ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Linear Ensembling of Time Series Ecological variables of COVID19

Shivaratri Narasimha Rao

Researcher Scholar, Department of Computer Science and Engineering

UCE(A),Osmania University, Hyderabad, Telangana, India.

Dr.S.China Ramu

Professor, Department of Computer Science and Engineering

CBIT(A), Hyderabad-75, Telangana,India.

Abstract

In the present era almost all the policy makers need assistance for performing decision making of ecological variables that tends to avoid spreading of COVID which can be get rid of and this paper will assist the process of forecasting which is the major disquiet of disease propagation. In this paper we present a model which comprises of a network to perform several predictions based on several factors using fuzzy logic to aggregate the responses for generating accurate predictions over multivariate time series data of India. Indeed the prediction errors provide us the decisive power to implement norms.

Keywords: Ensemble, Time series, COVID19, MAE, MSE, MAPE, RMSE

I. Introduction

In the very recent past we have identified that huge increase in infection of COVID19 Corona virus around the world and India lead to lockdown and curfews in most of the places in INDIA to break the chain of infection till 20 February 2022. This may be due to momentum of public during festival season as specially by February 5th 2021 3.95 Cr cases were registered and 5.2L people died due to COVD19 and 96.2% recovery rate was identified in India alone.

The Time series classification (TSC) is a kind of machine learning where the input vector features consists of real valued scenario that adds complexity layer problem with distinct characteristics of data over traditional algorithms [15].

The multi-variant time series classification comprises of par of observations denoted by $(x_1,\ldots,x_m)$ which is a discrete class of time series variables denoted by variable of class y with c possible values for mapping the space of possible inputs with d dimensions and m observations in a vector denoted by $X=<x_1,\ldots x_d>$ where $x_k=(x_{1,k},x_{2,k},\ldots,x_{m,k})$ denoted by jth observation with i dimension by using the scalar value of $x^{i,j,k}$.

Most of the ensemble predictions are performed for enhancing the predictions with exogenous variables to implement the distinct ensemble methods to perform selection in a strategy for selecting the appropriate model and for computing the weights by considering the meta models for integration of exogenous variables to implement distinct selection strategies for selecting the base models that pool by performing selection of best models based on the training error evaluation.

II. Literature Review

The study of propagation related to COVID-19 in India is applied using nonlinear autoregressive forecasting for the prevalence of virus as the authors modeled confirmed cases using the time series dataset to compare various regressive approaches like BATS [8] and TBATS [9] model and compared each of the features in the dataset and illustrated the results in the month of March by implementing the exponential smoothing using five forecasting rounds that cumulate the confirmed cases [1].

Another work proposed to implement the Holts Linear [2] model to forecast COVID19 cases in India using “Mean Absolute Percentage Error (MAPE)” [7] and the authors outperformed the model as the authors selected to perform the six prediction techniques for forecasting the cumulative cases as the proposed methodology is effective in predicting the sequential data [2].

An exponential growth is estimated using the ending points to verify the outbreak being applied as a variation to forecast the recovered and confirmed cases [3]. The recent advances in deep learning have revolutionized the healthcare industry as the various applications have been implemented and commercialized by incorporating the AI-driven healthcare providers who perform accurate diagnostics [4].

One of the researches propose a novel deep learning architecture which s trained over data obtained through 3D CT volume of lungs to split into various 2D patches and being fed as an encoding part for performing the feature extraction process in encoding the module into two sub-networks for segmentation joint classification and segmentation [5].

The classification part consists of feature embedding for the feature learning module to classify the patient severity determination [6]. The SEIR model to predict using fuzzy mathematical formulation for describing the disease transmission for every individual who passes through four states denoted by: susceptible (S), exposed (E), infectious (I), and recover (R) [10].

The proposed model denotes the peak was expected on July 24 2020 at worst and that the disease would disappear between September and November by implementing three phenomenological models: the generalized Exponential logistic growth model [11] the Exponential BATS model [12] and Exponential TBATS model [13] for performing the real-time forecast of the COVID19 increasing number of confirmed reported cases in Kerala state [14].

The findings are used to ascertain the Indian regions with potential vulnerability to weather-based spread by investigating the termination time of the outbreak in India using unique single peak SIR model to address the issue addresses using generalized logistic growth model to estimate the epidemic waves of the virus by traveling places are most often addressed [15].

