COVID-19 Cluster Identification and SVM Classifier Model Construction Using Global Healthcare and Socio-economic Features

Soumya Kanti Guha¹, Sandip Sandhukha¹ and Sougata Niyogi²*

1 Department of Computer Application, Dinabandhu Andrews Institute of Technology and Management, Patuli, Kolkata, India

2 Department of Medical Lab Technology, Dinabandhu Andrews Institute of Technology and Management, Patuli, Kolkata, India

*Correspondence

Dr. Sougata Niyogi

Department of Medical Lab Technology
Dinabandhu Andrews Institute of Technology and Management
Block-S, 1/406A, near Satyajit Ray Park, Baishnabghata Patuli Township, Patuli, Kolkata, West Bengal 700094

E mail: sniyogi10@gmail.com

ORCiD0000-0002-5257-9339

Phone: +91 8240983494

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.
Abstract

Coronaviruses of the human variety have been the culprit of global epidemics of varying levels of lethality, including COVID-19, which has impacted more than 200 countries and resulted in 5.7 million fatalities as of May 2022. Effective clinical management necessitates the allocation of sufficient resources and the employment of appropriately skilled personnel. The elderly population and individuals with diabetes are at increased risk for more severe manifestations of COVID-19. Countries with higher gross domestic product (GDP) typically exhibit superior health outcomes and reduced mortality rates. Here we suggest a predictive model for the density of medical doctors and nursing personnel for 134 countries using support vector machine (SVM). The model was trained on 107 countries and tested on 27, with promising results shown by Kappa statistics and ROC analysis. The SVM model used for predictions showed promising results with high level of agreement between actual and predicted cluster values.

Keywords

Coronaviruses, Mortality rates, Clinical management, Healthcare outcomes, Predictive modeling, Support Vector Machine
1. Introduction

Three major outbreaks caused by coronaviruses have affected the world: SARS in 2002, MERS in 2012, and COVID-19 in 2019. While mortality rates associated with pandemics can fluctuate depending on individual factors, such as age and overall health status, lethality of these viruses can also shift as medical knowledge and treatment methods advance [1-4, https://www.worldometers.info/coronavirus/]. Considering the situation World Health Organization (WHO) has affirmed COVID-19 a global public health emergency [5, 6] in March, 2020. World Health Organization (WHO) and the Centers for Disease Control and Prevention (CDC) have been actively providing guidance and recommendations to prevent the spread of these viruses and to minimize the risk of complications. Through adherence to the guidelines set forth by these organizations, individuals can play a crucial role in mitigating the transmission of these viruses, thereby safeguarding themselves and others from the deleterious effects of the pathogens [7]. Pneumonia is a prevalent cause of hospitalization, characterized by symptoms such as fever, cough, dyspnea, and chest pain. In severe instances, this condition can culminate in mortality among affected patients [8]. Rendering high-quality medical care to critically ill patients necessitates a robust and functioning medical infrastructure. Clinical management of COVID-19 patients mandates access to hospital beds, well-equipped mechanical ventilation devices, N95 respirators, isolation gowns, gloves, and trained medical professionals [9-11]. Higher COVID-19 federated death has found to be correlated with depleted hospital resources [12, 13]. COVID-19 has shown differential effects depending on disparate age groups [14, 15]. Elderly people (> 70 yrs.) are more susceptible to different health complications [16]. Diabetes is recognized as a substantial risk factor for severe illness and mortality among hospitalized COVID-19 patients [3, 17-20]. The 1918 influenza pandemic resulted in a reduction of up to 11.8 years in life expectancy at birth in the United States. Similarly, Ebola virus outbreak of 2014 had a substantial effect, reducing life expectancy by 1.6 to 5.6 years in Liberia. These clearly underscore the profound and catastrophic impact that pandemics can have on communities. By implementing measures to enhance accessibility to vaccines and treatments, it is feasible to diminish the impact of future pandemics and safeguard public health [21, 22]. The increased mortality rate associated with COVID-19 has the potential to result in a significant loss of life expectancy. COVID-19 has already claimed the lives of millions of people worldwide, and its
impact on life expectancy will depend on a number of factors, including the number of deaths, the age distribution of deaths, and trends in mortality from other causes [23]. Preston curve shows that, on average, people in wealthier countries have higher life expectancies and lower rates of adverse health outcomes and mortality. This relationship is largely due to the fact that wealthier countries tend to have better-developed health systems and more resources to invest in public health, medical research, disease prevention and treatment [24, 25].

