



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

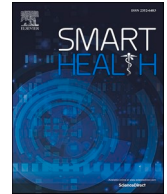
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Smart Health

journal homepage: www.elsevier.com/locate/smhl

Impact of COVID-19 on city-scale transportation and safety: An early experience from Detroit

Yongtao Yao^{a,*}, Tony G. Geara^b, Weisong Shi^a

^a Department of Computer Science, Wayne State University, Detroit, MI, 48202, USA

^b Department of Public Works, City of Detroit, Detroit, MI, 48216, USA

ARTICLE INFO

Keywords:

COVID-19
Data analysis
Prediction
Quarantine
Transportation networks
Traffic volume crashes
Social distancing weather
Daily cases Detroit

ABSTRACT

The COVID-19 pandemic brought unprecedented levels of disruption to the local and regional transportation networks throughout the United States, especially the Motor City—Detroit. That was mainly a result of swift restrictive measures such as statewide quarantine and lock-down orders to confine the spread of the virus and the rising number of COVID-19 confirmed cases and deaths. This work is driven by analyzing five types of real-world data sets from Detroit related to traffic volume, daily cases, weather, social distancing index, and crashes from January 2019 to June 2020. The primary goals of this work are: i) figuring out the impacts of COVID-19 on the transportation network usage (traffic volume) and safety (crashes) for the City of Detroit, ii) determining whether each type of data (e.g. traffic volume data) could be a useful factor in the confirmed-cases prediction, and iii) providing an early future prediction method for COVID-19 rates, which can be a vital contributor to life-saving advanced preventative and preparatory responses.

In addressing these problems, the prediction results of six feature groups are presented and analyzed to quantify the prediction effectiveness of each type of data. Then, a deep learning model was developed using long short-term memory networks to predict the number of confirmed cases within the next week. The model demonstrated a promising prediction result with a coefficient of determination (R^2) of up to approximately 0.91. Furthermore, six essential observations with supporting evidence are presented, which will be helpful for decision-makers to take specific measures that aid in preventing the spread of COVID-19 and protecting public health and safety. The proposed approaches could be applied, customized, adjusted, and replicated for analysis of the impact of COVID-19 on a transportation network and prediction of the anticipated COVID-19 cases using a similar data set obtained for other large cities in the USA or from around the world.

1. Introduction

The 2019 Novel Coronavirus (SARS-CoV-2), commonly known as COVID-19, has spread rapidly across the globe. As of December 7, 2020, over 66 million confirmed cases and 1.5 million deaths had been reported worldwide (WHO, 2020); meanwhile, the United States is one of the most affected nations in the world with around 15 million confirmed cases and 282 thousand deaths. This tragic spread of COVID-19 nationally has resulted in disparate impacts across states and cities. To slow the progression of COVID-19 and limit

* Corresponding author.

E-mail address: yongtaoyao@wayne.edu (Y. Yao).

<https://doi.org/10.1016/j.smhl.2021.100218>

Received 7 December 2020; Received in revised form 12 May 2021; Accepted 12 May 2021

Available online 14 September 2021

fatalities, public officials throughout Michigan had published a series of government directives that have been changed over time, starting with voluntary requests for stay-at-home and restrictions on large public gatherings, and then, statewide quarantine and lock-down orders. Nonetheless, essential travel activities continue to take place across Detroit, such as people's access to daily supplies, medical services, and other basic necessities of welfare and safety. These government directives inevitably affect various forms of travel activities and then impact transportation across Detroit significantly (Hu et al., 2020).

Traffic volume. Our work is based on the hypothesis that the objective, reliable, and continuous transportation data can reflect the degree of social distancing, i.e., the possibility of social activities and interpersonal communication to a certain extent, while many previous works provide the evidence proving that the social distancing measures enacted have led to control of COVID-19 (Ainslie et al., 2020; Briscese et al., 2020; Courtemanche et al., 2020). Therefore, we believe that traffic data can provide a basis for the current and incoming pandemic status, and it is meaningful to explore the changes in traffic patterns during the COVID-19 pandemic for a specific city.

Crashes. In addition, crash-related information, such as the total number of daily crashes, severity, and crash type can indirectly reflect traffic conditions (Abdel-Aty & Abdelwahab, 2004; Candefjord et al., 2016; Stutts et al., 2001). As COVID-19 has disrupted the way of living and hence the commuting habits of citizens, it is clear that changes in vehicle utilization might determine a change in the crashes' types and number. Since one of the focus points of this work is to explore the correlation between traffic volume data and the outbreak of COVID-19, we also collected and analyzed crash data from Detroit to identify the impacts of crash-related information on the confirmed-cases prediction.

Weather. Besides, through our literature review, an abundance of studies pointed out that weather factors, e.g., temperature (°C) and wind speed (mph) can contribute to the spread of COVID-19 (Gupta et al., 2020; Sahin, 2020; Tosepu et al., 2020). Inspired by these works, we also sought to determine whether weather could be a factor in the spread of this disease.

Social distancing index. Moreover, it is well known that maintaining social distancing can prevent the spread of COVID-19 disease and contain the number of casualties (Chen et al., 2020; Lewnard & Lo, 2020; Mohler et al., 2020; Olivera-La Rosa et al., 2020; Painter & Qiu, 2020; Singh & Adhikari, 2020), which is based on the assumption that the degree of social distancing is highly related to the spread speed of COVID-19. Therefore, we also collected social distancing-related data for Wayne County and Michigan State, to explore and test the correlation between social distancing and the severity of COVID-19 disease.

By considering the aforementioned data features, we then built an effective deep learning model using long short-term memory networks (LSTM) to predict the number of confirmed cases in Detroit. In order to provide statistical evaluation measures to quantify the prediction effectiveness of each type of data on the confirmed-cases prediction results, i.e., the performance of LSTM, we trained LSTM on six experiment groups with different features, then analyzed the prediction results of the six feature groups.

Our observations and prediction model are intended to help decision-makers to concentrate on suitable public health efforts and apply effective transportation management techniques to protect residents and improve safety for Detroiters. It must be noted that the presented statistical analysis approaches and the proposed prediction model were used on the Detroit-based data set as an example. The method could be applied, customized, adjusted, and replicated for analysis of the impact of COVID-19 on a transportation network and prediction of the anticipated COVID-19 cases using a similar data set obtained from other large cities from within the USA or from around the world.

Particularly, this paper sets out to answer the following questions:

- i. What are the sudden and drastic changes in overall temporal traffic patterns resulting from the outbreak of COVID-19?
- ii. Did traffic decrease and then recover evenly across all measured signals during COVID-19 as COVID-19 restrictions were being lifted?
- iii. What are the impacts of COVID-19 and the social distancing on the reasons behind crashes?
- iv. Can we leverage the traffic volume data, crash data, and other COVID-19 related information, such as the data on daily confirmed cases plus social distancing and combined with weather information, to predict the number of COVID-19 confirmed cases for the next week?

The rest of the paper is organized as follows. We first review previous works and introduce the research gap in Sec.2. Then, we elaborate on the data sets used for the experiments in Sec. 3. Sec. 4 describes experimental investigations for the questions presented in i) to iii), and Sec. 5 demonstrates the proposed algorithms to predict the number of COVID-19 confirmed cases as described in question iv). After presenting a discussion of the prediction results in Sec. 6, we conclude the work in Sec.7.

2. Related work

In this section, we review recently published works that focus on the COVID-19 confirmed-cases prediction or similar research triggered by the outbreak of COVID-19 in terms of transportation, weather, social distancing, and other aspects. To the best of our knowledge, prior works do not consider the effects of *traffic volume data* and *crash-related information* on the research of COVID-19 forecasts.

