*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

# Prediction and Distribution of Disease Using Hybrid Clustering Algorithm in Big Data

**Vinston Raja R[1], Deepak Kumar A[2], Prabu Sankar N[3], Senthamilarasi N[4], Dr. Chenni Kumaran J[5]**
[1]Assistant Professor, Information Technology, Panimalar Engineering College, Chennai, India.
rvinstonraja@gmail.com
[2]Assistant Professor, Computer Science and Engineering, St. Joseph's Institute of Technology, Chennai, India.
deepakkumar@stjosephstechnology.ac.in
[3]Assistant Professor, Department of Information Technology, Panimalar Engineering College, Chennai, India
n.prabusankar81@gmail.com
[4]Assistant Professor, Computer Science and Engineering, Sathyabama Institute of Science and Technology
senthamilarasi.n.cse@sathyabama.ac.in
[5]Professor, Professor, Department of Computer Science and Engineering,Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS)
drchennikumaran@gmail.com

**Abstract -** COVID disease plague of 2019 (COVID19) has made an overall health related crisis with a very high gamble of spreading and influencing the whole planet. In essentially every nation, new cases have been accounted. To identify all countries expanding number of tests, the manual clustering of COVID-19 and clinical infection information tests becomes tedious and requires profoundly talented work. As of late, a few calculations have been utilized for clustering clinical datasets deterministically; nonetheless, these definitions have not been powerful in gathering and investigating clinical infections. To rank and score more than 200 nations as indicated by COVID-19 cases and casualty in 2020 and contrast the outcomes with existing pandemic weakness forecast models and results produced by standard Data clustering scoring methods. Information clustering is a course of orchestrating comparative information into gatherings. A clustering algorithm bundles an informational collection into a few several clusters such an extent that the similitude inside a gathering is better compared to among clusters. This paper propose new Hybrid clustering algorithm KMHC in view of K-Means and Hierarchical Clustering calculation. This calculation KMHC, First isolated into every nation gatherings or fragments in light of the COVID patients count, Secondly grouping models has been made across the nations on the planet and across the states in India, and the presentation investigation is analyzed. This paper likewise center figures the future COVID count for India. By utilizing this outcome, set of nations which are having higher COVID count can be effortlessly pictured and the proper moves will be made to diminish the count.

**Keywords-** COVID-19, K-Means Clustering, Hierarchical Clustering, Prophet.

## I. INTRODUCTION

Covid illness (COVID-19) is a recently found Covid that causes an irresistible sickness. The globe is as of now managing the COVID-19 pandemic, which has a wide overall appropriation and, all the more significantly, has brought about countless deaths [1]. One of the new issues has been anticipating the spread of Covid and deciding the conveyance of COVID cases across nations. Anticipating the pandemic with extraordinary exactness will help different state run administrations in fostering a methodology to battle the infection's spread.

The field of information mining and information disclosure is arising as a new, essential exploration region with significant applications to science, designing, medication, business, and training. Information mining endeavors to figure out, investigate and carry out fundamental acceptance processes that work with the extraction of significant data and information from unstructured information [1]. Size of information bases in logical and clinical applications is immense where the quantity of records in a dataset can shift from approximately thousands to thousands of millions [2].For example, dataset tests gave by means of determination in the clinical field are expected for sickness investigation, and are broke down by a specialist or drug specialist to decide the phase of the infection. As the quantity of patients expands, additional time is expected to look at the examples. Consequently, an orderly technique is expected to naturally or semiautomatically assess the example dataset for every quiet. The information tests of ailments can be arranged by applying a precise strategy including grouping calculations. There has been inadequate examination to really bunch COVID-19 and other clinical sickness datasets

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

utilizing grouping calculations. As a special case, the concentrate in [14] presented a covering k-implies calculation for clinical applications, notwithstanding the restrictions of this calculation. Another review [15] analyzed the capability of stretching out troupe bunching techniques to the field of clinical diagnostics.

