Radiology:Cardiothoracic Imaging

Natural Language Processing and Machine Learning for Detection of Respiratory Illness by Chest CT Imaging and Tracking of COVID-19 Pandemic in the US

Ricardo C. Cury, MD[1,2,3], Istvan Megyeri, MBA[1], MS, Tony Lindsey, PhD [1], Robson Macedo, MD[1,2,3], Juan Batlle, MD[1,2,3], Shwan Kim MD[4], Brian Baker BS[4], Robert Harris, PhD[4], Reese H. Clark MD[5]

[1]MEDNAX Radiology Solutions, Sunrise, FL
[2]Department of Radiology, Herbert Wertheim College of Medicine, Florida International University, Miami, FL
[3]Miami Cardiac and Vascular Institute, Baptist Health South Florida, Miami, FL
[4]Virtual Radiologic, Eden Prairie, MN
[5]MEDNAX Center for Research, Education, Quality and Safety, Sunrise, FL

A preliminary version of this manuscript is available online through SSRN's First Look and Preprints with the Lancet, a place where journals identify content of interest prior to publication. These preprints are early stage research papers that have not been peer-reviewed or approved for publication.

Address all correspondence to:
Ricardo C. Cury, MD, FACR, FAHA, FACC, MSCCT
Professor and Chairman of Radiology
Herbert Wertheim College of Medicine - Florida International University
Chairman of Radiology and Director of Cardiac Imaging
Baptist Health South Florida
Miami Cardiac & Vascular Institute
Chief Medical Officer - MEDNAX Radiology
8900 N. Kendall Drive, Miami, FL 33176

1

Email: rcury@baptisthealth.net

Abstract

Background: Coronavirus disease 2019 (COVID-19) has spread quickly throughout the United States (US) causing significant disruption in healthcare and society. Tools to identify hot spots are important for public health planning. The goal of our study was to determine if natural language processing (NLP) algorithm assessment of thoracic computed tomography (CT) imaging reports correlated with the incidence of official COVID-19 cases in the US.

Methods: Using de-identified HIPAA compliant patient data from our common imaging platform interconnected with over 2,100 facilities covering all 50 states, we developed three NLP algorithms to track positive CT imaging features of respiratory illness typical in SARS-CoV-2 viral infection. We compared our findings against the number of official COVID-19 daily, weekly and state-wide.

Results: The NLP algorithms were applied to 450,114 patient chest CT comprehensive reports gathered from January 1st to October 3rd, 2020. The best performing NLP model exhibited strong correlation with daily official COVID-19 cases (r^2=0.82, p<0.005). The NLP models demonstrated an early rise in cases followed by the increase of official cases, suggesting the possibility of an early predictive marker, with strong correlation to official cases on a weekly basis (r^2=0.91, p<0.005). There was also substantial correlation between the NLP and official COVID-19 incidence by state (r^2=0.92, p<0.005).

Conclusion: Using big data, we developed a novel machine-learning based NLP algorithm that can track imaging findings of respiratory illness detected on chest CT imaging reports with strong correlation with the progression of the COVID-19 pandemic in the US.

KEYWORDS:

natural language processing; machine learning; big data; public health; viral outbreak; computed tomography; chest CT

ABBREVIATIONS:

Coronavirus disease 2019 (COVID-19); natural language processing (NLP); United States (US); computed tomography (CT); Reverse transcription polymerase chain reaction (RT-PCR); artificial intelligence (AI); machine learning (ML); Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE);

Key Points:

- Our work uses big data and a machine learning based natural language processing (NLP) model to rapidly identify and monitor imaging findings of respiratory illness by chest CT interpretation within the US during the COVID-19 pandemic.

- The NLP models indicated strong correlation with the number of new COVID-19 cases daily, weekly and on a state by state level.

- The NLP models could detect image findings of respiratory illness early in the pandemic, followed by the rise of new official COVID-19 cases in the US.

Summary Statement:

We developed a novel machine-learning based NLP algorithm to track respiratory illness detected on chest CT with strong correlation to the progression of the COVID-19 pandemic in the US.

Introduction

The pandemic of coronavirus disease 2019 (COVID-19)[1] has spread quickly throughout the United States (US) causing a significant disruption in healthcare[2,3] and society[4]. Several countries around the world, including the US, have been in shutdown for weeks to "flatten the curve" in order to mitigate COVID-19 disease spread and prevent local outbreaks that could threaten lives and overwhelm hospitals. Medical analytic tools that assess the early rise of possible infection and identify hot spots are important for public health planning and management. Inadequate availability of accurate testing, early identification and case tracking were failure points in the initial US[5] pandemic response , resulting in more than 1.5 million cases and over 90,000 deaths as of May 18, 2020 and progressing to more than 7.5 million cases and over 215,000 deaths as of October 3, 2020[6]. Reverse transcription polymerase chain reaction (RT-PCR) nasal swab is the reference standard molecular test for COVID-19 diagnosis confirmation: however, an important issue with RT-PCR test is the risk of false-negative or false-positive results[7]. In particular, studies have shown that RT-PCR can lead to false-negative results due to insufficient collection of material or due to testing at a stage of the disease with lower viral shedding[8]. There is increasing evidence supporting the importance of chest CT interpretation for suspected COVID-19 patients[9-11] and its role in the assessment of false negative RT-PCR results[8,12,13]. Although chest CT imaging is not recommended for mass initial triage, proper use of chest CT can serve as an important evaluation component in patients presenting with moderate to severe symptoms and in several specific scenarios, such as defining the severity of disease for determining admission and appropriate level of care, providing an alternative differential diagnosis, or evaluating patients with worsening symptoms who may have treatable complications of progressive COVID-19 (e.g., pulmonary embolism)[14,15]. Typical chest CT findings have been described in COVID-19 patients with high sensitivity (>90%)[13]. Lung

abnormalities during the early course of COVID-19 usually are peripheral multi-focal ground-glass opacities affecting both lungs and, as the disease progresses, air-space opacities and crazy paving (intra-lobular septal thickening) become more common CT findings[16]. These typical chest CT imaging findings can be helpful in early diagnosis of suspected cases and in evaluating the severity and disease extent.

Progressive respiratory failure due to severe acute respiratory syndrome is the primary cause of death in the COVID-19 pandemic[17]. Recently, work with artificial intelligence (AI) and machine learning (ML) demonstrates promising results in predicting COVID-19 spread and distinguishing it from other diseases such as pneumonia[18-20]. Robust AI and ML strategies utilizing big data capable of extracting image findings from scans performed throughout the US with search engines relying on natural language processing (NLP)[21] are needed. The goal of our study was to determine if a machine learning based NLP algorithm could perform syntactic analysis of chest CT imaging radiology reports to extract keywords for generating predictions, and to compare such predictions with officially reported COVID-19 cases and deaths in the US.

Methods

Using de-identified reports from our common imaging platform, which is interconnected with over 2,100 facilities throughout all 50 US states, we developed three NLP algorithms to track and display positive CT imaging features of respiratory illness and pulmonary findings that are typical of COVID-19. The NLP development was performed as part of an internal quality improvement project and to track and monitor the progression of respiratory illness by chest CT findings, specifically during a national pandemic. IRB was waived and patient consent was not required. This study was monitored and approved by our internal quality and safety committee. Our nationwide radiology practice provides interpretation of more than 20,000 examinations per

day and greater than 1,000 chest CT studies per day using the same common imaging platform and dictation system.

