

To use support vector machine regression, as in the classification problem the y_i is belonged to some class, for example, A1 and A2 in the case of regression and in support vector machine the $y_i \in$ here is the real number and other variables are same as for the classification problem (Yang *et al.*, 2005).

Regression methods are one of the powerful tools for the prediction or forecasting of a particular data. In this article we have constructed a simple average ensemble model by combining three popular regression models and predicted the number of typhoid and dengue cases for the northern region in India. In the previous studies most of the authors have applied the regression model particular to the infectious data sets and construct an ensemble model mostly with three prediction models (Yamana *et al.*, 2017; Buczak *et al.*, 2018).

1. Support Vector Regression
2. Linear Regression
3. Artificial Neural Network

We have also analyzed the data through the ARIMA model (Autoregressive Integrated Moving Average Model). ARIMA is also widely used prediction model for infectious diseases data set (Zhang *et al.*, 2016). We have constructed the simple average ensemble model based upon the performance of three regression models. The performance of the ARIMA model was not better in terms of the error metrics; we choose the best three regression models for the ensemble as the complexity issue is there. If we combine more number of models the proposed model will be complex.

Table 5. Performance by ARIMA model

Data Set	RMSE	MSE	MAE
Dengue	273.996	76902.502	170.405
Typhoid	16.796	297.761	18.417

Description of Proposed Ensemble Model

As mentioned before, there is a major challenge in the field of forecasting particularly in terms of forecasting of an incidence of a particular disease. There are numerous models available in the literature for the forecasting still there is an issue with the performance of an individual model. To overcome this limitation we propose an ensemble model. Here we describe the model as let $\widehat{X}^i$ be the set of independent variables where (i=1, 2, 3---n). The actual data set of a series is defined as $X = [x_1,x_2 - x_n]$ raise to power T being its forecast obtained from the i^t method.

Algorithm

Input: Set of Independent Variables $\widehat{X}^i$ (i=1, 2, 3---n).

Output: Prediction for the year 2018 $\acute{y}$.

1. Using the independent variables regress them to find the prediction using SVR, LR and ANN.
2. Regress values of all the three models and make an average of those such that $\acute{y} = (y_l + y_d + y_n)/3$
3. Compute the predicted values by weighted ensemble model.
4. Compare the individual forecast error with simple average ensemble model.

Generation of Regress Values

The Table 5 to 7 shows the regression values from all the models.

After applying all the three regressions models on the data and obtained the predicted value of the disease we average out these values as in Table 8.

Comparison of Forecast Accuracy

The Tables (10-13) shows the comparison of forecast accuracy in terms of root mean square error, mean scaled error and mean absolute error.

RESULTS

For the verification of our proposed ensemble method, two real time series data sets were used in this paper. These are number of dengue cases and number of typhoid cases in Chandigarh region all these data sets has been collected. The description of time series datasets is presented in Table 14.

Table 6. Regress Values (Support Vector Regression)

Months	Dengue	Typhoid	Months	Dengue	Typhoid
1	20.473836	25.495153	25	25.937462	22.193405
2	27.240693	25.608807	26	50.368814	21.558072
3	10.11226	30.683095	27	28.138764	29.499813
4	18.996226	24.199301	28	27.221034	24.481068
5	8.543517	22.838441	29	5.630744	22.838057
6	5.779967	31.487371	30	-10.2267	32.35578
7	27.221031	39.402606	31	23.7898	48.319137
8	27.11463	34.997756	32	16.82464	47.316709
9	31.982436	40.009651	33	18.044765	26.527713
10	56.021379	24.831812	34	70.88982	23.634982
11	92.600822	23.757521	35	93.139168	23.72701
12	29.150759	23.810865	36	37.210105	21.749079
13	20.67254	25.313676	37	18.533772	32.376546
14	42.410143	22.736777	38	51.493916	21.411574
15	-11.62781	34.321814	39	74.636737	26.705997
16	-3.771916	24.686221	40	23.165126	24.321493
17	16.458013	23.31925	41	15.034597	23.203163
18	-27.14255	33.319198	42	28.649569	40.31978
19	21.838203	47.799084	43	26.367936	39.034106
20	32.472615	43.347298	44	38.152853	41.973985
21	19.918867	32.232872	45	84.469388	41.679777
22	74.914162	23.336393	46	32.233349	22.676751
23	86.511498	24.113485	47	88.859957	23.973898
24	37.952166	21.596885	48	34.313901	22.402552

