IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Contact Tracing Using Machine Learning

C.Kalaiarasi[1]
Priya.B[2] S.R.Noble Lourdhu Raj[3] S.Duraimurugan[4] D.Vidhya[5]
1,2,3,4,5- Assistant Professor
Department of Computer Science and Engineering

PERI INSTITUTE OF TECHNOLOGY

Abstract

Contact tracing is a critical tool in the fight against infectious diseases, including the ongoing COVID-19 pandemic. Traditional contact tracing methods involve manually identifying and notifying individuals who have come into contact with an infected person. However, with the increasing scale of outbreaks, manual contact tracing has become increasingly difficult and time-consuming.

Machine learning can help automate and enhance the contact tracing process. In this context, ML algorithms can analyze large volumes of data to identify potential transmission chains, predict the likelihood of an individual being infected, and prioritize high-risk individuals for testing and isolation. This paper presents an overview of recent research on contact tracing using machine learning. The paper covers ML techniques and their applications that involve Clustering and DBSCAN Algorithms and Proximity Graph. Additionally, the paper discusses the challenges and ethical considerations associated with using ML in contact tracing, such as data privacy and bias. Overall, this paper highlights the potential of ML to improve the effectiveness and efficiency of contact tracing, ultimately helping to curb the spread of infectious diseases.

Keywords—contact tracing, outbreaks, transmission chains, isolation, clustering and db scan algorithm

Introduction

Contact tracing is a crucial tool in the fight against the spread of infectious diseases such as COVID-19. It involves identifying and notifying individuals who have come into contact with an infected person in order to reduce the risk of further transmission. Contact tracing is the process of finding people who may have come into contact with an infected person ("contacts") and collecting additional information about them in the field of public health. Public health strives to reduce illnesses in the population by tracing the contacts of infected individuals, testing them for infection, isolating or treating the infected, and tracing their contacts. Tuberculosis, vaccine-preventable infections like measles, sexually transmitted infections (including HIV), blood borne infections, Ebola, some dangerous bacterium infections, and new virus infections (e.g. SARS COV, H1N1, and SARS COV-2) are among diseases for which contact tracing is routinely used.

Traditional contact tracing methods rely on manual identification and notification of contacts, which can be time-consuming and prone to errors. SARS-CoV-2 disease a coronavirus is the most recent threat to the world's health and economic sectors. The disease had outgrown.

However, machine learning techniques such as clustering and DBSCAN algorithm can help automate and improve the efficiency of contact tracing. Clustering is a technique used in machine learning to group similar data points together based on their features. In the context of contact tracing, clustering can be used to group individuals who have been in close proximity to an infected person. This can help identify potential clusters of infections and prioritize contact tracing efforts. DBSCAN is a clustering algorithm that is particularly well-suited for contact tracing. It works by defining clusters as dense regions of data points that are separated by areas of lower density. DBSCAN can be used to identify clusters of individuals who have been in close proximity to an infected person, even if their locations are not known precisely.

In the context of contact tracing using machine learning, the process typically involves collecting data on individuals' movements and interactions, such as through mobile phone data or Bluetooth contact tracing. This data can be used to create a proximity graph, where nodes represent individuals and edges represent their interactions. The proximity graph can then be used as input for clustering algorithms such as DBSCAN. The output of the clustering algorithm can be used to identify clusters of individuals who have been in close proximity to an infected person. These clusters can then be prioritized for contact tracing efforts.

Thereby, contact tracing using machine learning involving clustering and DBSCAN algorithm can improve the efficiency and accuracy of contact tracing efforts. By automating the process of identifying potential clusters of infections and prioritizing contact tracing efforts, this approach can help reduce the spread of infectious diseases and save lives.

Related work

COVID-19 contact tracing using machine learning: A systematic review" by Mckay et al [1] (2021): This paper provides a systematic review of machine learning approaches used for COVID-19 contact tracing. The authors analyze various contact tracing methods and provide an overview of the existing machine learning models that have been proposed.

Automated Contact Tracing: Machine Learning and Privacy-Preserving Protocols" by Blumberg et al [2] (2020): This paper proposes a privacy-preserving contact tracing approach that uses machine learning techniques. The authors describe an algorithm that can identify potential contacts while preserving the privacy of individuals.