Gross domestic product (GDP) is a measure of a country's economic output and is used to estimate the aggregate amount spent on final goods and services. In theory, higher GDP should result in better living standards, including better education and public health programs, which can lead to improved life expectancy. However, the relationship between GDP and life expectancy is complex and can be influenced by many factors, such as income distribution, access to quality healthcare, and the prevalence of disease [25-30]. Early on in the COVID-19 pandemic, clinical classifications of the disease were primarily based on clinical signs and symptoms, as well as radiological findings from chest imaging. However, as our understanding of COVID-19 has evolved the classification of the disease and disease related outcome have become more nuanced and now incorporates additional factors, such as the presence of co-morbidities, disease severity, and the stage of the disease. By continuing to collect and analyze data on the disease, we can further improve our understanding of COVID-19 and enhance our ability to respond to future outbreaks [31, 32].

Our study found out that we can logically create 4 clusters out of 134 countries based on 6 parameters: Death Rate, Age older than 70, GDP per capita, Diabetes Prevalence, Hospital Beds per thousand, and Life Expectancy. The initial 6 parameters data was scaled and analyzed (dimensionality reduced) through PCA, later applied K-means. Further programmatically we developed an SVM model to train and test it on varied 6 parameter country data. The model predicted cluster values and the actual cluster values were compared using Kappa statistics. The ROC curve showed significant effectiveness of the SVM model.

2. Methods

2.1 Study design and data collection

The data collected from "Our World in Data" and “World Health Statistics 2020” provides valuable insights into the healthcare workforce and resources across different countries. This
information is useful for decision-makers and healthcare organizations in identifying areas for improvement and allocating resources where they are needed the most. Our study collected data for 211 countries, focusing on six parameters, namely Death Rate, Age older than 70 years, GDP per capita, Diabetes Prevalence, Hospital Beds per thousand, and Life Expectancy. However, it's important to note that the data has a cutoff date of May 12th, 2022.

2.2 Screening
The selection of data is a crucial step in any study as it impacts the reliability and final outcome of the results. For our study, we collected data for 211 countries, but only 158 had complete data on the selected six parameters. To reduce the influence of population size, we applied a population filter (Pf) to consider only countries with a population of over 1 million, resulting in 134 countries for our final analysis. Having a clear set of criteria for selecting data is important, although it does not guarantee the quality or accuracy of the research. By using a population filter, we aimed to minimize the risk of bias and increase the generalizability of our results, which can be confidently used to make informed decisions and recommendations.

2.3 K-means clustering
Clustering algorithms can help identify hidden features in random datasets, providing a structural identity to the data by grouping similar data points together. K-means is a popular unsupervised learning algorithm that can be easily implemented to cluster data points. In our study, we used K-means to analyze data from 134 countries based on six parameters. We applied three methods - Elbow, Silhouette, and Gap Statistic - to determine the optimal number of clusters (K). We found that K = 4 using these methods, and then normalized the data and performed Principal Component Analysis (PCA) before applying K-means. This allowed us to assign each of the 134 countries a cluster number from 1 to 4. We used the numpy, pandas, matplotlib, and scikit-learn Python packages to run the iterative algorithm for the Elbow, Silhouette, and Gap Statistic methods. While clustering doesn't guarantee the accuracy or quality of research, it provides a way to identify and group similarities in large datasets in an unsupervised manner.

2.4 Supervised Classification Model
Predictive modeling using classification is a wonderful mathematical process used to predict future events or outcomes by analyzing patterns in a given dataset. These classification methods programmatically give rise to various supervised machine learning algorithms. Support Vector
Machine (SVM) technique is one such example which ensures more effectiveness in high dimensional spaces yet being stable. Hence we chose non-linear SVM technique as it precisely uses hyper-planes (lines/decision boundaries) with various kernel functions to classify data that lack linear separation, hence offering better prediction and less error.