Table 7. Regress Values (Linear Regression)

Months	Dengue	Typhoid	Months	Dengue	Typhoid
1	183.38553	30.31281	25	197.58663	27.4885
2	161.22087	30.40169	26	178.91762	26.88218
3	107.28394	30.33061	27	113.8383	29.02709
4	49.65469	22.35546	28	57.1922	20.8564
5	22.57154	24.50233	29	15.58023	25.89275
6	90.69048	29.3115	30	83.58993	30.72365
7	114.34906	42.96291	31	021.49572	61.42949
8	176.6018	35.98115	32	71.40442	56.9027
9	127.6708	42.4731	33	185.89531	30.89346
10	146.70568	27.88946	34	152.6046	26.71628
11	149.33287	25.20736	35	149.54043	25.16608
12	178.25674	29.17321	36	187.76055	27.28311
13	184.04096	30.18245	37	160.22681	34.91859
14	172.90947	28.07707	38	179.79154	26.70837
15	57.90781	40.15049	39	128.58559	26.09416
16	19.6139	28.32994	40	53.47806	21.59507
17	35.57101	21.917	41	33.49546	22.32979
18	71.13666	33.20035	42	-12.32211	49.79854
19	50.77183	55.60709	43	117.62624	42.31115
20	117.39413	47.75631	44	134.65393	44.3237
21	166.99692	34.65196	45	111.50339	45.68845
22	154.24319	26.3904	46	158.28504	25.58656
23	147.03885	25.6636	47	147.91276	25.48979
24	188.63446	27.1093	48	184.37413	27.95659

Table 8. Regress Values (Neural Network)

Months	Dengue	Typhoid	Months	Dengue	Typhoid
1	118.479242	10.1961182	25	118.479242	32.3332204
2	118.479242	10.2051864	26	118.479242	32.3332686
3	118.479242	10.2381328	27	118.479242	32.3332721
4	118.479242	10.2388864	28	118.479242	32.3332721
5	118.479242	10.2864176	29	118.479242	32.3332722
6	118.479242	10.2876184	30	118.479242	32.3332722
7	118.479242	10.413522	31	118.479242	32.3332722
8	118.479242	10.4190062	32	118.479242	32.3332722
9	118.479242	10.4212564	33	118.479242	32.3332722
10	118.479242	10.5740867	34	118.479242	32.3332722
11	118.479242	10.5934228	35	118.479242	32.3332722
12	118.479242	11.2422239	36	118.479242	32.3332722
13	118.479242	11.7475205	37	118.479242	32.3332722
14	118.479242	12.4187948	38	118.479242	32.3332722
15	118.479242	12.5287582	39	118.479242	32.3332722
16	118.479242	22.851285	40	118.479242	32.3332722
17	118.479242	28.2147642	41	118.479242	32.3332722
18	118.479242	29.1769608	42	118.479242	32.3332722
19	118.479242	30.8762373	43	118.479242	32.3332722
20	118.479242	31.6611002	44	118.479242	32.3332722
21	118.479242	31.7512388	45	118.479242	32.3332722
22	118.479242	32.200315	46	118.479242	32.3332722
23	118.479242	32.2966802	47	118.479242	32.3332722
24	118.479242	32.3331801	48	118.479242	32.3332722

