Analysis of the Effectiveness of Face-Coverings on the Death Rate of COVID-19 Using Machine Learning

Ali Lafzi[1], Miad Boodaghi[2], Siavash Zamani[2], and Niyousha Mohammadshafie[3]

[1]Department of Agricultural and Biological Engineering, Purdue University, Indiana, USA

[2]School of Mechanical Engineering, Purdue University, Indiana, USA

[3]Department of Civil and Environmental Engineering, University of Pittsburgh, Pennsylvania, USA

Abstract

The recent outbreak of the COVID-19 shocked humanity leading to the death of millions of people worldwide. To stave off the spread of the virus, the authorities in the US, employed different strategies including the mask mandate (MM) order issued by the states' governors. Although most of the previous studies pointed in the direction that MM can be effective in hindering the spread of viral infections, the effectiveness of MM in reducing the degree of exposure to the virus and, consequently, death rates remains indeterminate. Indeed, the extent to which the degree of exposure to COVID-19 takes part in the lethality of the virus remains unclear. In the current work, we defined a parameter called the average death ratio as the monthly average of the ratio of the number of daily deaths to the total number of daily cases. We utilized survey data provided by New York Times to quantify people's abidance to the MM order. Additionally, we implicitly addressed the extent to which people abide by the MM order that may depend on some parameters like population, income, and political inclination. Using different machine learning classification algorithms we investigated how the decrease or increase in death ratio for the counties in the US West Coast correlates with the input parameters. Our results showed a promising score as high as 0.94 with algorithms like XGBoost, Random Forest, and Naive Bayes. To verify the model, the best performing algorithms were then utilized to analyze other states (Arizona, New Jersey, New York and Texas) as test cases. The findings show an acceptable trend, further confirming usability of the chosen features for prediction of similar cases.

1 Introduction

The recent pandemic of COVID-19 has affected millions of peoples worldwide and led to the tragic death of many innocent lives. The lack of a certain treatment at the beginning of pandemic traumatized populace and the only solutions were limited to preventive actions such as wearing face coverings, maintaining social distancing, washing hands, and self-quarantine. Owing to the high transmission rate, only in the US, the number of new daily cases increased from 6 to 22562 during March 2020 according to CDC (Center for Disease Control and Prevention) [1]. There is still extensive ongoing research about the possible factors being effective in the pace of this spread; as of now, scientists have declared that meteorological factors such as temperature, wind speed, precipitation and humidity are some of the important environmental parameters in this regard [2]. However, most of the parameters involved in the spread of COVDID-19 are out of our control. As a result, state officials began to impose legislative guidelines including mandatory use of masks and closure of businesses such as bars and restaurants. Shutting down different businesses has been sporadic due to its adverse economic impact, but obligatory face coverings order is still in effect across the US. In this respect, the effectiveness of facial masks gains further importance and requires scientific studies.

Presenting a model that can measure the effectiveness of the mask mandate orders can pave the way for governments to take decisive actions during pandemics. The experimental data in tandem with mathematical modelings can be utilized to study the effects of facial coverings on the spread of viral infections. A plethora of previous publications have tried to address the effectiveness of nonpharmaceutical interventions (NPIs) during pandemics, particularly for the spread of influenza [3,4]. Deterministic models have been widely used to study the effects of facial masks on the reproduction number R_0. Indeed, the face mask is taken into account by its role in reducing the transmission per contact [5]. The results of the deterministic model indicated that public use of face masks delays the influenza pandemic. On the other hand, some studies suggest that the use of a face mask does not have a substantial effect on influenza transmission and there is little evidence in favor of the effectiveness of facial masks [6,7]. As for the COVID-19, the efficacy of the facial mask in impeding the infectivity of the SARS-CoV-2 remains unclear. Having considered the effects of mask in reproduction number R_0, Li et al. [8] claimed that wearing face masks alongside the social distancing can flatten the epidemic curve. Other studies also pinpointed that public use of a facial mask may contribute to the reduction in spread of COVID-19 [9]. Despite these findings, the efficacy of face masks remains controversial.

The cardinal point that has not garnered enough attention is the relationship between the degree of exposure to the virus and its mortality rate. The idea that the severity of the symptoms correlates with the extent of exposure to the COVID-19 was presented by some researchers to justify the high death rate in healthcare workers [10]. Unfortunately, there is not a universal trend that can predict the relationship between the dose of the virus and the severity of the resulting symptoms. A study performed on the relationship between influenza and rhinovirus viral load, and the severity in the upper respiratory tract infections reported a different behavior for those viruses [11]. In fact, the results indicated that for influenza A and the rhinovirus, viral loads were not associated with hospitalization/ICU. On the other hand, for influenza B, viral load was higher in hospitalized/ICU patients. Furthermore, for Respiratory syncytial virus (RSV), viral load seems to correlate with the severity of symptoms as many studies in the literature suggest that a correlation exists [12–14]. The same controversy holds for the COVID-19. Recently, some studies have tried to investigate the severity of COVID-19 with its load, where they found that the load tightly correlates with the severity [15, 16]. However, another study suggests that no such a correlation exists [17].