2.1. Transportation and COVID-19

The travel restrictions put in place to reduce the spread of COVID-19 resulted in a sharp reduction in traffic throughout the United States. Some recent works explored the changes in the transportation mode. For example, in the work of (Hu et al., 2020). Hu et al.

Table 1
Analyzed metrics summary.

Data Category	#	Metrics	Definition
Traffic Volume	10	Bus, MotorizedVehicle, PickupTruck, ArticulatedTruck, SingleUnitTruck, Pedestrian, Motorcycle, Car, WorkVan, and Bicycle.	Traffic volume by classification: number of each transportation mode per day.
Daily Case	2	Confirmed cases, Confirmed death	Number of confirmed cases and death per day.
Weather Data	6	Rain precipitation, Snow precipitation Average temperature, Maximum temperature, Mini mum temperature Wind speed	Volume of precipitation of rain and snow per day. Value of average/maximum/minimum temperature per day. Average Speed of wind per day.
Crash	19	Total crashes Fatal, Serious, Minor, Possible, None Ped, Cyclist, YoungDriver(Under 24) Single motor vehicle, Head on, Head on left turn, Angle, Rear end, Rearend right turn, Sideswipe same, Sideswipe opposite, Backing, Other, UnknownNull or not entered	Number of total crashes per day. Severity of crashes (worst injury). Other reason for crash (Crashes with non-motorized roadway users and young drivers under 24 years of age). Type of crashes.
Social Distancing Related Data	21	Social distancing index % staying home Trips/person % out-of-county trips % out-of-state trips Miles/person Work trips/person Non-work trips/person New COVID cases % change in consumption COVID exposure/1000 people Unemployment claims/1000 people Unemployment rate % working from home COVID death rate New cases/1000 people Active cases/1000 people #days: decreasing COVID cases Testing capacity Tests done/1000 people % ICU utilizationImported COVID cases	An integer from 0*100. The higher value indicates a higher level social distancing. Percentage of residents staying at home. Average number of all trips taken per person per day. Percentage of all trips that cross county borders. Percentage of all trips that cross state borders. Average person-miles traveled on all modes per person per day. Number of work trips per person per day. Number of non-work trips per person per day. Number of COVID-19 daily new cases. % change in consumption from the pre-pandemic baseline. Number of residents already exposed to coronavirus per 1000 people. New weekly unemployment insurance claims/ 1000 workers. Unemployment rate updated weekly. Percentage of workforce working from home. % deaths among all COVID-19 cases. Number of COVID-19 daily new cases per 1000 people. Number of active COVID-19 cases per 1000 people. Number of days with decreasing COVID-19 cases. Ability to provide enough tests. Number of COVID-19 tests already completed per 1000 people. % ICU unites occupied with COVID-19 patients.

studied transportation modes during and after the COVID-19 pandemic using basic laws of traffic and mathematical analysis to explore scenarios of increased car commuting. Lacus et al. (Lacus et al., 2020) analyzed data on air traffic worldwide with the scope of analyzing the impact of the travel ban on the aviation sector as well as after changes in COVID-19 diagnostic criteria. Lau et al. (Lau et al., 2020) calculated the correlation of air traffic to the number of confirmed COVID-19 cases and determined the growth curves of cases before and after lock-down. Teixeira et al. (Teixeira & Lopes, 2020) presented clues on how bike-sharing can support the transition to a post-COVID-19 society.

2.2. Weather and COVID-19

Moreover, the weather factors including temperature (°C), humidity (%), wind speed (mph) are regarded as the factors that triggered the spread of COVID-19 in recent works. For example, through the Spearman-rank correlation test, Tosepu et al. (Tosepu et al., 2020) proved that among the minimum temperature, maximum temperature, average temperature, humidity, and rainfall, only the average temperature is significantly related to the COVID-19 pandemic in Jakarta Indonesia. Another experiment (Gupta et al., 2020) was conducted based on the daily new cases and weather information in 50 states in the United States, and clarify those weather parameters, i.e., temperature and absolute humidity will help classify the risky geographic areas in different countries. Besides, based on Spearman’s correlation coefficients, Mehmet et al. (Sahin, 2020) pointed out the highest correlations between wind speed (mph) with the outbreak of COVID-19. In addition, a few studies have claimed that warm weather can possibly slow down the global

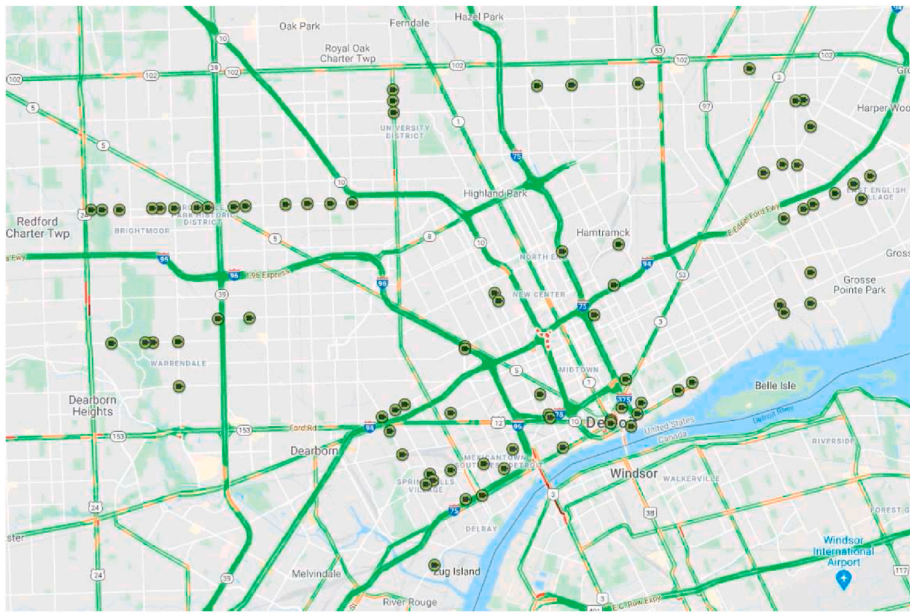


Fig. 1. The distribution of 73 intersections in Detroit.

pandemic of COVID-19 by considering nine cities in Turkey (Sahin, 2020).

2.3. Social distancing and COVID-19

In addition, a portion of the related work studied the relationship between social distancing and COVID-19. For instance, Singh et al. (Singh & Adhikari, 2020) focused on the age-structured impact of social distancing on the COVID-19 epidemic in India and presented a mathematical model of the spread of infection in a population that structured by age and social contact between ages. Courtemanche et al. (Courtemanche et al., 2020) pointed out that there would have been ten times greater spread of COVID-19 without shelter-in-place orders. The work of (Mohler et al., 2020) presented that social distancing and shelter-in-place has had some impact on crime and disorder, but only for a restricted collection of crime types and not consistently across places.

2.4. Other coronavirus-related research

To date, more and more researchers are aggressively shifting their focus to detect, predict, treat, and recover from COVID-19. Although the prior works presented promising results to address the coronavirus-related issue, they do not consider transportation, weather, or social distancing at the same time. For example, Qin et al. (Qin et al., 2020) proposed a novel model to predict the outbreak of COVID-19 in populations in affected areas based on the social media search indexes (SMSI) for dry cough, fever, chest distress, coronavirus, and pneumonia. Lacus et al. (Lacus et al., 2020) calculated the economic impact measured in terms of loss of GDP due to the aviation sector as well as the social impact due to job losses related to aviation and correlated sectors such as tourism and catering.

3. Critical dates and data description

In this section, we discuss all four types of data collected and analyzed for this study (shown in Table 1): (i) traffic volume data, (ii) daily cases number including daily confirmed cases and daily death number, (iii) weather information, (iv) social distancing-related data, and (v) crash data. For each category of data, we then elaborate on the corresponding data source and the embodied attributes, respectively.