The complete analysis in this paper was performed in R studio (R version 3.5). Forecasting performances of all the models are then evaluated in the terms of three well known errors: Root Mean Square Error (RMSE), Mean Scaled Error (MSE) and Mean Absolute Error (MAE).

$$RMSE = \frac{1}{N}\sqrt{\sum_{n=1}^{N}(P_n - Z_n)^2} \qquad (4)$$

$$MAE = \frac{1}{N}\sum_{n=1}^{N}|P_n - Z_n| \qquad (4.1)$$

$$MAPE = \tfrac{1}{N}\textstyle\sum_{n=1}^{N}|\frac{P_n - Z_n}{P_n}| \times 100 \qquad (4.2)$$

Here P_n and Z_n are actual forecast observations respectively; N is the total sample size. The values of all these errors are desired to be as low as possible for better forecast accuracy.

DISCUSSION

The main aim of this research is to construct a simple average ensemble model. To construct this we first calculate the regress values from the entire individual regression model followed by their combination in an ensemble model. In the second phase of building the model our goal is to minimize the prediction error such as Root Mean Square Error, Mean Absolute Error and Mean Square Error. Whenever a particular model has a higher prediction error relative to other models its weight in the ensemble is decreased and vice versa. Previously many researchers have been applied the regression models especially on the infectious diseases data sets (Adikari *et al.*, 2012; Adhikari *et al.*, 2014). We have combined the three applied models and modelled the data as mentioned in section 3.4 in manuscript. In our case the construction

Table 9. Average Values for all the dengue and typhoid

Months	Dengue	Typhoid
1	107.4462	22.22335
2	102.3136	22.314201
3	78.625149	23.958693
4	62.37672	19.139052
5	49.864766	19.401064
6	71.649898	23.887328
7	86.683112	31.076435
8	107.39856	27.28182
9	92.710825	31.117453
10	107.06877	21.19745
11	120.13764	19.951525
12	108.62891	21.294325
13	107.73092	22.132314
14	111.26628	20.591989
15	54.919748	28.480782
16	44.773743	21.528138
17	56.836088	20.705605
18	54.157783	28.222267
19	63.696426	42.909479
20	89.448664	39.333899
21	101.79834	31.960066
22	115.87886	26.77758
23	117.3432	27.329432
24	115.02196	27.013195
25	114.00111	27.338437
26	115.92189	26.92455
27	86.818768	30.286767
28	67.630825	25.890291
29	46.563405	27.021403
30	63.947491	31.804277
31	54.574934	47.360676
32	68.902768	45.517603
33	107.4731	29.918193
34	113.99122	27.561556
35	120.38628	27.075498
36	114.4833	27.121862
37	99.079942	33.209512
38	116.58823	26.817783
39	107.23386	28.377719
40	65.040811	26.08332
41	55.669767	25.95545
42	44.935568	40.817224
43	87.49114	37.892885
44	97.095343	39.543695
45	87.405328	39.900544
46	120.41122	26.865572
47	118.41732	27.265697
48	112.38909	27.564182

Table 10. Support Vector Regression

Data Sets	RMSE	MSE	MAE
Dengue	270.5655119	73205.69623	121.7640889
Typhoid	13.24346	175.3893	8.037682

Table 11. Linear Regression

Data Sets	RMSE	MSE	MAE
Dengue	262.67693	68999.1698	166.901987
Typhoid	13.49271	182.0533	9.551957

Table 12. Neural Network

Data Sets	RMSE	MSE	MAE
Dengue	255.9554772	65513.20631	163.961315
Typhoid	16.62243961	276.3054985	12.52777983

Table 13. Proposed Ensemble Model

Data Sets	RMSE	MSE	MAE
Dengue	13.0229212	169.5964785	8.938
Typhoid	3.6421932	13.2655713	2.248

Table 14. Description of Data Sets

Time Series	Description
Dengue	Number of registered dengue cases from (2014-2017)
Typhoid	Number of registered typhoid cases from (2014-2017)

of ensemble models shows a significant improvement. The data plays an important role in the part for these type of analysis same as in our case.