To unveil whether COVID-19 viral load is related to disease severity requires an in-depth study, which involves infecting volunteers with controlled doses of virus and monitoring their symptoms. However, experimental challenges in addition to the ethicality of these experiments make this type of studies very challenging at this point [10]. Although studies have not been convergent in whether nose [18] or mouth [19] is the primary site for COVID-19 infection, they underscored the importance of wearing a facial mask as a barrier to the virus spread. Additionally, although the protection level of different types of mask are different, wearing any mask even a cloth mask is better than wearing nothing at all, which can play a role in protection from the exposure to COVID-19 [20, 21]. Given the challenges of the experimental studies on the relationship between the extent of exposure and severity of COVID-19, one way to study whether the extent to which an individual is exposed to the COVID-19 correlates with the severity of the symptoms is to introduce a model that can capture changes in the mortality rate due to the wearing a facial mask. Indeed, if the ratio of the number of death to the number of cases decreases, this can support the hypothesis that there is a correlation between the viral load and the severity of symptoms. Thus, studying the effects of MM order on the mortality rate gains extra importance.

An ML analysis can be very useful to shed light on the possible correlation between the public use of mask and changes in the mortality rate. The success of implementing Machine Learning (ML) and Artificial Intelligence (AI) techniques in the previous pandemic has convinced researchers to use them as precious tools in fighting against the current outbreak [22]. ML and AI can be used for prediction and forecasting in different regions so that the corresponding health officials can take essential actions in advance [22]. In addition, this technology is capable of enhancing the prediction accuracy for screening both infectious and non-infectious diseases [23]. Six ML methods have been carried out to predict 1, 3, and 6 days ahead the total number of confirmed COVID-19 cases with error ranges of 0.87%–3.51%, 1.02%–5.63%, and 0.95%–6.90%, respectively, in 10 Brazilian states [24]. Moreover, an ML method like XGBoost model was capable of identifying 3 important biomarkers from 485 blood samples in Wuhan, China as the key mortality parameters [25]. ML algorithms also have been used to capture the correlation between the weather data, and COVID-19 mortality and transmission rates [26,27]. Additionally, ML has been utilized to study the effects of MM order on the number of daily cases, where no significant statistical difference was observed in the number of daily cases in state-wise analysis [28]. These studies confirm the strength of ML as a great tool to investigate the effects of MM order on mortality rates of COVID-19.

Another important factor regarding the effectiveness of MM order is society's adherence to the regulations. One study that tried to quantify public compliance with COVID-19 public health recommendations found notable regional differences in intent to follow health guidelines [29]. Some studies noticed a correlation between level of education and intent to voluntarily adhere to social distancing guidelines [29, 30]. However, not only the level of education but also level of income, race and political orientation can play a role in the adherence to the regulations [31]. Based on these findings, it's important to take into account the features that might be correlated with people's compliance with the MM order. Additionally, we will use a data based on the survey provided by New York times available on Github, which quantifies people's adherence to the MM order [32]. As a result, in this study, we will include factors that might play a role in people's adherence to the MM order as our input features.

In the proposed work, utilizing different ML classification algorithms, we aim to unveil how the change in the mortality rate correlates with certain features. The features will be chosen in a way that they can reflect abidance by MM order in different counties. We will use the data provided by CDC to find the average monthly number of COVID-19 cases. Additionally, the exact dates of the executive orders signed by the state officials are available for each state. To have appropriate unbiased data, similar to what Maloney et al. [28] has done in his study of the effect of mask mandate, we will be using the data for one month after and before the executive orders for each preventive measure for the three states in US West Coast. Indeed, with this data selection method, we limit the geographical region of the study to ensure that changes in the cases are highly attributed to the public use of masks rather than other factors such as environmental changes.

As a verification of the proposed work, the best performing algorithms are further chosen with the calculated hyper-parameters for testing four additional states (Arizona, New Jersey, New York and Texas). The findings demonstrate an acceptable accuracy scores, which justifies the correlation of the chosen features with the effect of COVID-19.

The rest of the paper is organized as follows. First, we will represent how our data was collected and arranged. Then we will explicate the ML methods we have used for our prediction. Finally, we will represent and compare the results obtained from different ML methods.

2 Methodology

In this section, we will explain the collected data and the ML algorithms used for the training and prediction.

2.1 Data

We defined the parameter of interest as the average ratio of the number of deaths to the total number of cases, referred to as the death ratio, which can be interpreted as a measure of the severity of the disease. The effective date of the executive orders by the governors, requiring mask mandate at all the counties in the three West Coast states of California, Oregon and Washington has been identified, which is publicly available [33]. We used the average death ratio one month before and after the order to study the mortality rate. The rationale behind this selection is to minimize the effects of other factors that might play role in changing the COVID-19 data. The raw dataset for the daily cases and deaths for all the US counties over time is extracted from the USAFACTS website [34], where county-level data is confirmed by the state and local agencies directly. After obtaining the daily values of death and case numbers for a month before and after the MM order, we divided the monthly average number of deaths by the monthly average number of cases for each county. Then we found the difference between the death ratio for one month before and after the MM order. Finally, we categorized the variation based on its sign to quantify whether the death ratio increases, decreases, or no change occurs. Out of the 130 samples, 47, 30, and 53 of them belong to the "decrease", "increase", and "no change" classes, respectively. We dropped the "no change" data as they all correspond to small counties, where there were zero reported COVID-19 cases and deaths, leaving 77 counties in total. Consequently, the two categories of increase (denoted by class 0) and decrease (shown by class 1) are remained for the prediction task. A histogram of the output classes is shown in the Fig. (1), which expresses that the data is not biased.

