



A novel cellular automata classifier for COVID-19 trend prediction

Pokkuluri Kiran Sree^{1*}, S. S. S. N. Usha Devi Nedunuri²

¹Department of Computer Science and Engineering, Shri Vishnu Engineering College for Women, West Godavari District, Andhra Pradesh, India, ²Department of Computer Science and Engineering, University College of Engineering, JNTU Kakinada, Andhra Pradesh, India

ABSTRACT

Introduction: China has witnessed a new virus Corona, which is named COVID-19. It has become the world's most concern as this virus has spread over the world at a higher speed; the world has witnessed more than one lakh cases and one thousand deaths in a span of few days.

Methods: We have developed a preliminary classifier with non-linear hybrid cellular automata, which is trained and tested to predict the effect of COVID-19 in terms of deaths, the number of people affected, the number of people being could be recovered, etc. This indirectly predicts the trend of this epidemic in India. We have collected the datasets from Kaggle and other standard websites.

Results: The proposed classifier, hybrid non-linear cellular automata (HNLCA), was trained with 23,078 datasets and tested with 6785 datasets. HNLCA is compared with conventional methods of long short-term memory, AdaBoost, support vector machine, regression, and SVR and has reported an accuracy of 78.8%, which is better compared with the cited literature. This classifier can also predict the rate at which this virus spreads, transmission within the boundary, and of the boundary, etc.

Keywords: COVID-19 prediction; cellular automata; non-linear CA

INTRODUCTION

The coronavirus sickness 2019 (COVID-19) flare-up starting in Wuhan, Hubei area, and China, agreed with Chunyun, the time of mass movement for the yearly spring festival. To contain its spread, China embraced uncommon across the country intercessions on January 23, 2020. These strategies included huge scope isolate, severe controls on movement, and broad observation of suspected

cases. In any case, it is obscure whether these strategies have affected the pestilence. We looked to show how these control measures affected the regulation of the pandemic. Cellular automata area technique that operates basically with fuzzy rules. We have done a feasibility study of what type of classifier is used to address this problem as the input is very dynamic, and the prediction rate is also low. We have applied non-linear cellular automata with hybrid rules to classify the data and predict the people infected probable deaths and probable recoveries, etc. Cellular automata are natural classifier that can classify dynamic data. Initially, we thought of applying a deep learning model to predict the effect of COVID-19, later we understood applying CA

*Corresponding author: Pokkuluri Kiran Sree, Department of CSE, Shri Vishnu Engineering College for Women, West Godavari District, Andhra Pradesh, India. E-mail: drkiranree@gmail.com

Submitted: 26 March 2020/Accepted: 29 March 2020

DOI: <https://doi.org/10.17532/jhsci.2020.907>



itself will give a better classifier when the dynamic rules are used.

Yang et al. (1) have studied the impact of COVID on China based on many statistical parameters. They have used a susceptible-exposed-infectious-removed model to predict the curve of propagation, which is very preliminary work. This work is not adoptable and about a limited area. Huet al.(2) have proposed an AI technique to forecast the moments of COVID-19. The authors have utilized the idle factors in the autoencoder and bunching calculations to gather the areas/urban communities for exploring the transmission structure. Yanet al. (3) have done basic research on the moments of the COVID virus. Zheet al.(4) have studied the primary reasons for this epidemic. We have done an extensive literature survey(5–8), on what type of classifier can be applied to this problem. After this study, we understood that cellular automata could be applied to this problem. The rules selected are non-linear and hybrid ones.

METHODS

As we know, cellular automata are a versatile classifier that can process dynamic data. The general rules of CA can be represented like equation 1.

$$T(n+1)=T(n-1)+T(n)+T(n+1) \quad (1)$$

The transition from one state to another state is defined by the rules, as described in the equation.

Wolfram (8) has defined, the rule number can be the decimal number equivalent of the next state as shown in Tables 1 and 2. In this context, when we consider three neighborhoods and two states for cellular automata, we can give 256 different next states.

$$\text{Complemented Rule 51: } q_i(t+1) = q_i(t) \quad (2)$$

$$\text{Non complemented 238: } q_i(t+1) = q_i(t)+q_{i+1}(t) \quad (2)$$

3.1 Example: The classifier working will be explained with the sample input(9).

Input: AAAGGGGACGTTTA

Embedding layer: 0.6,0.2,0.1,0.7, 0.8.....

Transitions: Processing by CA(Rule 204, 170, 250)

State1: 0.6 0.2 0.1

State 2: 0.6 0.2 0.1

State 3: 0.6 0.1 0.1

State 4:0.6 0.1 0.1

TABLE 1. Rules with neighborhood

Neighbors	111	110	101	100	011	010	001	000	Rule
R-51	0	0	1	1	0	0	1	1	51
R-238	1	1	1	0	0	1	1	1	238

TABLE 2. Rules without complementation

1.	250	q_{i-1}, q_i, q_{i+1}
2.	204	q_i
3.	240	q_{i-1}
4.	170	q_{i+1}
5.	0	0

Output: Basin with characteristics (the basin in the example is State:4).