The bunching is a class of information mining task in which calculations are applied to find fascinating information conveyances with regards to the hidden information space. The arrangement of bunches depends on the standard of boosting likeness between designs having a place with unmistakable groups. Similitude or closeness is normally characterized as distance work on sets of examples and in light of the upsides of the elements of these examples [3] [4]. Numerous assortment of grouping calculations have been arisen as of late. Different beginning stages and standards normally lead to various scientific categorizations of grouping calculations [5]. K-Means (KM) [6], K-Harmonic Means (KHM) [7], Fuzzy C- Means (FCM) [8][9], Spectral Clustering (SPC) [10][11], and a few other strategies established on the basis of previously stated methods[12] are currently commonly used partitional grouping approaches. The KM calculation is a famous partitional calculation. It is an iterative slope climbing calculation and the arrangement acquired relies upon the underlying grouping. Albeit the K-implies calculation had been effectively applied to numerous viable grouping issues, it has been shown that the calculation might neglect to merge to a worldwide least under specific circumstances [13].

The fundamental issue of K-implies calculation is that, the calculation will in general perform more terrible when the information are coordinated in more mind boggling and obscure shapes [14]. Also Hierarchical Clustering is a novel center based bunching calculation which utilizes the consonant midpoints of the good ways from every information [7]. It is exhibited that Hierarchical Clustering is basically uncaring toward the instatement of the focuses. In specific cases, Hierarchical Clustering essentially works on the nature of grouping results, and it is likewise appropriate for enormous scope informational collections [5]. Anyway existing grouping calculation requires numerous cycles (look over) the datasets to accomplish a combination [15-17], and the greater part of them are delicate to introductory circumstances, as a visual demonstration. This paper proposed KMHC-Hybrid clustering model which depends on K-Mean and Hierarchical Clustering approach.

## II. RELATED ALGORITHM REVIEWS

Protection Preserving Data Mining - PPDM procedures and systems were inspected and examination of benefits and impediments of various PPDM method is made which assists with distinguishing the open issues and future exploration patterns in PPDM. The current situation of Privacy saving information mining and some future exploration headings are proposed in [16]. Security safeguarding information mining methods by and by accessible are ordered in [17], and their benefits and burdens are underscored. The utilization of cryptographic methods on secure appropriated calculation in information mining is shown in [18]. An order progressive system is proposed in [19] alongside the point by point survey.

The total assignment to be settled is not one explicit calculation, but the group investigation is. It could be performed by a variety of calculations that differ fundamentally in their understanding of what constitutes a group and how to efficiently track them down. Gatherings with short distances between group members, a dense region of the information space, spans, or specialized factual disseminations are all common examples of bunches. As a result, clustering can be identified as a multi-objective advancement problem. The fitting bunching calculation and boundary settings (counting values, for example, the distance capacity to utilize, a thickness edge or the quantity of anticipated groups) rely upon the singular informational collection and planned utilization of the outcomes. Bunch examination as such is definitely not a programmed task, yet an iterative course of information disclosure or intuitive multi-objective improvement that includes preliminary and disappointment.

Hardly any bunching calculations accessible are recorded underneath:

- K-Means grouping
- EM bunching
- Farthest First bunching
- Sifted Clustering
- Progressive bunching
- Spider web Clustering
- Make thickness based bunching

The investigation of irresistible sickness spread instrument should be possible by many approaches like Susceptible-Infected-Recovered (SIR), powerless people, asymptomatic tainted, indicative contaminated, recuperated, and perished (SEIRD), and Susceptible, Exposed, Infectious, Hospitalized, Dead (SEIHRD). These models

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

utilize differential form frameworks or contrast conditions for portraying populace elements concerning contamination [20-25]. The SIR-model is probably the most straightforward model, and it isolated the whole populace into three gatherings or compartments: the quantity of vulnerable individuals (S), the quantity of irresistible individuals (I), and the quantity of taken out individuals[26]. In SIR-model, every individual is ordered as helpless, irresistible, or eliminated, where vulnerable shows sound individual and eliminated demonstrates either recuperated or dead because of the COVID cases. Every individual can change from one group or state (defenseless, irresistible, or eliminated) to other state. That is, every individual has an extraordinary Markov chain related to arrangement [27].