First we use scaling technique on the 6 parameter data and apply Principal Component Analysis (PCA) technique to reduce the data dimension. Once we receive the well-defined first two principal components of PC1 and PC2 we divide the 134 data set of PC1 and PC2 into test and training dataset using Pareto principle of 80:20 rule [33]. Hence we had 107 countries grouped under training set and 27 countries featuring under testing set. Next we developed a python program to implement a predictive model using SVM algorithm with Gaussian Kernel Radial Basis Function. The model was trained using training set (107 countries) to identify the correct clusters that a country originally belonged to. From K-means we already had the targets identified as cluster numbers 1 to 4 assigned to every country based on the previously discussed 6 parameter scale. Once training finished, we programmatically applied the testing set to test the model. The model predicted a cluster each out of 1 to 4 for all the 27 countries. Next we compared this predicted cluster number obtained from our SVM trained model with the actual cluster number identified by the K-means clustering technique earlier. We applied Kappa statistics to establish the substantial level of agreement between the actual and predicted cluster values for all the test data. The receiver operating characteristic (ROC) curve plotted for all 4 clusters showed promising outcome and usefulness of this SVM model in general.

2.5 Statistics

Welch's t-test is an appropriate test to use when the sample sizes are unequal and the variances between the two groups are also unequal. Most of our clusters are having unequal sample sizes. By using a 5% significance level and marking statistical significance with a p-value less than 0.001, Type I error can be minimized. In our study statistical significance was marked by ‘***’ with a ‘p’ value <0.001.
3. Results

3.1 K-means cluster analysis of 134 countries
COVID-19, since its inception has created a huge impact on global health. Total case fatality has escalated from one million on November 2020 to more than 6 million by April 2022. We selected 134 countries applying 6 parameter scale mentioned earlier and performed PCA and K-means clustering. Initially Elbow, Silhouette and Gap Statistics were performed to determine optimal cluster numbers and consistently it was found to be 4 (Fig 1A - D). Next we applied PCA first and then K-means clustering to classify these countries. Four distinct clusters were obtained namely cluster 1, 2, 3 and cluster 4 (Fig 2). In these four clusters (cluster 1, 2, 3 and 4) we had 51, 34, 38 and 11 countries listed respectively (Table 1).

3.2 Cluster wise distribution of divergent parameters
In our study we have considered a 6 parameter scale namely Death Rate, Age older than 70 years, GDP per capita, Diabetes Prevalence, Hospital Beds per thousand and Life Expectancy across 134 selected countries. Death Rate emphasizes the ratio of deaths to the population of a particular area or during a particular period of time. All the 134 countries are having divergent death in our study. The parameter of Age older than 70 years gives us a good knowledge of country wise older adults who have the risk of getting covid-19 due to their frail health. The parameter “Age older than 70 years” has comparatively high values for Cluster 3. The trend decreases with cluster 1, 2 and 4 respectively. Diabetes Prevalence parameter takes care of the strong probability of an individual experiencing co-morbidity during covid-19 infection. Diabetes Prevalence predominantly has high values for cluster 4, more than the maximum values of Cluster 1, 2 and 3. The number of hospital beds per thousand people is an important metric that provides insight into a country's healthcare system and its ability to provide adequate inpatient care. During the COVID-19 pandemic, the number of hospital beds per thousand people became even more important, as countries around the world struggled to provide enough hospital beds to care for the increasing number of patients with COVID-19. The parameter of Hospital beds per thousand shows maximum variance for cluster 3, followed by cluster1, 4 and 2. The median value of cluster 1 and 4 are similar. Cluster-2 suggests high overall data density. The parameter of Life Expectancy of any country establishes a statistical measure of the average time a citizen is expected to live, based on other demographic factors. Life Expectancy for all the 4 clusters is...
quite high though different. Data varies from 58 to 85 for all 134 countries. The median of Cluster 3 is more than 80, followed by Cluster 4 and 1, nearing 75. Cluster 2 has median value less than 65. GDP per capita gives us a comprehensive impression about a country’s medical infrastructure, existing expense and future vision of health care. GDP per capita is most varied in cluster 4 followed by cluster 3. The trend decreases with cluster 1 and 2 respectively. Cluster-1 suggests high overall data density. All cluster wise p-values for different parameters are given in Table-2.