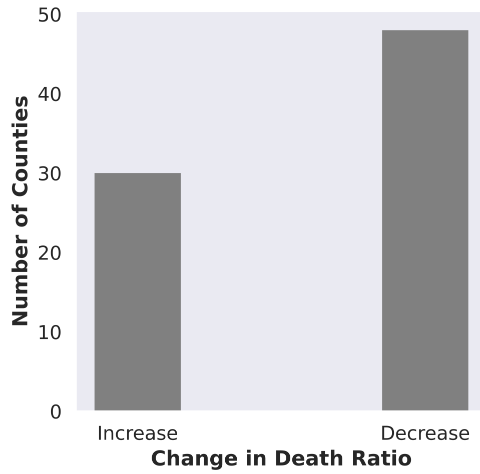


Figure 1: Histogram of change in death ratio for the three states

Since it is not known exactly what percentage of population follows the MM order and use face coverings, it is necessary to come up with features that can capture how likely is an individual to follow the recommended practice. For bridging this gap, four main features are chosen as primary indicators which are listed below:

1. County Population

2. Median Household Income

3. Political Inclination

4. Mask Usage based on New York Times Survey

Population in each county is obtained from the most recent surveys for the year 2019. The income level is the median household in each county in the years 2015-2019. The raw data for these features is all obtained from the US Census website [35]. The US Census measures the median income as the regular income received excluding other payments like tax, etc [36]. The data for the political inclination is constructed based on the 2020 US presidential election results [37]. This feature has been converted to the categorical type in a vectorized manner, i.e. the winner takes the value of 1 in the column, and the opponent takes 0 in their own. Furthermore, we used a survey data, provided by the New York Times, that quantifies the mask usage from 7/2/2020 to 7/14/2020 [32]. Since the survey timeline lies within the month after the MM order for all three studied states, it is valid to use its data for our purpose. Finally, we will try to establish an AI-based relationship between the features and the death ratios of the Pacific Coast states at the county level using 9 different classification algorithms, provided in section 2.2.

2.2 Methods

In this study, we have developed machine learning models to correlate the specified features mentioned in section 2.1 with the aim of shedding light on the relationship between adherence to mask mandate and mortality rate.

Classic ML methods of Logistic Regression [38] and Naive Bayes classifier [39] are used. In addition, ensemble learning-based models, Random Forest and Extra Trees, are also analyzed [40]. Moreover, the extreme boosting method, XGBoost is explored [41]. Other methods such as Support Vector Machine, K-Nearest Neighbors [42], Decision Trees [43], and Neural Network [44] are additionally used for prediction of effect of Mask Mandate on mortality rate.

It should be noted that for carrying out the analysis, the data is split into training and test sets, with a test size of 20%. A k-fold cross validation scheme with 5 folds has been used to evaluate the performance of each method on the validation set and tune its hyper-parameters with the classification accuracy as the metric accordingly. The hyper-parameter tuning is done using either grid search or random search for all the methods. A statistical summary of the final dataset for the purpose of binary classification is outlined in the table 1, which indicates a large difference between the orders of magnitudes of the features. Therefore, min-max and max-abs scaling have been used to transform the input features and output, respectively, before passing the data to the ML algorithms for training.

Table 1: Statistical summary of the final dataset before scaling. Columns are P:population, MI:median income, Dem:voted democratic, Rep:voted republican. Mask usage - N:never, R:rarely, S:sometimes, F:frequently, A:always. DR:change in death ratio between one moth before and after the corresponding MM order date

	P	**MI**	**Dem**	**Rep**	**Mask Usage**					**DR(%)**
					N	**R**	**S**	**F**	**A**	
Count	77	77	77	77	77	77	77	77	77	77
Mean	630413.5	66494.23	0.58	0.44	0.03	0.03	0.06	0.17	0.71	-0.47
Std	1297275	18484.92	0.5	0.5	0.02	0.03	0.03	0.05	0.09	2.83
Min	7208	43313	0	0	0.001	0	0.004	0.07	0.31	-12.9
25%	86085	53105	0	0	0.02	0.01	0.04	0.14	0.67	-1.4
50%	219186	62077	1	0	0.02	0.02	0.06	0.16	0.72	-0.44
75%	601592	74624	1	1	0.04	0.04	0.08	0.2	0.77	0.77
Max	10039110	124055	1	1	0.11	0.21	0.21	0.3	0.87	7.69

3 Results and Discussions

The change in death ratio from one month before to one month after the date of mandating face-covering in the three states is visualized for each county in Fig.(2). Two clusters of increase in death ratio can be seen, one near northern Washington, and one near central California. Our first intuition was that by increasing population, the chance of viral spread would increase, and therefore we expected to see a positive change in death ratio for more populated counties. However, as it can be seen from the map, there is an inherent randomness which defies our initial intuition about the spread mechanism. Further, it is shown that more counties experienced a decrease in death ratio one month after the usage of face-covering was mandated by each state, as shown in Fig. (1). Therefore, usage of face-covering is chosen as the main factor affecting the decrease of the change in death ratio. As explained previously, to quantify adherence to the mask mandate, other auxiliary features are chosen, namely, population, median income, and political inclination for each county.