Timeline of shutdown and data collection period. Knowing these specific dates are important for researchers to explore the changes in overall traffic volume patterns across Detroit pertaining to COVID-19 and the impact of these four types of data on the prediction of the confirmed cases. Note that in Detroit, the onset of COVID-19 was estimated to be on March 1st, the shutdown started on March 19th, and people started going back to work in the office on June 1st. However, the number of closed businesses in the downtown area and the number of people working from home, in general, continued to remain high.

Thus, we defined the shutdown period as going into effect from the 19th of March to the 1st of June 2020. We collected and analyzed transportation data from Detroit covering three time periods: i) before the first COVID-19 confirmed cases, ii) during the pandemic, and iii) after the release of the shutdown. In a nutshell, we collected and analyzed transportation data from 1/1/2019 to 6/30/2020—covering more than one and a half years. To keep data consistent, we also analyzed daily cases number, weather information, and social index data that spans the same time period.

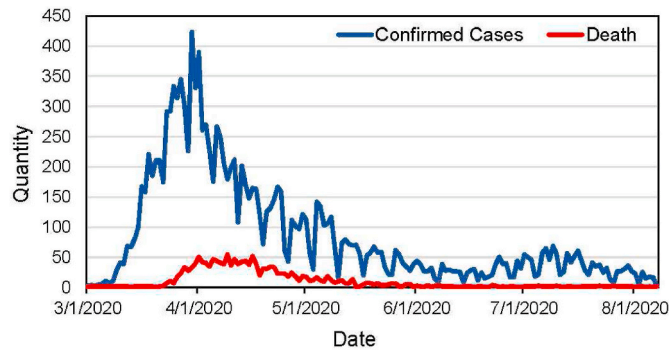


Fig. 2. The number of daily confirmed cases and death in Detroit.

3.1. Traffic volume data

More specifically, we collected and analyzed transportation data from 73 signalized intersection sites with advanced Remote Traffic Signal Management System (RTSMS) Level-II. Those locations have continuous data collection and analytics metrics through camera detection. They are all owned by the City of Detroit out of a total of 787 City-owned signals and another approximately 700 signals owned by other jurisdictions around the city. These signalized intersections are indispensable parts of urban traffic networks since around two-thirds of urban vehicle miles traveled on signal-controlled roads (McCracken, 1996).

Fig. 1 depicts the geographical distribution of the studied 73 signalized intersections, which are highlighted by the green dots. Those locations provided aggregated daily traffic volume data for the following 10 attributes—Bus, MotorizedVehicle, PickupTruck, ArticulatedTruck, SingleUnitTruck, Pedestrian, Motorcycle, Car, WorkVan, and Bicycle. Those attributes were later compounded into 6 as described in future sections. Additionally, it must be noted that in 2019, the number of Level-II RTSMS locations was limited to 25. However, to account for that discrepancy, the normalized average volume per intersection was used in this work.

3.2. Weather data

Since previous work research pointed out that weather factors, e.g., temperature (°C) and the wind speed (mph) might be a contributor to the spread of COVID-19, we also included weather information as one of the inputs for COVID-19 prediction. We collected weather data from the official website of the National Oceanic and Atmospheric Administration¹ and analyzed six weather-related attributes including i) rain precipitation, ii) snow precipitation, iii) average temperature, iv) maximum temperature, v) minimum temperature, and vi) average wind speed.

3.3. Daily cases

As to the daily cases data, we obtained and analyzed the daily cases data from Michigan's official Coronavirus dashboard,² including (i) the number of confirmed cases, and (ii) the number of reported deaths (shown in Fig. 2). More specifically, as to the number of confirmed cases, the number refers to the disease onset date; otherwise, either the specimen collection date of the first positive COVID-19 test or referral date is used. As for the number of reported deaths, the corresponding value represents the actual reported date of death, and 8 confirmed deaths did not have a valid date available and are not included in the collected data.

3.4. Social distancing information

We also collected and analyzed the social distancing-related attributes for Wayne County and Michigan State from the COVID-19 Impact Analysis Platform³ published by the University of Maryland. This data was not granular enough to account for specifically the City's boundaries, however, it included valuable data for the various input contributing to social distancing on both the County and State levels.

In particular, the COVID-19 Impact Analysis Platform defined and calculated the social distancing index from the six mobility indicators by the following equation: $\text{social distancing index} = 0.8 \times [\% \text{staying home} + 0.01 \times (100 - \% \text{staying home}) \times (0.1 \times \% \text{reduction of all trips compared with pre-COVID-19 benchmark} + 0.2 \times \% \text{reduction of work trips} + 0.4 \times \% \text{reduction of non-work trips} + 0.3 \times \% \text{reduction of travel distance})] + 0.2 \times \% \text{reduction of out-of-county trips}$. Here, the social distancing index can be regarded as a weighted summation of the percentages of staying home, reduction of all trips (such as work trips and out-of-country

¹ <https://www.usa.gov/federal-agencies/national-weather-service>.

² https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173--,00.html.

³ <https://data.covid.umd.edu/>.

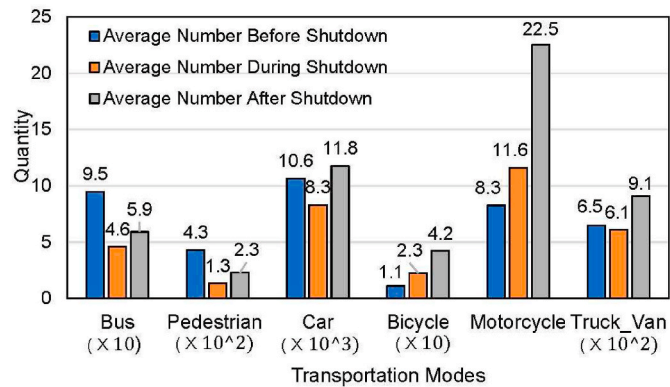


Fig. 3. Changes in the daily number of each transportation model before, during, and after quarantine in Detroit.

trips), and the reduction of travel distance. A higher social distance index score indicates fewer chances for close interpersonal interaction and reduced opportunities for the transmission of COVID-19. The weights are chosen based on i) the travel ratio of residents and tourists (e.g., about 20% of all trips are trips outside the county, leading to a choice of 0.8 for residents and 0.2 for out-of-county travel), and ii) what trips are considered more important (e.g., work travel is more important than non-work travel).}

3.5. Crash data

Next, to figure out the impact of COVID-19 and social distancing metrics on the rate and severity of crashes, we also collected 19 crash metrics for Detroit from the Michigan State Police (MSP) Traffic Crash Reporting System (TCRS),⁴ which contains the information related to the number of total crashes per day broken down by severity, other reasons, and types of crashes. Table 1 lists all collected and analyzed metrics in this study. The monitoring frequencies for all metrics are both once per day.

4. Statistical analysis

In this section, we conduct a statistical analysis of the collected metrics and present interesting observations related to the changes in traffic volume and pattern, correlation studies, and crashes statistics analysis.

4.1. Changes of traffic volume and pattern

4.1.1. Changes of traffic volume in 2020

We first sought to answer the question of “What’s the sudden spatial traffic patterns pertaining to COVID-19”. To answer this question, we first calculate the average number of each transportation mode among 73 signalized intersections for each monitoring day, then we further draw the average number of each mode during the period before, during, and after quarantine in Detroit (shown in Fig. 3). Out of the 10 collected metrics related to traffic counts at signalized intersections (shown in Table 1). We combined some of those metrics and consolidated the classifications down to six categories. For example, the value of “TruckVan” represents the total number of all recorded types of trucks and vans, while the value of “Car” indicates the total quantity of motorized vehicles and traditional cars.