It is widely known that lockdown measures leads to closure of educational institutes and workplaces [13] shut down of public transport and international travel [14] for avoiding of spread of virus or to contribute for suppressing COVID19 pandemic by implementing social distancing [12].

Furthermore most COVID forecast methods typically rely on limited features related to health sector indicators which are potential to exhibit similar pandemic patterns and additionally we train the proposed model by relaying on the multi variant time series data that comprises of a dataset consisting of cumulative daily number of cases to model the data over a time bound.

III. Proposed Methodology

3.1 Ensembles of multivariate classifiers

One of the most straightforward techniques is to implement the TSC algorithm over the multivariate data is by proposing the ensemble model over each of the dimension that exist independently as this approach is considered to be better for assessing the baseline and contrasting various MTSC classifiers to model the dimension over various dependencies. One of such accurate method is multivariate TSC is to perform “Hierarchical Vote Collective of Transformation-based Ensembles (HIVE-COTE)” that combines “Shapelet Transform Classifier (STC)” as the “Time Series Forest (TSF)” comprises of “Contractible Bag of Symbolic-Fourier Approximation Symbols (CBOSS)” along with the “Random Interval Spectral Ensemble (RISE)” using weighted probabilistic ensemble learning model. One of the simplest ways to construct a multivariate HIVE-COTE is by constructing each of the components independently over the ensemble.

Algorithm 1: Proposed Random Shapelet Forecast

Input: Training Set X_Train, n number of trees, len length of shapelet with lower, u upper bound and number of shapelets denoted by r

Output: ensemble generalized tree TREE=ST_1 to ST_n

Step 1: Start

Step 2: for I $\leftarrow$ 1 to n perform

Step 2.1: $I_i \leftarrow$ Sample (X_Train)

Step 2.2: $ST_i \leftarrow$ random_shapletTree(X_Train, I_i, u, r)

Step 2.3: TREE $\leftarrow$ TREE U ST_i

Step 3: return TREE

Step 4: Stop

Algorithm 2: random_shapletTree

Input: Training Set X_Train, Testing Set X_Test, n number of trees, d number of dimensions, l lower shapelet, u upper shapelet and number of shapelets denoted by r

Output: ST a random Shapelet

Step 1: Start

Step 2: k $\in$ [1, … , d]

Step 3: for I $\leftarrow$ 1 to r perform

Step 3.1: ran=rand(l,u)

Step 3.2: S ← S U sampleShapelet(X_Train, d, l, u, ran)

Step 3.3: t, S, k ← perform_bsetSplit (X_Train, X_Test, S)

Step 3.4: X_Train_L, X_Train_R ← distribute(X_Train, S, t, k)

Step 4: return [X_Train_L, X_Train_R]

Step 5: Stop

Algorithm 1 and 2 will generate the weak ensemble learners with generalized trees that grow at a rapid rate by introducing the variability among various constituent classifiers are implemented and. Algorithm 2 denotes that each fo the node is further generalized tree is generated with k dimensions along with X_Train represents training sent and X_Test represents the testing set and in step 3 splitting is performed which is returned to algorithm 1.

IV. Experimental Results

This paper s implemented using the covid19 dataset downloaded and the results generated using ecological variables are:

Table 1: 5 sample tuples of the dataset

Date	Cured	Deaths	Confirmed
2020-03-29	1	1	66
2020-03-30	1	1	71
2020-03-31	1	1	79
2020-04-01	1	3	96
2020-04-02	1	3	107

Table 2: Description of Dataset

	Cured	Deaths	Confirmed
Count	501.000000	501.000000	501.000000
Mean	243821.353293	1497.155689	260603.401198
Std	199246.672965	1151.878184	205507.293162
Min	1.000000	1.000000	66.000000
25%	45388.000000	519.000000	62703.000000
50%	261830.000000	1467.000000	272123.000000
75%	303601.000000	1741.000000	319054.000000
Max	638410.000000	3831.000000	650353.000000

Table 2 represents the statistical data related to the dataset comprising of attributes such as Cured, Deaths and Cured attributes.