We have considered various parameters for predicting the spread of the virus listed below.

1. Vu: Vulnerable people in the region number
2. InOu: Inward and outward movement of the people
3. Ti: Transmission rate of possible infections
4. Te: Transmission rate of possible exposure
5. Vo: Volatility of transfer per person wrt today
6. P: Total population in the regions
7. D: The probability of death
8. R: The probability to recover
9. E: The probability to get effected

Design of hybrid non-linear cellular automata (HNLCA)

Algorithm (HNLCA Algorithm)

Input: Nine parameters collected and DNA sequence

Output: Prediction(rate of death, infected, and curable)

Step 1: Embedding the sequence and preprocess the parameters.

Step 2: Use the rules as Stated 1,2, and 3.

Step 3: Feature identification.

Step 4: Construct the HNLCA classifier as represented in section 3.3.

Real-time processing of HNLCA

The input from (9) was processed that looks like the structure. The output of the classifier is 0.4, 0.4, and 0.8, which shows that the DNA sequence processed

has a chance of 0.4 infected, 0.4 death, and 0.8 chance of recovery in Figure 1.

RESULTS

The datasets are collected from Kaggle, which consists of state, region, latitude, longitude, date, confirmed, deaths, and recovered, as represented in (10-13). Figure 2 shows the comparison of the proposed classifier with the standard classifiers. We have

observed that the error rate is very less in HNLCA, which is 11.8% better compared to SVR. HNLCA reports an accuracy of 78.8, which is better than long short-term memory.

Figure 3 represents how HNLCA processed has different parameters and can predict the rate of changes in death, infected, and recovery rate. HNLCA was applied to a small village Chodavaram

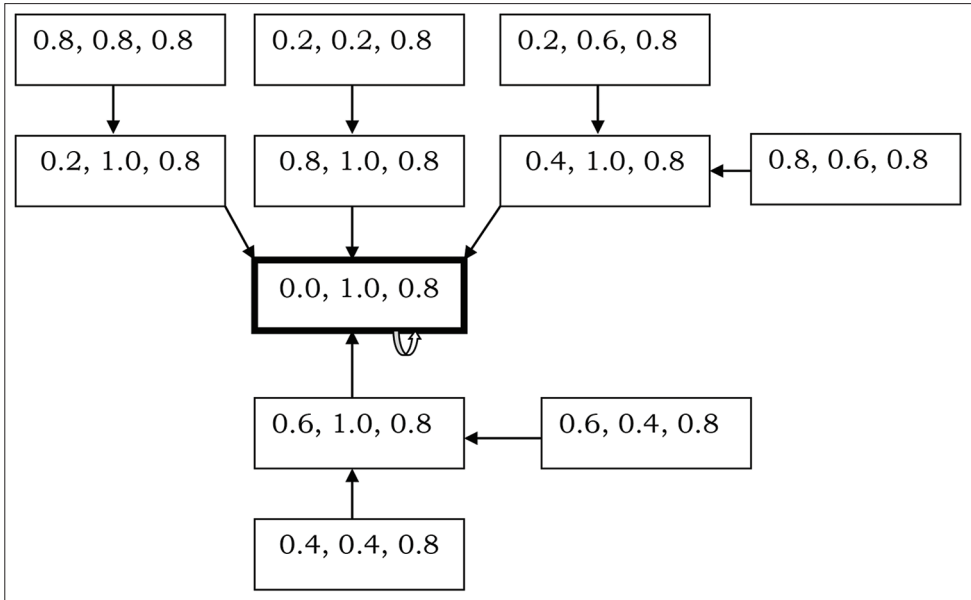


FIGURE 1. Real-time processing of hybrid non-linear cellular automata.

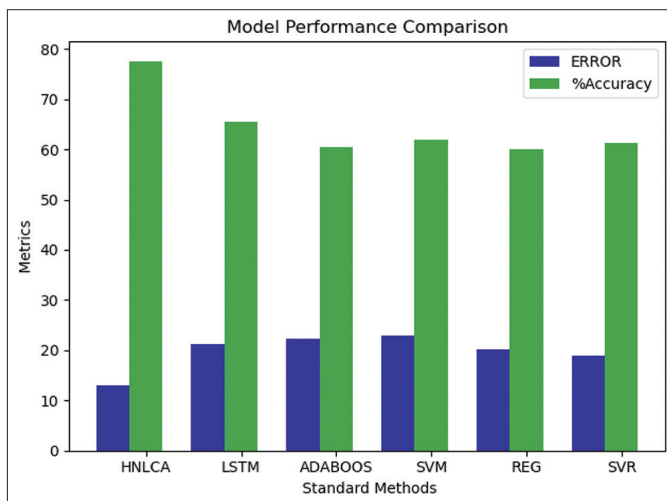


FIGURE 2. Performance of HNLCA and comparisons (HNLCA: Hybrid non-linear cellular automata, LSTM: Long short-term memory, AdaBoost: Adaptable boosting, SVM: Support vector machine, REG: Regression, SVR: Support vector regression).