The Figures (2-5) shows the result on dengue, typhoid a time series data sets. Our proposed ensemble method is compared with individual forecast of these models. The obtained forecast accuracy obtained for all the models are depicted in the Table (10-12). The presented results in Table 13 shows that our proposed ensemble method has provided lowest forecast errors among all the individual fitted models. It is quite evident significant improvement in forecasting accuracies can be achieved by employing the proposed ensemble method.

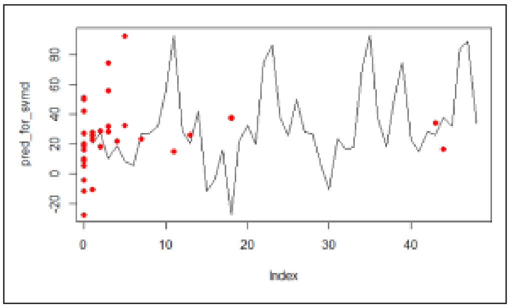

Figure 2. Support Vector Regression (Dengue).

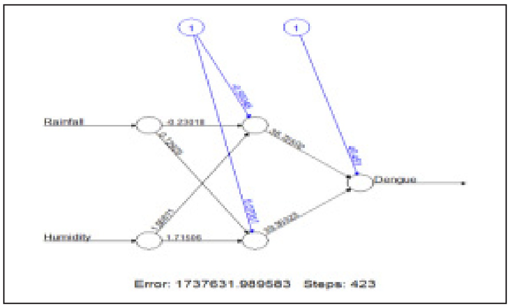

Figure 3. Neural Network (Dengue).

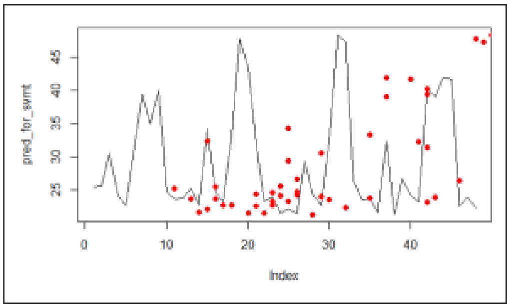

Figure 4. Support Vector Regression (Typhoid).

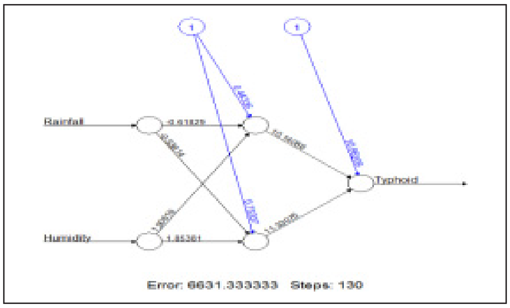

Figure 5. Neural Network (Typhoid).

Table 15. Forecast Table from Proposed Model (2018)

Months	Dengue	Typhoid
January-18	59.5	22.4
February-18	77.0	16
March-18	67.7	17.6
April-18	25.5	15.3
May-18	16.4	15.4
June-18	5.7	30.3
July-18	48.3	27.4
August-18	57.9	29
Setember-18	48.2	29
October-18	80.9	16
November-18	78.9	16.4
December-18	72.8	16.7

CONCLUSION

Combining the forecast from different models efficiently reduces the forecast errors and hence provides the considerably higher accuracy. Over the years, many authors have proposed various forecast models in the literature. In this paper we propose an ensemble model of simple average. The proposed ensemble method considers the individual forecast and the model is constructed with three very well known Neural Network, Support Vector Regression and Linear Regression models. The finding demonstrates that the proposed model outperforms as compared to already available applied models on the basis of forecast accuracy. In future the efficiency of suggested method can be further explored.

Acknowledgement. The author gratefully acknowledges the support from Integrated Diseases Surveillance Programme, Chandigarh for providing the database. We thank Director Health and Services, Chandigarh Dr. G Dewan for giving the approval for the collection of data of Chandigarh region.

