ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

THE SIGNIFICANCE OF DYNAMIC COVID-19 DASHBOARD IN FORMULATING SCHOOL REOPENING STRATEGIES

Feby Artwodini Muqtadiroh[1], Eko Mulyanto Yuniarno[1,2], Supeno Mardi Susiki Nugroho[1,2], Muhammad Reza Pahlawan[3], Riris Diana Rachmayanti[4], Tsuyoshi Usagawa[5], Mauridhi Hery Purnomo[1,2*]

[1]Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[2]Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
[3]Sekolah Tinggi Ilmu Ekonomi Indonesia (STIESIA), Surabaya 60118, Indonesia
[4]Health Promotion and Behavioral Science, Universitas Airlangga, Surabaya 60115, Indonesia
[5]Department of Computer Science and Electrical Engineering, Kumamoto University, Kumamoto 8608555, Japan

Corresponding email: hery@ee.its.ac.id[*]

Abstract

Experiments conducted with the COVID-19 dataset have predominantly concentrated on predicting cases fluctuating and classifying lung-related diseases. Nevertheless, the consequences of the COVID-19 pandemic have also spread to the education sector. To safeguard educational stability in response to the remote learning policy, we leverage authentic COVID-19 datasets alongside school information across 154 sub-areas in Surabaya City, Indonesia. Our focus is predicting the dynamic within these sub-areas where schools are located. The outcomes of this study, by incorporating the recurrent neural network of long- and short-term memory (RNN-LSTM) architecture and refined hyperparameters, effectively enhanced the predictive model's performance. The findings are showcased on a dashboard, visually representing the transmission of COVID-19 in schools across each sub-area. This information serves as a basis for informed decisions on the safe reopening of schools, aiming to mitigate the decline in education quality during the challenging pandemic.

Keywords: COVID-19, Schools reopen safely, Dashboard, Recurrent Neural Network, LSTM, Decision-making.

Received: 28-02-2024 | **Revised:** 18-03-2024 | **Accepted:** 19-03-2024
DOI: https://doi.org/10.23887/janapati.v13i1.76017

INTRODUCTION

The global devastation caused by the COVID-19 pandemic is due to the aggressive nature of the SARS-CoV-2 virus, known for its lethality and high transmissibility. Beyond the 18.2 million deaths caused by COVID-19 from the pandemic's onset to the end of 2021, the broader impact is significantly more than just numbers [1]. Various global public health strategies are necessary to enhance the surveillance of both current and forthcoming pandemics. Furthermore, more studies are essential to facilitate the tracking of COVID-19 transmission and forecast its trajectory. Studies from diverse regions worldwide have developed numerous models to anticipate the spread of COVID-19 [1]–[5].

Furthermore, world institutions provide tools in the form of dashboards for tracking the movement of COVID-19 in real-time, along with the advantages of the features presented. However, there are also some drawbacks. Since the beginning of the COVID-19 pandemic, numerous studies have utilized dashboards primarily to visually represent the global, national, or regional distribution of COVID-19 cases. Even prestigious institutions like John Hopkins University [6], the World Bank [7], the WHO [8], and the New York Times [9] have developed dashboards to display the statistical progression of COVID-19 to the public. Nevertheless, they have not fully explored the potential utility of these dashboards. Unfortunately, this approach did not guarantee the reliability of the information presented in the dashboards.

The challenges posed by COVID-19 extend beyond the vulnerabilities of global health infrastructure and the effectiveness of regional prevention measures [10]. They also include the increased worldwide attention to student learning loss, resulting from lockdown measures and mobility restrictions to curb the spread of the virus [11]–[13]. Learning setbacks represent a novel crisis caused by the widespread closure of schools in response to the severity of COVID-19 transmission, particularly in densely populated areas, as recommended by UNICEF [14].

The closure of schools has been recommended as a non-pharmaceutical measure in pandemic management due to its capacity to lower infection transmission among children, school personnel, and their contacts. Nevertheless, considering schools' diverse roles in society, prolonged closures are expected to result in more significant impacts [15]. Moreover, global education is facing challenges, leading to the emergence of several studies aimed at preserving the quality and standard of education [16], [17]. This includes analyses and perspectives specific to Indonesia [18], [19].

According to Mustafa [15], prior studies on school closures have primarily focused on their epidemiological impacts, leaving a gap in understanding the wide-ranging effects of such closures. There is a pressing need for studies to examine these extensive effects, including the crucial issue of whether schools should reopen during a pandemic crisis, as highlighted by Kekić and Miladinović in 2016 [20].

Having observed the global response to preserve education amidst uncertainties, ensuring the sector's timely reopening is vital. Consequently, technology plays a significant role in addressing this need. This technology should not only predict future COVID-19 events but also assist in monitoring the dynamic patterns to aid in making well-informed decisions regarding educational settings.

In crisis situations, decision-makers require comprehensive, precise, and thorough information. Nowadays, technology offers numerous options for generating, gathering, and analyzing data. During the COVID-19 pandemic, the abundance of information and the demand for a more comprehensive emergency response approach led to the development of dashboard technology [21]. While numerous dashboards have been created to address COVID-19, they primarily focus on the public health sector [22], neglecting the education sector, which significantly impacts learning loss, particularly in developing countries [23].

We discovered that existing dashboards primarily concentrate on displaying COVID-19 case distribution. There is a lack of connectivity between dynamic COVID-19 data and school data to provide tailored recommendations for school reopening decisions. This is crucial since there is a correlation between the fluctuation in COVID-19 spread and the implementation of movement restrictions, including school closures, aimed at curbing the virus's transmission, which is on the rise [24].

As a result, our study focuses on developing a dashboard capable of presenting COVID-19 movement data, serving as a foundation for determining safe reopening strategies during the pandemic. Our dashboard provides COVID-19 incidence statistics over time alongside extensive school-related data crucial for reopening, all of which is processed using an artificial intelligence system.

We investigated COVID-19 real-case data sourced primarily from March 23, 2020, to October 31, 2022, covering 154 sub-areas in Surabaya. We obtained primary data from the official Surabaya municipal government website, https://lawancovid-19.surabaya.go.id, which provides daily updates on positive confirmations, recoveries, mortality rates, hospitalizations, suspects, and probable cases.

This extensive dataset is then carefully processed to create a predictive model that provides reliable insights into the COVID-19 trend. Therefore, our dashboard can serve as an analytical tool, providing precise information and improving decision-making processes. We optimize the deep learning-based predictive model, specifically the Recurrent Neural Network with Long Short-Term Memory (RNN-LSTM) algorithm, to enhance public and decision-makers' confidence in utilizing our dashboard for effective action planning. We chose the RNN-LSTM algorithm because it effectively predicts time series data [25], [26]. Additionally, the RNN processes data inputs gradually based on time-lags or timesteps [27], implying that the current data at time t_n is influenced by previous data points such as t_{n-1}, t_{n-2}, …, t_{n-k}. This characteristic of RNN mirrors the growth pattern of the virus driving COVID-19, where preceding data heavily impact recent infection cases.

LSTM models have demonstrated significant accuracy in forecasting time-series data through finely tuned hyperparameter configurations [27]. By adeptly selecting time-lagged features, LSTM models effectively capture the nuanced characteristics of time-series data, resulting in decreased Root Mean Square Error (RMSE) for both short-term and long-term predictions. Therefore, this study also underscored the importance of optimizing hyperparameters for the RNN-LSTM architecture before the data output will be the input for the dashboard system.

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

While previous studies have predominantly focused on neural network optimization by examining two to nine hyperparameters [28]–[30], our experiment explored various scenarios, as detailed in the Experimental Setup section. Consequently, we empirically investigated the effects of specific parameters and the significance of certain hyperparameters. Hyperparameter tuning is crucial for improving the model's performance and stability, especially when dealing with extensive and diverse datasets [31], [32].

Consequently, this study made significant contributions in the following areas:

- Analysis of extensive time-series data to predict the dynamics of COVID-19 as a basis to formulate the appropriate decision-making.
- Optimization of models based on various scenarios of hyperparameter tuning on RNN-LSTM to ensure the presentation of reliable information visualized in the dashboard for both the public and decision-makers.