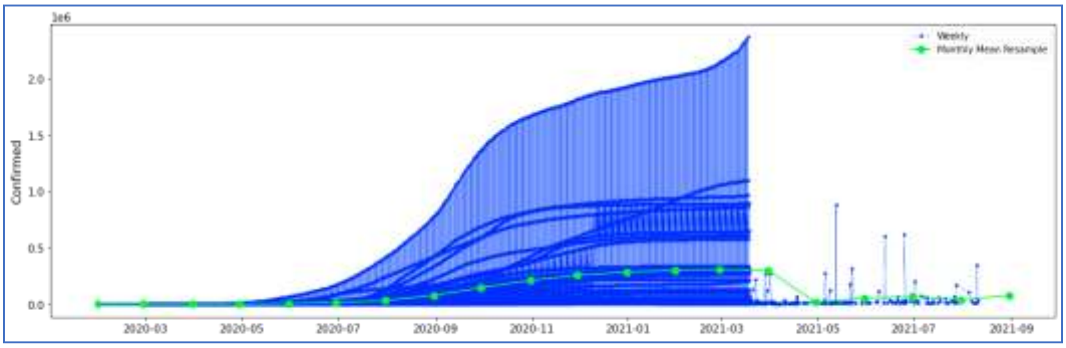

Figure1: Prediction of Confirmed cases in India

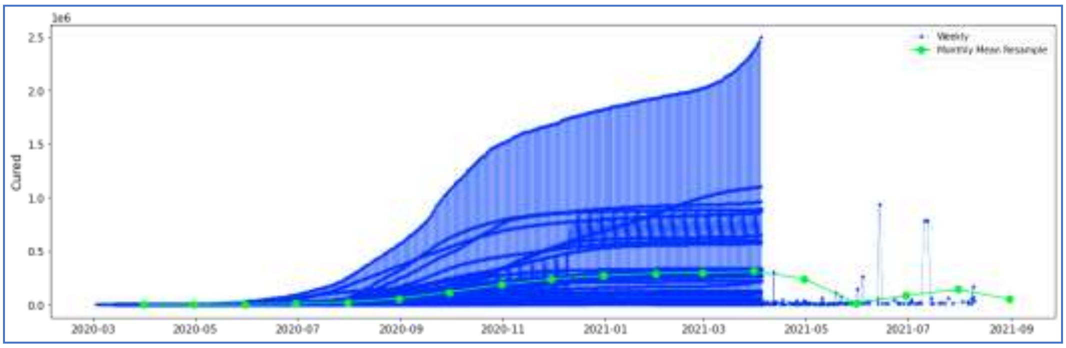

Figure2: Prediction of Cured cases in India

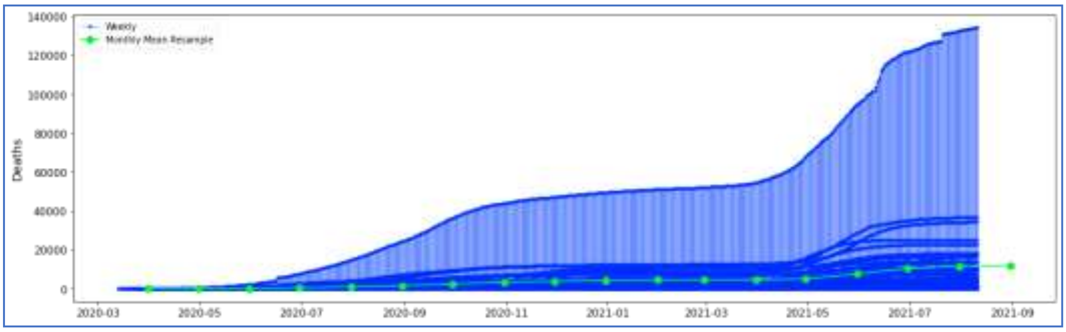

Figure3: Prediction of Deaths cases in India

Figures 1 to Figure 3 represents the prediction values generated based on the proposed algorithm