It can be seen that the number of buses, pedestrians, and cars both showed declining trends during the shutdown period and then increased after the ending of the shutdown. On the contrary, it is notable that the quantities of Bicycles and Motorcycles increased sharply even after the statewide quarantine order. For example, the average number of Bicycles per day increased to approximately $2\times$ its original value and even increase by approximately $4\times$ during and after the quarantine, respectively. Similarly, the average number of Motorcycle surged to approximately $3\times$ of the number before shutdown numbers.

Observation 1. Although cars are still the predominant transportation mode, biking and motorcycling have demonstrated a transition in usage and are showing increased popularity (up to $4\times$ of the previous volume) in the post-COVID-19 urban mobility for Detroit.

These numbers do not account for the seasonal weather impact on Bike and Motorcycle usage.

This observation may be attributed to the fact that cycling can be an alternative mode of transport as it can be compatible with social distancing regulations and allow for short individual trips. It must be noted that this exercise does not account for the seasonal weather impact on Bike and Motorcycle usage which coincided with the onset of COVID-19. Although literature exploring the role of cycling in previous epidemics is rare, it is recognized that one of the factors leading to the rise of e-bikes in China was the 2002–2004

⁴ <https://milogintp.michigan.gov/mdot-waps6/crash/>.

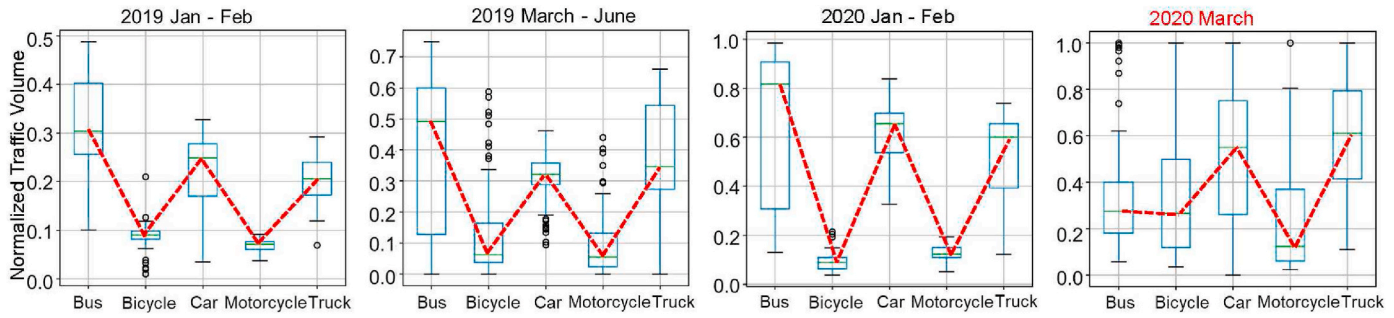


Fig. 4. The change of traffic patterns from 2019 to 2020 pertaining to COVID-19.

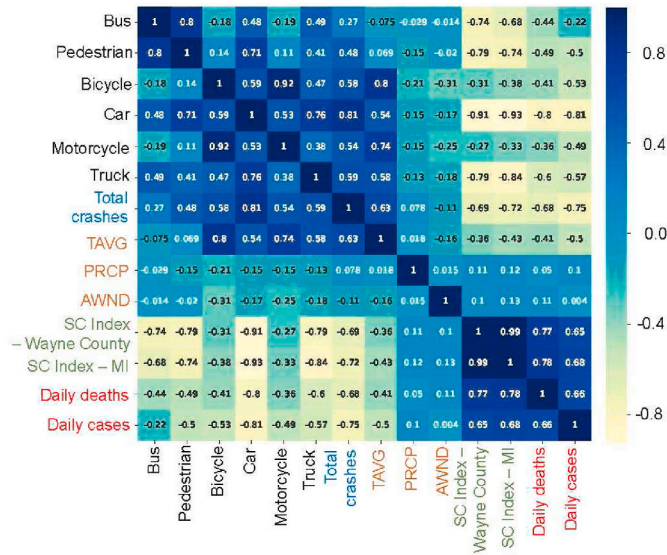


Fig. 5. The value of correlation coefficients. TAVG represents the average temperature per day. PRCP indicates the daily rain precipitation. AWND stands for the daily average wind speed, and SC index represents the social distancing index.

SARS outbreak as people tried to avoid overcrowded public transportation services (Simha, 2016; Weinert et al., 2007). Additionally, the same pattern was also observed in New York City (Teixeira & Lopes, 2020), showing some evidence of a modal transfer from some subway users to the bike-sharing system.

Observation 2. The total number of trucks and vans was almost the same before and during the shutdown. An increase of approximately 40% is observed in the post-shutdown numbers.

The above truck-related observation might be attributed to: i) with the outbreak of the epidemic, a portion of Detroit citizens are turning to online shopping methods. However, due to the shutdown policies, some industry-related trucks and vans stopped running, leading to a relatively stable volume of trucks and vans during the quarantine period, and ii) after the shutdown period, although there are no specific rules to limit the transportation of trucks and vans, it is expected that the increased delivery demand was maintained in addition to the increased demand from the reopening of industrial activity and road construction.

4.1.2. Changes of traffic pattern between 2019 and 2020

Then, to the answer “how did traffic patterns change from 2019 to 2020 considering the impact of COVID-19”, we explored the temporary distribution of each transportation mode’s quantity among 2019 and 2020, including bus, bicycle, car, motorcycle, and truck. More specifically, we divided the collected traffic data set into four groups based on the temporary periods, i.e., 2019 Jan through Feb 2019 March through June 2020 Jan through Feb, and 2020 March through June (the outbreak time period of COVID-19). For each group of data sets, we calculated the normalized traffic volume of each transportation mode and draw the corresponding box plots (Fig. 4). Fig. 4 displays the minimum, first quartile, median, third quartile, and maximum value of the normalized traffic quantities. We use red straight lines to connect the median value of each type of transportation volume, and the shape of the red lines on a single sub-figure of Fig. 4 can provide the general traffic pattern information. It can be seen that the red lines of sub-figures titled “2019 Jan–Feb”, “2019 March–June”, and “2020 Jan–Feb” both show the “W” shape, i.e., the median value of bicycle and motorcycle is significantly lower than the value of bus. However, when it comes to “2020 March–June”, this “W” shape is disappeared since the median value of bicycles is increased while the median value of buses is declined markedly compared with other sub-figures.

Observation 3. Comparing the traffic volume data of 2019 and 2020, the traffic pattern of 2020 March–June (outbreak period of COVID-19) is significantly different from the patterns of 2019 and the first two months of 2020 (before the outbreak of COVID-19).

4.2. Correlation study

To further explore the correlation between each type of collected metrics, such as the correlation between the number of new confirmed cases with the daily traffic volume, we conduct correlation coefficient analysis and joint distribution analysis and present the statistic correlation analysis results.