Numerical methodologies have been successfully used to learn the population aspects of corona cases in a few countries. Kucharski et al. [28] introduced a one of best spread numerical model based on information from corona cases. This assessment took into account all incidents of Wuhan, China, till 2020 March. Ndairou et al.[29] used an independent numerical technique to investigate the spread of corona in Wuhan, China. Prem et al. [30] used numerical demonstrating to do another review that efforts on regulating issue with corona of city in china. Hellewell et al. [31] used the detachment strategy to examine the potent technique of corona virus sickness in another work. In the Diamond Princess Cruises Ship, Mizumoto at al. [32] examined the transmission elements of corona. To focus on the concept of irresistible diseases. Liu et al. [33] discovered the mathematical value of corona case critical proliferation number in order to focus on the

For Canada, South Korea, United Kingdom, France, India, and the Chakraborty and Ghosh [35] proposed a cross breed autoregressive integrated moving normal Wavelet-based estimation of the quantity of day wise admitted instances. Ribeiro et al. [36] used stochastic-based relapse technique to estimate the assignment of time series anticipating with 2, 4, and 7 days of the COVID-19 aggregate certified instances in 10 states of Brazil with a high day by day occurrence.

Yang et al. [37] established a numerical approach based on an altered defenceless exposed irresistible recuperated (SEIR) technique to study COVID-19 plague progression in China. Under various mediation processes, they forecast the level and timing of the COVID-19 plague pinnacles. Peng et al. [38] to examine the COVID-19 scourge in China. They've predicted the expression fact and possible completion time of confirmed COVID cases for five different districts. The corona virus spreading elements in Wuhan and out parts in china were predicted using the traditional SEIR model by Li et al. [39-42]. They looked at how lockdowns affected the COVID-19 transmission elements in various cities.

COVID-19 has been the subject of previous and recent discussions, with an analysis of its dissemination and future expectations exhibiting a wide range of strategies. However, there are a variety of degrees to which the COVID-19 data can be revealed using allotment-based grouping algorithms.

## III. PROPOSED METHEDOLOGY

Information Science strategies of K-Means and Hierarchical grouping have been applied to frame the cluster among nations and states. It can deal with both straight out and nonstop information. Information Science is an interdisciplinary field that consolidates software engineering, math, measurements and space information. Utilizing information science calculations, intermingling is ensured. Bunching for the world - gathering nations on the planet as indicated by the COVID count and Clustering for India - Grouping states in India as per the COVID count [2]. Utilizing Prophet, gauging on time series information is performed, to foresee the future COVID count of India [1,2], Figure-1 illustrations the system model of proposed system.In Data importing, input document for doing the clusters among the nations on the planet, is a csv record taken from world wellbeing association dataset. The information record for doing the clusters among the states in India is additionally a csv document taken from covid19.india dataset. Whenever information have been gathered, it has been brought into the R climate for grouping and forecast. In information pre-handling, trait determination, normalization and standardization capacities [3] will be applied. In normalization, crude information is changed into normal, justifiable configuration. In property determination, hold just the characteristics which is influencing the investigation and it isn't important to hold every one of the traits for doing the analysis. In Normalization, mean of the quality will be zero and usual deviation will be one. The preprocessed information is sent to the K-Means clustering calculation, presently it frames the groups among the nations on the planet utilizing the COVID patient's count and structures the gatherings among the states in India utilizing COVID patient's count [3,4]. Progressive grouping calculation shaped the bunches among the nations and states among the India, utilizing the COVID patient’s count. As far as bunch in Figure 1

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

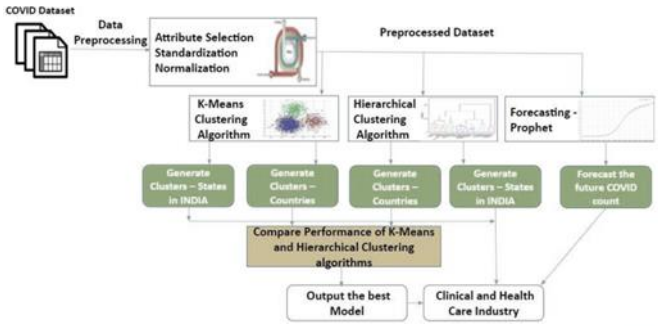

Figure 1. Proposed Hybrid system model

Development quality, the exhibition of the two calculations is analyzed. For the quality estimations, parts, for example, between amount of squares, inside amount of squares, absolute amount of squares and agglomerative coefficient have considered. Utilizing Prophet, the future Coronavirus count for India is determined. The progressive grouping Model has shaped a quality bunch when contrasted with K-implies for this world wellbeing association's COVID informational collection.