Ethical Approval:
All applicable guideline were followed with reference number E-151913.

Competing interests:
None declared.

REFERENCES

Adhikari, R. & Agrawal, R.K. (2012). Forecasting strong seasonal time series with artificial neural networks. *Journal of Scientific and Industrial Research* **10**: 657-666.

Adhikari, R. & Agrawal, R.K. (2012). Combining multiple time series models through a robust weighted mechanism. *Proceedings of the 1st International Conference on Recent Advances in Information Technology* 455-460.

Adhikari, R. & Agrawal, R.K. (2014). A combination of artificial neural network and random walk models for financial time series forecasting. *Neural Computing and Applications* **24**: 1441-1449.

Ahmed, I., Raza, F.A., Iqbal, M. & Ashraf, M. (2017). Dengue virus serotypes and epidemiological features of dengue fever in faisalabad, Pakistan. *Tropical Biomedicine* **34**: 928-935.

Anish, C.M., Majhi, B. & Majhi, R. (2016). Development and evaluation of novel forecasting adaptive ensemble model. *The Journal of Finance and Data Science* **3**: 188-201.

Barrow, D.K., Crone, S.F. & Kourentzes, N. (2010). An evaluation of neural network ensembles and model selection for time series prediction. *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Black, R.E., Cisneros, L., Levine, M.M., Banfi, A., Lobos, H. & Rodriguez, H. (1985). Case-control study to identify risk factors for paediatric endemic typhoid fever in Santiago, Chile. *Bulletin of the World Health Organization* **63**: 899-904.

Bu, H., Hao, J., Guan, J. & Zhou, S. (2018). Predicting Enhancers from Multiple Cell Lines and Tissues across Different Developmental Stages Based On SVM Method. *Current Bioinformatics* **13**: 655-660.

Buczak, A.L., Baugher, B., Moniz, L.J., Bagley, T., Babin, S.M. & Guven, E. (2018). Ensemble method for dengue prediction. *PLoS ONE* **13**: e0189988.

Dewan, A.M., Corner, R., Hashizume, M. & Ongee, E.T. (2013). Typhoid Fever and Its Association with Environmental Factors in the Dhaka Metropolitan Area of Bangladesh: A Spatial and Time-Series Approach. *PLoS Neglected Tropical Diseases* **7**: 12-15.

Kothari, A., Pruthi, A. & Chugh, T.D. (2008). The burden of enteric fever. *Journal of Infection in Developing Countries* **2**: 253-259.

Kumar, C.K.K. & Jalota, G. (2012). A Note on Implementing Recurrence Quantification Analysis for Network Anomaly Detection. *Defence Science Journal* **62**: 112-116.

Lal, V., Gupta, S., Gupta, O. & Bhatnagar, S. (2012). Forecasting incidence of dengue in Rajasthan, using time series analyses. *Indian Journal of Public Health* **56**: 281.

Luby, S.P., Faizan, M.K., Fisher-Hoch, S.P., Syed, A., Mintz, E.D., Bhutta, Z.A. & Mccormick, J.B. (1998). Risk factors for typhoid fever in an endemic setting, Karachi, Pakistan. *Epidemiology and Infection* **120**: 129-138.

Mermin, J.H., Villar, R., Carpenter, J., Roberts, L., Samaridden, A., Gasanova, L. & Mintz, E.D. (1999). A Massive Epidemic of Multidrug Resistant Typhoid Fever in Tajikistan Associated with Consumption of Municipal Water. *The Journal of Infectious Diseases* **179**: 1416-1422.

Mutheneni, S.R., Morse, A.P., Caminade, C. & Upadhyayula, S.M. (2017). Dengue burden in India: Recent trends and importance of climatic parameters. *Emerging Microbes and Infections* **6**: 8.