We believe that our experiment-based dashboard system, developed using RNN-LSTM on COVID-19 data, can significantly contribute to future research on endemic situations, epidemics, and other potential pandemics that may cause disruptions in various sectors. To present our methodology and findings, this paper is structured as follows: Section 2 describes the Methods, explaining the dataset and experimental setup, followed by Section 3, which discusses the Results and Discussion, and lastly, the Conclusion.

METHOD

The urgency of data visualization on the dashboard aligns with the evolving nature of each variant of the COVID-19 virus, emphasizing the importance of this issue. The unpredictable fluctuations in COVID-19 dynamics during the pandemic, which may recur in the future, pose challenges to the education system's quality. To address these issues, we have developed a dashboard system that tackles two main concerns: COVID-19 and learning loss.

We are optimistic that our designed dashboard will not only guide educational strategies, but will also help other sectors prevent crises during outbreaks. To this purpose, we offer a framework (Figure 1) for building an AI-powered dashboard system using the RNN-LSTM deep learning algorithm. The resulting insights will be visualized in the dashboard, aiding in decision-making processes, particularly in determining the optimal time for reopening schools amid pandemic crises.

Public skepticism due to the abundance of COVID-19 data and availability reports should not be ignored. The discrepancy between changing case data and policy implementation highlights the need for the public to stay informed through various media, which can be overwhelming.

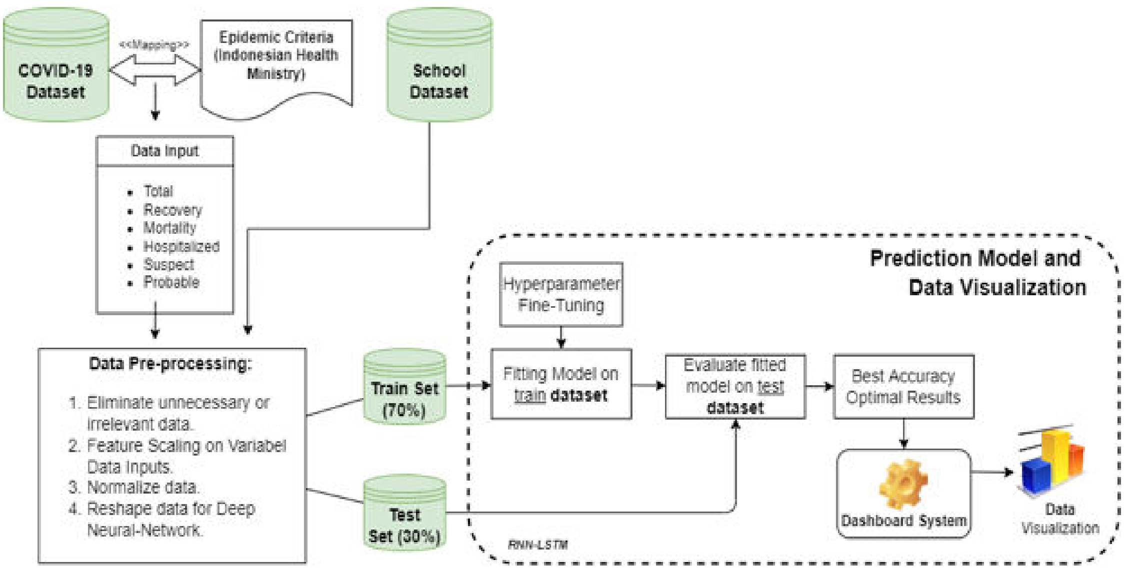


Figure 1. Optimized RNN-LSTM Model to Generate Dashboard System as Decision-Making Material

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

Hence, integrated, relevant, and interactive data visualization becomes paramount to effectively communicate the dynamic COVID-19 information and its future prediction for each sub-area. This urgency drives us to propose recommendations regarding school reopening, presented in the dashboard format. Figure 1 illustrates our conceptual framework, based on a deep neural network utilizing an artificial intelligence-driven approach, specifically the optimized RNN-LSTM model.

It is crucial to emphasize that a fundamental element of developing a dashboard system is ensuring the integrity and reliability of the data. Therefore, we elaborate on our extensive dataset, aligned with government epidemiological standards and its correlation with school-related data in this segment. Additionally, we provide a comprehensive overview of the experimental settings in the subsequent subsection.

Dataset

Table 1. Structure of the COVID-19 Dataset

Date (yyyy-mm-dd)	**City**	**Area**	**Sub-Area**	**TC**	**RC**	**MC**	**HC**	**SUSP**	**PROB**
2020-03-23	South Surabaya	Jambangan	Karah	0	0	0	0	0	0
2020-03-23	South Surabaya	Jambangan	Jambangan	0	0	0	0	0	0
2020-03-23	South Surabaya	Jambangan	Pagesangan	0	0	0	0	0	0
2020-03-23	South Surabaya	Wonocolo	Margorejo	0	0	0	0	0	0
2020-03-23	South Surabaya	Wonocolo	Bendul Merisi	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…
2021-08-15	East Surabaya	Gubeng	Gubeng	406	361	14	31	0	2
2021-08-15	East Surabaya	Gubeng	Airlangga	479	417	12	50	1	0
2021-08-15	East Surabaya	Rungkut	Kedung Baruk	475	431	19	25	0	2
…	…	…	…	…	…	…	…	…	…
2022-10-31	Central Surabaya	Bubutan	Bubutan	591	572	17	2	0	0
2022-10-31	Central Surabaya	Simokerto	Tambakrejo	689	658	30	1	0	0
2022-10-31	Central Surabaya	Simokerto	Simokerto	843	820	21	2	0	0
…	…	…	…	…	…	…	…	…	…
2022-10-31	South Surabaya	Wonocolo	Siwalan kerto	1089	1055	34	0	0	0
2022-10-31	South Surabaya	Jambangan	Kebonsari	627	611	13	3	0	0

Note:
- TC = Total Confirmation
- RC = Recovery Confirmation
- MC = Mortality Confirmation
- HC = Hospitalized Confirmation
- SUSP = Suspect
- PROB = Probable

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

```
Algorithm 1 Exploratory Data Analysis
Input: daily cases data
Output: free of waste data distribution
Initialization: data scrapping from https://lawancovid-19.surabaya.go.id/
Data Organization
   data segregation: accumulative sequential data (per two or three weeks)
   set input: X =(X1, X2, X3, …Xn)

Anomaly Check
   if X[accumulated]=X[accumulated].astype(float).astype(int) then
         perform
   end if
  if Xdate-clean = Xdate.dropna(); Xdate-clean.info() then
         delete new data (inconsistent)
   end if
  if Xdate[Xdate.isnull().any(axis=1)] then
         delete empty field (feature)
  end if

Outlier Data Check
  if X<accumulated[xi1, xi2, …, xin, xj1, xj2, …, xjn, …, xnn ] then
         use previous data field
  end if
  if Xdaily<0 then
         delete data field
  end if
       get new data distribution
  return X
```

We possess two distinct datasets. The first dataset encompasses information about the spread of COVID-19 in Surabaya City, Indonesia, covering 154 sub-areas. Acquired from the website https://lawancovid-19.surabaya.go.id, we collected primary data from March 23, 2020, to October 31, 2022. Utilizing Python-based code for web scraping, we retrieved a substantial dataset comprising 146,299 rows of data with fields including Date, City, Area, Sub-Area, Total Confirmation, Recovery Confirmation, Mortality Confirmation, Hospitalized Confirmation, Suspect, and Probable. The COVID-19 dataset we possess differs from other datasets primarily focused on predicting the virus's spread across countries [33], [34], provinces, or cities [35]. These predictions are intended for authorities to make decisions to curb the virus's broader transmission. However, such analyses typically adhere to single coverage, overlooking the varied nature of the virus spread within a city. Recognizing that a city comprises multiple districts or areas, and each area further consists of several sub-areas, we acknowledge that society's activities primarily revolve around the most minor environment—the sub-area. Consequently, the character of COVID-19 spread is expected to differ among sub-areas. To address this, we utilize a dataset containing information on daily COVID-19 cases in 154 sub-areas within one of Indonesia's largest cities, Surabaya. By closely observing the dynamics of COVID-19 in each sub-area, our results aim to be more specific, providing a basis for making relevant and appropriate decisions.