As a preliminary analysis, political inclination, based on the result of the 2020 presidential election, is chosen as the focal criterion to categorize the data for changes of death ratio for all three states, as presented in Fig. (3). Fig. 3(a) shows that in general, communities that voted republican in presidential election of 2020 were affected worse compared to democratic counties. Further, a noticeable correlation is observed between average median income and the change of death ratio, presented in Fig. 3(b). It is shown that, on average, the communities with less median income experienced a positive change in death ratio, meaning more mortality rate regardless of their political inclination. The strongest correlation however is observed by considering county population, shown in Fig. 3(c). The counties with fewer residents were affected more adversely by the pandemic compared to high-population counties. The counter intuitive relation between population and change in death ratio further corroborates necessity of inclusion of the two other supplementary features.

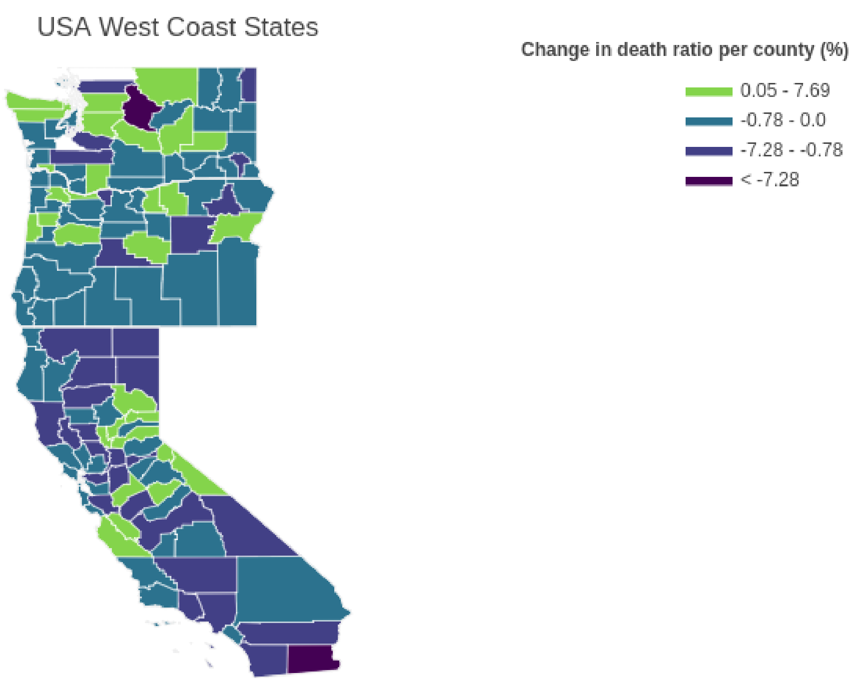
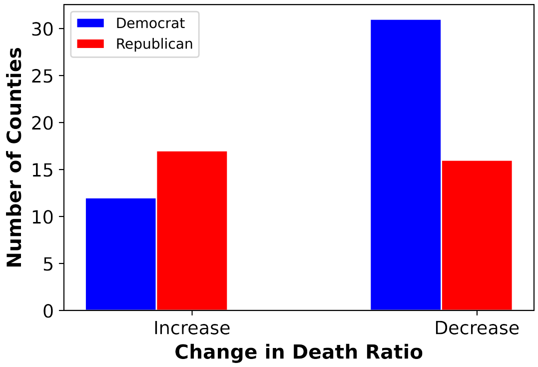
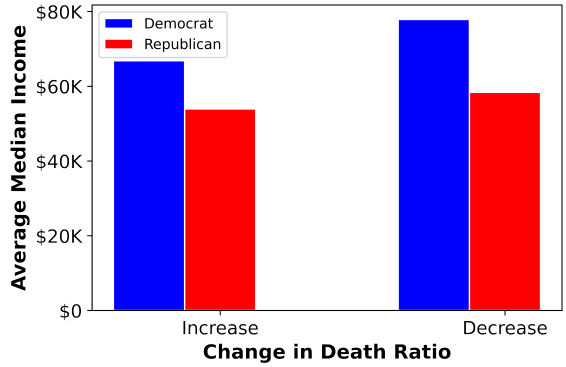


Figure 2: Change in death ratio in US West Coast states counties

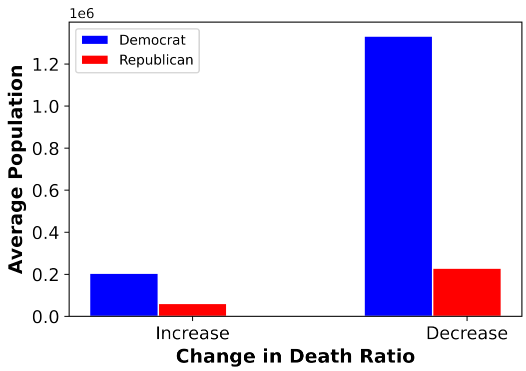
To have an initial assessment of the variation of percent change in the death ratio, we plotted the percent death ratio as functions of population, median income, and percent of the population that frequently uses mask, which has a relatively high correlation coefficient. Fig 4 a-c shows no detectable pattern between parameters of interest and death ratio. As a result, it is not possible to predict the value of change in the death ratio using regression. On the other hand, as we will show, converting changes to categories of increase and decrease



(a)



(b)



(c)

Figure 3: Visualization of the combined data for California, Oregon and Washington. Change in death ratio and (a) representation of number of counties (b) median income (c) average population, based on political inclination.

would pave the way for capturing the status of change. A summary of the overall death rates in the months before and after the mask mandate order for the 3 states is presented in table 2. It can be observed that change in death ratio significantly decreases in California and Washington, but slightly increases in Oregon. This suggests an intrinsically complex pattern between the death rate as the desired output and the selected inputs. According to a recent study, there is a number of factors attributing to the possibility of a person to follow or not follow the health guidelines set by the state officials [31]. Three features among these parameters plus the mask usage as the fourth feature have been used to conduct the current study.