To collect data used to train the NLP models, a list of keywords was selected to identify possibly relevant reports for each model. For the first NLP model, named "Viral Pneumonia NLP", the keywords used were "ground-glass opacities", "crazy paving", "viral pneumonia", "viral infection", "covid", "sars-cov", and "coronavirus". The second NLP model, named "Imaging Findings NLP," used the keywords "bilateral ground-glass opacities", "bilateral hazy opacities" and "crazy paving". The third NLP model, named "COVID NLP," used the keywords "viral pneumonia", "atypical pneumonia", "viral infection", "covid", and "coronavirus". We selected the above three NLP models in order to compare a more general NLP algorithm including findings for viral pneumonia features ("Viral Pneumonia NLP"); one more specific NLP for typical imaging findings present in patients with COVID-19 ("Imaging Findings NLP") and the last NLP to include specific words in the interpretation of free text reports indicating viral pneumonia and COVID-19 incidence (COVID NLP). All Chest CT reports from Jan 1, 2020 to March 15, 2020 were screened for these keywords as a training dataset and any report containing them was extracted. The "Findings" and "Impressions" sections of these clinical reports were split into sentences and pre-processed to remove non-alphanumeric characters. For each NLP model, these sentences were labeled as positive or negative for the pathology. Sentences were also labeled as negative if they contained any of the following keywords: tree-in-bud, air-space trapping, centrilobular, pleural effusion, atelectasis, and masses. Once sentences for a model were labeled, they underwent similar NLP methodology to prior published work from our group[22]. The keyword extraction NLP uses an unsupervised machine learning approach, i.e. the algorithm doesn't require training on a corpus nor any pre-defined rules, dictionary, or thesaurus.

Instead, statistical features from the text itself are used and as such can be applied to large reports easily without re-training. The pipeline for extracting key phrases was to preprocess the radiologist report document to remove less informative common words and punctuation. These three NLP models were then wired into our common imaging platform. A filter was first applied to screen for Chest CT reports containing any keyword relevant to an NLP model; when a keyword was detected, the NLP model was applied to that report to determine whether sentence context was positive or negative for the given pathology. Similar to training, the NLP models extract the Findings and Impressions report sections and pre-process the sentences prior to analysis. These NLP models were then applied to all clinical reports that passed through our system from January 1, 2020 to October 3rd, 2020 and the results were saved in our database.

All Chest CT imaging studies were performed as part of clinical care in patients following typical indications from the American College of Radiology for the use of chest CT. Structured reporting is used in our common imaging platform. Radiologists received educational training regarding the typical imaging findings present in patients with suspected or confirmed Covid-19 infection. For data visualization and analysis, we used Microsoft Power BI, Power View (Microsoft Corporation, Redmond, WA) and R (R Foundation). Power View engine utilized machine learning to synchronize the data and display the data in real time and this software was used to create a temporal data display with a Geo map of the United States (Video 1). GITHUB (subsidiary of Microsoft which provides hosting for software development and data repository) served as the data repository. Three authors (RCC, IM and TL) had full access to the data and performed the analysis.

In order to track the rate of reported COVID-19 cases and deaths in the United States over time, we used the same data repository that is being used by the COVID-19 Data Repository by the

Center for Systems Science and Engineering at Johns Hopkins University or "JHU CSSE COVID-19 Data", which has been previously validated as a credible source to track new Covid-19 cases and deaths[23] and is available at the following website: https://github.com/CSSEGISandData/COVID-19.

Data sources and normalization methodology

Data from several sources, including information from our common imaging platform, developed NLP models and JHU CSSE COVID-19 Data, were acquired and merged as a requisite input for our covid-19 prediction algorithm. R Studio desktop version 1.3.1093 IDE was utilized to develop R markdown source code. JHU and common imaging platform data sources in the form of csv files were read and stored as dataframe structures consisting of 2 million rows and 20 columns. A series of date format conversions were applied to resolve temporal differences and duplicate patient records were identified and removed. Additionally, absent common imaging platform patient county information was generated by mapping zip codes to (county, state) pairs.

A normalization methodology was established to mitigate test result misrepresentation due to nonuniform exam occurrence across states. A country benchmark was established by dividing the total Chest CT Imaging read count by the US 2019 census population and applying a "per million" scaling factor. Local ratios were computed by dividing total state Chest CT reads by state population. A weight index was computed for each state and multiplied by state findings to determine normalized representation of adjusted reads. Additional information is provided on the supplemental material tab 3 regarding the normalization process, variables and formula.

Forecast and Future prediction of new cases

Finally, we utilized two third-party commercially software available for Microsoft Power BI users, to generate forecast models based on our NLP algorithms in order to predict new daily COVID-19 cases on a prospective basis.

The first forecast model, which was developed by MAQ software and is available to Microsoft Power BI users (https://appsource.microsoft.com/en-us/product/power-bi-visuals/wa104381845?tab=overview), utilized linear regression and neural network to observe, analyze and learn from the following inputs: NLP models developed by our team and historical data from the JHU database of historical official COVID-19 cases (independent variables) to predict future new daily cases (dependent variable). Since the model provides capability to adjust learning model parameters, variable weights and biases are continuously optimized using a machine learning Gradient descent algorithm to minimize output error of model predictions. The team trained the machine learning predictive model for one month to ensure data input accuracy and predictive model performance.

The second forecast model utilizes the forecasting TBATS method to model time series data on a prospective basis. TBATS uses the R statistical programming language and is an extension of Power BI visualization. It is downloaded from (https://github.com/Microsoft/PowerBI-visuals-forcasting-tbats/tree/full). The acronym TBATS indicates key model features, i.e. Trigonometric seasonality, Box-Cox transformations for heterogeneity, ARMA errors for short-term dynamics, Trend, Seasonal components. The model's main aim is to forecast time series with complex seasonal patterns using exponential smoothing. The forecasting TBATS model handled historical daily volume of COVID-19 cases from the JHU database and the developed NLP models by our team as inputs and monthly seasonality with no constraints to create detailed, long-term

forecasts. The TBATS predictive model was trained for one month to ensure data input accuracy and the predictive model performance.

Statistical Analysis

First, we calculated a correlation matrix using Pearson's correlation (r) to demonstrate the correlation among the number of cases per day over time for the different NLP models and compared them against the number of official cases per day over time and deaths per day over time of COVID-19 in the US. Then, we performed linear regression analysis using R (Version 4.0.0) to correlate the number of cases per day over time for the different NLP models as independent variables and compared them against the number of official cases per day over time and deaths per day over time of COVID-19 in the US as dependent variables. A similar analysis was performed aggregating the data on a weekly basis over time and displayed the data as a figure. We also calculated the correlation coefficient between the number of cases detected by our NLP models and the number of new official COVID-19 cases on a state level, including all 50 states, Washington D.C. and Puerto Rico. We used Data Analysis Expressions (DAX) and R for language visualization. We used the Microsoft Excel analysis tool pack VBA to build the correlation matrix, for the correlation analysis and ANOVA to test model significance.

Results:

The three NLP algorithms were applied to 450,114 chest CTs performed from January 1st, 2020 to October 3rd, 2020. There were 107,120 positive cases (23.8%) flagged by the "viral pneumonia NLP", 22,267 cases flagged by the "imaging findings NLP" (4.9%) and 21,202 cases flagged by the "COVID NLP" (4.7%). The correlation matrix is presented in Table 1. We demonstrate that the "viral pneumonia NLP", "imaging findings NLP" and "COVID NLP" had a

correlation (r) of 0.37, 0.81 and 0.91 with the official number of COVID-19 cases, respectively, with overall higher correlation for the more specific models ("viral pneumonia NLP" < "imaging findings NLP" < "COVID NLP"). The "imaging findings NLP" and "COVID NLP" had a correlation of 0.29 and 0.49 with the official number of COVID-19 deaths. There was no correlation between the "viral pneumonia NLP" and number of COVID-19 deaths. We can also observe that there was a very good correlation between the "imaging findings" and "COVID NLP" with correlation coefficient of 0.91 and a moderate correlation between "viral pneumonia" and "Imaging Findings NLP" with correlation coefficient of 0.63.