4.2.1. Correlation coefficients

Correlation analysis is one of the most frequently used statistical methods. Here, the correlation coefficient represents the degree of linear association between two variables (Lee Rodgers & Nicewander, 1988; Taylor, 1990). In this work, we calculated the correlation coefficients between i) transportation modes (e.g., bus, pedestrian, bicycle, car, motorcycle, and truck), ii) total crashes, iii) weather

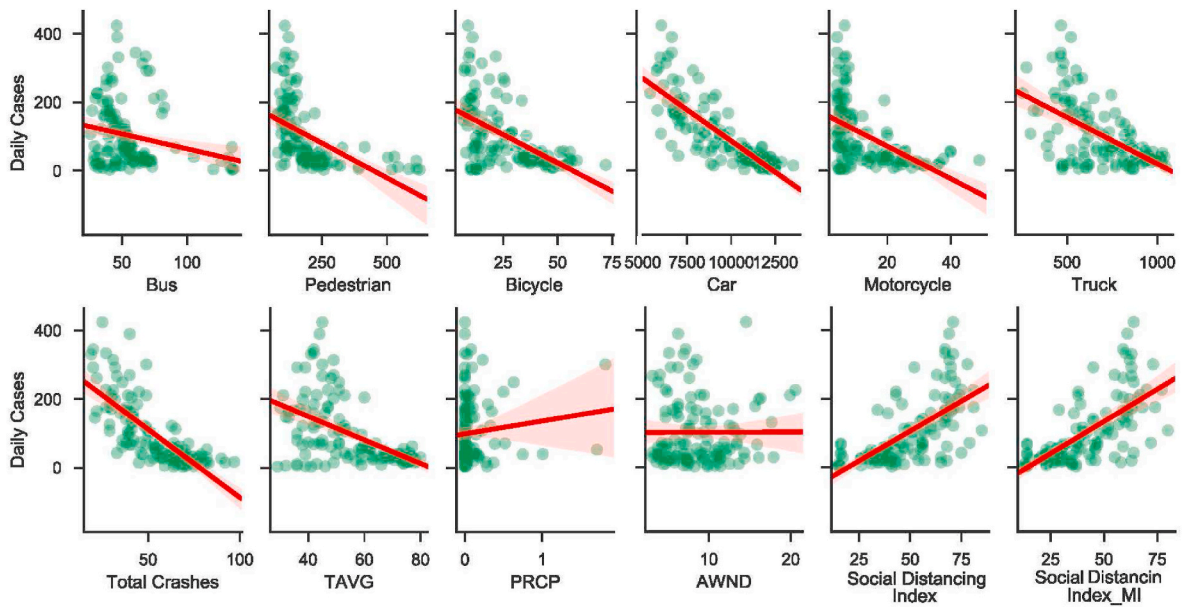


Fig. 6. Joint distribution of various relevant attributes. The greater the slope of the red line, the stronger the linear relationship between the two attributes.

info (e.g., average temperature, rain precipitation, and daily average wind speed), iv) social distancing index of Wayne County and Michigan State, and v) daily cases (e.g., daily confirmed cases and daily deaths). The formula for the correlation coefficient is defined as follows.

$$\rho_{x,y} = \frac{E[XY] - E[X]E[Y]}{\sqrt{E[X^2] - (E[X])^2} \sqrt{E[Y^2] - (E[Y])^2}} \tag{1}$$

The correlation coefficient is a statistic metric that measures the linear correlation between two attributes X and Y, with the value range of [-1, +1]. A value of +1 is the total positive linear correlation, 0 is no linear correlation, and -1 is the total negative linear correlation. The higher correlation coefficient represents the higher correlation between the two attributes. Fig. 5 presents the values of the calculated correlation coefficients between every two attributes based on Equation (1) and we use the gradient color from yellow to blue to indicate negative correlations to positive correlations.

Observation 4. Daily cases, i.e., daily confirmed cases and daily death, are highly related to:

- 1 the number of transportation volume, especially for the cars (-0.81 and -0.8 respectively),
- 2 total crashes (-0.75 and -0.68 respectively),
- 3 social distancing index at the Wayne County level (-0.65 and -0.77 respectively) and Michigan State level (-0.68 and -0.78 respectively), and
- 4 the average temperature (-0.5 and -0.41 respectively).

Note that although previous work revealed the wind speed (mph) is one of the factors that triggered the spread of COVID-19 (Sahin, 2020) in this work, we do not find the high linear correlation between the number of daily cases with the wind speed and other weather factors such as rain precipitation in Detroit based on our collected data set.

4.2.2. Joint distribution

Based on the correlation coefficients, we can roughly know the correlation degree between each pair of attributes, but we cannot decisively derive the specific reason for that relationship. Therefore, we calculated and drew the joint distributions (Sklar, 1973) for select pairs of attributes. This demonstrates the intuitive quantitative relationship between variables (linear/non-linear, or whether there is a more obvious correlation). Most importantly, joint distributions allow us to identify the relationship between multiple attributes.

To be concrete, for $\{X = x, Y = y\}$, we found all elements in the sample space that satisfy these two values. These elements formed a subset of the sample space, and the probability of this subset was the joint probability of $[P(X = x, Y = y)]$. $[\rho(x, y) = P(X = x, Y = y)]$ is called joint PMF (joint probability mass function). The joint probability can be regarded as the probability when two events occur at the same time. Event A is $[X = x]$, and event B is $[Y = y]$, which is $[P(A \cap B)]$.

Fig. 6. presents the joint distribution between the daily confirmed cases with other attributes such as transportation volume, total crashes, weather, and social distancing index from 2020 March to June. Green points display the specific value of the selected

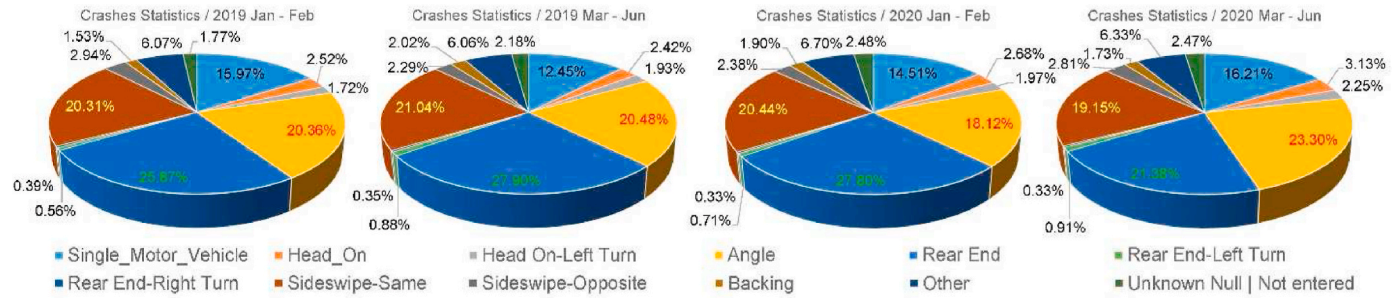


Fig. 7. Distribution by crash type percentage for 2019 to 2020.

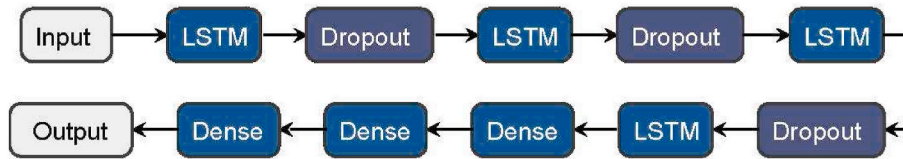


Fig. 8. The structure of the LSTM.

attributes, and red straight lines are used to represent the linear fit results between the pairs of two attributes, i.e., the linear fit result of the daily confirmed case and the number of buses per day. The greater the slope of the red line, the stronger the linear relationship between the two attributes.

Based on the joint distribution of the daily confirmed cases and the rain precipitation shown in Fig. 6, we can deduce the reason why they are *not linearly correlated*: no matter how large the value of daily cases is, the value of PRCP is very small, i.e., the rainfall in Detroit from March to June of 2020 is very small, resulting in a very small range of rainfall. Under these circumstances, the value of the correlation coefficient between the daily cases and PRCP is low, indicating a low *linear correlation*.

4.3. Crashes statistics

With the same principle, we then analyze crash data covering 2019 Jan to Feb 2019 Mar to Jun 2020 Jan to Feb, and 2020 Mar to Jun. Our goal was to identify the percent distribution of the different crash types, which is shown in Fig. 7.