So this model alongside the COVID patient's count will be given to the clinical and medical services industry to play it safe ahead of time [4] The hybrid clustering algorithm KMHC steps to get final result first process is importing COVID Dataset and Data Preprocessing into algorithm, Second K- Means Algorithm and Hierarchical Algorithm analysis the dataset for clustering countries and states, Thirty hybrid clustering algorithm comparison K- Means and Hierarchical Clustering Algorithms for best output and Finally forecasting future COVID count for clinical and health care industry using Prophet.

### *A. Data importing and reprocessing*

Information is accessible in any document design like .txt, .csv, .xlsx, .spss and so forth Information must be stacked in to R climate for investigation. Whenever information have been separated from the document it ought to be put away in an information outline. Libraries essential for anticipating must be stacked into the R climate [5,6]. For determining the future qualities in the time series information, prophet package needs to introduced and stacked into the program. Introduce a few additional packages for ensuing activities like, dplyr package utilized for applying the SQL question to separate the India dataset from the world, ggplot2 Package to give some more perception to the result and lubricate package to change over the date stamp in the configuration of YYYYMMDD [7]. Information pre-handling is a mining approach that involves changing over crude information into an organization that can be perceived. Clamor, missing qualities, and irregularity are altogether normal issues with crude information. Figure 2. Shows proposed hybrid algorithm. Certifiable information is much of the time fragmented, conflicting, and inclined to mistakes.

1. **Start**
2. **Instate** Dataset DN, K, Cp1, Cp2, . . . , Cy,
   Current Pass=1;
   Where D is dataset, N is size of informational collection,
   K is number of bunches to be framed,
   Cp1, Cp2, . . . , CK are bunch focuses.
   Current Pass is the all-out no. of sweeps over the dataset.
3. **Do allocate** the n information focuses to the nearest Ci;
   **If CurrentPass%2==0**
   Recomputed Cp1, Cp2, . . . , Cy utilizing Mean capacity;
   **Else**
   Recomputed Cp, Cp, . . . , Cy utilizing Arithmetic Mean capacity;
   Increment Current Pass by one.
   until no adjustment of Cp, Cp, . . . , Cy;
4. **return** Cp1, Cp2, . . . , Cy;
5. **End**

Figure 2. Proposed Hybrid Algorithm

The missing qualities issues must be settled by straightforward measurable procedures. Preprocessing information is a time tested way for settling such issues. Accordingly, the mining aftereffects of crude information are pre-handled to build the information quality, and the proficiency interaction is moved along. It isn't important to hold every one of the traits for doing the examination; we can hold just the qualities which are influencing the investigation. Information normalization is the method involved with changing over comparative information gained in various configurations into a solitary organization that empowers for simpler correlations, cooperative exploration, and enormous scope examination [7, 8]. Standardization is a scaling procedure that movements and rescales values to make them range somewhere in the range of 0 and 1. Min-Max scaling is one more name for it.

### *B. K-Means algorithm – Clustering countries and states*

Clustering is a famous exploratory information examination instrument for acquiring a comprehension of the information's design. It is the assignment of distinguishing subgroups in information so items inside a similar subgroup are basically the same while relevant elements inside different groups are extremely divergent. Bunching is the method involved with gathering things into bunches whose

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:*** *23 July 2023* ***Revised:*** *20 September 2023* ***Accepted:*** *01 October 2023*

individuals are connected here and there.

Let α = {α1, α2, α3… … ..αn} be the arrangement of important items and β = {β1, β2, β3… … .βc} be the arrangement of focuses the concept of unstoppable sickness. By developing numerical displays, Fanelli and Piazza [34] researched and predicted the idea of this virus distribution in China and other countries till 2020 March.

1) Choose 'γ' group focuses indiscriminately.

2) Calculate the distance between every main item and the focal point of each group.

3) Assign the information highlight the group community with the most limited distance among it and any remaining bunch habitats.