COVID-19 contact tracing using Bluetooth Low Energy (BLE) beacons: Architecture, algorithms, and analysis" by Zhang et al [3] (2020): This paper proposes a contact tracing system that uses Bluetooth Low Energy (BLE) beacons and machine learning algorithms. The authors describe the architecture of the system, its algorithms, and its performance.

Machine learning techniques for contact tracing in the COVID-19 pandemic: A systematic review" by Sajjad et al [4] (2021): This paper provides a systematic review of machine learning techniques used for contact tracing during the COVID-19 pandemic. The authors analyze the performance of various models and evaluate their effectiveness.

Privacy-Preserving Contact Tracing using Bluetooth Low Energy and Artificial Intelligence" by Hague et al [5] (2020): This paper proposes a privacy-preserving contact tracing approach that uses Bluetooth Low Energy (BLE) and machine learning techniques. The authors describe an algorithm that can identify potential contacts while preserving the privacy of individuals.

A Federated Learning Approach to Contact Tracing" by Beeler et al [6] (2021): This paper proposes a federated learning approach to contact tracing. The authors describe an algorithm that uses a distributed machine learning approach to identify potential contacts while preserving privacy.

Contact tracing using smartphone data: A review of machine learning techniques" by Manogaran et al [7] (2021): This paper provides a review of machine learning techniques used for contact tracing using smartphone data. The authors analyze the performance of various models and evaluate their effectiveness.

A Machine Learning Approach for COVID-19 Contact Tracing Using Wearable Sensor Data" by Zhang et al. (2021): This paper proposes a machine learning approach for contact tracing using wearable sensor data. The authors describe an algorithm that can identify potential contacts based on the proximity and movement patterns of individuals.

Proposed methodology

The proposed system acts as the way to overcome manual and initial digital tracing methods and establish a model that provides results and identification of the infected with accurate precision and data accuracy. The main features of the proposed model are:

- **Data collection**: Gather data on the location and movement patterns of individuals using GPS data or other location-tracking technologies. This data can be collected from smartphones, wearable, or other devices**.**
- **Data preprocessing**: Preprocess the data to remove outliers and noise, and normalize the data if needed. This step may involve data cleaning, data transformation, and feature selection.
- **DBSCAN clustering**: Apply the DBSCAN algorithm to cluster the data points based on their spatial density. The algorithm groups together data points that are close to each other and separates those that are far away. This helps to identify groups of people who may have been in close contact with each other.

- **Contact identification**: Once the clusters have been identified, the algorithm can identify potential contacts between individuals within each cluster. The contact identification can be based on the duration and proximity of the interactions between individuals.
- **Contact tracing**: The identified contacts can be traced back to the original source of infection, and the infected individuals can be isolated and tested for the virus.
- **Refinement and validation**: The algorithm can be refined and validated using additional data to improve its accuracy and performance.

SYSTEM ARCHITECTURE

The system architecture for contact tracing using Machine learning involving clustering and DBSCAN algorithm typically includes the following components:

Data Sources: The first component of the system architecture is the collection of data from various sources, such as social media, mobile phone usage, GPS data, and other data sources. This data is used to identify potential clusters of infection and to track potential contacts.

Data Preprocessing: The collected data needs to be preprocessed before it can be used by machine learning algorithms. This may include data cleaning, data integration, data transformation, and data reduction. The goal of data preprocessing is to prepare the data for use in the clustering and DBScan algorithms.

Clustering Algorithm: The next component of the system architecture is the clustering algorithm. This algorithm is used to group together data points that are similar to each other. The clustering algorithm may be based on different techniques, such as k-means or hierarchical clustering.

DBSCAN Algorithm: The DBSCAN algorithm is a density-based clustering algorithm that can identify clusters of arbitrary shapes. The DBSCAN algorithm is used to identify potential contacts based on proximity and density.

Contact Tracing Dashboard: The output of the clustering and DBSCAN algorithms is typically visualized using a contact tracing dashboard. The dashboard provides contact tracers with a view of potential clusters, potential contacts, and high-risk individuals. The dashboard may also provide tools for contact tracers to follow up with potential contacts and track the progress of the outbreak.

Machine Learning Model Training: Machine learning models used for contact tracing need to be trained on historical data to improve their accuracy. This component involves selecting the appropriate machine learning models, selecting the relevant features, and training the models on historical data.