Our dataset includes more data fields compared to commonly used public reference datasets, such as the Center for Disease Control and Prevention [36], which comprises data fields like Hospitalizations, Deaths, Emergency Department Visits, and Test Positivity, and the Kaggle dataset [37], which includes data fields Confirmed, Deaths, and Recovered.

The collected primary data must be aligned with epidemiological variables or indicators established by the Indonesian government [38], consisting of 15 items. These epidemiological variables are derived from global epidemiological standards (Table 1). However, we cannot incorporate all variables due to the limited scope of our available data components.

Due to limitations associated with internal person-by-person case data and hospital management data, our prediction model relies on six variables extracted from our dataset. They are Total Confirmation, Recovery Confirmation, Mortality Confirmation, Hospitalized Confirmation, Suspect, and Probable. These six variables will eventually become features of our RNN-LSTM model. To obtain high-quality input data from our extensive primary data set, we implemented Algorithm 1 as a crucial stage in exploratory data analysis, which involves the removal of unclean data containing abnormalities, errors, and inconsistencies.

Following this, we conducted a correlation analysis to examine the relationship between these six variables, aiming to ensure their independence. A higher degree of independence suggests that the data contains distinct information, each of which can contribute to effective RNN-LSTM-based prediction modeling. We performed a correlation test, and the results, displayed in Figure 2, indicate that the variables passed the test with a correlation coefficient value of less than 0.68 ($|r| < 0.68$).

Table 2 represents the input data we have developed after successfully completing the exploratory data analysis process and correlation test. Our data set comprises 146,299

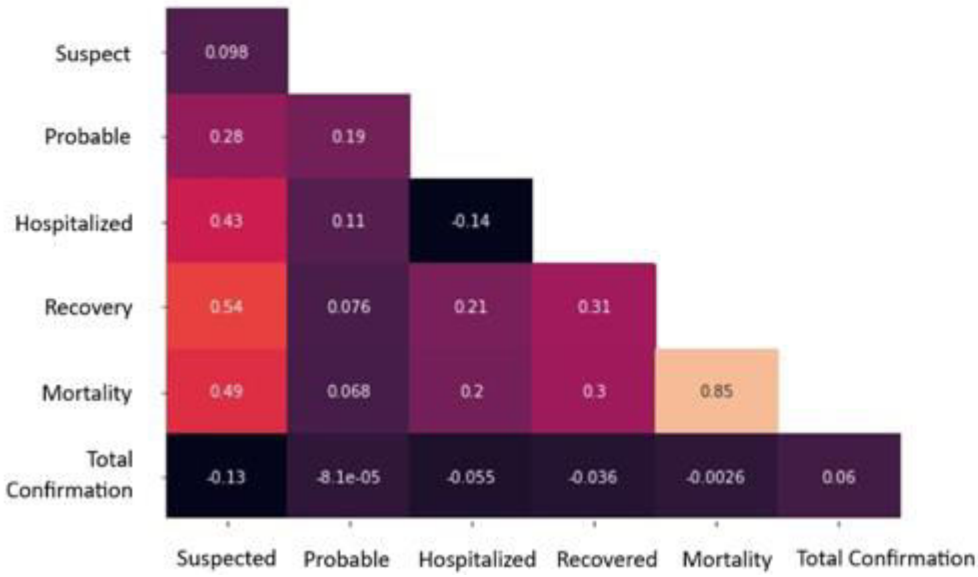


Figure 2. Correlation Test Results

Table 2. Epidemiological Variables

Item	Epidemiological Variables
1	The decrease in positive cases over the past two weeks from the peak
2	Reduction in probable and suspect cases over the past three weeks from the peak
3	The decline in deaths from positive cases over the past three weeks from the peak
4	The decline in deaths from probable and suspect cases over the past three weeks from the peak
5	The decrease in the number of positive cases undergoing hospital treatment over the past two weeks from the peak
6	The decline in the number of probable and suspect cases undergoing hospital treatment over the past two weeks from the peak
7	Increase in the number of recoveries from positive cases
8	Increase in the number of completed monitoring and supervision for probable and suspect
9	Reduction in the incidence rate of positive cases per 100,000 population
10	Reduction in the death rate per 100,000 population
11	Effective reproduction number less than 1
12	Increase in the number of specimen examinations over two weeks
13	A positivity rate of less than 5%
14	The number of beds in the isolation room at the referral hospital can accommodate more than 20% of the number of positive COVID-19 patients
15	The number of beds in referral hospitals can accommodate more than 20% of the number of

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

rows, providing daily case information from March 23, 2020, to October 31, 2022, for each sub-area within Surabaya, divided into the Southern, Eastern, Central, and Northern parts. Dividing areas and sub-areas is crucial for our study, as we believe each sub-area exhibits unique characteristics of COVID-19 spread.

This approach allows us to avoid generalizing our research findings on school openings at the city level, as outcomes may differ across sub-areas due to varying distribution dynamics. Our modeling considers the ratio of area sizes, ensuring that data variables are divided according to each sub-area size. By doing so, we can obtain more precise and discernible results, which will serve as a solid basis for making well-informed decisions.

On the other hand, our second dataset pertains to school information sourced from the official website of the Indonesian Central Bureau of Statistics, https://surabayakota.bps.go.id/. This dataset encompasses details from 1,466 schools and includes fields such as Area, School ID, School Name, Address, Sub-Area, Status (Public, Private), and Level (Primary school, Middle school, and High school).

Data on all schools in Surabaya at various levels and scattered throughout all sub-areas makes it simple for us to display them all in the dashboard system. This allows us to monitor each school within each sub-area, ensuring that the optimal school opening time aligns with the progression of COVID-19 spread.

As experts and global organizations indicated, the fluctuation of virus transmission has led to interruptions in school activities, impacting the quality of education during home-based learning. Our model integrates both datasets to address this issue, creating a framework for safely reopening schools during the pandemic. This framework assists affected school stakeholders in determining the optimal time to resume school activities safely. Therefore, those two datasets are crucial in constructing our dashboard system. When the prediction results illustrate the dynamics of COVID-19 spread in each sub-area, the dashboard system provides information on all schools within the desired time frame.

Experimental Setup

In our study, all COVID-19 data is segmented by area and sub-area to establish a ratio, ensuring that case trends in each sub-area are contextually relevant or possess unique characteristics based on the specific area or location. Normalization is then applied to the daily input of COVID-19 case data,

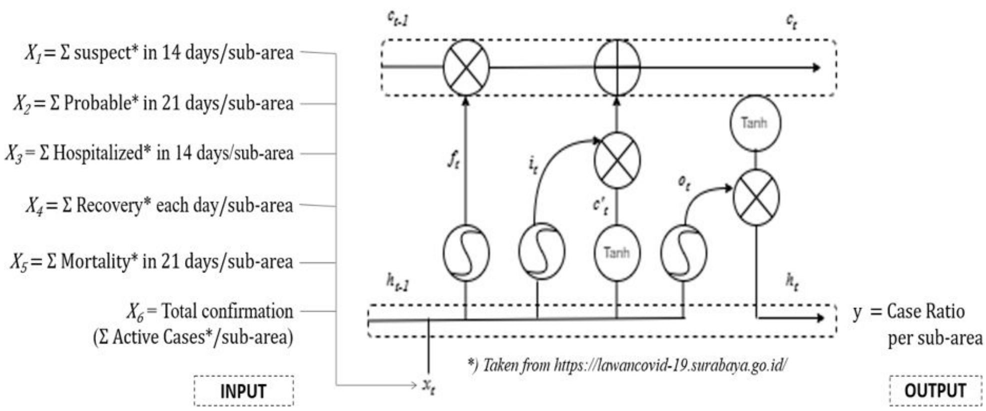


Figure 3. RNN-LSTM architecture

Figure 4. Data Accumulation in Bi-weekly (X_1 and X_3) and Tri-weekly (X_2 and X_5)

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

facilitating the processing of case distribution data within the deep neural network of the RNN-LSTM. Figure 3 illustrates the operation of our RNN-LSTM intelligence-driven approach, receiving clean and error-free data inputs through the feature extraction process. The inputs are:

- $X_1 = \Sigma$ Suspect* in 14 days/sub-area
- $X_2 = \Sigma$ Probable* in 21 days/sub-area
- $X_3 = \Sigma$ Hospitalized* in 14 days/sub-area
- $X_4 = \Sigma$ Recovery* each day/sub-area
- $X_5 = \Sigma$ Mortality* in 21 days/sub-area
- X_6 = Total Confirmation

Those six inputs are applicable to each sub-area. As a result, we generated 154 prediction models based on the total number of sub-areas observed. The final accuracy measurement yields the average calculation. Additionally, we can identify which sub-areas exhibit the highest accuracy, which we will briefly address in the subsequent section. Our primary focus remains on developing a dashboard system, where we will visually present the information in accordance with our research objectives after obtaining the best results.