4) Recalculate the new group community as follows:

where 'γi' is the quantity of relevant items in the ith group. [9,10]

$$\beta_i = (1/\gamma_i) = \sum_{j=1}^{\gamma_i} \alpha_i$$

5) Recalculate the distance grid and the new group communities that have been found.

6) Stop assuming no information focuses was reassigned; in any case, begin once again at stage 3.

The Euclidean distance technique is utilized to work out the distance between the information parts.

$$d = |A - B| = \sqrt{\sum_{I=1}^{N} [A_I - B_I]^2}$$

### *C. Hierarchical Algorithm – Clustering countries and states*

The information is assembled into a tree of gatherings in a progressive bunching process. Each datum point is treated as a different bunch in progressive grouping. The objective of progressive bunching is to make a progression of settled groups in a various leveled request. This progressive system is graphically addressed by a Dendrogram [8] (A Dendrogram is a tree-like realistic that counts the quantity of consolidations or parts in a series), which is a modified tree that portrays the request wherein factors are combined. Consider every main informative element as a different Cluster right away, and afterward join the bunch's closest couples at each stage. At first, every informational index is viewed as a different substance or bunch. The groups converge with different bunches in each cycle until just one bunch remains. The Hierarchical Clustering Algorithmic Steps: [8,9,10]

1. Start by assigning everything to its own bunch, to such an extent that assuming there are N things, there will be N groups, each with only one thing. Permit the distances (similitudes) across bunches to approach the distance (likenesses) between the things in each group.
2. View as the most comparative (nearest) bunch pair and consolidation them into a solitary group, bringing about one less group.
3. Work out the distances (similitudes) between every one of the past bunches and the new group.
4. Stages 2 and 3 ought to be rehashed until all things are assembled into a solitary N-size group. Accept six things for instance (a, b, c, d, e, and f). Everything has two qualities that can be estimated (x1, x2). Each item ought to be allocated to its own bunch.

Distance framework: Use the Euclidean Distance recipe to figure the distance between the items and structure the distance lattice.

$$dist\big((\alpha, \beta), (a, b)\big) = \sqrt{(\alpha - a)^2 + (\beta - b^2)}$$

### *D. Forecasting future COVID count for India using prophet*

Time Series Analysis is an approach to examining and learning the conduct of datasets over a period. Also, it helps in learning the conduct of the dataset by plotting the time series object on the chart. Facebook's Core Data Science group distributed Prophet as open source programming. [11,12] Prophet is a period series information anticipating process that fits non-direct examples with yearly, week after week, and day by day irregularity, as well as occasion impacts. It's best for time series with articulated occasional impacts. Prophet is pardoning of missing information and pattern movements, and it typically handles exceptions well. It could be downloaded from CRAN and PyPI. By projecting costs, deals, or climate, it helps organizations in learning the conduct of their items. It utilizes a decomposable time series model that comprises of three essential parts: pattern, irregularity, and occasions [10,11,12]

$$z(t) = x(t) + y(t) + z(t) + €t\, z(t) \quad (4)$$

Forecast *x(t)* - pattern work (x(t)) y(t) - shows changes that happen consistently (e.g., week by week and yearly irregularity) z(t) - the ramifications of excursions on possibly unpredictable timetables €t - blunder term represents any startling changes that the model doesn't represent. A pattern is a change being developed either in addition or decrement course. Irregularity is an element of time series object that happens at a specific time/season and changes the pattern. Occasions are a time-frame that changes a great deal to the business. It can create a gain or misfortune relying on the business. Prophet is used in an assortment of Facebook applications to produce precise gauges for arranging and objective setting. In by far most of conditions, Prophet beats

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

any remaining methodologies. With no manual exertion, Prophet delivers a sensible figure on messy information. Exceptions, missing information, and huge changes in your time series are not an issue for Prophet. The Prophet technique furnishes clients with various choices for calibrating [12,13] and changing estimates.

## IV. RESULTS AND DISCUSSION

Experiments are done to check the proposed model performance. These analyses are performed on a server with 8 GB RAM and 5i Processor. Algorithm is implemented using JAVA programming language and datasets are stored in memory. The stack memory of JAVA is exceptionally small. Hence the investigations utilize as it were of unique dataset for example 1500 records to perform examination.