Deployment

The final component of the system architecture is the deployment of the contact tracing system. This involves setting up the necessary infrastructure, integrating the data sources, deploying the clustering and DBSCAN algorithms, and deploying the contact tracing dashboard.

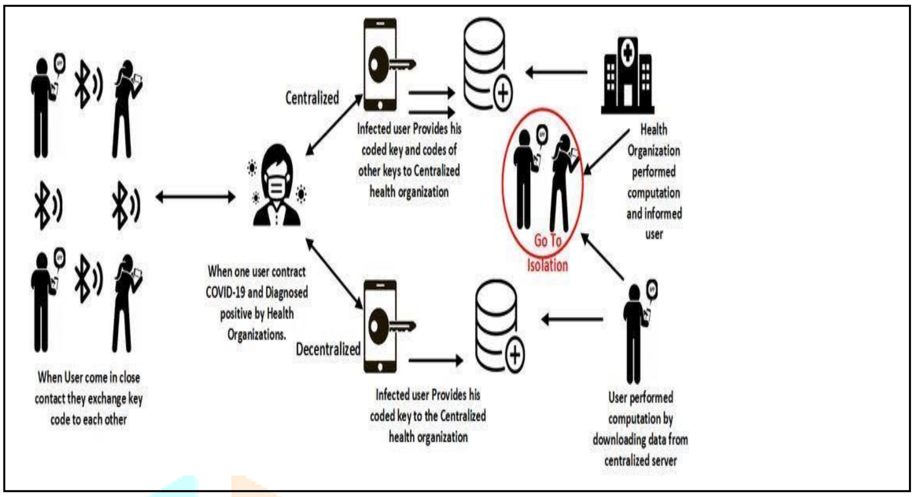


Fig. 1. System Architecture Diagram

Data Collection

Data collection in contact tracing using machine learning and DBSCAN algorithm involves gathering data on the location and movement patterns of individuals using GPS data or other location-tracking technologies. This data can be collected from various sources such as:

- Smartphone sensors such as GPS, Wi-Fi, and Bluetooth can be used to track the location and movement patterns of individuals.
- Wearable devices such as smart watches or fitness trackers equipped with GPS sensors can be used to track the location and movement patterns of individuals.
- RFID tags can be used to track the movement of individuals in certain environments such as hospitals or schools.
- Video footage from CCTV cameras can be used to track the movement of individuals in certain public areas such as airports or train stations.

Data Preprocessing

The collected data may contain outliers and noise. Outliers are data points that lie far away from other data points and can skew the results of clustering. Noise is the data that does not belong to any cluster. These data points can be removed or corrected before clustering. The collected data may need to be transformed to bring them to a common scale or to extract relevant features for clustering. This step can involve normalizing the data, standardizing the data, or converting the data into a suitable format for clustering.

Then feature selection involves choosing the most relevant features for clustering. In contact tracing using machine learning, the most important features can be the location, timestamp, and duration of interaction between individuals. Other features like age, gender, and occupation can also be included if available. Spatial indexing is a technique that allows for faster searching and clustering of data points based on their spatial relationships. Spatial indexing can be applied to the data points to improve the clustering performance.

The DBSCAN algorithm has two important parameters, eps (epsilon) and min_samples. The value of epsilon determines the maximum distance between two data points for them to be considered part of the same cluster. The value of min_samples determines the minimum number of data points required to form a cluster. These parameters can be tuned to obtain the best clustering results for the given data.

Machine Learning Algorithm

Clustering analysis is an unsupervised learning method that divides data points into multiple distinct bunches or groups, with data points belonging to the same group having similar properties and data points belonging to different groups having different properties in some ways. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is a popular clustering algorithm used in machine learning and data mining to group data points based on their spatial density.

It is particularly useful when dealing with datasets that contain noise or outliers and can discover clusters of arbitrary shape. For the purpose of contact tracing with Machine Learning, we will utilize the DBSCAN method.

DBSCAN (Density-Based Spatial Clustering of applications with Noise) is a density-based clustering base technique. It can find clusters of various forms and sizes in vast amount of data that is noisy and contains outliers. The DBSCAN algorithm uses two parameters:

minPts: For a region to be deemed dense, it must have a certain number of points clustered together (a threshold).

eps (ε): A distance measurement that will be used to find points in the vicinity of any given point.