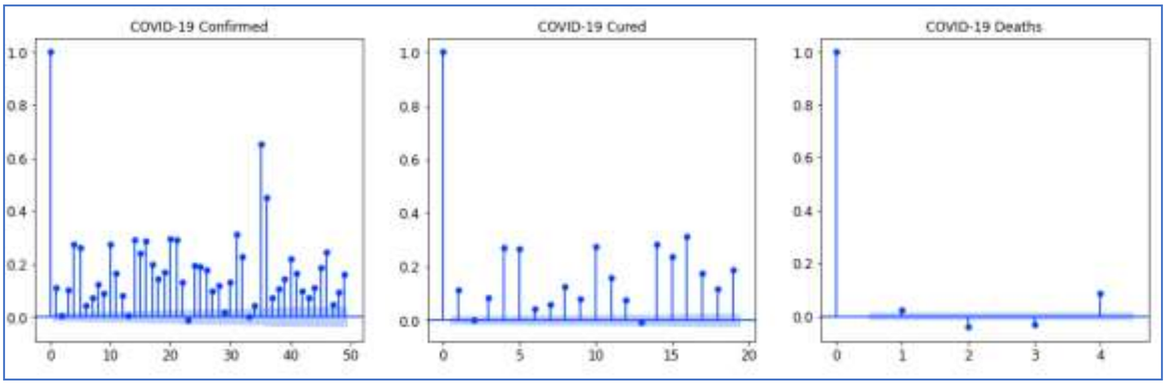

Figure 4. Correlations generated for the data

And Figure 4 represents the correlation visualized

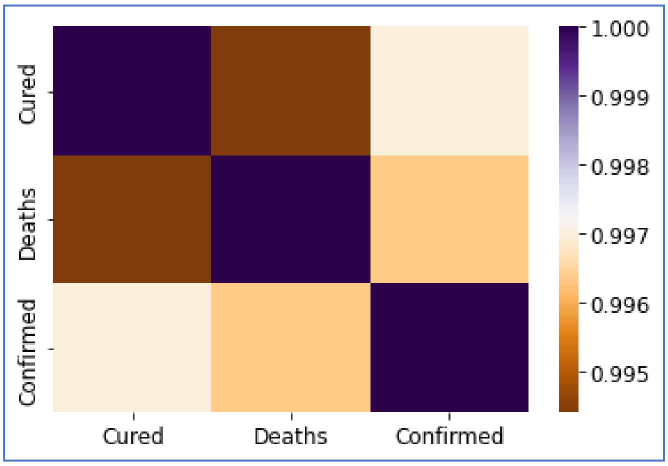

Figure 5. Heat map generated

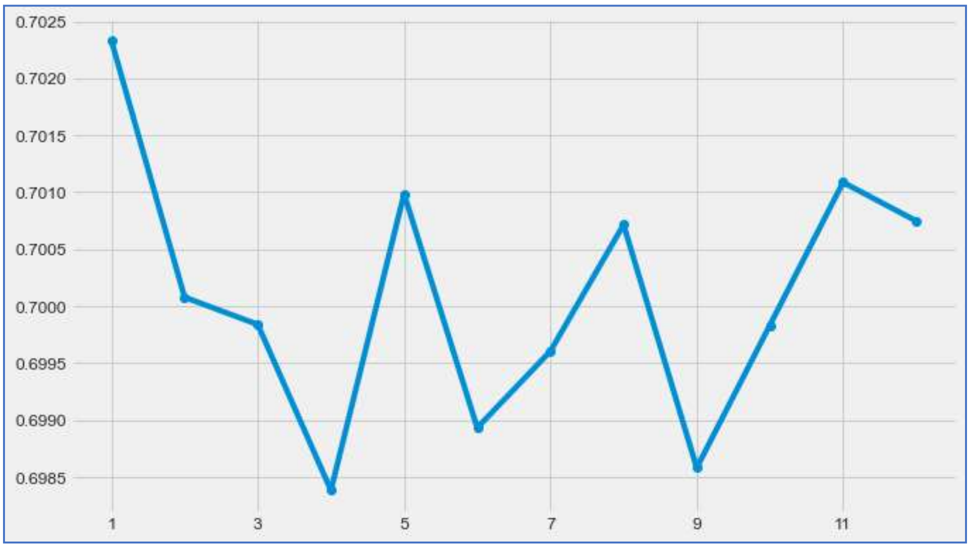

Figure 6. Implementation results of the proposed algorithm

The above figure that is figure 6 represents the result obtained by implementing the proposed algorithm for predicting the model.