Using the obtained data, the combined effect of features is analyzed on the death ratio. Then

Table 2: Total death rates in the month before and after the corresponding date of the mandatory face coverings executive order in each state

State	**1 month before MM order**	**1 month after MM order**	**Change** (%)
California	63.13	32.67	-48
Washington	28.16	21.15	-25
Oregon	38.03	39.14	+3

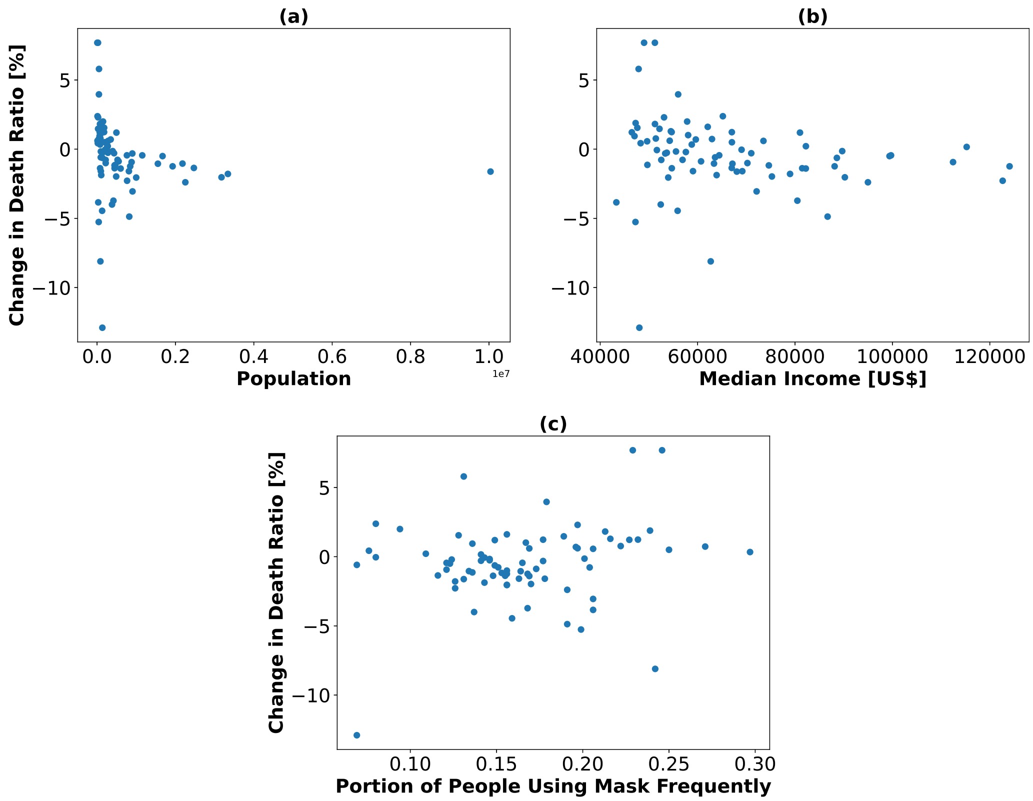


Figure 4: Scatter plot of the percent change in the death ratio as a function of a) population b) median income c) percent people frequently using mask.

the performance of each algorithm is evaluated for test and train sets. The effect of each feature on the change of death ratio is visualized by the correlation heatmap provided in the figure 5. Each row of the correlation matrix is an appropriate indicator of how correlated that feature is with change in death ratio. A more negative value implies that increase of that specific feature is positively correlated by a decrease in change of death ratio. For instance, increase in population, median income, and votes for democratic party would result in a decrease in change of death ratio. On the other hand, the positive correlation for republican votes leads to a higher change of positive increase in death ratio. An interesting observation is the disordered correlation pattern for mask usage. It can be seen that, as one expects, increasing the number never and rarely mask users is positively correlated with change in death ratio. However, the data associated with frequently mask users have resulted in a positive correlation value. Such erratic correlation behavior necessitates inclusion of other features in the analysis.

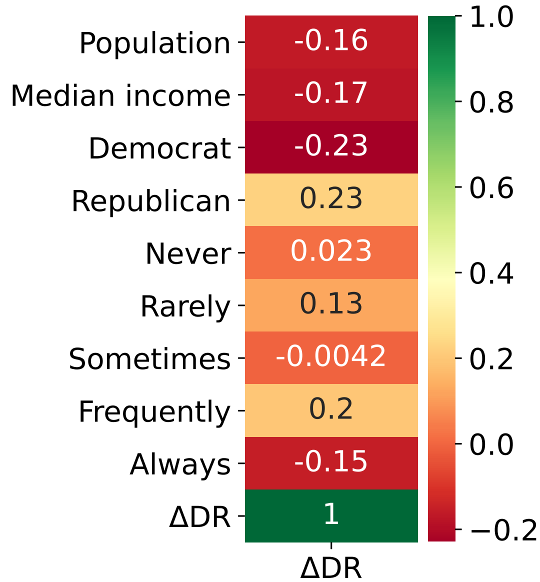


Figure 5: Correlations between the features and the output

All implemented algorithms in this study are capable of providing us with high classification accuracy i.e, to predict whether a county has experienced a decrease in its death ratios or an increase. As provided in Table (3), it can be seen that in general, most of the algorithms have relatively high accuracy scores for both training and test sets. The lowest accuracy comes from neural net algorithm with a score of 63% for the test set. This could be a result of the low sample data set. In general, neural network would incrementally increase in accuracy by providing more training data set. In our case, we were limited by the existing data.

