

EPIDEMIC PROGNOSIS: COMPARATIVE PERFORMANCE OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR PREDICTING VIRUS TRANSMISSION DYNAMICS

Faulinda Ely Nastiti*^{1,3}, Shahrulniza Musa¹, and Eiad Yafi², Marta Ardiyanto³

Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Malaysia¹

School of Computer Science University of Technology Sydney Sydney, Australia²

Information System Department, Universitas Duta Bangsa Surakarta, Indonesia³

*Correspondence Email: nastiti.faulinda@s.unikl.edu.my

ABSTRACT

The transmission of viral diseases, such as COVID-19, influenza, and other viral strains, poses a substantial worldwide challenge. In the context of health, it is necessary to possess a comprehensive comprehension, meticulous examination, and precise anticipation of the dissemination of this infectious disease.

Nonetheless, the presence of diverse data characteristics among different nations poses a considerable obstacle in the development of prediction models for assessing the transmission, mortality, and recovery rates in Indonesia. Understanding the intricacies of viral transmission poses a significant hurdle because to the fluctuating nature of the generalization rate, which is contingent upon country-specific data. The research entailed a comparison of different predictive models, including Random Forest, Simple Linear Regression (SLR), Gaussian Naive Bayes, Multi-Layer Perceptron (MLP), H2O, and Long Short-Term Memory (LSTM), with the purpose of predicting viral transmission. The evaluation metrics encompass MAE, RMSE, and MAPE. The outcomes of the examination of comparison models will aid in identifying the most suitable model for forecasting the transmission of the virus, encompassing the rates of recovery, death, and positive cases, within the specific setting of Indonesia. This work has significance in elucidating the inherent trade-off between efficiency and accuracy within the realm of dynamic data modeling, specifically in the context of COVID-19 viral data.

KEYWORDS

virus, predictive models, time series, machine learning, deep learning



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International

INTRODUCTION

The transmission of viral pathogens among individuals, including but not limited to COVID-19, Influenza, and various other viral strains, is a significant issue of worldwide significance (Hofmeister, Seckler, and Criss 2021; Ripoll and Wilkinson 2023). Within the realm of health, it is vital to possess a comprehensive comprehension, meticulous analysis, and accurate prognostication (Newport, Malhotra, and Widera 2020) of the dissemination of this particular infection (Ripoll and Wilkinson 2023). The accurate forecast of virus propagation is of utmost importance in effectively addressing a worldwide epidemic.

Given the homogeneity of virus (Tolou, Nicoli, and Pisano 2004) distribution data in the time series domain (Martinez-Hernandez et al. 2020), it is imperative to acknowledge the heterogeneity of data features across various countries (Li, Zhao, and Liu 2020). This heterogeneity poses a significant challenge in developing a predictive model for positive transmission, mortality rate, and recovery for Indonesia.

Comprehending the dynamics of virus transmission poses a formidable challenge due to the inherent variability in the generalization rates exhibited by each model (Thompson et al. 2018), which is contingent upon the specific country-level data under consideration. These dynamics encompass a myriad of factors, such as daily fluctuations, non seasonality, and long-term trends (Mahmoud 2020). Prediction models that fail to account for these intricate dynamics have the potential to yield substantial inaccuracies.

In the realm of time series data analysis pertaining to human-to-human virus transmission, a multitude of models have been proposed and investigated (Yadav and Rana 2022). There are several models available for processing time series data, including Random Forest (Ong et al. 2018), Simple Linear Regression (SLR), Gaussian Naive Bayes, Multi-Layer Perceptron (MLP) (Ibrahim 2020), H2O and Long Short-Term Memory (LSTM). Nevertheless, it is imperative to acknowledge that the efficacy and efficiency of said models may not inherently yield optimal accuracy metrics when applied to the prediction of Covid-19 transmission patterns within the Indonesian context.

Hence, it becomes imperative to conduct a comprehensive analysis of comparative models in order to ascertain the optimal and dependable model capable of effectively handling the diverse fluctuations observed in the Covid-19 viral data specifically within the Indonesian context. The act of conducting a comparison enables the researcher to evaluate the inherent trade-offs between efficiency and accuracy.