3.3 SVM classifier model construction and optimization

Our study uses a dataset with 6 features from 134 countries, and follows the Pareto principle by using an 80/20 split for training and testing. This rule, based on wealth distribution, is useful for explaining human and machine phenomena. The train-test split data (107 countries and 27 countries) is useful in estimating the performance of our SVM model when asked to make predictions on data not used to train the model (Fig 4A). We have used Gaussian Kernel Radial Basis Function (RBF) to add radial basis to improve the transformation. To trade off misclassification of training examples against simplicity of the decision surface, we have kept “C” as 1. A low C makes the decision surface smooth, while a high C aims at classifying all training examples correctly. We keep the Gamma hyper parameter as 1, to keep optimum curvature at decision boundary. Our SVM model is plotted across Principal Component 1 and 2 generated out of our data. In the plot four distinct colors represent 4 clusters based on training data set. The dots represent the testing dataset predictions. The colors of the dots represent the actual cluster they original belong to. The location of the dots shows the cluster where the SVM model has predicted the data to be part of (Fig 4B). Kappa statistics beautifully deals with data that are the result of a judgment, not a measurement. In our study the kappa statistics value of 0.777 lies within the range of [0, 1] showing substantial or strong level of agreement between the actual and predicted cluster values for all the test data (Fig 5A). 11 countries get correctly identified for Cluster 1, 6 countries for Cluster 3 and 4 countries for cluster 2. Two data each from cluster 1, 3 and 4 get wrongly identified. The receiver operating characteristic (ROC) curve plotted for all the 4 clusters show the area > 80 % showing promising outcome and usefulness of this SVM model in general (Fig 5B). ROC for Cluster 2 comes out as a perfect classifier with area equal to 1.00, followed by Cluster 1 with area 0.90. Next is Cluster 3 with area equal to 0.85. Cluster 4 has the least area of 0.80.
4. Discussion

Any disease like COVID-19 which has global impact with deadly outcome always attribute to a comprehensive range of factors. The various factors influencing the spread and severity of the disease can be gathered through various sources such as health records, surveys, and data analysis. Initially we used K-means clustering analysis and PCA (principal component analysis) to group countries into 4 distinct clusters based on a 6 parameter scale. By grouping countries into clusters, based on similar characteristics, it becomes easier to identify patterns and relationships between the variables.