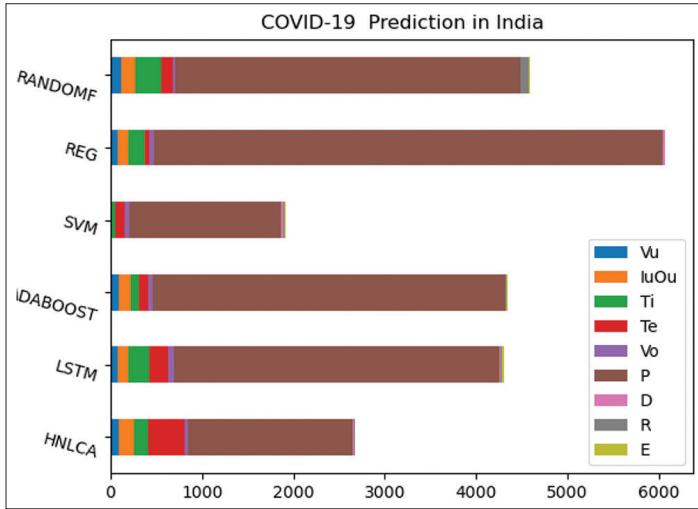


FIGURE 3. COVID-19 predictions with different parameters. (Vu: Vulnerable people in the region number, luOu: Inward and outward movement of the people, Ti: Transmission rate of possible infections, Te: Transmission rate of possible exposure, Vo: Volatility of transfer per person wrt today, P: Total population in the regions, D: The probability to death, R: The probability to recover, E: The probability to get affected).

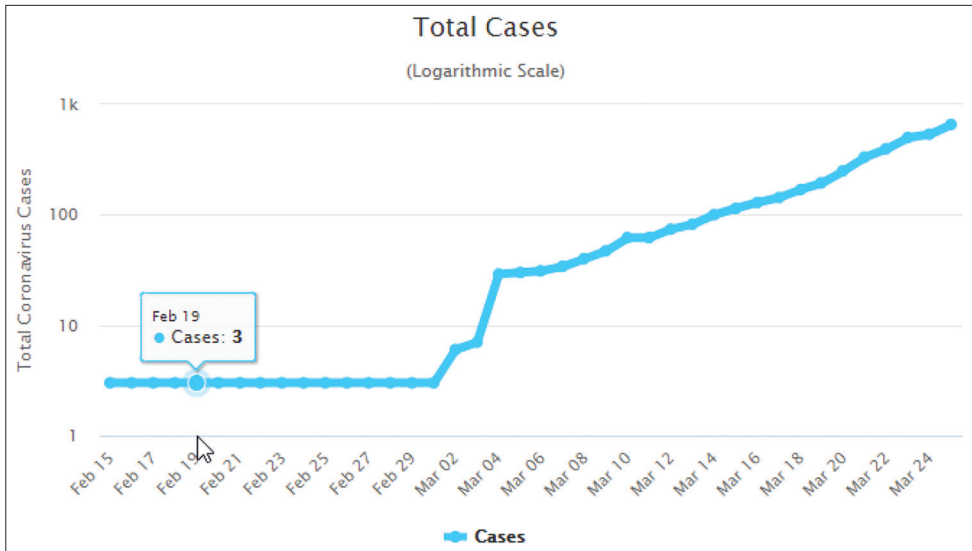


FIGURE 4. COVID-19 cases prediction in India.

in Andhra Pradesh, where the number of people residing is 1799, vulnerable people in the region number are 911, inward and outward movement of the people are 162, the transmission rate of possible infections 154, and the transmission rate of possible exposure 406. The classifier predicts that the volatility of transfer per person today probability is 0.022, the probability to death 0.0055, The probability to

recover 0.0072, and The probability to get effected is very less. Figure 4 predicts the total number of cases in India every day. This graph shows that the number of people affected is gradually increasing every day, and on March 24 represents the maximum number of cases reported. Figure 5 shows the number of deaths predicted for the past 15 days. The number of deaths from February 15 to March 10