RESEARCH METHOD

The methodology employed in this study is rooted in a data-driven paradigm, wherein the analysis is predicated upon empirical data procured from esteemed sources from reputable institutions such as the Ministry of Health of the Republic of Indonesia - specifically the Satuan Gugus Covid-19 and the Meteorology, Climatology, and Geophysical Agency of the Republic of Indonesia.

The primary aim of this study is to conduct a comparative analysis of different prediction models, specifically Random Forest, Simple Linear Regression (SLR), Gaussian Naive Bayes, Multi-Layer Perceptron (MLP), H2O, and Long Short-Term Memory (LSTM). The objective is to evaluate their performance in forecasting the transmission of the virus among individuals using time series data obtained from two governmental ministries.

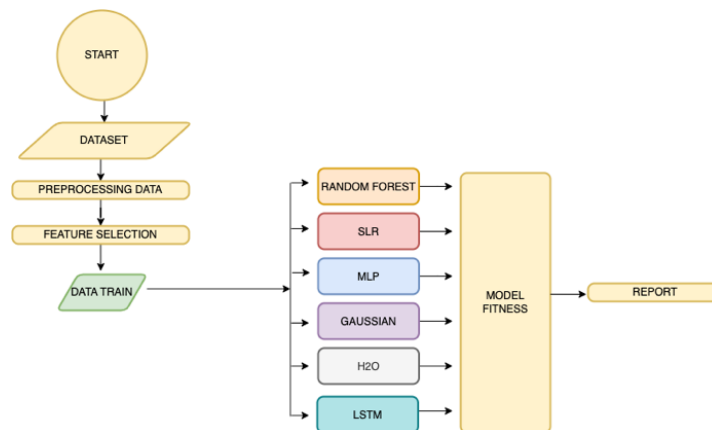


Figure 1. Research Method Pipeline

During the preprocessing stage, the data cleaning process is executed with the objective of addressing missing values, outliers, and transforming time series data into a format suitable for subsequent analysis.

The Covid-19 dataset procured from the two ministries encompasses a total of 82 variables. Consequently, in order to streamline the dataset and identify the most pertinent features, a meticulous feature selection analysis is conducted, employing the method of correlation analysis. The features that have been chosen for analysis are those which exhibit a correlation value exceeding 0.8 with respect to the target features, specifically recovery cases, fatalities, and positive cases.

(A) Random Forest

The utilization of machine learning techniques The Random Forest algorithm use an ensemble of decision trees to make Covid-19 predictions (Ganesh and Vani 2022). The utilization of bootstrapping sampling in the Random Forest technique (Conceição et al. 2021) enables the training of individual decision trees on a subset of the training data, while also considering a random subset of features at each node (Wang et al. 2020).

Every decision tree is trained by the random selection of dataset attributes. The inclusion of randomization in this process serves to introduce variability and mitigate the risk of trees exhibiting excessive similarity.

The Random Forest technique involves the aggregation of several decision trees in order to enhance the accuracy and reliability of estimations, surpassing the capabilities of individual decision tree algorithms (Conceição et al. 2021). The Random Forest technique integrates the forecasts of many decision trees in order to provide the ultimate forecast .

(B) Simple Linear Regression (SLR)

Simple Linear Regression (SLR) is a fundamental methodology in the field of predictive analytics.

$$y = \beta_0 + \beta_1x + \varepsilon$$

The purpose of this tool is to construct a correlation between two variables, wherein one variable is considered independent (x) and the other is considered dependent (y). The intercept refers to the constant coefficient in the linear equation (β) (Kim, Jung, and Choung 2020). When examining the transmission of viruses, some independent variables to consider include time, meteorology, and vaccine. The

dependent variable in this context would be the number of persons who have been infected.

(C) The Multi-Layer Perceptron (MLP)

MLP exhibits a multitude of advantageous characteristics, rendering it a highly sought-after variant of artificial neural networks within the realm of machine learning and deep learning. MLP possesses the remarkable capability to effectively capture intricate relationships between input and output variables (Musa and Aliyu 2019), primarily due to the inclusion of a hidden layer that incorporates a nonlinear activation function. This enables the system to address challenges that are beyond the scope of linear or other rudimentary models.