To clarify the input data aspect, it is essential to note that our data is accumulative. Not all input data is daily; some information is gathered every two or three weeks, which should be considered when analyzing the data. Particularly for input variables X_1 and X_3, data processing occurs bi-weekly, and X_2 and X_5 occur tri-weekly in alignment with epidemiologic variables. Therefore, this procedure involves the initial set of data spanning from the first day to the fourteenth day, followed by subsequent iterations covering consecutive days, including tri-weekly data cases, as illustrated in Figure 4.

Following the explanation, the daily COVID-19 dataset we provide (DOI: 10.17632/smpzjpdzwr.1) can be readily organized to meet the requirements of observations and experiments, similar to the treatment of data variables X_1, X_2, X_3, and X_5 in our experiment. This allows for dynamic sorting of our COVID-19 dataset based on the specific requirements of daily, weekly, bi-weekly, and tri-weekly observations, and so forth.

The subsequent stage involves partitioning the data into training and testing sets with a ratio of 70:30. During the modeling phase, 70% of the complete training dataset is utilized for adjustments, incorporating subsequent hyperparameter fine-tuning, including:

1. Timesteps: 5, 10, 15
2. Hidden Layer Nodes: (5, 10, 15), (10, 20, 30), (15, 30, 45)
3. Epochs: 100, 200
4. Batch Size: 4, 8
5. Optimizer: Stochastic Gradient Descent (SGD), RMSprop, Adam

The remaining 30% of the dataset will be reserved for the testing process to evaluate the performance of the best model.

In standard, untuned circumstances, our prediction model employs random parameters that contain default values supplied by the Python library [39], specifically *keras.layers.LSTM is* as follows:

```
keras.layers.LSTM(
    units,
    activation="tanh",
    recurrent_activation="sigmoid",
    use_bias=True,
    kernel_initializer="glorot_uniform",
    recurrent_initializer="orthogonal",
    bias_initializer="zeros",
    unit_forget_bias=True,
    kernel_regularizer=None,
    recurrent_regularizer=None,
    bias_regularizer=None,
    activity_regularizer=None,
    kernel_constraint=None,
    recurrent_constraint=None,
    bias_constraint=None,
    dropout=0.0,
    recurrent_dropout=0.0,
    seed=None,
    return_sequences=False,
    return_state=False,
    go_backwards=False,
    stateful=False,
    unroll=False,
  )
```

After identifying the most effective predictive model, we compared its performance before and after optimization. To validate the model, we conducted a focus group discussion (FGD) with experts in the epidemiology field from Universitas Airlangga in Surabaya, Indonesia. Our prior study comprehensively explained how expert validation works [40]. Consequently, we can confidently present our findings, showcased in a dashboard tool that offers valuable insights and information to facilitate more informed decision-making among the general public and policymakers.

As a result, the dashboard reveals which schools can be safely opened or closed at the specified time. This system enables conclusions based on tested data patterns, facilitating decision-making. The dashboard system discussed in the subsequent section visualizes intelligent insights, categorizing school openings into controllable and vulnerable categories according to expert formulation derived from the outcomes of RNN-LSTM intelligent modeling. Thus, the factors determining opening and closing are as follows:

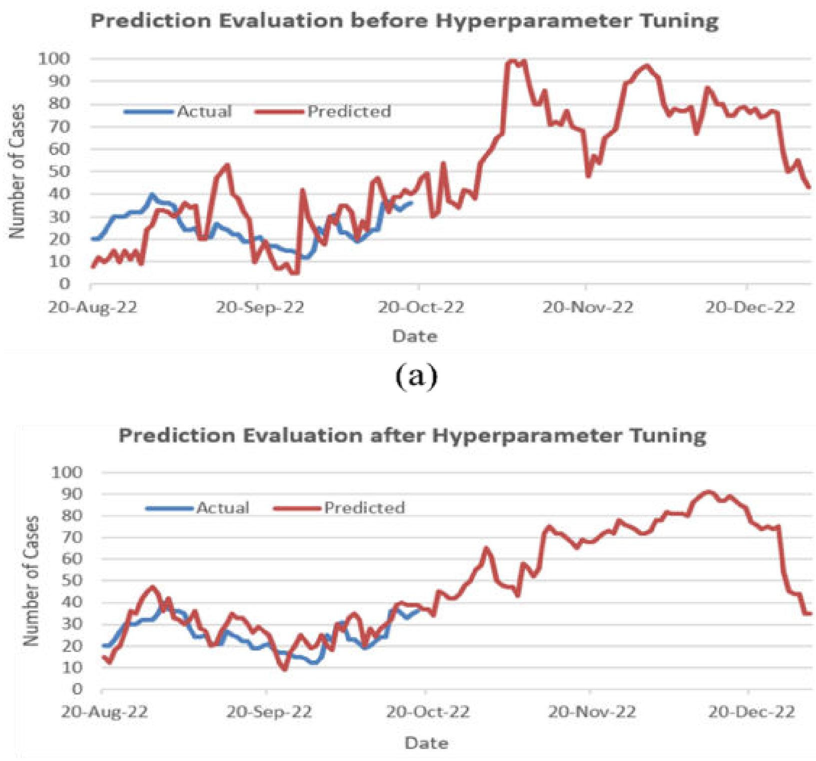
- Controllable: Refers to areas where data shows a decrease in the number of suspects, a decline in mortalities, and a diminishing trend in confirmed cases. These conditions signify readiness for school reopening.
- Vulnerable: Describes areas with an increase in suspected cases, a high death rate, and a low recovery trend. Under these circumstances, schools are advised against reopening

Our experimental results can be more confidently showcased, as they will be exhibited in a dashboard designed to function as an analytical tool. This dashboard will provide information and insights, ultimately fostering more responsible decision-making among the public and decision-makers.

We use a data visualization dashboard to guide decisions, particularly on safe school reopening during the pandemic. This process involves employing RNN-LSTM for analysis.

By refining the hyperparameters, the highest model accuracy was attained from the Adam optimizer, incorporating five (5) nodes in the hidden layer, an epoch duration of 100, a timestep of 5, and a batch size of 4. Employing RNN-LSTM and adjusting parameters for fine-tuning, the model's performance exhibited a notable increase in average accuracy by 5.28% (from 92.44% to 97.32%).

The overall accuracy is determined by calculating the average values across the entire sub-area. The Waru Gunung Sub-area holds the highest accuracy levels, as shown in Figure 5. This improvement suggests better COVID-19 trend predictions, as the model aligns more closely with actual values during the post-optimization phase. This is evident in Figure 5b, where the convergence of data movement is observed, with the blue trend line and the red one side by side.