Observation 5. When comparing the 2020 crash type percentages between before the outbreak of COVID-19 and those during the pandemic period, a clear crash type distribution shift is observed:

- 1 Angle crashes became the most common moving up from third place (18.12%~23.30%),
- 2 Rear-end crashes moved from first to second place (27.80%~21.38%),
- 3 Sideswipe crashes slightly decrease but maintained third place (20.44%~19.15%), and
- 4 Single-vehicle crashes increased slightly and maintained fourth place (14.51%~16.21%).

Data key: (2020 Jan–Feb~2020 March–June).

By observing the crash type percentage distribution change, it can be seen that rear-end crashes which used to be the most predominant crash type were overtaken by angle crashes. This observation is indicative of a potentially problematic change in driver behavior: more distraction and speeding, and less adherence to traffic rules and controls. Additionally, rear-end crashes are usually more prevalent in high-volume areas, which is not the case during COVID-19 when the volumes are much lower. A deeper dive into the changes in the total crash numbers and the crash rate is needed due to the unexpected crash behavior during the COVID-19 pandemic. Those analyses were not completed or discussed as part of this paper.

5. COVID-19 confirmed cases prediction

In this section, we aim to conduct a further study on the influence of traffic volume data, weather information data, crash data, and social distancing index information on the confirmed-cases prediction. Our ultimate goal is to build a suitable deep learning model to predict the number of confirmed cases based on (i) traffic volume data, (ii) daily cases number including daily confirmed cases and daily death number, (iii) weather information, (iv) social distancing-related data, and (v) crash data.

5.1. Problem formulation and solution

Problem definition.

We formulate the problem of predicting the number of COVID-19 confirmed cases as a regression problem. Specifically, we use $T = \{input_i\}_{i=1}^n$ to represent our training data set, in which $input_i \in I$ denotes all input features, i.e., the 58 features present in Table 1 of Sec. 3. Our goal is to employ the best method to learn the function f , which minimizes the loss function $l(h(input); groundtruth)$, a measurement of the difference between the desired output and the actual output of the current model, such that the trained model is able to predict the number of confirmed cases over a specific prediction horizon with high performance. Besides, we choose 21 days as our monitoring window, and we aim to predict confirmed cases for the next 7 days.

Deep learning model selection. Recently, machine learning methods have been applied with success in regression tasks. We tackle the confirmed-cases prediction problem using Long Short-Term Memory Networks (LSTM) (Hochreiter & Schmidhuber, 1997; dos Santos Lima et al., 2017) since it has become highly successful learning models for both classification and regression problems across diverse domains (Basak et al., 2019; Hong et al., 2019; Lu et al., 2019, 2020). Specifically, LSTM is a type of recurrent neural networks (RNNs) with the capability of processing sequences of sequential data sets. After being proposed by Hochreiter and Schmidhuber (Hochreiter & Schmidhuber, 1997), LSTM has been proved the ability to address long-term back-propagating issues. It includes a memory cell and a gating mechanism, which allows it to decide what is kept in the memory cell, and how the new input data

Table 2
Input features for six experimental groups.

Group	#of Metrics	Traffic Volume	Daily Case	Weather Data	Crash	Social Distancing Related Data	Social Distancing Related Data_MI
A	79	✓	✓	✓	✓	✓	✓
A-T	69	×	✓	✓	✓	✓	✓
A-W	73	✓	✓	×	✓	✓	✓
A-SD	58	✓	✓	✓	✓	✓	✓
A-SDM	58	✓	✓	✓	✓	✓	×
A-C	60	✓	✓	✓	×	✓	✓

contributes to what is already in the memory cell. Fig. 8 depicts the structure of the LSTM model that we deployed for the confirmed-cases prediction. In particular, after conducting a grid search on the values of hyperparameters to find the best combinations, we build LSTMs with 4 layers and 128 nodes. The LSTM model consists of 4 normal LSTM layers and 3 dense layers. The prior LSTM layer outputs a 3D array as input for the subsequent layer, where the dropout is applied between two LSTM layers to prevent the network from overfitting. Since the output of an LSTM layer under dropout are randomly subsampled, it is able to reduce the capacity of the network during the training phase. Lastly, 3 dense layers work together to get the final output. A dense layer is a fully connected layer, *i.e.*, all the neurons in a layer are connected to those in the next layer.

Effective measurements. To be able to design the best prediction method, we need some metrics to accurately measure the wellness of our prediction approaches. To begin with, we use some commonly used measures for our study: coefficient of determination (R^2), mean square error (MSE), and the root mean square error (RMSE), which both are the commonly used evaluation metrics for the regression problem (Anastassopoulou et al., 2020). R^2 score is widely used to indicate the fit of the machine learning model, *i.e.*, the higher value, the better fit result generated by the model. The maximum value of R^2 is 1 (ideal case), and it may be a negative value with a range of $(-\infty, 1]$. MAE measures the average magnitude of the errors in a set of predictions without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. RMSE is a quadratic scoring rule that also measures the average magnitude of the error. It's the square root of the average squared differences between prediction and actual observation.

Suppose the input data (ground truth) is $y = \{y_1, y_2, \dots, y_N\}$, and the prediction result is noted as $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N\}$. MSE and RMSE are defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (2)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

Specifically, the formula to calculate R^2 is defined as follows:

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (5)$$

where TSS (total sum of squares) is the difference between all samples and the mean value, which is $N \times$ of the variance. Besides, RSS (residual sum of squares) is the sum of the squares of all sample errors, which is $N \times$ times the MSE. When the predicted value of all samples is the same as the true value, RSS is 0, so R^2 equal to 1 (ideal case).

5.2. Model creation

In this subsection, we introduce the experimental hardware configuration and the used packages, then we give a detailed explanation of why we conduct experiments on six experimental groups, and what's the precise format of our experimental input.

Experimental setup. In this work, we adopt NVIDIA GPU Workstation as our experiment platform, which is powerful hardware with high-quality components ($4 \times$ GeForce RTX 2080 Ti graphics cards) with Intel Xeon E5-2690 v4 (CPU), 2.6 GHz of frequency, 14 cores, 64 GB memory, and installed with Ubuntu 16.04.6 LTS (operating system). NVIDIA GPU Workstation is capable of delivering the cluster-level performance for even the demanding applications (Morozov et al., 2011; Spiga & Giroto, 2012). The models learned in this paper are implemented in Python, using TensorFlow 1.13.1 (Abadi et al., 2016), Keras 2.1.5 (Gulli & Pal, 2017), and scikit-learn libraries (Pedregosa et al., 2011) for model building.

Experimental groups. To show the impact of traffic volume data, weather-related metrics, crash data, and social distancing related data on the confirmed-cases prediction, we conduct experiments on six experimental groups. Our first step is to combine all five categories of 58 features present in Table 1 of Sec. 3 to train models using LSTM methods, and we label this group as A Group (A

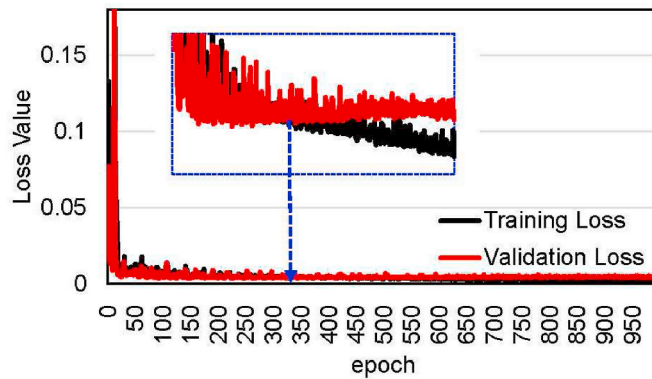


Fig. 9. The validation loss reaches minimum value at 340.