Despite the lack of sufficient training data set, Naive Bayes, Random Forest, and XGBoost have an accuracy of 94%. The selected hyper-parameters for XGBoost and Random Forest classifiers are shown in table 4. The random search method has been done to tune these hyper-parameters for XGBoost, and grid search is used for Random Forest. Naive Bayes does

Table 3: Train and test accuracies for all the studied algorithms.

Algorithm	Train Accuracy	Test Accuracy
Support Vector Machine	0.82	0.81
Decision Tree	1.00	0.81
KNN	0.74	0.69
Logistic Regression	0.79	0.75
Neural Net	0.75	0.63
Extra Trees	0.93	0.81
Naive Bayes	0.7	0.94
Random Forest	1.00	0.94
XGBoost	0.98	0.94

not have any important hyper-parameter because of which, it has the capability of being generalized well. Random Forest and XGBoost also have the popularity of rarely over-fitting the data. These reasons could be why these three algorithms have outperformed the others.

Table 4: Model Parameters for XGBoost and Random Forest. Columns of XGBoost - CSbT:column sample by tree, G:gamma, LR:learning rate, MD:max depth, NE:number of estimators, S:subsamples, RS:random state. Columns of Random Forest - MD:maximum depth of the tree, MF:number of features for best split, MSS:minimum number of samples to split an internal node, NE:number of estimators.

Extreme Gradient Boosting						
CSbT	G	LR	MD	NE	S	RS
0.9605	0.4735	0.0975	4	119	0.6232	27

Random Forest			
MD	MF	MSS	NE
7	2	2	10

Using the calculated hyper-parameters from the best performing algorithms, it would be possible to predict effect of similar viral illnesses in future. To verify the legibility of the proposed work, the best performing algorithms (Naive Bayes, Random Forest, and XGBoost), were chosen with the computed hyper-parameters to process the data for four additional states, namely, Arizona, New Jersey, New York, and Texas. For choosing states for testing purposes, three main criteria were considered: (*i*) availability of data provided by NY Times survey (*ii*) population (*iii*) versatility of death rate ratio. The NY Times mask usage survey is only available for the time period of interest, July 2nd-14th; therefore, the month after the corresponding MM order should contain this period for validity of our analysis. The chosen states all have high population. Lastly, Arizona, New Jersey, and New York all experienced a negative change of death ratio, while Texas suffered significant losses in the month after the MM was placed, as shown in the Table (5). Inclusion of cases with extreme positive and negative change of death ratio was done deliberately to assay functionality of the selected algorithms. The accuracy score for the processed algorithms on these four states are presented in the Table (6).

Table 5: Total death rates in the month before and after the corresponding date of the mandatory face coverings executive order for test states

State	**1 month before MM order**	**1 month after MM order**	**Change** (%)
Arizona	45.06	38.88	-14
New Jersey	910.12	126.04	-86
New York	240.14	113.73	-53
Texas	197.30	608.66	+208

Table 6: Accuracy results for the four states of Arizona, New Jersey, New York, and Texas.

Algorithm	**Test Accuracy**
Naive Bayes	0.76
Random Forest	0.68
XGBoost	0.69

It should be noted that the results of the three west coast states were chosen as training data set. The entire data from the four states is treated as test data set. Hence, it is expected for the accuracy score to drop for testing the additional states. However, the trend of high accuracy for train and test data sets, signifies the existence of a pattern between the chosen features and the change in death ratio.

For instance, against the common belief that highly populated areas might experience harsher effect of COVID-19, in the west coast of the United States, the areas with lower population endured worse conditions. Further, the result of this work would further signify the importance of political leadership in guiding communities and ensuring the well-being of the general public. Additionally, such a modeling approach could be used to optimize distribution of services and media coverage for possible future adversities. A possible solution for decreasing effect of future pandemics such as COVID-19 would be improving media coverage and public knowledge, especially in more vulnerable areas.

4 Conclusion

In this body of work, we have analyzed the effect of mask covering on the intensity of spread of the COVID-19 virus by considering the death ratio at the county level to be the primary indicator. To bridge the gap between level of adherence to mask mandate, four main features are used as input data, population, income, political inclination, and the results of the survey on mask usage from New York Times. The change in the death ratio is used as the metric to quantify the effectiveness of face-coverings on the COVID-19 spread. After extracting and refining the data-set from reliable sources, we analyzed the information using 9 different algorithms. Among all the methods used, Random Forest, XGBoost, and Naive Bayes had the best performance with a classification accuracy of 94%. The high performing algorithms, with the computed hyper-parameters, are then used to process four additional states, Arizona, New Jersey, New York, and Texas entirely used as test data set. The acceptable accuracy results for the large test case, further verifies legibility of the chosen features as influential criteria for modeling purposes. The obtained hyper-parameters for these models (except for Naive Bayes) can now be used to predict future conditions of the spread of the virus.