(b)

Figure 5. Model Evaluation, (a) data trend before fine-tuning, (b) data trend after fine-tuning in Sub-Area of Waru Gunung

RESULT AND DISCUSSION

Table 3. Outcome of Model Optimization (average value)

Measurement Items	Before Tuning	After Tuning
RMSE (%)	7.56	2.68
Accuracy (%)	92.44	97.32
Running Time (seconds)	7.81	4.26

Table 4. Partial Examples of Information in Tabular Form Containing School Opening Predictions

Date	Prediction 1=opening 0=closing	Area	School ID	School Name	Sub-Area	Status	Level
2020-03-23	1	Bulak	20532905	SDN BULAK RUKEM I-258	Bulak	Public	Primary school
2020-03-23	1	Bulak	20532869	SD TRI GUNA BHAKTI	Bulak	Private	Primary school
2020-03-23	1	Bulak	20532693	SMKS TRI GUNA BHAKTI SURABAYA	Bulak	Private	High school
2020-03-29	0	Wonokromo	20532665	SMP BRAWIJAYA SAKTA I	Ngagel Rejo	Private	Middle school
2020-03-29	1	Dukuh Pakis	20533415	SDN PRADAH KALIKENDAL III-530	Pradah Kali Kendal	Public	Primary school
2020-03-30	1	Pakal	20532605	SMP WACHID HASYIM 7 SURABAYA	Benowo	Private	Middle school
2020-04-03	0	Wonokromo	20532710	SMKS TUNAS WIJAYA SURABAYA	Darmo	Private	High school
2021-04-03	1	Gubeng	20532564	SMPN 6 SURABAYA	Gubeng	Public	Middle school
2020-04-03	0	Wonokromo	20533258	SDN SAWUNGGALING VIII-389	Sawunggaling	Public	Primary school
2020-05-01	0	Wonokromo	20571616	SMPN 48 SURABAYA	Ngagel Rejo	Public	Middle school
2020-05-01	1	Tandes	20532600	SMP TRI KARYA	Manukan Wetan	Private	Middle school
2021-03-12	1	Bubutan	20539228	SMP WACHID HASYIM 2	Jepara	Private	Middle school
2021-03-12	0	Bubutan	60720879	MIS DARUSSALAM	Bubutan	Private	Primary school
2021-05-17	1	Gayungan	20532237	SMAN 15 SURABAYA	Dukuh Menanggal	Public	High school
2022-08-25	0	Bulak	20531942	SD ANGKASA	Sukolilo Baru	Private	Primary school
2022-08-25	1	Semampir	20532583	SMPN 11 SURABAYA	Ujung	Public	Middle school
2022-08-25	0	Gayungan	20532548	SMPN 22 SURABAYA	Gayungan	Public	Middle school
2022-08-25	0	Gayungan	20532173	SMKS DHARMA BHAKTI SURABAYA	Menanggal	Private	High school

The comprehensive performance is evident in Table 3, which displays the average value of all sub-areas, consisting of information regarding the model measurement based on the Root Mean Square Error RMSE, accuracy, and running time (in seconds).

The subsequent process generates the outcome of optimized RNN-LSTM models by merging two datasets to predict the movement of COVID-19 within distinct sub-areas, denoting controllable and vulnerable conditions for school resumptions, as outlined in Table 4. Key components of the table encompass Date, Prediction (opening or closing), Area, School ID, School Name, Address, Sub-Area, School Status, and School Level.

In Table 4, the data field or prediction column represents the outcome of the RNN-LSTM optimization process. We assign a flag to our prediction model, with the value 1 representing opening and 0 representing closing. The 'opening' condition signifies a controllable scenario, implying stable COVID-19 dynamics. This allows sub-areas classified as controllable to operate schools safely during the pandemic. Controllable and vulnerable conditions depend on the number of suspects, mortality rates, and confirmed cases.

Table 4 contains 1,130,580 data entries, covering observations from March 23, 2020, to October 31, 2022. The dashboard system visually displays this information to track COVID-19 movement in each sub-area over time. The dashboard also provides information about schools appropriate for opening (controllable condition) or closure (vulnerable condition). The dashboard we have developed offers dynamic visualization capabilities to provide clear and precise information on the movement of COVID-19 in each specific sub-area where schools are situated, ensuring ease of understanding.

Our dashboard can be an essential guide to support the decision-making process concerning a safe school reopening during the pandemic in a display encompasses the following information:

- Page 1: Overview - Summarizes the fluctuation of case numbers in Surabaya, providing general insights from the COVID-19 dataset within the specified time frame and particular sub-areas for monitoring (Figure 6).
- Page 2: Distribution of Cases - Illustrates sub-areas with the highest case numbers, periodic case numbers, and trends (Figure 7).
- Page 3: Impact of COVID-19 Variants on the Case Growth - Presents details on the proportion of each variant (delta and omicron) in each area and sub-area. It delineates the impact of these variants on fluctuations in case numbers, deaths/mortality, and recoveries (Figure 8).
- Page 4: Recommendations for School Openings - Displays recommendations based on the increasing and decreasing COVID-19 cases (Figure 9).

The prediction model of RNN-LSTM we proposed and the fine-tuning hyperparameter reveal discernible patterns that contribute to decisions about the safe reopening of schools. Schools situated in areas anticipated to be vulnerable will remain closed. However, if the situation is deemed controllable, schools in those areas will be safe to reopen amidst the ongoing pandemic.

Notably, we found that the number of schools opened shows a negative correlation with the growth in case numbers; whenever cases increase, the number of schools opened decreases. Conversely, a substantial decrease in case numbers permits the opening of more schools.

Our developed dashboard incorporates various components, such as geographical maps, search functionalities, and data visualization through tables, charts, and graphs. Area, Sub-Area, and Date Range determine the filtering options within our dashboard. A crucial aspect of our dashboard system is the recommendations provided by the prediction model regarding school openings and closures. These recommendations dynamically adjust based on anticipated observation time and the specific area or sub-area under observation.

Our experimental model intends to be valuable and offers novel perspectives for extended studies, not only in the context of the recent pandemic but also for future outbreaks or situations affecting health systems at local, regional, or global scales. This model's applicability extends beyond the education sector, reaching various industries. By leveraging advanced data processing techniques, error identification, and optimized deep learning models, we aim to enhance our approach towards an expert system. This is achieved by developing a dashboard system that simplifies monitoring through visual information and user-friendly features, making it accessible to a wide range of users.