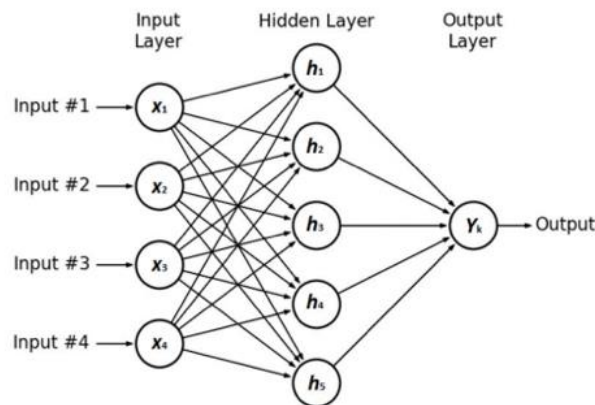


Figure 2. Multi-Layer Perceptron Diagram

The equation used for calculating the output of the output layer is as follows (Bhattacharjee and Bhattacharja 2019),

$$y_k(t) = \varphi \left[\sum_{j=1}^n x_{jk}(t) * y_{jk}(t) - \theta_k \right]$$

y_k = Final output, φ = Activation function, θ_k = Threshold, n = Number of neurons

(D) Gaussian Naive Bayes

Gaussian Naive Bayes or Gaussian Process Regression (GPR) leverages the Gaussian distribution as an integral component of their models. By incorporating this distribution, these algorithms are able to generate predictions and quantify uncertainty in a robust manner (Jain et al. 2022).

$$p(x_i|y_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}}$$

(E) H2O Deep Learning

H2O Deep Learning employs Deep Feedforward Neural Networks. The artificial neural network is composed of many layers, often including an input layer, a hidden layer, and an output layer. This network gradually analyzes data by means of neurons. Deep Learning framework has been specifically developed to provide rapid

and effective parallel processing. This enables the efficient training of deep neural networks on big datasets with accelerated processing (Bisandu et al. 2021).

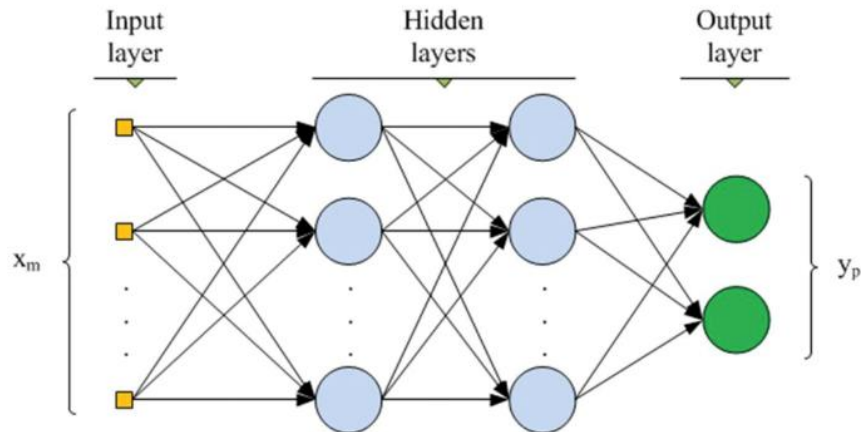


Figure 3. Deep Feedforward Neural Networks Diagram

(F) Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks have been specifically engineered to address the inherent issue of information loss encountered in conventional Recurrent Neural Networks (RNNs).