Daily Correlation

Based on the correlation matrix, we selected the two best NLP models ("imaging findings NLP" and "COVID NLP") to perform linear regression analysis to compare with the number of official cases and deaths of COVID-19. All the details of the regression analysis are displayed in the Supplementary appendix. The "COVID NLP" had a strong correlation with the number of official COVID-19 cases per day over time (r^2=0.82; p-value<0.005). The "imaging findings NLP" had a good correlation with the number of COVID-19 cases (r^2=0.66; p-value<0.005). The "viral pneumonia NLP" had a weak correlation with the number of COVID-19 cases (r^2=0.13).

Weekly correlation

We selected the two best performing NLPs ("imaging findings NLP" and "COVID NLP") to display the data in week by week from the beginning of the year to compare the progression of the models with the number of official COVID-19 cases and deaths (Figure 1). The NLP models had a material increase in number of cases during the first wave starting in early March (week 11) and peaking in late March/early April (week 14). There was a strong correlation with the

12

"COVID NLP" when compared to the number of new official COVID-19 cases on a weekly basis (r^2=0.91, p<0.005). It is important to note that the early rise of cases detected by both NLP models occurred before the rise of new official COVID-19 cases during the first wave.

State correlation

There was an average of 540 chest CT studies analyzed in our platform per 1 million residents per state during the study period. The correlation coefficient was 0.88 for the "COVID NLP" and 0.89 for the "imaging findings NLP" when compared to new official COVID-19 cases on a state level, including all 50 US states, Washington DC and Puerto Rico (Figure 2). There was a strong correlation with the "COVID NLP" when compared to new COVID-19 cases by state (r^2=0.92, p<0.005). The top states in number of cases identified by our more specific NLP model (COVID NLP) were: Texas, California, Florida, New York and Georgia, as displayed in Figure 2.

Temporal correlation

Finally, we tracked the progression of new COVID-19 cases in all 50 states using a real-time interactive US map model displaying the temporal relationship between the COVID-19 pandemic and our NLP models. This interactive real-time temporal model was created using a machine-learning algorithm with live input from our NLP models connected with our imaging platform and a direct feed from the Johns Hopkins University Center for Systems Science and Engineering data repository (Figure 3 panel and Video 1).

Forecast and future prediction of new cases

The first forecast model based on machine learning is presented in Figure 4 and can predict the next 10 days. As an example, as of November 20th, 2020 the model was predicting 190,858 new COVID-19 cases on November 26th, 2020 (CI = 167,197 - 205,330). The second forecast model

utilizes forecasting methodology to model time series data with complex seasonal patterns using exponential smoothing with no constraints to create detailed, long-term forecasts. The second model is presented in Figure 4B and can predict a longer period with appropriate confidence intervals. As an example, as of November 20th, 2020 the model was predicting 166,412 new COVID-19 cases on December 19th, 2020 (CI = 125,920 - 206,903). The actual JHU COVID-19 number of cases on December 19th was 193,947, which is within the confidence interval.

Discussion

The COVID-19 pandemic is challenging our society and economy in an unprecedented way. Overall, the US response to contain COVID-19 may not have been as effective as other countries. This may have been due to insufficient or delayed testing and lack of alternative monitoring tools near the beginning of the pandemic[5]. Early warning and detection may represent a critical opportunity for the US to track the rate of respiratory illness and quickly institute policies to prevent or at least mitigate a future outbreak.

Strategies such as testing, contact tracing and isolation of people who test positive will also be essential to successfully reopening state economies and keeping them open. Moreover, we believe it will also be fundamental to have tools to rapidly identify and track the geographically-disproportionate emergence of respiratory illness and thereby identify hot spots of infection. Accelerated insights can be derived from aggregated data using machine learning and NLP[21,24] The present work uses big data and a machine learning based NLP model to rapidly identify and monitor rates of respiratory illness as identified by chest CT imaging, based on key imaging findings previously reported in patients with COVID-19 infection. This work is distinctive because our common imaging platform is connected to over 2,100 facilities representing all 50 states. This provides an unparalleled opportunity to gather big data in real time that can be an

14

accurate representation of regional and even local trends across the country. One of the main findings of our work is that the NLP models were able to detect imaging findings of respiratory illness before the rise of new official COVID-19 cases during the beginning of the pandemic. This is an interesting observation demonstrating the potential for this surveillance algorithm to flag respiratory illness as an early predictor of new COVID-19 cases and subsequent attendant mortality. Moreover, the NLP models had a strong correlation with the number of official new COVID-19 cases on a weekly level and on a state level.

The three NLP models used reflected a wide range of scenarios that could provide guidance to radiology departments in providing meaningful insights to Hospitals and communities they serve. The first NLP model "viral pneumonia NLP" was a more general algorithm for viral infection including CT findings that were suggestive of viral infection, however not necessarily viral infections that were specifically suggestive of COVID-19 and were quite prevalent even during the pre-pandemic season. The second NLP model "imaging findings NLP" focused on CT findings that were suggestive specifically of COVID-19. The third NLP model "COVID NLP" represented words of an interpretation of the radiologist as the findings being suggestive of viral infection or highly suggestive of COVID-19. It is interesting to note that the "COVID NLP" model was the best performing model with all correlations performed and highlights the ability of Radiologists to summarize their findings in the impression and provide additional insights to clinical care. Other key words, as presented by CO-RADS[25], should be considered for further correlation in future NLP studies and this study serves as a baseline for such studies.

Real-time epidemiological data are critical to managing different aspects of a pandemic. For instance, this data can help public health authorities to forecast demand/surge models, which may thereby allow public or private organizations to quickly reposition resources or reallocate

personnel. This is corroborating data that should be used in combination with other indicators, such as the officially reported number of newly positive laboratory tests, disease-related hospitalizations and disease-related deaths. Because our data is a marker of respiratory illness and of findings typical for viral pneumonia, it can serve as an additional indicator to predict the need for emergency medical services, hospital staff, hospital beds, personal protective equipment and ventilator equipment, among others. Most cases of COVID-19 are mild[26]; for this reason early identification of a small excess number of severe and critical cases is crucial for planning hospital resource allocation. The extent and magnitude of imaging findings have been recently reported to correlate with worse outcomes and prognosis[15,27].

Our NLP model output may have included non-COVID viral pneumonia cases such as influenza, adenovirus and other atypical pneumonias. However, we can clearly see the rapid rise of respiratory illness detected by our NLP models starting in early March when the COVID-19 pandemic started to progress rapidly in the US. Therefore, while the absolute numbers could have included non-COVID-19 pneumonia cases, the relative increase and decrease in the total number of Chest CT studies demonstrating respiratory illness and viral pneumonia may be very helpful to detect trends during an epidemic.

Limitations

As expected, we do not have access to all patients who have chest CT imaging in the United States, however our sample size is representative and provides an inclusive sample of all 50 states. Another limitation is the lack of a gold standard to confirm COVID-19 status on a patient undergoing chest CT, since an HL7 interface to provide laboratory results for the patients was not available in all cases. The COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University is a credible source of data for new cases and

deaths due to COVID-19 and the US data is provided on a city level by the Centers for Disease Control and Prevention. Nevertheless, the utility of an early warning system using chest CT findings may in fact shine in the absence of laboratory data, as chest CT abnormalities can identify a regional spike in respiratory illness before a virus has even been isolated or, if already isolated, if viral testing is not yet widely available.

Conclusion

In conclusion, we developed a novel machine-learning based NLP algorithm to track respiratory illness imaging findings of viral pneumonia detected on chest CT and displayed as real-time data with strong correlation to the progression of the COVID-19 pandemic in the US. This nationwide surveillance algorithm has the potential to help health care entities and public health authorities develop strategies against COVID-19 and other similar pandemics in the future. Future work will be required to further validate predictive forecast models based on NLP findings.