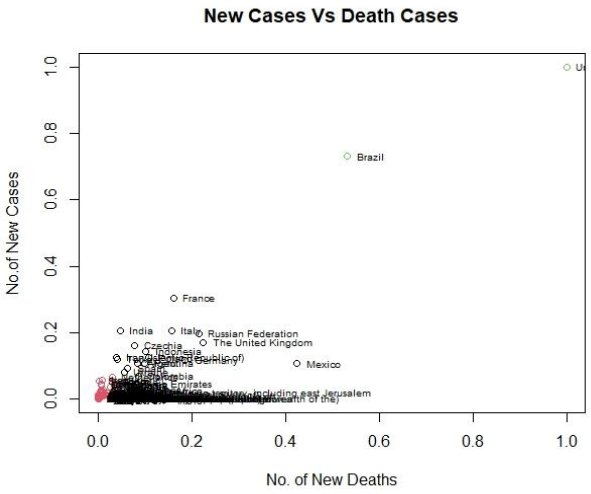


Figure 3. Cases (New vs Death)

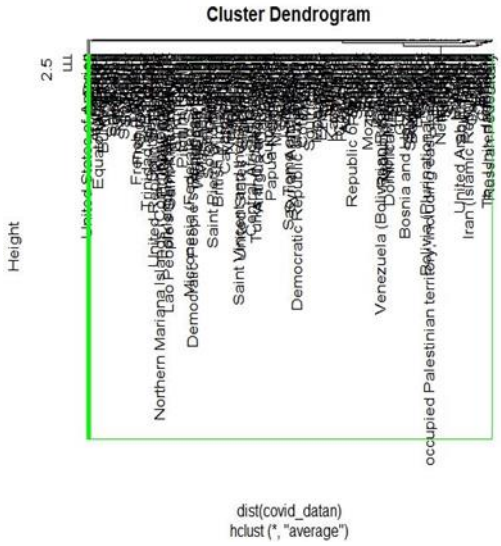


Figure 4. Dendrogram (Cluster)

The clustering utilizing K-Means Algorithm and Hierarchical Algorithm of COVID-19 information of April: 2020-2021 has more than lakh passages with date, revealed cases, recuperated, new cases, and deaths. The groups are shifted from five to fifty five in isolated runs of both the calculations. The groups are kept at least 5 and a most extreme from five to fifty five.

The ideal number of bunches is gotten in light of the calculations' computations, as referenced in the past segment. The groups of COVID-19 made depend on 1) date 2) announced cases, 3) recuperated, 4) new cases and 4) deaths.

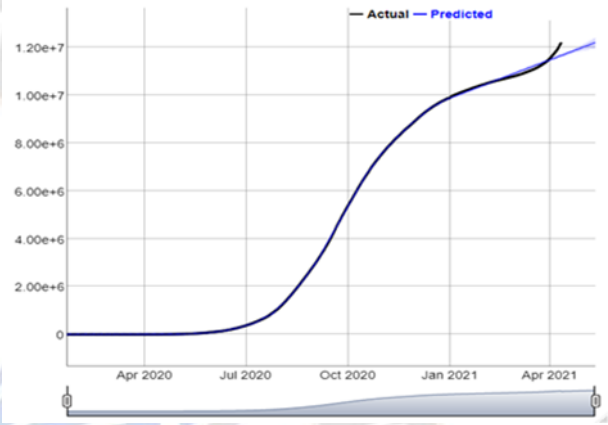


Figure 5. Predicted Forecast

Afterward, the legitimacy lists to see all the COVID-19 groups' quality are dissected in light of the file, index, capacity, and CE records. Figure 3 shows the new case and Death detail worldwide , Figure 4. Dendrogram (Cluster) shows death details of various countries and Figure 5 shows Predicted of COVID-19

## V. CONCLUSION AND FUTURE DIRECTION

The main procedure of this work to detect the covid clusters of countries which is having complex, average and lesser and corona cases between the countries on the earth and between the states in India. Proposed KMHC model generates the clusters. The aftereffects of the trials show that the nature of KMHC merges the advantages of Hierarchical algorithm and advantages of K-Means grouping calculation. It additionally introduced the patterns and conjectures of future Coronavirus include in India. This work closes mixture clustering calculation KMHC is the best model for grouping the Coronavirus informational collection got from world wellbeing association. These half breed grouping models, as well as the future Coronavirus count, are basic for clinical and medical care experts to go to proper lengths early. Thus, the medical care area's weight can be decreased by expecting future occasions and creating techniques and strategies. In future some additional time

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

series calculation can be incorporated to come by ideal outcome.