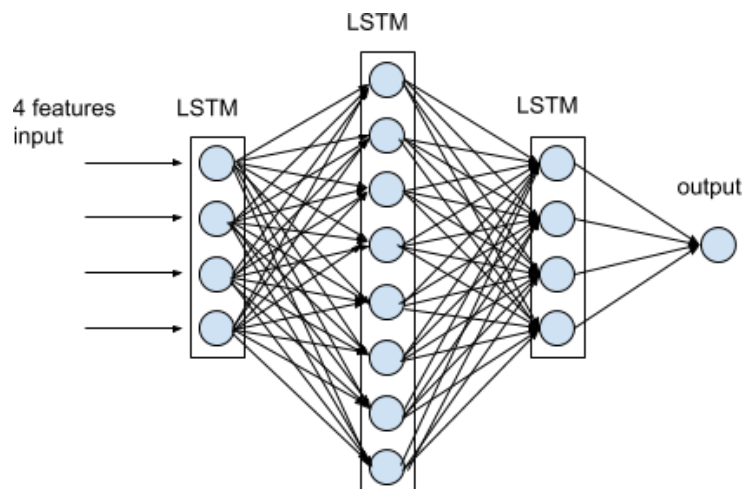


Figure 4. Deep Feedforward Neural Networks Diagram

The system under consideration exhibits the presence of long-term memory cells, which possess the remarkable capability to retain and store information over extended periods of time, thereby establishing a temporal sequence of stored data. The Long Short-Term Memory (LSTM) architecture is characterized by its three fundamental gates: the forget gate, the input gate, and the output gate. These gates play a crucial role in regulating the flow of information within the LSTM network, enabling it to effectively capture and retain long-range dependencies in sequential data. The forget gate determines which information from the previous time step should be discarded, while the input gate controls the incorporation of new information into the memory cell. Lastly, the output gate governs the selective output of relevant information from the memory cell. Together, these gates empower the LSTM The

aforementioned gates serve as pivotal components in the regulation of information flow, both ingress and egress, to and from the memory cells (Masum et al. 2020).

(G) Performance Matrix

Hence, the aforementioned error can be effectively employed as a means to assess the veracity and precision of the model's predictive capabilities. The optimization of model performance is directly correlated with the minimization of error. Consequently, a smaller error magnitude signifies a superior fitting effect (Musa and Aliyu 2019). The mean absolute error (MAE), rooted mean squared error (RMSE), and mean absolute percentage (MAPE) as a prevalent metric for evaluating forecast accuracy within the realm of time series analysis is widely acknowledged (Ali et al. 2020; Alsuwaiket 2020; Masum et al. 2020; Yadav and Rana 2022).

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x_i|$$

$$RSME = \sqrt{\frac{1}{N} \sum_{i=1}^n (x_i - x_i)^2}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - x_i}{x_i} \right| \times 100\%$$

A comparative analysis was conducted to elucidate the outcomes of model evaluation in order to ascertain the optimal model for predicting the propagation of the virus, encompassing prognostications pertaining to recovery, mortality, and positive cases.

RESULT AND DISCUSSION

The primary objective of this study is to assess and juxtapose the efficacy of various predictive models in forecasting three pivotal variables pertaining to the transmission of the virus among the human population. These variables encompass the number of individuals who have successfully recuperated, the count of fatalities, and the tally of confirmed positive cases. The present study encompassed an empirical investigation wherein diverse modeling tools and techniques were employed. These encompassed the utilization of the Weka Application in conjunction with the Random Forest algorithm, Single Linear Regression, Multilayer Perceptron (MLP), and Gaussian Naive Bayes models. Furthermore, the utilization of RapidMiner is employed for the purpose of processing H2O Deep Learning. Additionally, the implementation of LSTM, which stands for Long Short-Term Memory, is executed utilizing the Python programming language in conjunction with the TensorFlow framework.

(A) A Comparative Analysis Of Model Performance In Predicting Cases Of Recovery

Table 1. Model Performance Analysis In Predicting Cases Of Recovery

| METHOD | MAPE | RMSE | MAE |
|------------------------|--------|--------------------|------------------|
| Random Forest | 0.9761 | 4927964296937754.5 | 3150408232562883 |
| Single Linear Regesion | 0.1563 | 13343145202103170 | 9512100451829778 |
| Multi-Layer Perceptron | 0.9552 | 4295573302031879 | 3285326721029128 |
| Gaussian Naive Bayes | 0.4005 | 12390881276790422 | 8558012786676220 |
| H2o Deep Learning | - | 163558 | - |
| LSTM | 1.24 | 2682.2 | 1215.6 |

The Random Forest model has a Mean Absolute Percentage Error (MAPE) of around 0.9761, indicating an average relative error of approximately 97.61%. A significantly high root mean square error (RMSE) value signifies a substantial issue within this particular model. The Mean Absolute Error (MAE), which exhibits a substantial magnitude, signifies that this model possesses a notable discrepancy in its ability to accurately forecast the data.

The Single Linear Regression model has a Mean Absolute Percentage Error (MAPE) of around 0.1563, signifying an average relative error of approximately 15.63%. A significantly high root mean square error (RMSE) is indicative of subpar precision performance. The Mean Absolute Error (MAE), which exhibits a substantial magnitude, suggests that this model possesses a notable discrepancy in its ability to accurately forecast the data.

The MLP model exhibits a Mean Absolute Percentage Error (MAPE) of around 0.9552, indicating that the average relative error is approximately 95.52%. A substantial root mean square error (RMSE) value suggests the presence of an issue within this particular model. A high Mean Absolute inaccuracy (MAE) value suggests that the model exhibits a substantial level of inaccuracy.