In addition, we employed an SVM classifier model to make predictions about which cluster each country would belong to. SVM is a supervised learning algorithm commonly used for classification tasks. By using the SVM model based on the 6 parameters, we were able to accurately predict which cluster each country belonged to, showing a high level of agreement between the actual and predicted cluster values. To evaluate the performance of our classifier model, we used the ROC curve, which plots the true positive rate against the false positive rate. The high ROC area values of 1.00, 0.90, 0.85, and 0.80 for the different clusters suggest that our classifier had a high degree of accuracy in assigning countries to their respective clusters. Overall, our study demonstrates the utility of clustering and SVM classification in identifying factors that influence the spread and severity of diseases like COVID-19. Life expectancy and death rate are two significant measures that describe the overall health of a population. Our research indicates that these two parameters are inversely related - higher life expectancy shows lower death rate, while lower life expectancy demonstrates higher death rate. Our study also showed that life expectancy and hospital beds per thousand people have a positive correlation, although it is not a very strong one. Life expectancy is influenced by various factors such as genetics and economic status [34]. Adequate access to healthcare can have a positive impact on life expectancy. On the other hand, the availability of a high number of hospital beds per thousand people can ensure that individuals have access to essential medical care, reducing the risk of death [35]. However, other factors such as the quality of medical care and the general health of the population can also affect death rates, and so hospital beds per thousand should not be considered the sole indicator of a country's death rate. Older adults, particularly those over the age of 70, are more prone to health complications associated with COVID-19, and may require hospitalization [16]. A high number of hospital beds per thousand can provide better support and care to older adults in a country. The
death rate and age of 70 in a country are related as older individuals have a higher mortality rate due to health complications. However, our research showed an exception to this due to factors such as a strong GDP per capita, sufficient number of hospital beds per thousand, and high life expectancy. Diabetes is a chronic condition that can increase the risk of health complications and death, especially if it's not properly managed. Studies show that individuals with diabetes are more susceptible to severe illness from COVID-19 and other infectious diseases [18]. Although diabetes prevalence and death rate are not directly related, they can provide a broader understanding of a population's health when used together. Diabetes prevalence, life expectancy, and age of 70 in a country can be related as higher diabetes prevalence is more common in older adults, particularly those over the age of 70. Higher rates of diabetes can lead to health complications and decreased life expectancy, while areas with higher life expectancy may have lower rates of diabetes. However, the relationship between diabetes prevalence and life expectancy can be impacted by other factors such as access to healthcare and healthy lifestyle practices. The relationship between diabetes prevalence and GDP per capita is complex and depends on several factors, including lifestyle and government policies. Generally, a higher GDP per capita indicates better access to medical services, which may lower diabetes prevalence [36]. However, lifestyle factors such as unhealthy diets and lack of physical activity, which is often associated with higher GDP, can contribute to higher diabetes prevalence in some countries. Further studies and analysis are required to better understand the relationship between diabetes prevalence and GDP per capita. GDP per capita and hospital beds per thousand can serve as indicators of a country's health system and its ability to provide medical services. Typically, higher GDP per capita is associated with better medical infrastructure and more resources to provide better health services, including more hospital beds. This can actually reduce the death burden associated with pandemics like COVID-19.

5. Conclusion
There are several factors that have been found to influence the spread and severity of COVID-19. We used K-means cluster analysis and principal component analysis to group countries into 4 distinct clusters based on 6 parameters and further developed SVM classifier model to make predictions. Results showed a high degree of accuracy in assigning countries to their respective clusters. Life expectancy and death rate were found to be inversely related and positively correlated with hospital beds per thousand people. Diabetes prevalence was related to life
expectancy and older age, and its relationship with GDP per capita was complex and dependent on several factors. GDP per capita and hospital beds per thousand were indicators of a country's health system and ability to provide medical services. In conclusion, the study sheds light on the various factors that affect the spread and severity of COVID-19, highlighting the importance of access to healthcare and medical services in reducing death rates.

Author contributions
SN conceptualized the idea and designed the study. SKG performed the experiments. SS helped in the construction and optimization of SVM model. SN and SKG analyzed the data and wrote the manuscript. SN emended it and approved the final version.

Data Availability Statement
The data related to the study are available from the corresponding author upon reasonable request.

Ethical approval
Not Applicable.