References:

1.   World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19. March,11th 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

2.   Tanne JH, Hayasaki E, Zastrow M, Pulla P, Smith P, Rada AG. Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. BMJ 2020;368:m1090.

3.   Nelson R. COVID-19 disrupts vaccine delivery. Lancet Infect Dis 2020;20:546.

4.   Scott R. Baker NB, Steven J. Davis, and Stephen J. Terry. COVID-INDUCED ECONOMIC UNCERTAINTY. NATIONAL BUREAU OF ECONOMIC RESEARCH 2020.

5.   Schneider EC. Failing the Test - The Tragic Data Gap Undermining the U.S. Pandemic Response. N Engl J Med 2020.

6.   Worldmeter COVID-19 Coronavirus Pandemic. May 18th, 2020. https://www.worldometers.info/coronavirus/

7.   Wang Y, Kang H, Liu X, Tong Z. Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. J Med Virol 2020.

8.   Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing. Radiology 2020:200343.

9.   Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. Lancet Infect Dis 2020;20:425-34.

10.   Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. AJR Am J Roentgenol 2020:1-7.

11.   Bao C, Liu X, Zhang H, Li Y, Liu J. Coronavirus Disease 2019 (COVID-19) CT Findings: A Systematic Review and Meta-analysis. J Am Coll Radiol 2020.

12.   Huang P, Liu T, Huang L, et al. Use of Chest CT in Combination with Negative RT-PCR Assay for the 2019 Novel Coronavirus but High Clinical Suspicion. Radiology 2020;295:22-3.

13.   Ai T, Yang Z, Hou H, et al. Correlation of Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. Radiology 2020:200642.

14.   American College of Radiology Guidelines for the Use of Chest Radiograph and Computed Tomography for Suspected Covid-19 Infection. March 22th, 2020. https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection.

15.   Zhao W, Zhong Z, Xie X, Yu Q, Liu J. Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study. AJR Am J Roentgenol 2020;214:1072-7.

16.   Bai HX, Hsieh B, Xiong Z, et al. Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. Radiology 2020:200823.

17.   Ackermann M, Verleden SE, Kuehnel M, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. N Engl J Med 2020.

18.   Li L, Qin L, Xu Z, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. Radiology 2020:200905.

19.   Zheng N, Du S, Wang J, et al. Predicting COVID-19 in China Using Hybrid AI Model. IEEE Trans Cybern 2020.

20.   Alimadadi A, Aryal S, Manandhar I, Munroe PB, Joe B, Cheng X. Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics 2020;52:200-2.

21.   Hirschberg J, Manning CD. Advances in natural language processing. Science 2015;349:261-6.

22.   Harris RJ, Kim S, Lohr J, et al. Classification of Aortic Dissection and Rupture on Post-contrast CT Images Using a Convolutional Neural Network. J Digit Imaging. 2019 Dec;32(6):939-946.

23.   Dong E, Du H, Gardner L. An interactive web-based dasboard to track COVID-19 in real time. Lancet Inf Dis. 20(5):533-534

24.   Bragazzi NL, Dai H, Damiani G, Behzadifar M, Martini M, Wu J. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. Int J Environ Res Public Health 2020;17.

25.   Prokop M, van Everdingen W, van Rees Vellinga T, et al.; COVID-19 Standardized Reporting Working Group of the Dutch Radiological Society. CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19-Definition and Evaluation. Radiology. 2020 Aug;296(2):E97-E104.

26.   Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020.

27.   Colombi D, Bodini FC, Petrini M, et al. Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. Radiology 2020:201433.

Table

Table 1. Correlation matrix demonstrating correlation among the variables.

Cases by Day	NLP Covid	NLP Image Finding	NLP Viral Pneumonia	Daily cases	Daily deaths
NLP Covid	1				
NLP Image Finding	0.91	1.00			
NLP Viral Pneumonia	0.46	0.63	1.00		
Daily cases	0.91	0.81	0.37	1.00	
Daily deaths	0.47	0.29	NA	0.51	1.00

Figures
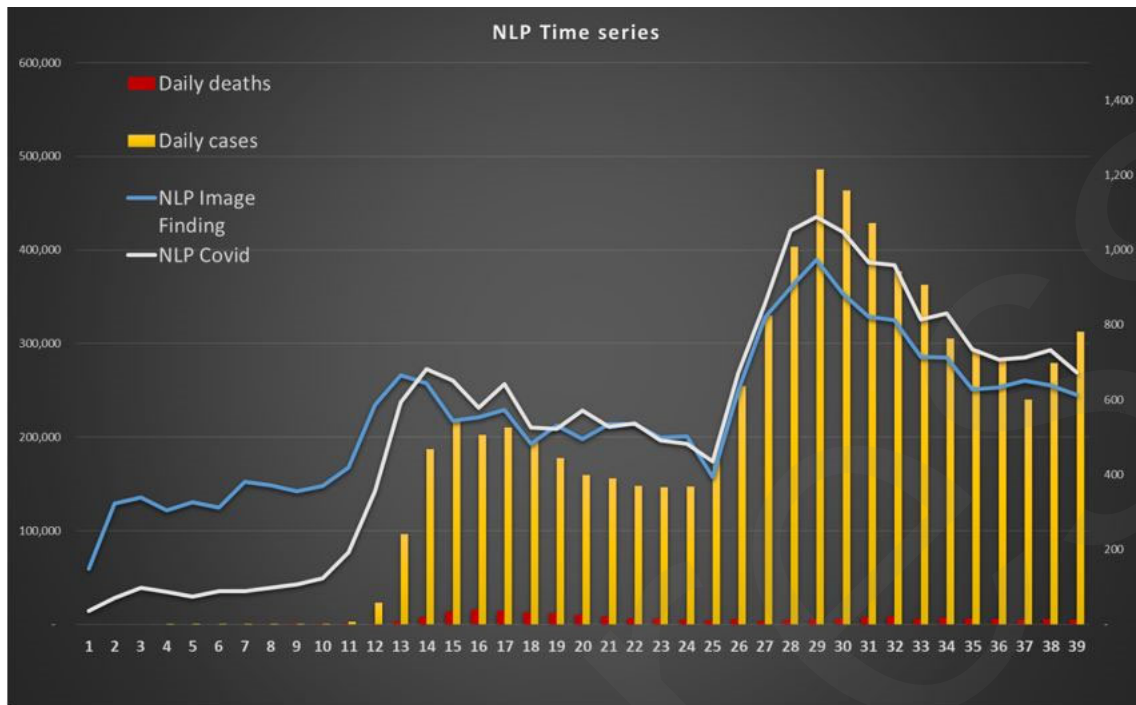
Figure 1. Weekly time series - Correlation of the two NLP models (Imaging findings NLP and COVID NLP) versus number of official COVID-19 cases per week: Temporal course on a week by week basis demonstrating the relationship between the progression and early rise of the NLP positive Chest CT studies, followed by the increase in the number of official COVID-19 cases and subsequent increase in COVID-19 deaths. There was a strong correlation with the COVID NLP when compared to the number of official COVID-19 cases on a weekly basis (r^2=0.91, p<0.005).
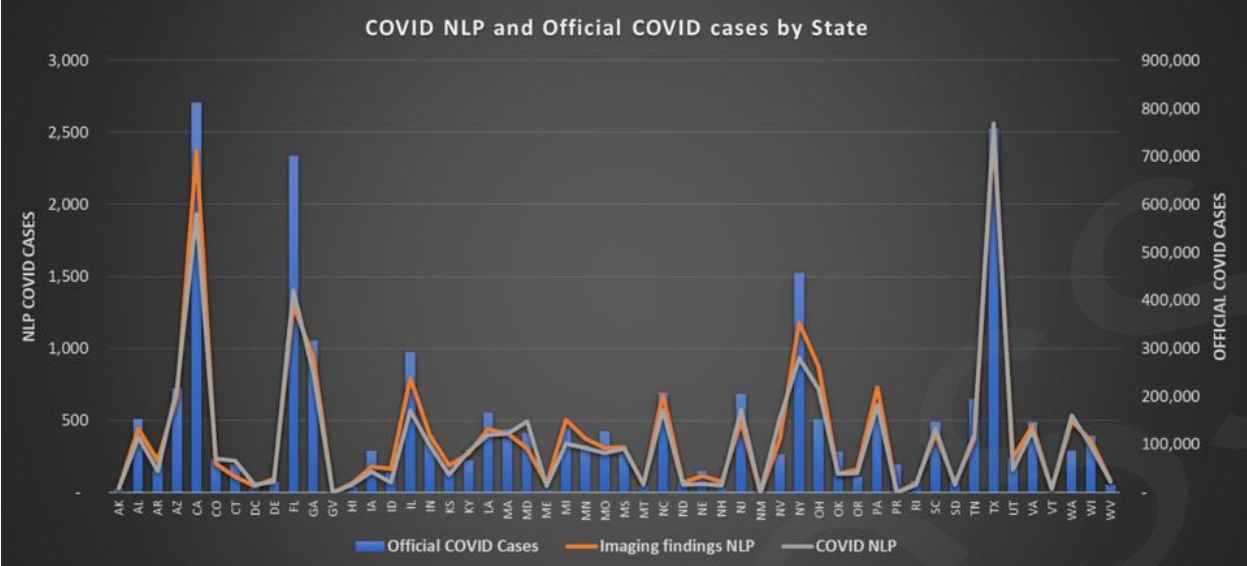
Figure 2. Official COVID-19 cases and NLP models by state: Correlation between the number of cases detected by the NLP models and the number of official COVID-19 cases on a state level. All 50 US states, Washington D.C. and Puerto Rico are included in the analysis. There was a strong correlation with the COVID NLP model when compared to new COVID-19 cases by State (r^2=0.92, p<0.005).