Phakhounthong, K., Chaovalit, P. & Jittamala, P. (2018). Predicting the severity of dengue fever in children on admission based on clinical features and laboratory indicators: Application of classification tree analysis. *BMC Pediatrics* **18**: 1-9.

Potts, J.A., Thomas, S.J. & Srikiatkhachorn. (2010). Classification of dengue illness based on readily available laboratory data. *American Journal of Tropical Medicine and Hygiene* **83**: 781-788.

Ram, P.K., Naheed, A., Brooks, W.A., Hossain, M.A., Mintz, E.D., Breiman, R.F. & Luby, S.P. (2007). Risk factors for typhoid fever in a slum in Dhaka, Bangladesh. *Epidemiology and Infection* **135**: 458-465.

Ramachandran, V.G., Roy, P., Das, S., Mogha, N.S. & Bansal, A.K. (2016). Empirical model for calculating dengue incidence using temperature, rainfall and relative humidity: a 19-year retrospective analysis in East Delhi, India. *Epidemiology and Health* **e2016052**.

Science, C. (2003). Dimensionality Reduction for Indexing Time Series Based on the Minimum Distance *Work* **711**: 697-711.

Shabana, I.I., Bouqellah, N.A. & Zaraket, H. (2017). Investigation of viral and bacterial enteropathogens of diarrheic sheep and goats in Medina, Saudi Arabia. *Tropical Biomedicine* **34**: 944-955.

Shaikh, M.S. & Dote, Y. (1999). An Empirical Application of Linear Regression Method and Fir Network for Fault **12**: 57-63.

Shanthi, S. & Kumar, D. (2012). Prediction of blood glucose concentration ahead of time with feature based neural network. *Malaysian Journal of Computer Science* **25**: 136-148.

Sharma, P.K., Ramakrishnan, R., Hutin, Y., Manickam, P. & Gupte, M.D. (2009). Risk factors for typhoid in Darjeeling, West Bengal, India: Evidence for practical action. *Tropical Medicine and International Health* **14**: 696-702.

Sitepu, M.S., Kaewkungwal, J., Luplerdlop, N., Soonthornworasiri, N., Silawan, T., Poungsombat, S. & Lawpoolsri, S. (2013). Temporal patterns and a disease forecasting model of dengue hemorrhagic fever in Jakarta based on 10 years of surveillance data. *Southeast Asian Journal of Tropical Medicine and Public Health* **44**: 206.

Sur, D., Ali, M., Von Seidlein, L., Manna, B., Deen, J.L., Acosta, C.J. & Bhattacharya, S.K. (2007). Comparisons of predictors for typhoid and paratyphoid fever in Kolkata, India. *BMC Public Health* **7**: 1-10.

Wongkoon, S., Jaroensutasinee, M. & Jaroensutasinee, K. (2012). Assessing the temporal modelling for prediction of dengue infection in northern and northeastern, Thailand. *Tropical Biomedicine* **29**: 339-348.

Xuan, L.T.T., Van Hau, P., Thu, D.T. & Toan, D.T.T. (2014). Estimates of meteorological variability in association with dengue cases in a coastal city in northern Vietnam: An ecological study. *Global Health Action* **7**: 1-7.

Yamana, T.K., Kandula, S. & Shaman, J. (2017). Individual versus superensemble forecasts of seasonal influenza outbreaks in the United States. *PLoS Comput Biol* **13**: e1005801.

Yang, S.H., Liang, C.K. & Hsieh, C.Y. (2005). Watermarking MPEG-4 2D mesh animation with time-series analysis. *Journal of Information Science and Engineering* **21**: 341-359.

Zhang, Y., Wang, T., Liu, K., Xia, Y., Lu, Y., Jing, Q. & Lu, J. (2016). Developing a Time Series Predictive Model for Dengue in Zhongshan, China Based on Weather and Guangzhou Dengue Surveillance Data. *PLoS Neglected Tropical Diseases* **10**: 1-17.