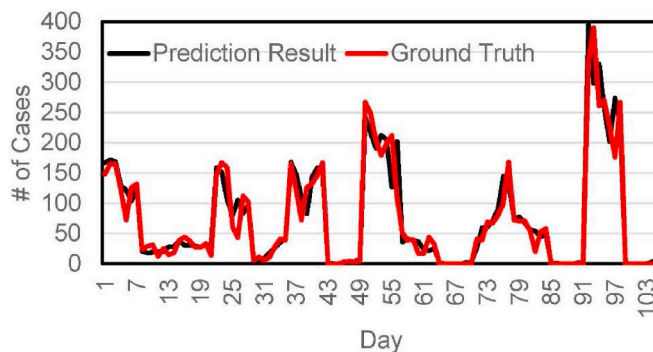


Fig. 10. Prediction result and ground truth.

represent all). Then, we exclude all traffic volume metrics but keep the left features, and we denote it as A-T Group. Similarly, we exclude weather information but keep other features, and we get the A-W Group. Since we have two levels of social distancing index, *i. e.*, social distancing index of Wayne County (denoted as SD), and social distancing index of Michigan state (marked as SDM), we delete SD and SDM to get A-SD group and A-SDM group, respectively. Finally, in order to figure out the impact of crash data (noted as C) on the confirmed-cases prediction, we get the A-C group. Table 2 shows the input features for A Group, A-T Group, A-W Group, A-SD Group, A-SDM, and A-C Group.

5.3. Training and validation methodology

Training and validation methodology. Next, we provide a high-level description of our methodology before delving into the details. We use 5-fold cross-validation (Kohavi et al., 1995), which is a validation technique to assess the predictive performance of machine learning models, judge how models perform to an unseen data set (testing data set) (Rodriguez et al., 2010) and avoid the over-fitting issue during the training phase. More specifically, our data set is randomly partitioned into five equal-sized sub-samples. At a time, we take one sub-sample as the testing data set and take the remaining four sub-samples as the training data set. Then, we fit a model on the training data set, evaluate it on the testing data set, and calculate the evaluation scores. After that, we retain the evaluation scores and discard the current model. The process is then repeated five times with different combinations of sub-samples, and we use the average of the evaluation scores as the final result for each method.

First, we need to determine the hyperparameters of our models—an important aspect of building effective deep learning models. To be concrete, we use the hold-out method (Kim, 2009) to split up our training phase data set further into the parameter training process and the validation process (80% and 20% of the training phase data set respectively), and the validation is an unbiased evaluation of a model fit on the training dataset when tuning parameters. Then, we conduct a grid search on these values of parameters to find the best combination that achieves the highest performance.

Avoiding overfitting of the models. Another important factor during the training process is epoch (Graves & Schmidhuber, 2005), which indicates the number of iterations of processing the input data set during the training process. With a higher value of epoch, the error on training data will reduce further; however, at a crucial tipping point, the network begins to over-fit the training data. Hence, finding the best value of the epoch is essential to avoid overfitting. Fig. 9 shows the change in the value of the training and validation loss functions (the smaller, the better) as the epoch increases. Initially, the values of the two loss functions are decreasing with higher epoch; but after 340 epochs, the value of the validation loss function slowly increases (higher than the training loss), which

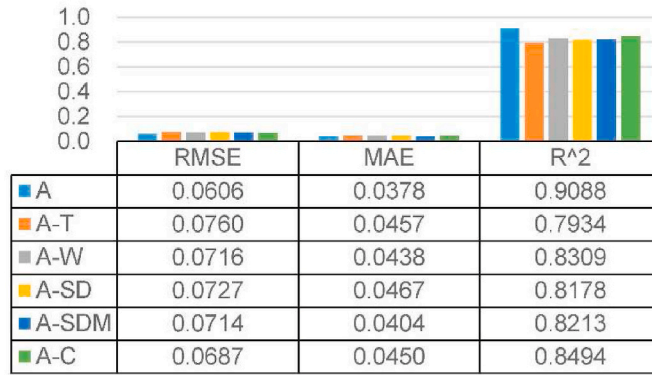


Fig. 11. Model prediction quality with six experiment groups of A (all selected features), A-T (without traffic volume related features), A-W (without weather-related features), A-SD (without social distancing related index data on the Wayne County level), A-SDM (without social distancing related index data on the Detroit level), and A-C (without crash-related features).

Effectiveness	Traffic Volume	>	Social Distancing Related Data	>	Weather	>	Crash		
↓ R ²	A-T	<	A-SD	<	A-SDM	<	A-W	<	A-C
↑ RMSE	A-T	>	A-SD	>	A-SDM	≈	A-W	>	A-C

Fig. 12. Effectiveness comparison between four types of data in terms of prediction improvement. The upward/downward arrow direction indicates that the higher/lower the value of the evaluation index, the better the predictive performance of the index.

indicates the over-fitting issue. Therefore, we choose 340 epochs for LSTM to avoid overfitting of the models.

6. Results and discussion

In this section, we present and analyze the sensitivity of LSTM toward different feature groups. Our discussion includes supporting evidence and reasoning to explain observed trends, and implications of observed trends for the authorities and decision-makers on taking specific measures for Detroit.

6.1. Prediction results and ground truths

Fig. 10 presents the confirmed-cases prediction results, which is denoted by the black curve. In order to conduct an intuitive comparison between the prediction results with the ground truth, we also include a red curve to represent the value of ground truth. It can be seen that our prediction result is very close to the ground truth, which motivates us to get the statistic evaluation results.

6.2. Evaluation results and observations

Next, we present the key prediction quality measures for the six experiment groups (Fig. 11). Note that among the three evaluation metrics, i.e., R² RMSE, and MAE, R² is a more intuitive and objective performance indicator of the fitting effect in the regression problem.

Therefore, we treat R² as our primary evaluation metric. Finally, we make several interesting observations as follows:

1. We observe that A group performs the best across all experiment groups, i.e., achieving the highest (around 0.91) R² score and lowest MAE and RMSE. This observation verifies our hypothesis that i) traffic volume data, ii) weather features, iii) social distancing-related data, and iv) crash information are both useful and helpful for improving the effectiveness of confirmed-cases prediction.
2. Considering the difference of the R² score between A group and the other five groups, we observe that: (1) A-T group achieves the lowest highest score, i.e., there is the biggest effectiveness difference between A group and A-T group, which proves that deleting traffic volume data could result in the most significant adverse effect on the confirmed-cases prediction, i.e., traffic volume data is critical for the improvement of confirmed-cases prediction.
3. Similar to the above observation, we get the conclusion on the effectiveness comparison between traffic volume data, social distancing related data, weather data, and crash data in terms of prediction improvement (shown in Fig. 12)—adding these four

types of data can both improve the prediction performance, and traffic volume data is more effective compared with social distancing related data, then followed by weather features, and crash data seems has least impacts on the prediction performance.

Observation 6. Besides the daily case data and social distancing index, the data related to traffic volume, crashes, and weather can both be good indicators for COVID-19 confirmed-cases prediction. The traffic volume data is very useful information regarding the prediction improvement.

7. Conclusion

In this work, we collected and analyzed five types of data sets including traffic volume, daily cases, weather information, crash features, and social distancing-related data. Important observations with supporting evidence and analysis are presented to provide practical implications for authorities and decision-makers on taking preventive actions for Detroit. In terms of crashes, there was a clear change in crash percentage distribution by type. During COVID-19 there was a significant increase in angle crashes which are typically more severe and indicative of more severe driver-behavior-related issues. In terms of correlations, daily cases *i.e.*, daily confirmed cases and daily death is highly related to i) the number of transportation volume, especially for the cars, ii) total crashes, iii) social distancing index at the Wayne County level and Michigan level, and iv) the average temperature. Additionally, we have trained an accuracy deep learning model, which shows the effectiveness of predicting COVID-19 confirmed cases for the next week, *i.e.*, 7 days, with R^2 up to approximately 0.91. The prediction quality is tested on six experiment groups, and the prediction results also proved that adding traffic volume data, social distancing related metrics, weather information, and crash features could both improve the prediction performance.