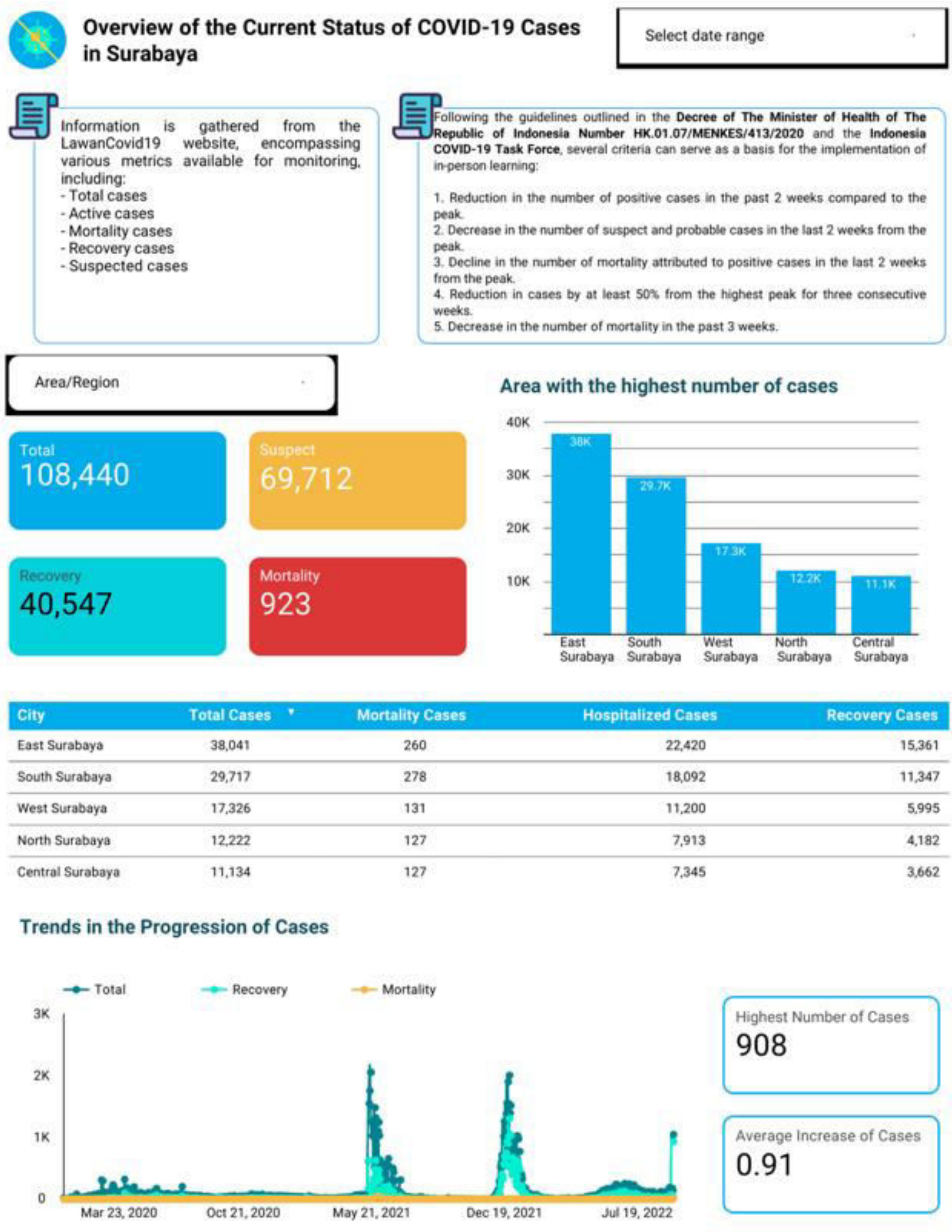


Figure 6. Dashboard Data Visualization: Page 1 – Overview

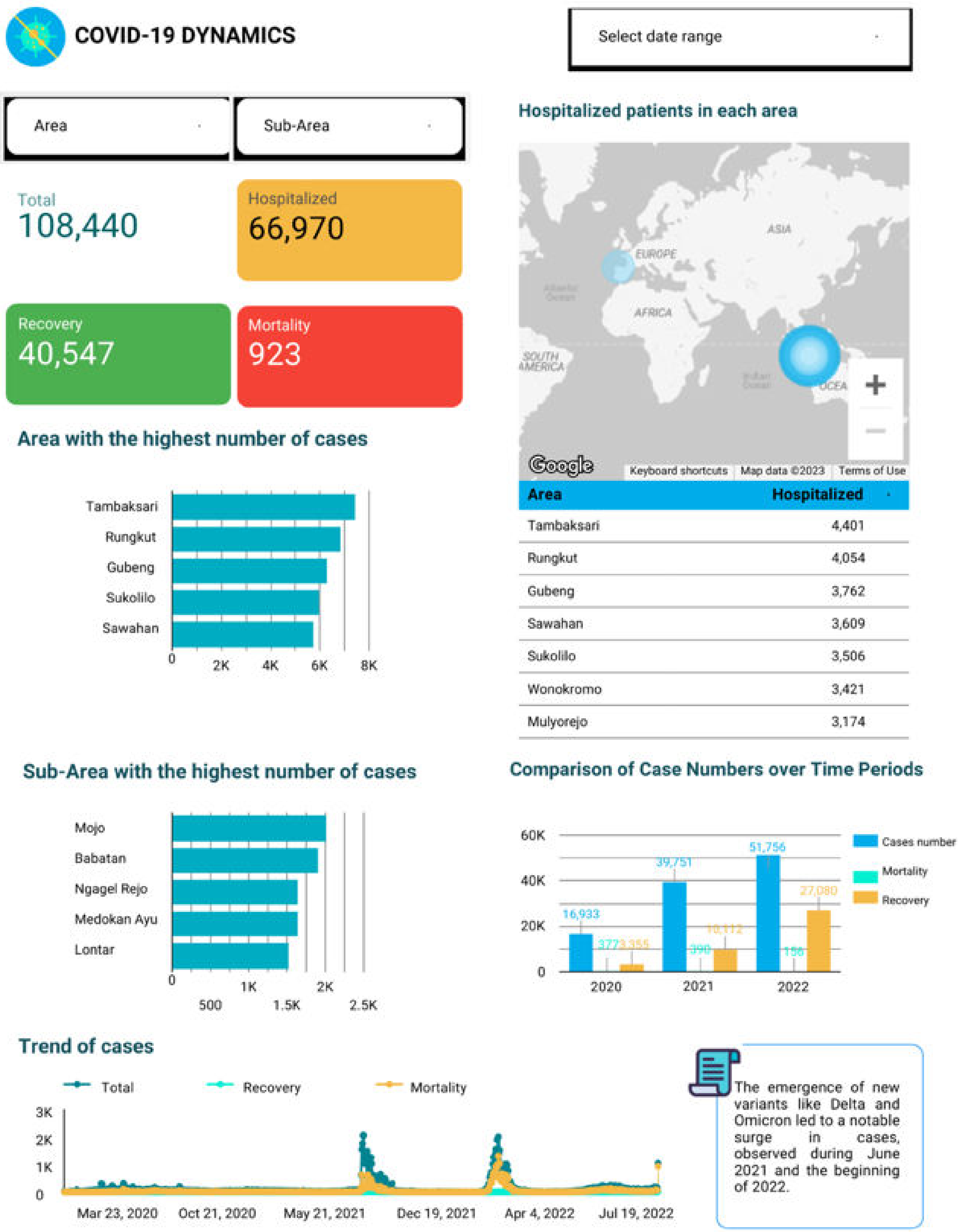


Figure 7. Dashboard Data Visualization: Page 2 – Distribution of Cases

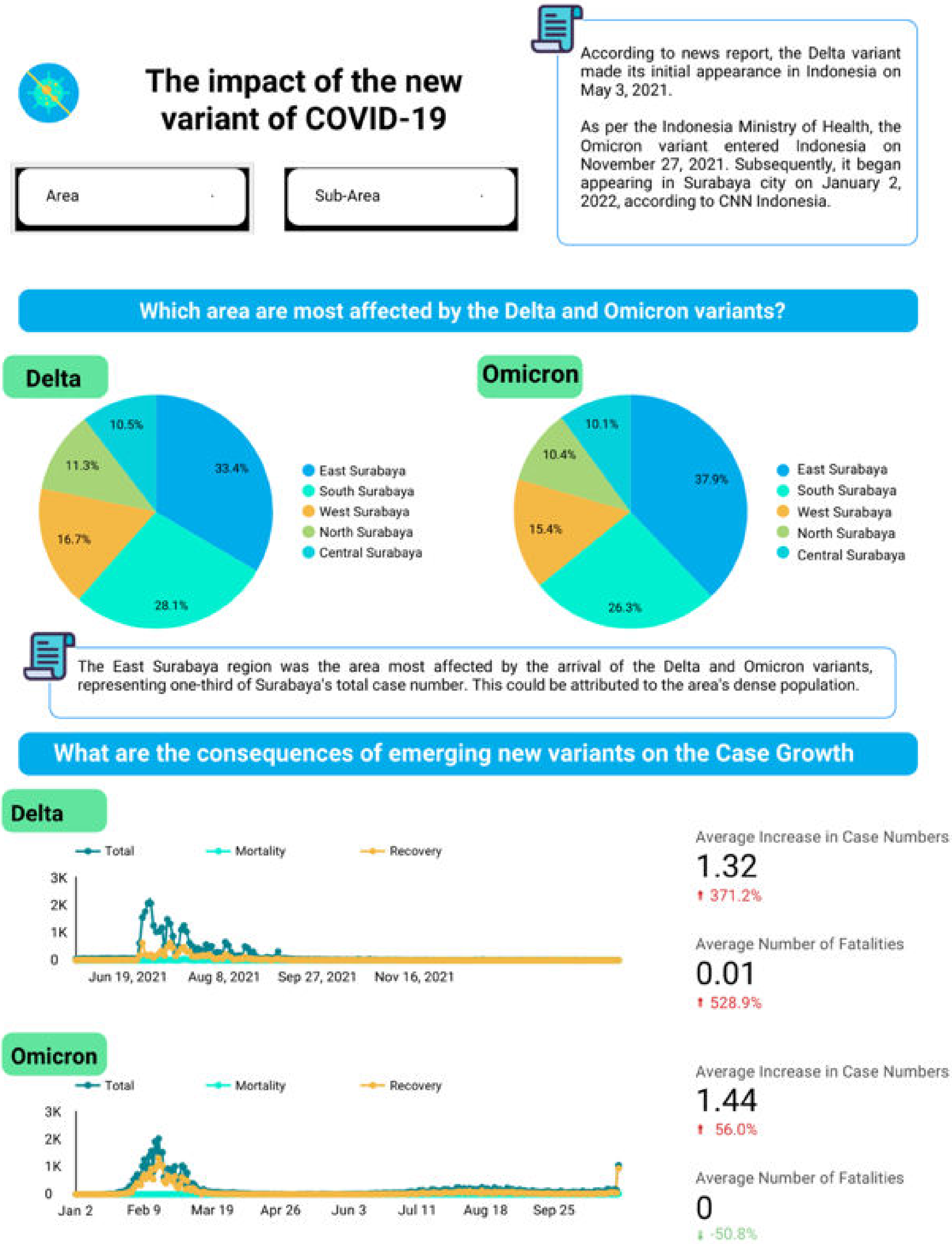


Figure 8. Dashboard Data Visualization: Page 3 – Impact of COVID-19 Variants on the Case Growth

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 1, March 2024

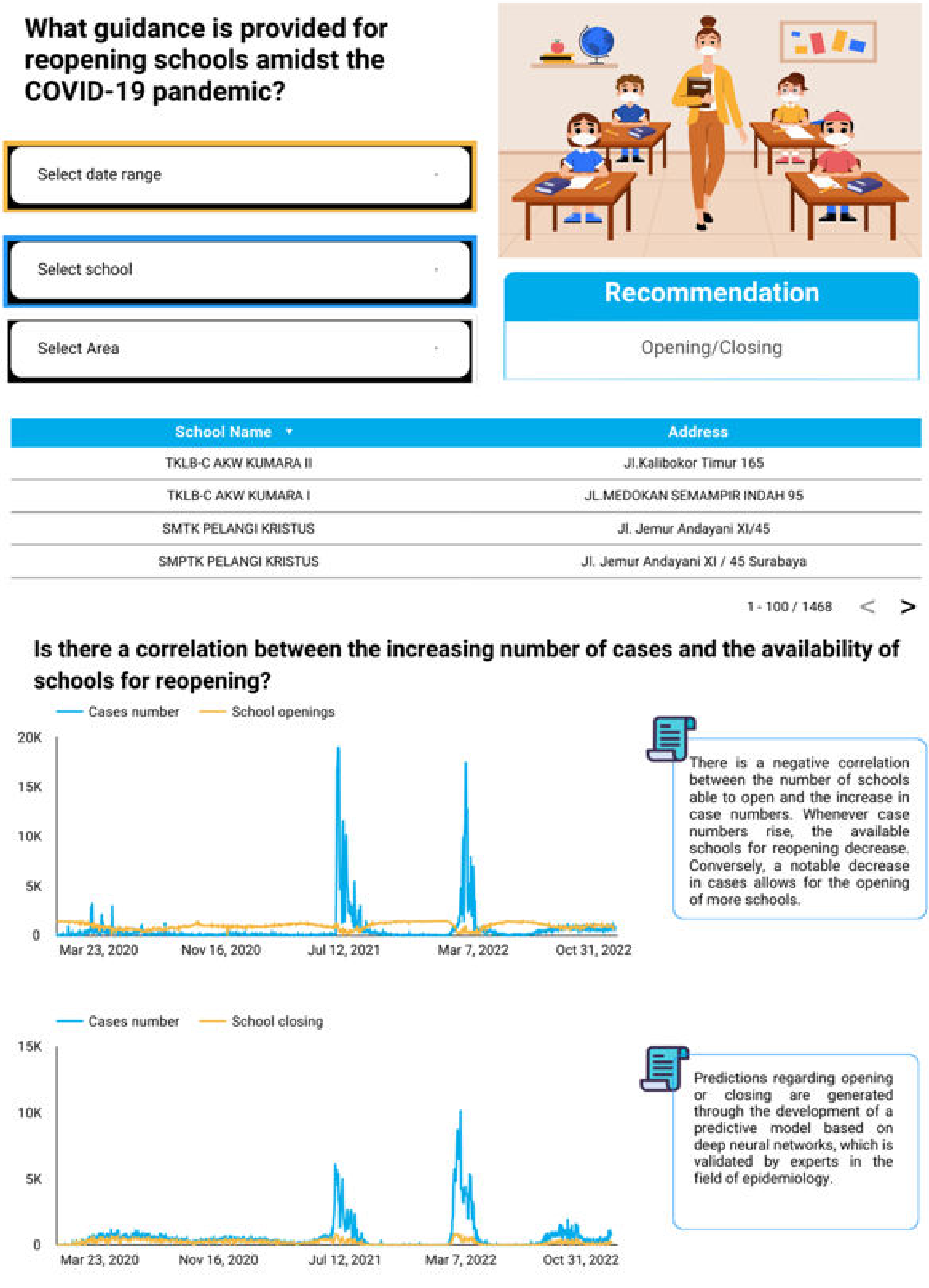


Figure 10. Dashboard Data Visualization: Page 4 – Recommendations for School Openings

CONCLUSION

Our study successfully integrates COVID-19 datasets with school information in Surabaya City, Indonesia, utilizing the RNN-LSTM algorithm to predict COVID-19 transmission dynamics in areas surrounding schools across 154 sub-areas.

School opening and closing conditions are determined by categorizing COVID-19 dynamics into controllable and vulnerable situations. As shown in Table 4, each school in each sub-area will be classified as 1 (opening) or 0 (closing) using RNN-LSTM modeling. RNN-LSTM modeling offers insights into identifying controlled conditions. The pandemic was influenced by reduced suspects, decreased mortality, and declining trends in confirmed cases, signifying readiness for school reopening. Conversely, vulnerable conditions warn that schools cannot be opened initially as the COVID-19 spread continues to increase.

Through hyperparameter refinement, we achieved a notable 5.28% increase in accuracy (97.32%), enhancing the model's ability to predict COVID-19 trends. The developed dashboard provides dynamic visualization capabilities, facilitating informed decision-making regarding safe school reopening strategies during the pandemic by presenting comprehensive data on COVID-19 trends alongside relevant school information. This contribution highlights the importance of data-driven approaches in addressing the multifaceted impacts of the pandemic on the education sector, aiming to mitigate disruptions and safeguard educational stability amidst unprecedented challenges.