## REFERENCE

[1]. Anastasiya Doroshenko., "Analysis of the Distribution of COVID-19 in Italy Using Clustering Algorithms", IEEE International Conference on Data Stream Mining & Processing, pp:325-328, 2020.

[2]. Ginalber L.O.Serra, Daiana Gomes, "Machine Learning Model For Computational Tracking and Forecasting the COVID-19 Dynamic Propagation",IEEE Journal of Biomedical and Health Informatics, 2021.

[3]. Vishan Kumar Gupta, et.a., " Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model", IEEE Big Data Mining and Analytics - Volume: 4, Issue: 2, June 2021.

[4]. Rahmad Kurniawan et.al., " Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries", IEEE International Conference on Cyber and IT Service Management, 2020.

[5]. Valerio Bellandi; Paolo Ceravolo; Samira Maghool; Stefano Siccardi, "A Comparative Study of Clustering Techniques Applied on Covid-19 Scientific Literature" IEEE International Conference on Internet of Things: Systems, Management and Security, 2020.

[6]. C. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan China", Lancet, vol. 395, no. 10223, pp. 497-506, 2020.64

[7]. Remigio Ismael Hurtado Ortiz; Juan Carlos Barrera; Katherine Michelle Barrera., "Analysis model of the most important factors in Covid-19 through data mining, descriptive statistics and random forest", IEEE International Autumn Meeting on Power, Electronics and Computing, pp.4-6, 2020.

[8]. N. Zhu et al., "A novel Coronavirus from patients with Pneumonia in China 2019", N. Engl. J. Med., vol. 382, no. 8, pp. 727-733, 2020.

[9]. Pınar Cihan, "Fuzzy Rule-Based System for Predicting Daily Case in COVID-19 Outbreak", IEEE International Conference on Multidisciplinary Studies and Innovative Technologies, 2020.

[10]. S. S. Chenar and Z. Deng, "Development of genetic programming-based model for predicting oyster norovirus outbreak risks", Water research, vol. 128, pp. 20-37, 2018.

[11]. Shreyas Selur Arun, Ganesh Neelakanta Iyer, "On the Analysis of COVID19 - Novel Corona Viral Disease Pandemic Spread Data Using Machine Learning Techniques", 4th International Conference on Intelligent Computing and Control Systems, 2020.

[12]. Othman Istaiteh, Tala Owais, Nailah Al-Madi, Saleh Abu-Soud, "Machine Learning Approaches for COVID-19 Forecasting", International Conference on Intelligent Data Science Technologies and Applications, 2020.

[13]. Amit Bhati, Anurag Jagetiya, "Prediction of COVID-19 Outbreak in India Aopting Bhilwara Model of Containment", 5th International Conference on Communication and Electronics Systems, 2020.

[14]. S. Khanmohammadi, N. Adibeig, S. Shanehbandy, An improved overlapping k-means clustering method for medical applications, Expert Syst. Appl. 67 (2017).

[15]. B.A. Hassan, T.A. Rashid, Datasets on statistical analysis and performance evaluation of backtracking search optimisation algorithm compared with its counterpart algorithms, Data Br. 28 (2020) 105046.

[16]. M.B. Malik, M.A. Ghazi and R. Ali, "Privacy Preserving Data Mining Techniques: Current Scenario and Future Prospects", Proceedings of 3rd International Conference on Computer and Communication Technology, pp. 26-32, 2012.

[17]. Y.A. Alsahib, S. Aldeen, M. Salleh and M. Razzaque, "A Comprehensive Review on Privacy Preserving Data Mining", SpringerPlus, Vol. 4, pp. 694-705, 2015.

[18]. Benny Pinkas, "Cryptographic Techniques for Privacy-preserving Data Mining", ACM SIGKDD Explorations Newsletter, Vol. 4, No. 2, pp. 12-19, 2002.

[19]. Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yucel Saygin and Yannis Theodoridis, "State-of-the-Art in Privacy Preserving Data Mining", ACM SIGMOD Record, Vol. 33, No. 1, pp. 50-57, 2004.