Figure 3 panel and Video 1. Temporal progression using a Machine-Learning based geomap comparing the rise overtime of the Viral pneumonia NLP (more general NLP) and COVID NLP (more specific NLP) with the number of official COVID-19 cases (Period January 1st, 2020 to October 3rd, 2020).
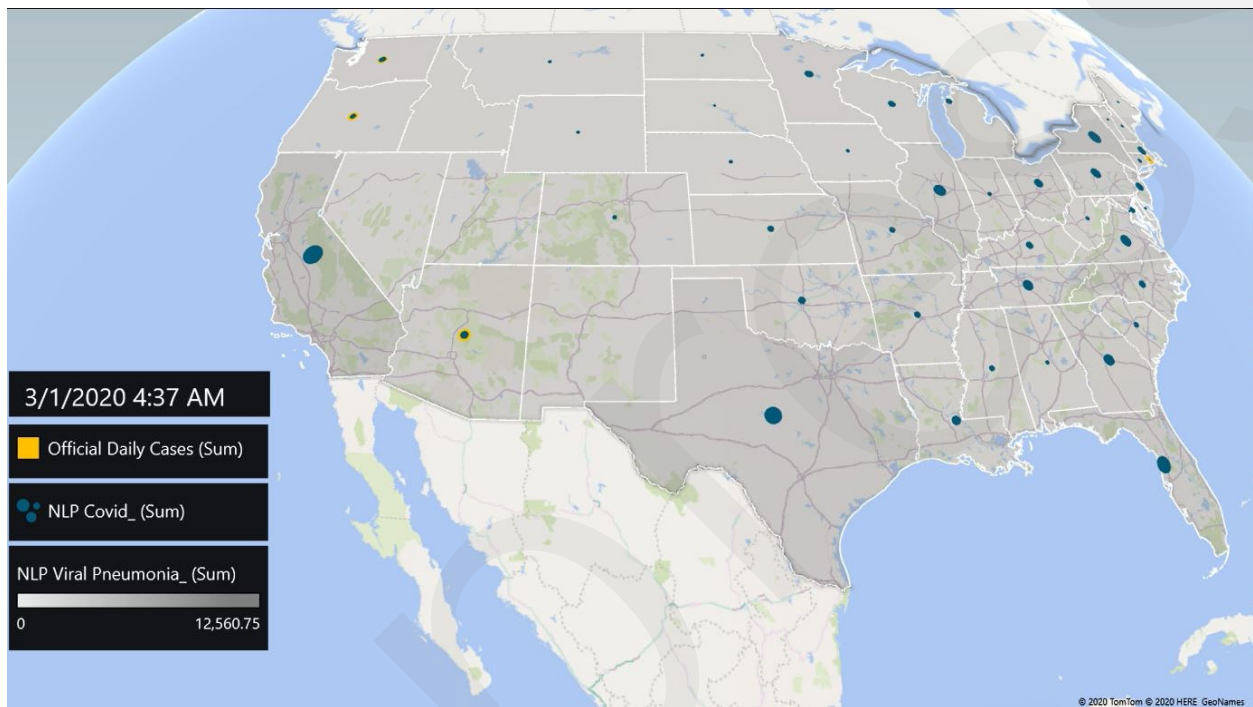
Figure 3A - March 1st snapshot: note below the progression of the Viral pneumonia NLP cases (gray areas in the states) which correlates with the flu season and the appearance of the COVID NLP cases (blue circles), as well as slightly increase in the number of official COVID-19 cases (yellow circles).
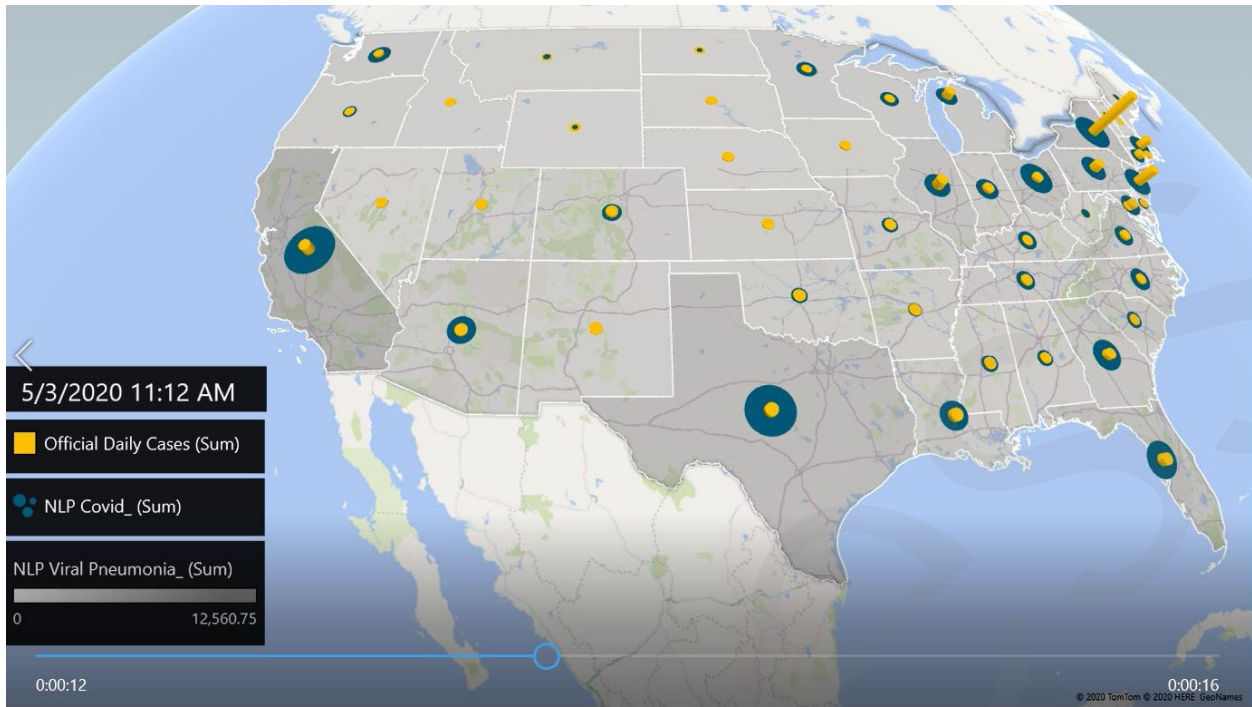
Figure 3B - May 3rd snapshot: note below continuous increase of the COVID NLP cases (blue circles enlarging) and now we see a material increase in the number of official COVID-19 cases (yellow circles growing vertically), particularly in New York and the northeast of the US.
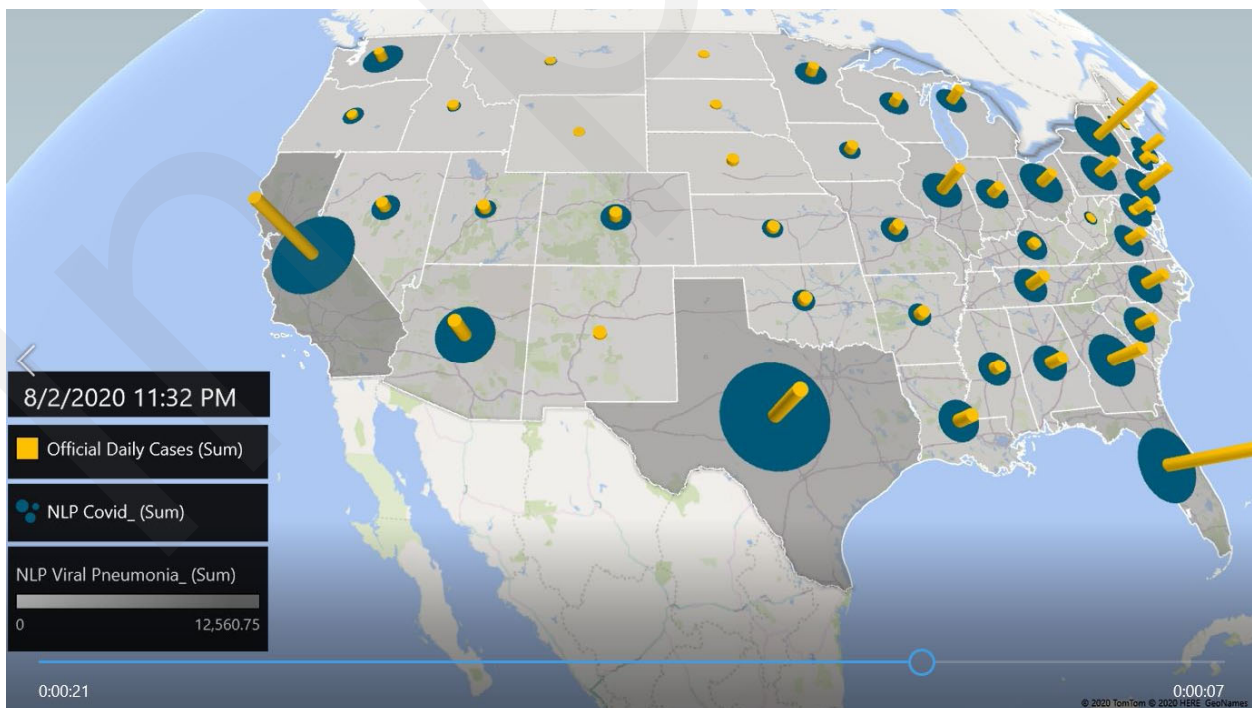
Figure 3C - August 2nd snapshot: note below the continuous increase of the COVID NLP cases particularly in Florida, California and Texas (blue circles enlarging) and a continuous material increase in the number of official COVID-19 cases (yellow circles growing vertically). Please note the correlation between the size of the blue circles and the height of the yellow vertical bars in each state.
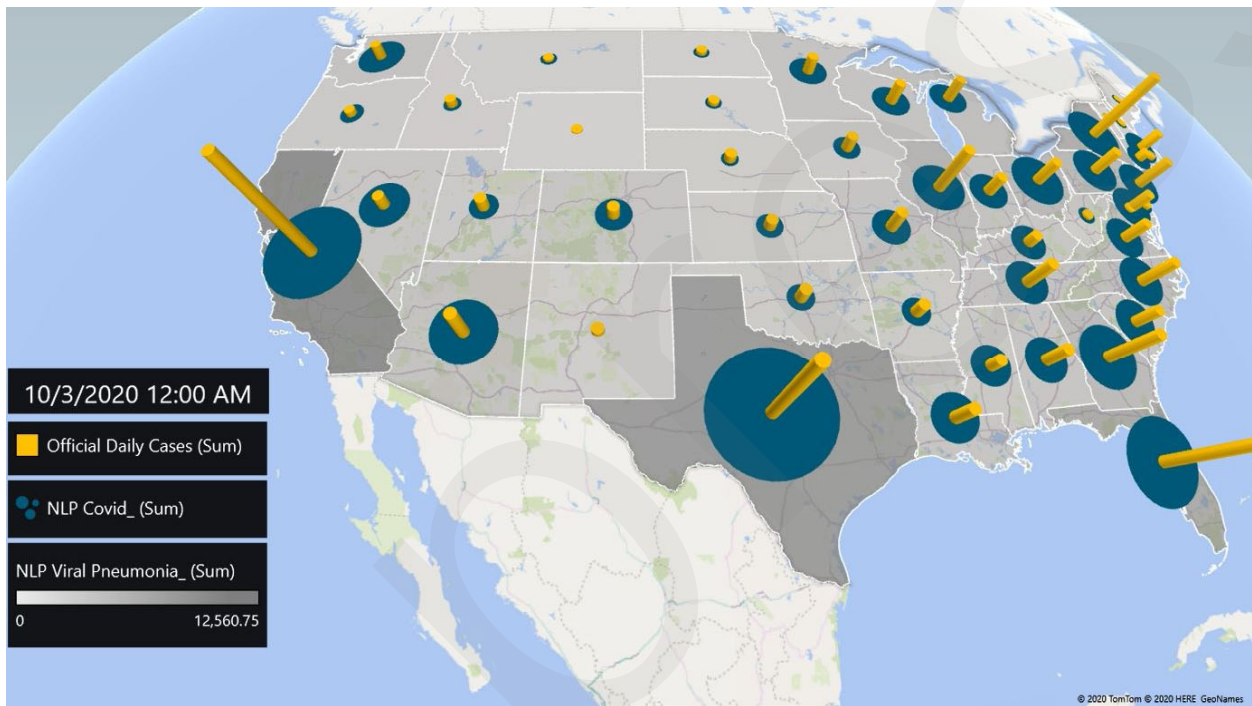