The Gaussian Naive Bayes model exhibits a Mean Absolute Percentage Error (MAPE) of around 0.4005, indicating an average relative error of approximately 40.05%. A significantly high root mean square error (RMSE) is indicative of subpar precision performance. The magnitude of the Mean Absolute Error (MAE), which is notably high, suggests that this model exhibits a substantial level of inaccuracy in its predictions of the data.

The H2O Deep Learning model exhibits a root mean square error (RMSE) of around 163.558, signifying a significantly reduced level of error when compared to the other models presented in the table. This observation suggests that the model has a comparatively elevated degree of accuracy in forecasting data.

The LSTM model has a Mean Absolute Percentage Error (MAPE) of around 1.24, indicating an average relative error of approximately 124%. The model has a high level of accuracy, as seen by the comparatively tiny root mean square error (RMSE). The tiny Mean Absolute inaccuracy (MAE) value suggests that this model exhibits a lower level of inaccuracy in comparison to the majority of the other models shown in the table 1.

(B) A Comparative Analysis Of Model Performance In Predicting Cases Of Death

Table 2. Model Performance Analysis In Predicting Cases Of Death

| METHOD | MAPE | RMSE | MAE |
|------------------------|-------------|--------------------|------------------|
| Random Forest | 0.9695 | 4415879457735281.5 | 3132359493189716 |
| Single Linear Regesion | 0.2323 | 9737012988225092 | 1287493155962131 |
| Multi-Layer Perceptron | 0.9551 | 6194481494646637 | 5212064805798128 |
| Gaussian Naive Bayes | 0.4368 | 11915386815010498 | 9115686885464880 |
| H2o Deep Learning | - | 124637 | - |
| LSTM | 2.846 | 375.752 | 375.7518 |

The outcomes of the training process entail a comparative evaluation of model performance in prognosticating mortality instances. Tabel 2, observed that the H2O Deep Learning Model exhibits a relatively diminished Root Mean Square Error (RMSE), thereby indicating a commendable degree of accuracy in predicting death cases. Regrettably, the absence of MAPE values precludes us from acquiring insights into the relative error expressed in percentage terms.

The LSTM model exhibits a notable MAPE value of 2.846, signifying a substantial average relative error expressed in percentage terms. Nevertheless, the remarkably low values of the root mean squared error (RMSE) and mean absolute error (MAE) suggest that the model exhibits a negligible absolute error rate within this particular context.

The Single Linear Regression model exhibits a commendably low Mean Absolute Percentage Error (MAPE) value of 0.2323, which suggests a notable level of precision in forecasting mortality occurrences. Furthermore, the observed values of the root mean squared error (RMSE) and mean absolute error (MAE) suggest a commendably low level of inaccuracies in the predictive models.

The comparative analysis reveals that both the Random Forest and Multi-Layer Perceptron (MLP) models exhibit marginally elevated Mean Absolute Percentage Error (MAPE) values in comparison to the Single Linear Regression model. However, it is noteworthy that these aforementioned models also demonstrate a higher Root Mean Square Error (RMSE) metric.

The Gaussian Naive Bayes model exhibits a notable MAPE value of 0.4368, suggesting a relatively elevated relative error rate expressed in percentage terms.

(C) A Comparative Analysis Of Model Performance In Predicting Cases Of Positive

Table 3. Model Performance Analysis In Predicting Cases Of Positive

| METHOD | MAPE | RMSE | MAE |
|------------------------|--------|--------------------|------------------|
| Random Forest | 0.9651 | 4318324962837836.5 | 3026664359365676 |
| Single Linear Regesion | 0.2323 | 11717353835759028 | 8579073446639973 |
| Multi-Layer Perceptron | 0.9685 | 4180986283368301 | 3322297094861063 |
| Gaussian Naive Bayes | 0.4258 | 10915705645415416 | 7959994448679431 |
| H2o Deep Learning | - | 184959 | - |
| LSTM | 1.12 | 6881.7 | 447.75 |

Tabel 3, the H2O Deep Learning model demonstrates a low root mean square error (RMSE), suggesting a comparatively accurate prediction of positive cases. Regrettably, the absence of MAPE values precludes us from obtaining any insights into the relative inaccuracy expressed in percentage terms.