Acknowledgments

This work is supported in part by the National Science Foundation (NSF) Award #2027251.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI)*, 16 pp. 265–283.
- Abdel-Aty, M. A., & Abdelwahab, H. T. (2004). Predicting injury severity levels in traffic crashes: A modeling comparison. *Journal of Transportation Engineering*, 130, 204–210.
- Ainslie, K. E., Walters, C. E., Fu, H., Bhatia, S., Wang, H., Xi, X., Baguelin, M., Bhatt, S., Boonyasiri, A., Boyd, O., et al. (2020). Evidence of initial success for China exiting covid-19 social distancing policy after achieving containment. *Wellcome Open Research*, 5.
- Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLoS One*, 15, Article e0230405.
- Basak, S., Dubey, A., & Bruno, L. (2019). Analyzing the cascading effect of traffic congestion using lstm networks. In *2019 IEEE international conference on big data (big data)* (pp. 2144–2153). IEEE.
- Briscese, G., Lacetera, N., Macis, M., & Tonin, M. (2020). *Compliance with covid-19 social-distancing measures in Italy: The role of expectations and duration*. Technical Report National Bureau of Economic Research.
- Candefjord, S., Buendia, R., Caragounis, E.-C., Sjoqvist, B. A., & Fagerlind, H. (2016). Prehospital transportation decisions for patients sustaining major trauma in road traffic crashes in Sweden. *Traffic Injury Prevention*, 17, 16–20.
- Chen, H., Xu, W., Paris, C., Reeson, A., & Li, X. (2020). *Social distance and sars memory: Impact on the public awareness of 2019 novel coronavirus (covid-19) outbreak*. medRxiv.
- Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong social distancing measures in the United States reduced the covid-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed covid-19 cases across the United States. *Health Affairs*, 10–1377.
- Graves, A., & Schmidhuber, J. (2005). Framework phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks*, 18, 602–610.
- Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.
- Gupta, S., Raghuvanshi, G. S., & Chanda, A. (2020). *Effect of weather on covid-19 spread in the us: A prediction model for India in 2020* (p. 138860). Science of The Total Environment.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 1735–1780.
- Hong, J., Wang, Z., & Yao, Y. (2019). Fault prognosis of battery system based on accurate voltage abnormality prognosis using long short-term memory neural networks. *Applied Energy*, 251, 113381.
- Hu, Y., Barbour, W., Samaranyake, S., & Work, D. (2020). *Impacts of covid-19 mode shift on road traffic*. arXiv preprint arXiv:2005.01610.
- Iacus, S. M., Natale, F., Santamaria, C., Spyrtatos, S., & Vespe, M. (2020). *Estimating and projecting air passenger traffic during the covid-19 coronavirus outbreak and its socio-economic impact* (p. 104791). Safety Science.
- Kim, J.-H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis*, 53, 3735–3745.
- Kohavi, R., et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *International joint conference on artificial intelligence* (Vol. 14, pp. 1137–1145). IJCAI.
- Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in wuhan on containing the covid-19 outbreak in China. *Journal of travel medicine*, 27, taaa037.
- Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42, 59–66.
- Lewnard, J. A., & Lo, N. C. (2020). Scientific and ethical basis for social-distancing interventions against covid-19. *The Lancet. Infectious diseases*, 20, 631.
- Lu, S., Luo, B., Patel, T., Yao, Y., Tiwari, D., & Shi, W. (2020). *Making disk failure predictions smarter! In 18th {USENIX} conference on file and storage technologies ({FAST} 20)* (pp. 151–167).
- Lu, S., Yao, Y., & Shi, W. (2019). Collaborative learning on the edges: A case study on connected vehicles. In *2nd {USENIX} workshop on hot topics in edge computing*. HotEdge 19.
- McCracken, J. (1996). Demonstration project 93-making the most of today's technology. *Public Roads*, 59.

- Mohler, G., Bertozzi, A. L., Carter, J., Short, M. B., Sledge, D., Tita, G. E., Uchida, C. D., & Brantingham, P. J. (2020). Impact of social distancing during covid-19 pandemic on crime in los angeles and indianapolis. *Journal ofCriminal Justice*, 101692.
- Morozov, I. V., Kazennov, A., Bystryi, R., Norman, G. E., Pisarev, V., & Stegailov, V. V. (2011). Molecular dynamics simulations of the relaxation processes in the condensed matter on GPUs. *Computer Physics Communications*, 182, 1974–1978.
- Olivera-La Rosa, A., Chuquichambi, E. G., & Ingram, G. P. (2020). Keep your (social) distance: Pathogen concerns and social perception in the time of covid-19. *Personality and Individual Differences*, 166, 110200.
- Painter, M., & Qiu, T. (2020). *Political beliefs affect compliance with covid-19 social distancing orders*. Available at: SSRN 3569098.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal ofmachine learning research*, 12, 2825–2830.
- Qin, L., Sun, Q., Wang, Y., Wu, K.-F., Chen, M., Shia, B.-C., & Wu, S.-Y. (2020). Prediction of number of cases of 2019 novel coronavirus (covid-19) using social media search index. *International journal ofenvironmental research andpublic health*, 17, 2365.
- Rodriguez, J. D., Perez, A., & Lozano, J. A. (2010). Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32, 569–575.
- Sahin, M. (2020). *Impact of weather on covid-19 pandemic in Turkey* (p. 138810). Science of The Total Environment.
- dos Santos Lima, F. D., Amaral, G. M. R., de Moura Leite, L. G., Gomes, J. P. P., & de Castro Machado, J. (2017). Predicting failures in hard drives with lstm networks. In *Proceedings ofthe 2017 Brazilian conference on intelligent systems (BRACIS)* (pp. 222–227). IEEE.
- Simha, P. (2016). Disruptive innovation on two wheels: Chinese urban transportation and electrification of the humble bike. *Periodica Polytechnica Transportation Engineering*, 44, 222–227.
- Singh, R., & Adhikari, R. (2020). *Age-structured impact of social distancing on the covid-19 epidemic in India*. arXiv preprint arXiv:2003.12055.
- Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, 9, 449–460.
- Spiga, F., & Giroto, I. (2012). phiGEMM: a CPU-GPU library for porting quantum espresso on hybrid systems. In *2012 20th euromicro international conference on parallel, distributed and network-based processing* (pp. 368–375). IEEE.
- Stutts, J. C., Reinfurt, D. W., Staplin, L., Rodgman, E., et al. (2001). *The role of driver distraction in traffic crashes*.
- Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal ofdiagnostic medical sonography*, 6, 35–39.
- Teixeira, J. F., & Lopes, M. (2020). The link between bike sharing and subway use during the covid-19 pandemic: The case-study of New York's citi bike. *Transportation Research Interdisciplinary Perspectives*, 6, 100166.
- Tosepu, R., Gunawan, J., Effendy, D. S., Lestari, H., Bahar, H., Asfian, P., et al. (2020). *Correlationbetween weather and covid-19 pandemic injakarta, Indonesia* (p. 138436). Science ofThe Total Environment.
- Weinert, J., Ma, C., & Cherry, C. (2007). The transition to electric bikes in China: History and key reasons for rapid growth. *Transportation*, 34, 301–318.
- WHO. (2020). *Weekly operational update on covid-19 (7 december 2020)*. <https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19-7-december-2020>.