Figure 3D - October 3rd snapshot: note below the continuous increase of COVID NLP cases (blue circles enlarging) and a continuous material increase in the number of official COVID-19 cases (yellow circles growing vertically), including the mid-west of the US. Please note the correlation between the size of the blue circles and the height of the yellow vertical bars in each state.
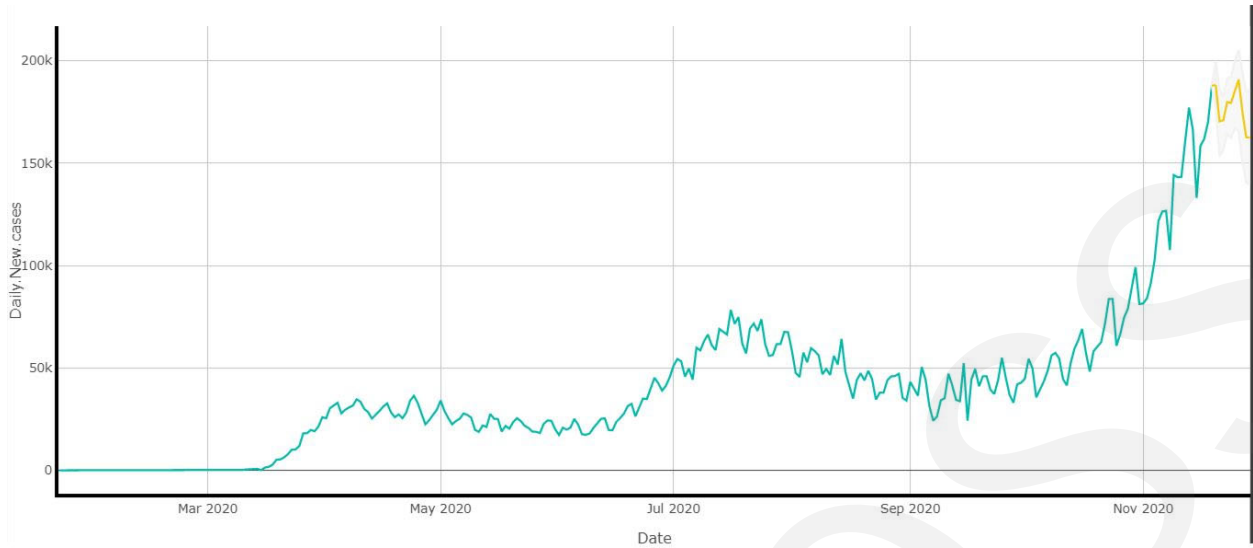
Figure 4A – First forecast model based on an artificial neural network to learn from the NLP models and historical data (independent variables) to predict future new daily COVID-19 cases (dependent variable).
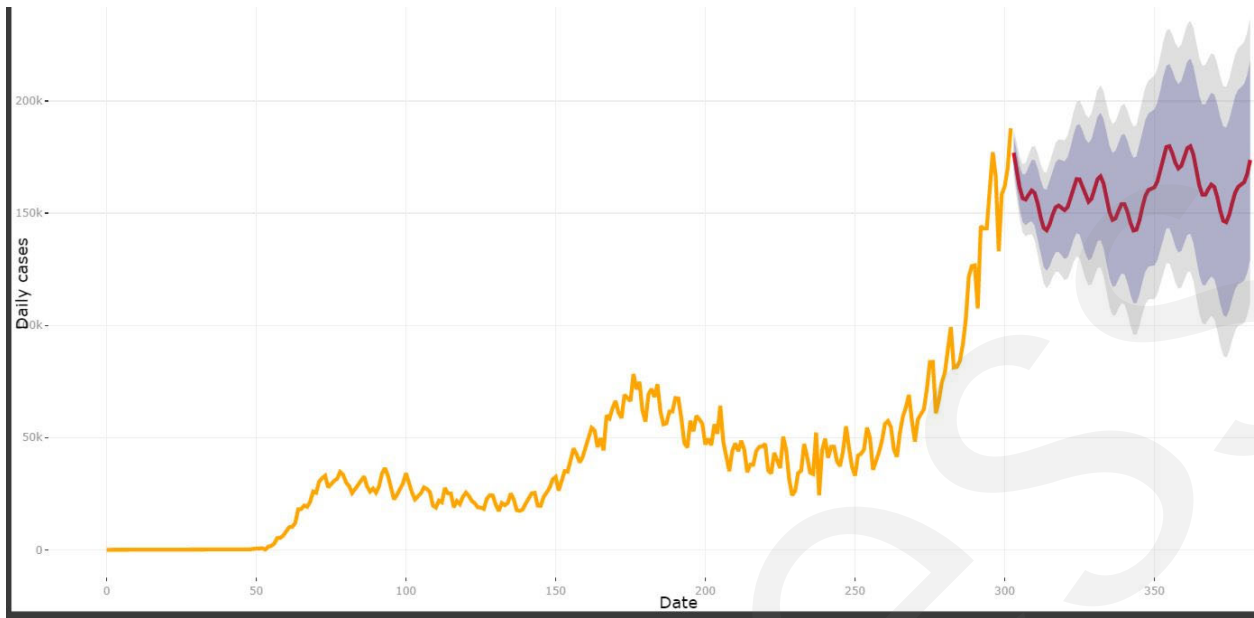
Figure 4B – Second model utilizes forecasting methodology to model time series data with complex seasonal patterns using exponential smoothing with no constraints to create detailed, long-term forecasts**.**