It is shown that, in most of the counties, there exist a connection between adherence to the mask mandate and change in death ratio. The findings of this work emphasizes importance of immediate legislative action on well-being of societies. It is hoped that the findings of this work, highlight importance of socioeconomic and political settings on behavior of different communities, which as portrayed could be complex and counter-intuitive. For instance, if the mask mandate had been issued earlier, with better implementation procedures along with effective incentives targetted at specific communities, more people would be encouraged to abide by the issued ordinance, and consequently, fewer individuals and families would have become the victim of the pandemic.

References

[1] centers for Disease Control and Prevention, "Previous u.s. covid-19 case data." `https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/previouscases.html`.

[2] M. H. Shakil, Z. H. Munim, M. Tasnia, and S. Sarowar, "Covid-19 and the environment: A critical review and research agenda," *Science of the Total Environment*, p. 141022, 2020.

[3] A. E. Aiello, R. M. Coulborn, T. J. Aragon, M. G. Baker, B. B. Burrus, B. J. Cowling, A. Duncan, W. Enanoria, M. P. Fabian, Y.-h. Ferng, *et al.*, "Research findings from nonpharmaceutical intervention studies for pandemic influenza and current gaps in the research," *American journal of infection control*, vol. 38, no. 4, pp. 251–258, 2010.

[4] P. Saunders-Hastings, J. A. Crispo, L. Sikora, and D. Krewski, "Effectiveness of personal protective measures in reducing pandemic influenza transmission: A systematic review and meta-analysis," *Epidemics*, vol. 20, pp. 1–20, 2017.

[5] N. C. Brienen, A. Timen, J. Wallinga, J. E. Van Steenbergen, and P. F. Teunis, "The effect of mask use on the spread of influenza during a pandemic," *Risk Analysis: An International Journal*, vol. 30, no. 8, pp. 1210–1218, 2010.

[6] J. Xiao, E. Y. Shiu, H. Gao, J. Y. Wong, M. W. Fong, S. Ryu, and B. J. Cowling, "Non-pharmaceutical measures for pandemic influenza in nonhealthcare settings—personal protective and environmental measures," *Emerging infectious diseases*, vol. 26, no. 5, p. 967, 2020.

[7] B. Cowling, Y. Zhou, D. Ip, G. Leung, and A. E. Aiello, "Face masks to prevent transmission of influenza virus: a systematic review," *Epidemiology & infection*, vol. 138, no. 4, pp. 449–456, 2010.

[8] T. Li, Y. Liu, M. Li, X. Qian, and S. Y. Dai, "Mask or no mask for covid-19: A public health and market study," *PloS one*, vol. 15, no. 8, p. e0237691, 2020.

[9] V. C. Cheng, S.-C. Wong, V. W. Chuang, S. Y. So, J. H. Chen, S. Sridhar, K. K. To, J. F. Chan, I. F. Hung, P.-L. Ho, *et al.*, "The role of community-wide wearing of face mask for control of coronavirus disease 2019 (covid-19) epidemic due to sars-cov-2," *Journal of Infection*, 2020.

[10] S. Caddy, "Coronavirus: does the amount of virus you are exposed to determine how sick you'll get?." `https://theconversation.com/coronavirus-does-the-amount-of-virus-you-are-exposed-to-determine-how-sick-youll-get-135119`.

[11] A. Granados, A. Peci, A. McGeer, and J. B. Gubbay, "Influenza and rhinovirus viral load and disease severity in upper respiratory tract infections," *Journal of Clinical Virology*, vol. 86, pp. 14–19, 2017.

[12] E. T. Martin, J. Kuypers, J. Heugel, and J. A. Englund, "Clinical disease and viral load in children infected with respiratory syncytial virus or human metapneumovirus," *Diagnostic microbiology and infectious disease*, vol. 62, no. 4, pp. 382–388, 2008.

[13] M. Houben, F. Coenjaerts, J. Rossen, M. Belderbos, R. Hofland, J. Kimpen, and L. Bont, "Disease severity and viral load are correlated in infants with primary respiratory syncytial virus infection in the community," *Journal of medical virology*, vol. 82, no. 7, pp. 1266–1271, 2010.

[14] J. P. DeVincenzo, C. M. El Saleeby, and A. J. Bush, "Respiratory syncytial virus load predicts disease severity in previously healthy infants," *The Journal of infectious diseases*, vol. 191, no. 11, pp. 1861–1868, 2005.

[15] Y. Liu, W. Liao, L. Wan, T. Xiang, and W. Zhang, "Correlation between relative nasopharyngeal virus rna load and lymphocyte count disease severity in patients with covid-19," *Viral immunology*, 2020.

[16] J. Fajnzylber, J. Regan, K. Coxen, H. Corry, C. Wong, A. Rosenthal, D. Worrall, F. Giguel, A. Piechocka-Trocha, C. Atyeo, *et al.*, "Sars-cov-2 viral load is associated with increased disease severity and mortality," *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020.

[17] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, "Temporal dynamics in viral shedding and transmissibility of covid-19," *Nature medicine*, vol. 26, no. 5, pp. 672–675, 2020.