We believe in the effectiveness of our developed approach employing RNN-LSTM prediction modeling. This method, meticulously fine-tuned with numerous parameters and presented through a dashboard interface, can be valuable for guiding decision-making in unforeseen or potentially life-threatening scenarios arising from virus transmission. Not only do pandemics exist, but endemic conditions, epidemics, and other hazardous conditions are considered.

In future work, continuous refinement of the predictive model and expansion of analysis to include additional factors influencing COVID-19 transmission dynamics and school reopening decisions can further enhance the dashboard's utility and relevance. Further developments to the dashboard system will also focus on the interaction design paradigm to secure the dashboard’s sustainability and benefits to the public, not only in the present but also in the future. Furthermore, the methodologies and insights gained from our study can be generalized to various risky sectors to contribute to national resilience and effective decision-making in response to outbreak crises.

ACKNOWLEDGMENT

The Indonesian government has funded this research through the LPDP Scholarship (Lembaga Pengelola Dana Pendidikan). A partnership between Universitas Airlangga, Indonesia, and Kumamoto University, Japan, also supports this work.

REFERENCES

[1] H. Wang *et al.*, “Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21,” *Lancet*, vol. 399, no. 10334, pp. 1513–1536, 2022, doi: 10.1016/S0140-6736(21)02796-3.