Conflicts of interest
Authors declare no potential conflicts.
References


Figure 1: Selection of number of clusters. Elbow (A), Silhouette (B & C) and Gap Statistic (D) to find probable number of clusters (K).
Figure 2: K-means clustering data. K-means clustering on PCA components was done with 134 countries using different parameters.
Figure 3: Cluster wise distribution of different parameters. Death rate (A), Age older than 70 yrs. (B), Diabetes Prevalence (C), Hospital beds per thousand (D), Life expectancy (E) and GDP per capita (F) were plotted for all the clusters.
Figure 4: Supervised machine learning approach to classify selected countries. Workflow for Support Vector Machine (SVM) (A). Decision surface showing different class boundaries using the model and test dataset (B).
Figure 5: Performance of SVM-based classifier model. Confusion matrix showed that most countries in test dataset were classified precisely (A). Receiver Operating Characteristics (ROC) curves for the SVM classifier exemplify the predictive powers of Class 1, 2, 3 and 4 respectively. Areas under the curves (AUCs) of the four ROC curves are indicated accordingly (B).
<table>
<thead>
<tr>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tajikistan</td>
<td>Eritrea</td>
<td>Estonia</td>
<td>Mauritius</td>
</tr>
<tr>
<td>Timor</td>
<td>Central African Republic</td>
<td>Latvia</td>
<td>Qatar</td>
</tr>
<tr>
<td>Nicaragua</td>
<td>Djibouti</td>
<td>Finland</td>
<td>Bahrain</td>
</tr>
<tr>
<td>Gabon</td>
<td>Equatorial Guinea</td>
<td>New Zealand</td>
<td>Oman</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>Liberia</td>
<td>Uruguay</td>
<td>Kuwait</td>
</tr>
<tr>
<td>Jamaica</td>
<td>Niger</td>
<td>Belarus</td>
<td>United Arab Emirates</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>Benin</td>
<td>Slovenia</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>Gambia</td>
<td>Norway</td>
<td>Singapore</td>
</tr>
<tr>
<td>El Salvador</td>
<td>Burundi</td>
<td>Lithuania</td>
<td>Egypt</td>
</tr>
<tr>
<td>Botswana</td>
<td>Burkina Faso</td>
<td>Ireland</td>
<td>Malaysia</td>
</tr>
<tr>
<td>Albania</td>
<td>Togo</td>
<td>Croatia</td>
<td></td>
</tr>
<tr>
<td>Algeria</td>
<td>Guinea</td>
<td>Denmark</td>
<td></td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>Mali</td>
<td>Sweden</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>Haiti</td>
<td>Slovakia</td>
<td></td>
</tr>
<tr>
<td>Libya</td>
<td>Tanzania</td>
<td>Bulgaria</td>
<td></td>
</tr>
<tr>
<td>Mongolia</td>
<td>Madagascar</td>
<td>Switzerland</td>
<td></td>
</tr>
<tr>
<td>North Macedonia</td>
<td>Eswatini</td>
<td>Israel</td>
<td></td>
</tr>
<tr>
<td>Armenia</td>
<td>Yemen</td>
<td>Hungary</td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>Laos</td>
<td>Austria</td>
<td></td>
</tr>
<tr>
<td>Honduras</td>
<td>Cameroon</td>
<td>Greece</td>
<td></td>
</tr>
<tr>
<td>Moldova</td>
<td>Ghana</td>
<td>Belgium</td>
<td></td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>Malawi</td>
<td>Portugal</td>
<td></td>
</tr>
<tr>
<td>Costa Rica</td>
<td>Cambodia</td>
<td>Czechia</td>
<td></td>
</tr>
<tr>
<td>Panama</td>
<td>Mozambique</td>
<td>Canada</td>
<td></td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>Uganda</td>
<td>Romania</td>
<td></td>
</tr>
<tr>
<td>Nepal</td>
<td>Sudan</td>
<td>Australia</td>
<td></td>
</tr>
<tr>
<td>Lebanon</td>
<td>Zambia</td>
<td>Netherlands</td>
<td></td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>Zimbabwe</td>
<td>Japan</td>
<td></td>
</tr>
<tr>
<td>Paraguay</td>
<td>Kenya</td>
<td>Ukraine</td>
<td></td>
</tr>
<tr>
<td>Guatemala</td>
<td>Afghanistan</td>
<td>Poland</td>
<td></td>
</tr>
<tr>
<td>Morocco</td>
<td>Ethiopia</td>
<td>Argentina</td>
<td></td>
</tr>
<tr>
<td>Bolivia</td>
<td>Myanmar</td>
<td>South Korea</td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td>Pakistan</td>
<td>Spain</td>
<td></td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>South Africa</td>
<td>Italy</td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td></td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Tunisia</td>
<td></td>
<td></td>
<td>Germany</td>
</tr>
<tr>
<td>Ecuador</td>
<td></td>
<td></td>
<td>France</td>
</tr>
<tr>
<td>Bangladesh</td>
<td></td>
<td></td>
<td>Russia</td>
</tr>
<tr>
<td>Iraq</td>
<td></td>
<td></td>
<td>United States</td>
</tr>
</tbody>
</table>

**Table 1:** List of clustered countries.
## Table 2: List of statistical parameters for different conditions. Welch's test was performed to determine statistical significance.