Supplementary Material –

The summary output for the linear regression equations are provided below:

1. Predict the number of new official COVID-19 cases based on the COVID NLP

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.907353798
R Square	0.823290915
Adjusted R Square	0.822645991
Standard Error	9004.648387
Observations	276

ANOVA

	df	SS	MS	F	Significance F
Regression	1	1.03509E+11	1.03509E+11	1276.57109	3.9721E-105
Residual	274	22216931766	81083692.57		
Total	275	1.25726E+11			

|	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-5864.157681	1060.784033	-5.52813532	7.5404E-08	-7952.480381	-3775.834981
NLP Covid	424.1324239	11.87077221	35.72913509	3.972E-105	400.7629139	447.501934

Liner regression analysis

NLP Covid Line Fit Plot
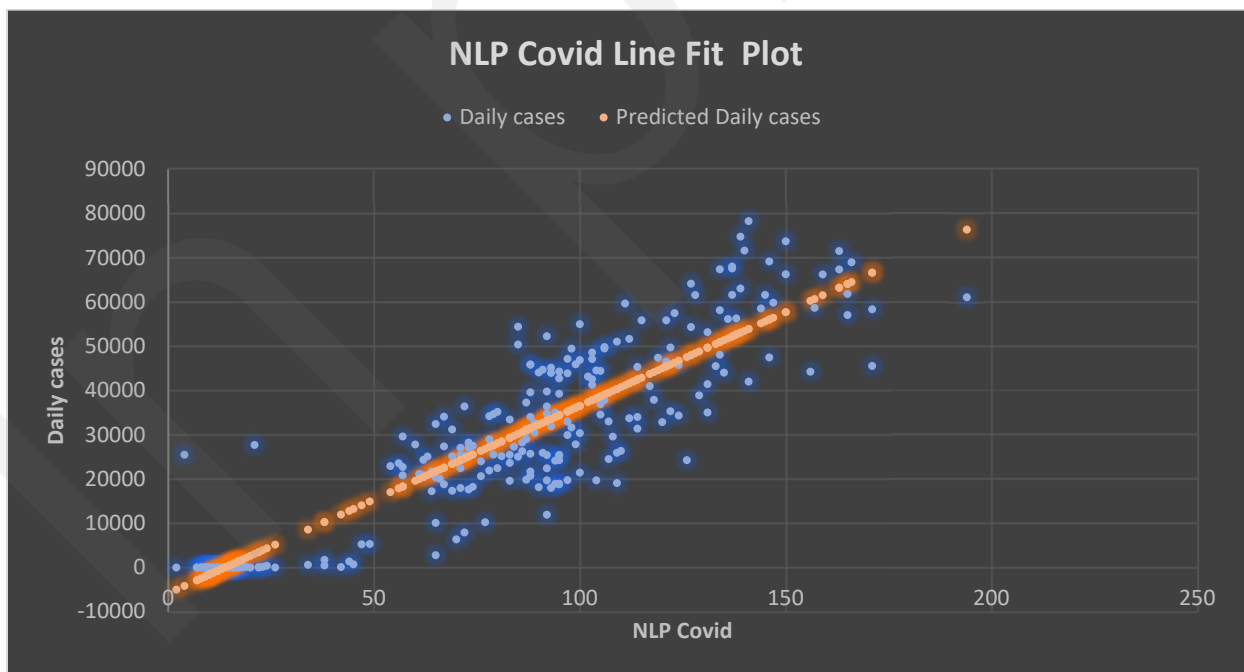
2. Predict the number of new official COVID-19 cases based on the Imaging findings NLP