[20]. WHO. Coronavirus disease (COVID-2019) situation reports n.d. 21. Worldometers. COVID-19 Coronavirus Pandemic n.d.

[21]. Panda S.K., Ravichandran C., Hazarika B. Results on system of Atangana–Baleanu fractional order Willis aneurysm and nonlinear singularly perturbed boundary value problems. Chaos Solitons Fractals.2021;142:110390. doi: 10.1016/j.chaos.2020.110390.

[22]. Ravichandran C., Logeswari K., Jarad F. New results on existence in the framework of Atangana–Baleanu derivative for fractional integro-differential equations. Chaos Solitons Fractals. 2019;125:194–200.

[23]. Logeswari K, Ravichandran C NK. Mathematical model for spreading of COVID-19 virus with the Mittag–Leffler kernel. Numer Methods Partial Differ Equ2020.

[24]. Hethcote H.W. The mathematics of infectious diseases. SIAM Rev. 2000;42(4):5 99–653. doi: 10.1137/ S00361445 00371907.

[25]. Bonyah E, Khan MA, Okosun KO, Islam S. A theoretical model for Zika virus transmission. PLoS One 2017;12:e0185540.

[26]. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. Proc R Soc London Ser A. 1927;115:700–721

[27]. Aleman D.M., Wibisono T.G., Schwartz B. Proc. 2009 winter Simul. Conf., IEEE. 2009. Accounting for individual behaviors in a pandemic disease spread model; pp. 1977–1985.

*International Journal on Recent and Innovation Trends in Computing and Communication*
*ISSN: 2321-8169 Volume: 11 Issue: 10*
*DOI: https://doi.org/10.17762/ijritcc.v11i10.8469*
***Article Received:** 23 July 2023 **Revised:** 20 September 2023 **Accepted:** 01 October 2023*

[28]. Kucharski A.J., Russell T.W., Diamond C., Liu Y., Edmunds J., Funk S. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis. 2020.

[29]. Ndaïrou F., Area I., Nieto J.J., Torres D.F.M. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos Solitons Fractals. 2020; 135:1z09846. doi: 10.1016/ j. chaos.2020. 109846.

[30]. Prem K., Liu Y., Russell T.W., Kucharski A.J., Eggo R.M., Davies N. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. Lancet Public Heal.

[31]. Hellewell J., Abbott S., Gimma A., Bosse N.I., Jarvis C.I., Russell T.W. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. doi: 10.1016/S2214-109X(20)30074-7.

[32]. Mizumoto K, Chowell G. Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020. Infect Dis Model 2020;5 :264 – 70.Liu Y., Gong M., Liu S., Pan Y., Huo Y. Effects of blood glucose on vaspin secretion in patients with gestational diabetes mellitus. Gynecol Endocrinol. 2021;37(3):221–224. 35. Fanelli D., Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals. 2020;134:1–12. doi: 10.1016/ j.chaos.2020. 109761

[33]. Chakraborty T., Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. Chaos Solitons Fractals. 2020;135:109850. doi: 10.1016/j.chaos.2020.109850.

[34]. Ribeiro M.H.D.M., da Silva R.G., Mariani V.C., Coelho L.dos.S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos Solitons Fractals. 2020;135:109853. doi: 10.1016/j.chaos.2020.109853.

[35]. Yang Z., Zeng Z., Wang K., Wong S.-S., Liang W., Zanin M. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J Thorac Dis. 2020;12(3):165–174.

[36]. Peng L, Yang W, Zhang D, Zhuge C, Hong L. Epidemic analysis of COVID-19 in China by dynamical modelling; 2020.

[37]. Li X., Zhao X., Sun Y. The lockdown of Hubei Province causing different transmission dynamics of the novel coronavirus (2019-nCoV) in Wuhan and Beijing. MedRxiv. 2020.

[38]. Wu J.T., Leung K., Leung G.M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020;395(10225):689–697.

[39]. Tang B, Wang X, Li Q, Bragazzi NL, Tang S, Xiao Y, et al. Estimation of the transmission risk of the 2019- nCoV and its implication for public health interventions. J Clin Med 2020;9:462.23.