The Single Linear Regression model has a low Mean Absolute Percentage Error (MAPE) score of 0.2323, suggesting a relatively good level of accuracy in its ability to forecast positive instances. Nevertheless, a significantly large root mean square error (RMSE) suggests that this particular model exhibits a substantial degree of imprecision in its predictions.

The Multi-Layer Perceptron (MLP) model has a very elevated Mean Absolute Percentage Error (MAPE) of 0.9685 and a substantial Root Mean Square Error (RMSE), suggesting a notable level of inaccuracy in forecasting positive instances.

The Random Forest and Gaussian Naive Bayes models exhibit somewhat higher Mean Absolute Percentage Error (MAPE) values compared to the Single Linear Regression model, while demonstrating lower Root Mean Square Error (RMSE) values.

The LSTM model has a slightly elevated Mean Absolute Percentage Error (MAPE) of 1.12, denoting a considerably higher relative error percentage. However, the comparatively lower Root Mean Square mistake (RMSE) and Mean Absolute Error (MAE) values suggest that the model demonstrates a reduced mistake rate in this particular scenario.

CONCLUSION

Based on the previously mentioned test's findings, the LSTM (Long Short-Term Memory) model emerges as the priority model. This is mostly due to the LSTM model's notable characteristic of exhibiting a comparatively low MAPE (Mean Absolute Percentage Error) number. This suggests that the LSTM model is capable of generating estimations that are quite accurate in terms of relative error rates.

The LSTM model has low values for both the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), namely 6,881.7 and 447.75, respectively. This finding suggests a reduced level of prediction error in actual units, which has significance in such circumstances.

In addition, Long Short-Term Memory (LSTM) exhibits the capability to effectively capture and represent temporal dependencies and patterns within time series data. In the domain of forecasting the propagation of viruses in human populations, whereby temporal dependencies play a crucial role, Long Short-Term Memory (LSTM) models have demonstrated exceptional proficiency in capturing temporal patterns.

Long Short-Term Memory (LSTM) possesses the capability to comprehend and manipulate sequential data, a prevalent attribute observed in time series data. The Long Short-Term Memory (LSTM) model possesses the capacity to dynamically adjust its network architecture throughout the training process, hence enabling the model to acquire a deeper understanding of intricate patterns within the dataset.

REFERENCES

- Ali, M., D. M. Khan, M. Aamir, U. Khalil, and Z. Khan. 2020. "Forecasting COVID-19 in Pakistan." *Plos One*.
- AlsawaiKet, Mohammed A. 2020. "Predicting the COVID-19 Spread, Recoveries and Mortalities Rates in Saudi Arabia Using Ann." *Journal of Theoretical and Applied Information Technology* 98(23):3643–53.
- Bhattacharjee, Indronil, and Pryonti Bhattacharja. 2019. "Stock Price Prediction: A Comparative Study between Traditional Statistical Approach and Machine Learning Approach." *2019 4th International Conference on Electrical Information and Communication Technology, EICT 2019* (April 2020). doi: 10.1109/EICT48899.2019.9068850.
- Bisandu, Desmond Bala, Mohammed Salih Homaid, Irene Moulitsas, and Salvatore Filippone. 2021. "A Deep Feedforward Neural Network and Shallow Architectures Effectiveness Comparison: Flight Delays Classification Perspective." *Proceedings of the 5th International Conference on Advances in Artificial Intelligence*.
- Conceição, Marcos Reinan Assis, Luís F. Mendonça, Carlos Alexandre Domingos Lentini, André T. Cunha Lima, José Marques Lopes, Rodrigo Nogueira de Vasconcelos, Mainara Biazati Gouveia, and Milton José Porsani. 2021. "SAR Oil Spill Detection System through Random Forest Classifiers." *Remote. Sens.* 13:2044.
- Ganesh, K., and B. Vani. 2022. "An Efficient COVID-19 Pandemic Survival Analysis to Compare Random Forest and Support Vector Machine for Classifying Performance in Censored Data." *ECS Transactions* 1(107):12993. doi: 10.1149/10701.12993ecst.