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.8101					
R Square	0.6563					
Adjusted R Square	0.6550					
Standard Error	12,558.95					
Observations	276					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	8.25088E+10	8.25088E+10	5.23111E+02	1.72610E-65	
Residual	274	4.32172E+10	1.57727E+08			
Total	275	1.25726E+11				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-23720.7145	2331.1715	-10.1754	0.0000	-28309.9977	-19131.4313
NLP Image Finding	625.1602	27.3334	22.8716	0.0000	571.3499	678.9704

NLP Image Finding Line Fit Plot
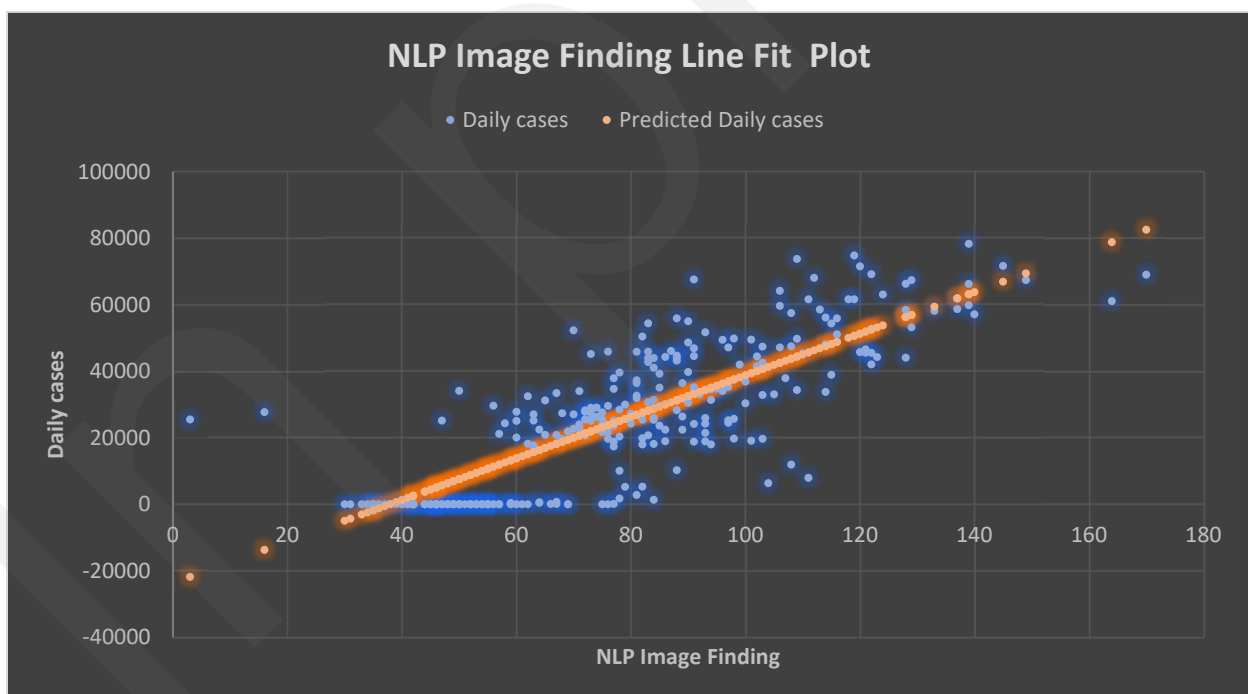
3. **Normalization process and adjusted reads:** Mitigation of patient case tally misrepresentation due to nonuniform exam occurrence across states.
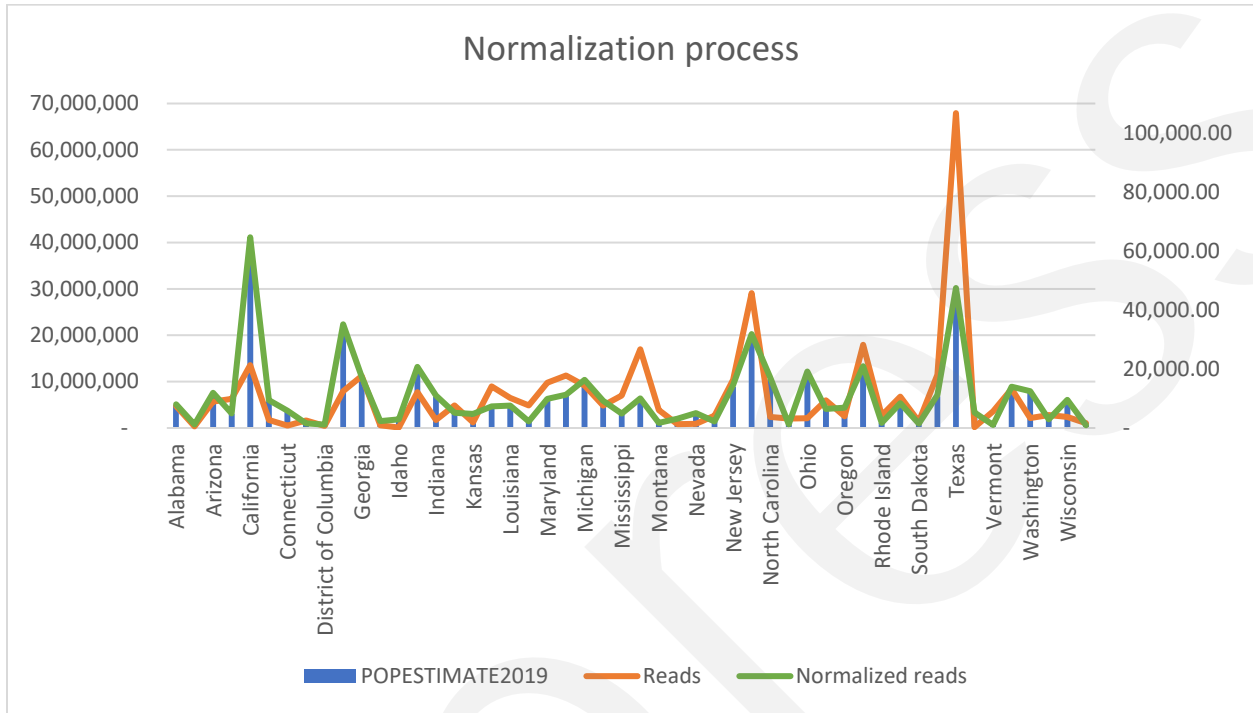
Normalization process variables and definitions:

Measure population = Total population

Measure reads = Total CT Reads performed per state

Measure benchmark = Measure reads/Measure population

Measure weight = Measure benchmark/Measure reads

Measure findings = Imaging Findings NLP by State

Measure normalization =Measure weight * Measure findings

The following formula was used for the normalization process:

$$w_{S_i} = \frac{ratio_{S_i}}{B}$$

$$AdjF_{S_i} = F_{S_i} \times w_{S_i}$$

where w_{S_i}, $ratio_{S_i}$: the state index and ratio, B: country benchmark, F_{S_i}, $AdjF_{S_i}$:imaging findings and

adjusted normalized reads by state.