[18] Y. J. Hou, K. Okuda, C. E. Edwards, D. R. Martinez, T. Asakura, K. H. Dinnon III, T. Kato, R. E. Lee, B. L. Yount, T. M. Mascenik, *et al.*, "Sars-cov-2 reverse genetics reveals a variable infection gradient in the respiratory tract," *Cell*, vol. 182, no. 2, pp. 429–446, 2020.

[19] N. Huang, P. Perez, T. Kato, Y. Mikami, K. Okuda, R. C. Gilmore, C. D. Conde, B. Gasmi, S. Stein, M. Beach, *et al.*, "Integrated single-cell atlases reveal an oral sars-cov-2 infection and transmission axis," *medRxiv*, 2020.

[20] Y. Goh, B. Y. Tan, C. Bhartendu, J. J. Ong, and V. K. Sharma, "The face mask how a real protection becomes a psychological symbol during covid-19?," *Brain, behavior, and immunity*, 2020.

[21] S. K. Sharma, M. Mishra, and S. K. Mudgal, "Efficacy of cloth face mask in prevention of novel coronavirus infection transmission: A systematic review and meta-analysis," *Journal of education and health promotion*, vol. 9, 2020.

[22] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for covid-19 (sars-cov-2) pandemic: A review," *Chaos, Solitons & Fractals*, p. 110059, 2020.

[23] S. Agrebi and A. Larbi, "Use of artificial intelligence in infectious diseases," in *Artificial Intelligence in Precision Health*, pp. 415–438, Elsevier, 2020.

[24] M. H. D. M. Ribeiro, R. G. da Silva, V. C. Mariani, and L. dos Santos Coelho, "Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil," *Chaos, Solitons & Fractals*, p. 109853, 2020.

[25] L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, *et al.*, "An interpretable mortality prediction model for covid-19 patients," *Nature Machine Intelligence*, pp. 1–6, 2020.

[26] Z. Malki, E.-S. Atlam, A. E. Hassanien, G. Dagnew, M. A. Elhosseini, and I. Gad, "Association between weather data and covid-19 pandemic predicting mortality rate: Machine learning approaches," *Chaos, Solitons & Fractals*, vol. 138, p. 110137, 2020.

[27] L. K. Shrivastav and S. K. Jha, "A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of covid-19 in india," *Applied Intelligence*, pp. 1–13, 2020.

[28] M. J. Maloney, N. J. Rhodes, and P. R. Yarnold, "Mask mandates can limit covid spread: Quantitative assessment of month-over-month effectiveness of governmental policies in reducing the number of new covid-19 cases in 37 us states and the district of columbia," *medRxiv*, 2020.

[29] R. P. Lennon, S. M. Sakya, E. L. Miller, B. Snyder, T. Yaman, A. E. Zgierska, M. T. Ruffin, and L. J. Van Scoy, "Public intent to comply with covid-19 public health recommendations," *HLRP: Health Literacy Research and Practice*, vol. 4, no. 3, pp. e161–e165, 2020.

[30] S. Sathianathan, L. J. Van Scoy, S. M. Sakya, E. Miller, B. Snyder, E. Wasserman, V. M. Chinchilli, J. Garman, and R. P. Lennon, "Knowledge, perceptions, and preferred information sources related to covid-19 among healthcare workers: Results of a cross sectional survey," *American Journal of Health Promotion*, p. 0890117120982416, 2020.

[31] B. D. Weiss, M. K. Paasche-Orlow, *et al.*, "Disparities in adherence to covid-19 public health recommendations," *HLRP: Health Literacy Research and Practice*, vol. 4, no. 3, pp. e171–e173, 2020.

[32] N. Y. Times, "Mask-wearing survey data." https://github.com/nytimes/covid-19-data/tree/master/mask-use.

[33] A. Markowitz, "State-by-state guide to face mask requirements." https://www.aarp.org/health/healthy-living/info-2020/states-mask-mandates-coronavirus.html.

[34] USAFACTS, "Usa coronavirus cases and deaths." https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/state/oregon.

[35] USCensus, "United states census bureau." https://www.census.gov.

[36] USCensus, "United states census bureau." https://www.census.gov/library/visualizations/interactive/2014-2018-median-household-income-by-county.html.

[37] POLITICO, "Us 2020 presidential election results." https://www.politico.com/2020-election/results/president/.

[38] V. K. Ayyadevara, "Pro machine learning algorithms," *Apress: Berkeley, CA, USA*, 2018.

[39] W. Richert, *Building machine learning systems with Python*. Packt Publishing Ltd, 2013.

[40] O. Steinki and Z. Mohammad, "Introduction to ensemble learning," *Available at SSRN 2634092*, 2015.

[41] W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting hard rock pillar stability using gbdt, xgboost, and lightgbm algorithms," *Mathematics*, vol. 8, no. 5, p. 765, 2020.

[42] I. Gad and D. Hosahalli, "A comparative study of prediction and classification models on ncdc weather data," *International Journal of Computers and Applications*, pp. 1–12, 2020.

[43] Priyanka and D. Kumar, "Decision tree classifier: a detailed survey," *International Journal of Information and Decision Sciences*, vol. 12, no. 3, pp. 246–269, 2020.

[44] L. Deng and Y. Liu, *Deep learning in natural language processing.* Springer, 2018.