- Hofmeister, Anne M., James M. Seckler, and Genevieve M. Criss. 2021. "Possible Roles of Permafrost Melting, Atmospheric Transport, and Solar Irradiance in the Development of Major Coronavirus and Influenza Pandemics." *International Journal of Environmental Research and Public Health* 18(6):1–24. doi: 10.3390/ijerph18063055.
- Ibrahim, S. 2020. "Performance Evaluation of Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF): COVID-19 Spread and Death Contributing Factors." *International Journal of Advanced Trends in Computer Science and Engineering* 9(1):625–31. doi: 10.30534/ijatcse/2020/8791.42020.
- Jain, Shipra, Anjali Dhall, Sumeet Patiyal, and Gajendra P. S. Raghava. 2022. "IL13Pred: A Method for Predicting Immunoregulatory Cytokine IL-13 Inducing Peptides." *Computers in Biology and Medicine* 143:105297. doi: <https://doi.org/10.1016/j.compbiomed.2022.105297>.
- Kim, Seon Woo, Donghwi Jung, and Yun Jae Choung. 2020. "Development of a Multiple Linear Regression Model for Meteorological Drought Index Estimation Based on Landsat Satellite Imagery." *Water (Switzerland)* 12(12). doi: 10.3390/w12123393.
- Li, Xin, Zebin Zhao, and Feng Liu. 2020. "Big Data Assimilation to Improve the Predictability of COVID-19." *Geography and Sustainability* 1:317–20.
- Mahmoud, R. H. Al. 2020. "Covid-19 Global Spread Analyzer: An ML-Based Attempt." *Journal of Computer Science* 16(9):1291–1305. doi: 10.3844/jcssp.2020.1291.1305.
- Martinez-Hernandez, Francisco, Elaine Luo, Kento Tominaga, Hiroyuki Ogata, Takashi Yoshida, Edward F. Delong, and Manuel Martínez-García. 2020. "Diel Cycling of the Cosmopolitan Abundant Pelagibacter Virus 37-F6: One of the Most Abundant Viruses in Earth." *Environmental Microbiology Reports*.
- Masum, Mohammad, Hossain Shahriar, Hisham M. Haddad, and Md Shafiul Alam. 2020. "R-LSTM: Time Series Forecasting for COVID-19 Confirmed Cases with LSTMbased Framework." *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020* 1374–79. doi: 10.1109/BigData50022.2020.9378276.
- Musa, Aminu, and Farouq Aliyu. 2019. "Performance Evaluation of Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF)." *2019 2nd International Conference of the IEEE Nigeria Computer Chapter, NigeriaComputConf 2019* (June 2022):1–5. doi: 10.1109/NigeriaComputConf45974.2019.8949669.
- Newport, Kristina Braine, Sonia Malhotra, and Eric W. Widera. 2020. "Prognostication and Proactive Planning in COVID-19." *Journal of Pain and Symptom Management* 60:e52–55.
- Ong, Janet, Xu Liu, Jayanthi Rajarethinam, Suet Yheng Kok, Shaohong Liang, Choon Siang Tang, Alex Richard Cook, Lee Ching Ng, and Grace Yap. 2018. "Mapping Dengue Risk in Singapore Using Random Forest." *PLoS Neglected Tropical Diseases* 12.

- Ripoll, Santiago, and Annie Wilkinson. 2023. "The Role of Social Science in Influenza and SARS Epidemics BT - Handbook of Social Sciences and Global Public Health." Pp. 1–21 in, edited by P. Liamputtong. Cham: Springer International Publishing.
- Thompson, Robin N., Chris Wymant, Rebecca A. Spriggs, Jayna Raghvani, Christophe Fraser, and Katrina A. Lythgoe. 2018. "Link between the Numbers of Particles and Variants Founding New HIV-1 Infections Depends on the Timing of Transmission." *Virus Evolution* 5.
- Tolou, Hugues J., Jean Nicoli, and Marina Pisano. 2004. "Homogeneity of Yellow Fever Virus Strains Isolated During an Epidemic and a Post-Epidemic Period in West Africa." *Virus Genes* 14:225–34.
- Wang, Qingyong, Yun Zhou, Weiping Ding, Zhiguo Zhang, Khan Muhammad, and Zehong Cao. 2020. "Random Forest with Self-Paced Bootstrap Learning in Lung Cancer Prognosis." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16:1–12.
- Yadav, Ameet, and Chhavi Rana. 2022. "A Sytematic Study of Covid-19 Prediction Models of India." Pp. 1–45 in *Research Square*.