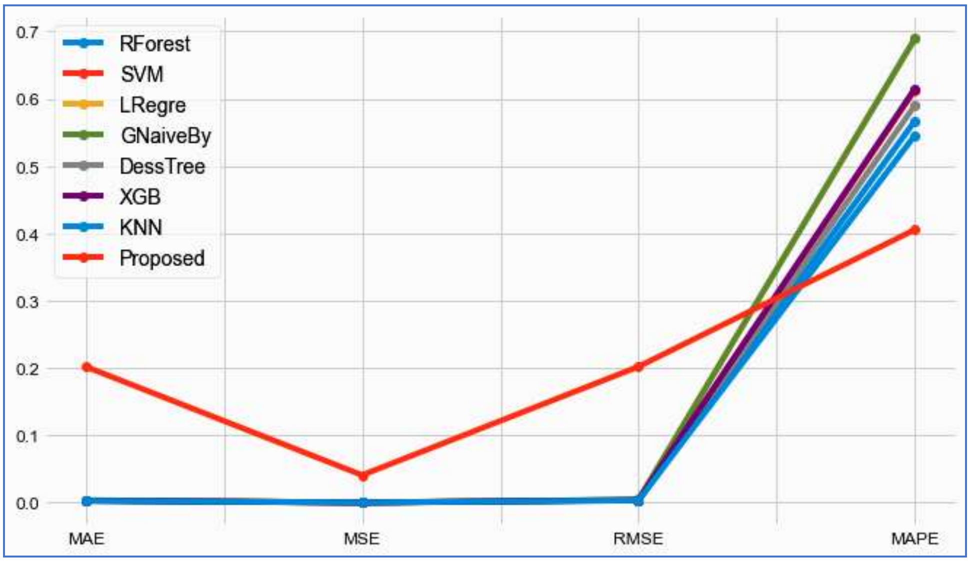

Figure 7. Evaluation of proposed model using the rest of ensemble models

The above figure that is Figure 7 represents the MAE error rate, MSE and RMSE error rates are better than that of the rest of the algorithms that are compared with the proposed algorithm. Hence based on figure 7 we can say that the proposed methodology is better than that of the rest of the ensemble methods.

V. Conclusion and Feature work

In this paper we have presented a linear time series algorithm over ecological variables for performing ensembling by comparing with most of the present day renounced algorithms and the results represent that the proposed algorithm Random Shapelet results are a lot better than the rest of the algorithms. As a part of feature work we want to implement the proposed algorithm on image data also.

References:

[1]. 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.

[2]. 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.

[3]. 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.

[4]. Mohammed Ali Shaik and Dhanraj Verma, (2020), Deep learning time series to

forecast COVID-19 active cases in INDIA: A comparative study, 2020 IOP Conf. Ser.:Mater.Sci.Eng. 981 022041, doi.org/10.1088/1757-899X/981/2/022041
[5]. Kothari, A., Pruthi, A., & Chugh, T. D. (2008). The burden of enteric fever. Journal of Infection in Developing Countries 2: 253–259.
[6]. 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.
[7]. 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.
[8]. Imran Qureshi, Burhanuddin Mohammad, Mohammed Abdul Habeeb and Mohammed Ali Shaik, (2020), Mathematical model for implementing Non Linear Time Series Forecasting using Deep Learning, 2020 IOP Conf. Ser.: Mater. Sci. Eng. 981 022021, doi.org/10.1088/1757-899X/981/2/022021
[9]. 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.
[10]. 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.
[11]. Mohammed Ali Shaik, "Time Series Forecasting using Vector quantization", International Journal of Advanced Science and Technology (IJAST), ISSN: 2005-4238, Volume-29, Issue-4 (2020), Pp.169-175.
[12]. 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.
[13]. 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.
[14]. Mohammed Ali Shaik, "A Survey on Text Classification methods through Machine Learning Methods", International Journal of Control and Automation (IJCA), ISSN: 2005-4297, Volume-12, Issue-6 (2019), Pp.390-396.
[15]. 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.
[16]. Kumar, C., K., K., & Jalota, G. (2012). A Note on Implementing Recurrence Quantification Analysis for Network Anomaly Detection. Defence Science Journal 62: 112–116.