[2] M. F. Aslan, M. F. Unlersen, K. Sabanci, and A. Durdu, “CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection,” *Appl. Soft Comput.*, vol. 98, p. 106912, 2021, doi: 10.1016/j.asoc.2020.106912.

[3] Mohammad H. Tayarani Najaran, “Applications of artificial intelligence in battling against covid-19: A literature review,” *Chaos, Solitons and Fractals*, vol. 142, no. January, pp. 1–31, 2020.

[4] S. Dutta and S. K. Bandyopadhyay, “Machine learning approach for confirmation of COVID-19 cases: positive, negative, death and release,” *Iberoam. J. Med.*, vol. 2, no. 3, pp. 172–177, 2020, doi: 10.53986/ibjm.2020.0031.

[5] C. C. John, V. Ponnusamy, S. Krishnan Chandrasekaran, and N. R, “A Survey on Mathematical, Machine Learning and Deep Learning Models for COVID-19 Transmission and Diagnosis,” *IEEE Rev. Biomed. Eng.*, vol. 14, no. 8, 2021, doi: 10.1109/RBME.2021.3069213.

[6] J. Hopkins, “COVID-19 Dashboard,” *Corona Virus Resource Center*, 2022. https://coronavirus.jhu.edu/map.html (accessed Feb. 07, 2022).

[7] D. for G. (D4G) Team, “Interactive COVID-19 Household Monitoring Dashboard,” *The World Bank*, 2021. https://www.worldbank.org/en/data/interactive/2020/11/11/covid-19-high-frequency-monitoring-dashboard (accessed Jan. 05, 2022).

[8] WHO, "WHO Coronavirus (COVID-19) Dashboard," *World Health Organization*, 2020. https://covid19.who.int/ (accessed Oct. 20, 2021).

[9] J. Allen *et al.*, "Coronavirus in the U.S.: Latest Map and Case Count," *New York Times*, 2020. https://www.nytimes.com/interactive/2021/us/covid-cases.html (accessed Nov. 02, 2021).

[10] R. Hao *et al.*, "Control strategies and their effects on the COVID-19 pandemic in 2020 in representative countries," *J. Biosaf. Biosecurity*, vol. 3, no. 2, pp. 76–81, 2021, doi: 10.1016/j.jobb.2021.06.003.

[11] H. Patrinos, E. Vegas, and R. Carter-Rau, "An Analysis of COVID-19 Student Learning Loss," *Oxford Res. Encycl. Econ. Financ.*, no. May, 2023, doi: 10.1093/acrefore/9780190625979.013.893.

[12] R. Donnelly and H. A. Patrinos, "Learning loss during Covid-19: An early systematic review," *Prospects*, vol. 51, no. 4, pp. 601–609, 2022, doi: 10.1007/s11125-021-09582-6.

[13] ADB, "Learning and Earning Losses from COVID-19 School Closures in Developing Asia," *Asian Dev. Outlook*, no. April, pp. 1–25, 2021.

[14] N. Reuge *et al.*, "Education response to COVID 19 pandemic, a special issue proposed by UNICEF: Editorial review," *Int. J. Educ. Dev.*, vol. 87, p. 102485, 2021, doi: 10.1016/j.ijedudev.2021.102485.

[15] Nasir Mustafa, "Impact of the 2019 – 20 coronavirus pandemic on education," *Int. J. Heal. Prefer. Res.*, pp. 1–36, 2020.

[16] United Nations, "Policy Brief: Education during COVID-19 and beyond," 2020. [Online]. Available: https://reliefweb.int/report/world/policy-brief-education-during-covid-19-and-beyond-august-2020.

[17] UNESCO, "COVID-19 Education Response: Preparing the reopening of schools," no. May, pp. 1–29, 2020, [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000373401.

[18] N. Butcher, N. N. Khairina, C. Kumala, and S. Loots, "The Struggle Against COVID-19 in Indonesian Education: Responses, requirements, and policy needs for learning recovery," 2021. [Online]. Available: www.worldbank.org.

[19] UNICEF, "Indonesia Case Study: Situation Analysis on the Effects of and Responses to COVID-19 on the Education Sector in Asia," 2021. doi: 10.1007/978-1-4615-5667-1_5.

[20] D. Kekić and S. Miladinović, "Functioning of Educational System During an Outbreak of Acute Infectious Diseases," no. November, 2016.

[21] K. Meesters, "The Challenge and Value of Dashboard Development During the COVID-19 Pandemic," in *IEEE Global Humanitarian Technology Conference (GHTC) 2022*, 2022, pp. 9–16.

[22] J. Everts, "The dashboard pandemic," *Dialogues Hum. Geogr.*, vol. 10, no. 2, pp. 260–264, 2020, doi: 10.1177/2043820620935355.

[23] UNICEF Indonesia, "The Impact of the COVID-19 Pandemic on Children ' s Learning in Indonesia," 2021.

[24] UNESCO, UNICEF, World Bank, and World Food Programme, "Framework for reopening schools," p. 5, 2020, [Online]. Available: https://www.unicef.org/media/68366/file/Framework-for-reopening-schools-2020.pdf.

[25] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020, doi: 10.1016/j.physd.2019.132306.

[26] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, pp. 3285–3292, 2019, doi: 10.1109/BigData47090.2019.9005997.

[27] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, 2018, doi: 10.3390/en11071636.

[28] S. Sanders and C. Giraud-carrier, "Informing the Use of Hyperparameter Optimization Through Metalearning," in *2017 IEEE International Conference on Data Mining*, 2017, no. 1, pp. 1051–1056, doi: 10.1109/ICDM.2017.137.

[29] T. Hinz, S. Magg, and S. Wermter, "Speeding up the Hyperparameter Optimization of Deep Convolutional Neural Networks," *Int. J. Comput. Intell.*

Appl., vol. 17, no. 2, pp. 1850008-1-1850008–15, 2018, doi: 10.1142/S1469026818500086.

[30] T. N. Fatyanosa and M. Aritsugi, “Effects of the Number of Hyperparameters on the Performance of GA-CNN,” *Proc. - 2020 IEEE/ACM Int. Conf. Big Data Comput. Appl. Technol. BDCAT 2020*, pp. 144–153, 2020, doi: 10.1109/BDCAT50828.2020.00016.

[31] F. Hutter, L. Kotthoff, and J. Vanschoren, *Automated Machine Learning. Methods, Systems, Challenges*. 2020.

[32] A. Sinha and M. Rathi, “COVID-19 prediction using AI analytics for South Korea,” *Appl. Intell.*, 2021, doi: 10.1007/s10489-021-02352-z.

[33] O. Castillo and P. Melin, “Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic,” *Chaos, Solitons and Fractals*, vol. 140, 2020, doi: 10.1016/j.chaos.2020.110242.

[34] A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, “Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study,” *Chaos, Solitons and Fractals*, vol. 140, no. January, pp. 1–12, 2020.

[35] F. W. Wibowo and Wihayati, “Prediction Modelling of COVID-19 on Provinces in Indonesia using Long Short-Term Memory Machine Learning,” *J. Phys. Conf. Ser.*, vol. 1844, no. 1, 2021, doi: 10.1088/1742-6596/1844/1/012006.

[36] CDC, “COVID Data Tracker,” *Department of Health and Human Services*, 2021. https://covid.cdc.gov/covid-data-tracker (accessed Mar. 31, 2023).

[37] K. DEVAKUMAR, “COVID-19 Dataset,” *Kaggle*, 2021. https://www.kaggle.com/datasets/imdevskp/corona-virus-report (accessed Apr. 10, 2023).

[38] C.-19 T. F. Indonesia, “Analysis of Recommendations Towards a Productive and Safe to COVID-19 in Indonesia,” 2020.

[39] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[40] F. A. Muqtadiroh, T. Usagawa, R. D. Rachmayanti, S. M. S. Nugroho, E. M. Yuniarno, and M. H. Purnomo, “Rules Determination Based on Time-Series Data to Classify Unsupervised Cases Based on Fuzzy Expert System,” *Int. J. Intell. Eng. Syst.*, vol. 16, no. 3, pp. 258–268, 2023, doi: 10.22266/ijies2023.0630.20.