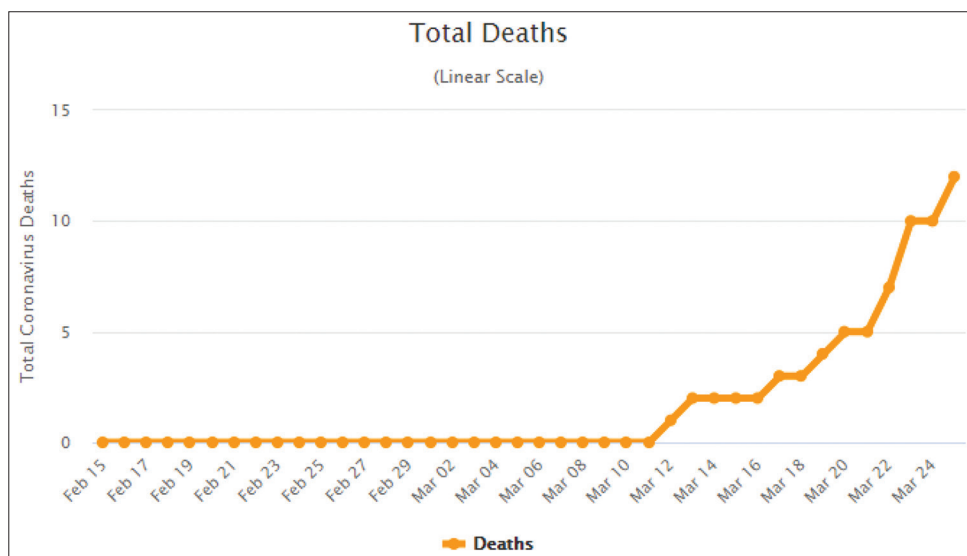


FIGURE 5. COVID-19 death prediction in India.

was zero, and after that, there is consistent growth in the number of deaths.

CONCLUSION

We have successfully developed as a novel classifier HNLCA, which can process real-time datasets to predict the number of deaths and cases in the near future. This can also predict the number of people that can recover, the rate at which this virus spreads, transmission within the boundary and of the boundary, etc. The classifier is thoroughly trained and tested. The developed classifier is adaptable to various dynamics inputs and can process data of different lengths. Even though this is a preliminary work to predict trends of COVID, we have developed the first and novel classifiers with dynamic rules with cellular automata. This work can be future extended with incorporating fuzzy sets augmented with CNN, which will give the best performance. The average accuracy of 78.8% is reported, which is considerable at this moment.

REFERENCES

1. Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020;12(2):165-74.
2. Hu Z, Ge Q, Li S, Jin L, Xiong M. Artificial Intelligence Forecasting of COVID-19 in China. 2020. Available from: <http://www.arXiv.Org>q-bio>arXiv:2002.07112>.
3. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA* 2020; ahead of print, <https://doi.org/10.1001/jama.2020.2565>
4. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med* 2020; online first, [https://doi.org/10.1016/S2213-2600\(20\)30076-X](https://doi.org/10.1016/S2213-2600(20)30076-X)
5. Kiran SP, Babu IR. Identification of protein coding regions in genomic DNA using unsupervised FMACA based pattern classifier. *IJCSNS* 2008;8(1):305.
6. Kiran SP, Babu IR, Devi U. Investigating an artificial immune system to strengthen protein structure prediction and protein coding region identification using the cellular automata classifier. *Int J Bioinform Res Appl* 2009;5(6):647-62. <https://doi.org/10.1504/ijbra.2009.029044>.
7. Pokkuluri KS, Babu IR. AIX-MACA-Y multiple attractor cellular automata based clonal classifier for promoter and protein coding region prediction. *J BioinformIntell Control* 2014;3(1):23-30. <https://doi.org/10.1166/jbic.2014.1071>.
8. Wolfram S. Universality and complexity in cellular automata. *Physica D* 1984;10(1-2):1-35. [https://doi.org/10.1016/0167-2789\(84\)90245-8](https://doi.org/10.1016/0167-2789(84)90245-8).
9. Hunter C, Wei X. Coronavirus sequence assembly isolated from 2019/2020 Wuhan Outbreak Patient. Available from: https://www.ncbi.nlm.nih.gov/nuccore/LR757996?fbclid=IwAR1LXMPn5NSi6iNpz72cQYA2JwsK4gEvlMLp-yk_9KRTBG9sbVoezxwA0. (last access: 24.March 2020)
10. Available from: https://www.laconicml.com/predict-infected-people-coronavirus-python/?fbclid=IwAR2mbc_0jRfJRsuotKzfk_T4WSFYZl7mR5wG34ufeliGI-OrZHonkJ65s1k. (last access: 24. March 2020)
11. Available from: <https://www.github.com/datasets/covid-19/commit/3e531107d-ed2feed441d39804b2e82cd14d88e7>. (last access: 24 March 2020)
12. Available from: <https://www.kaggle.com/imdevskp/corona-virus-report>. (last access 24 March 2020)
13. Available from: <https://www.worldometers.info/coronavirus/coronavirus-cases>. (last access 24 March 2020)