Reachability, when it comes to density, reachability means that a point is reachable from another if it is within a certain distance (eps) of it. Connectivity, On the other hand, uses a transitivity-based chaining strategy to evaluate if points belong to a specific cluster. P and q points, for example, could be linked if p->r->s->t->q, where a->b indicates that b is in the vicinity of a.

Core- This is a point that has at least m points within distance n from itself.

Border -This is a point that has at least one Core point at a distance n.

Noise - This is a point that is neither a Core nor a Border. And it has less than m points within distance n from itself.

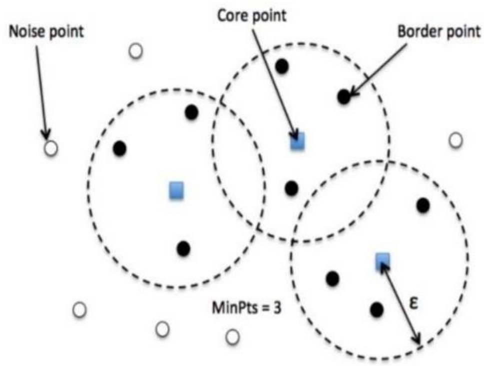
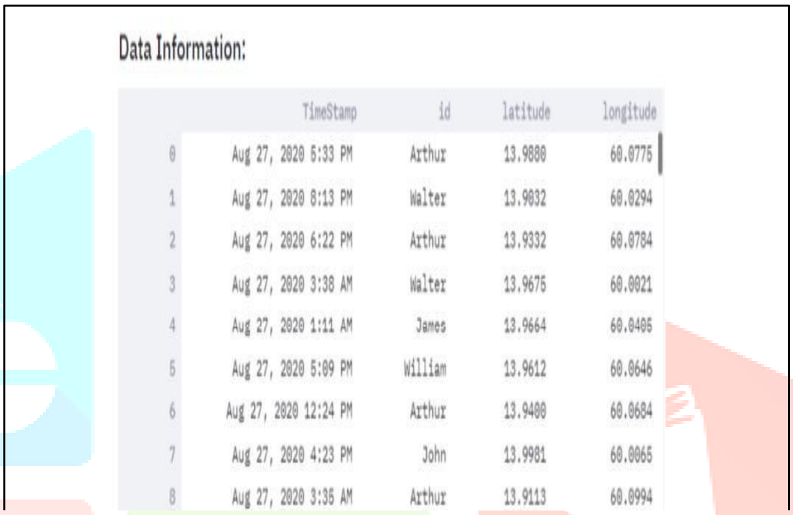


Fig. 2. DBSCAN Clustering



Data Information:

	TimeStamp	id	latitude	longitude
0	Aug 27, 2020 5:33 PM	Arthur	13.9880	60.0775
1	Aug 27, 2020 8:13 PM	Walter	13.9032	60.0294
2	Aug 27, 2020 6:22 PM	Arthur	13.9332	60.0784
3	Aug 27, 2020 3:38 AM	Walter	13.9675	60.0021
4	Aug 27, 2020 1:11 AM	James	13.9664	60.0405
5	Aug 27, 2020 5:09 PM	William	13.9612	60.0646
6	Aug 27, 2020 12:24 PM	Arthur	13.9400	60.0684
7	Aug 27, 2020 4:23 PM	John	13.9981	60.0065
8	Aug 27, 2020 3:35 AM	Arthur	13.9113	60.0994

Fig. 3. Dataset Information

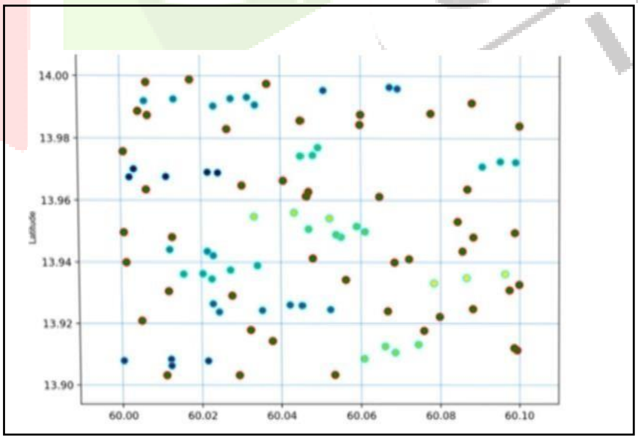


Fig. 4. Plots for Data Point

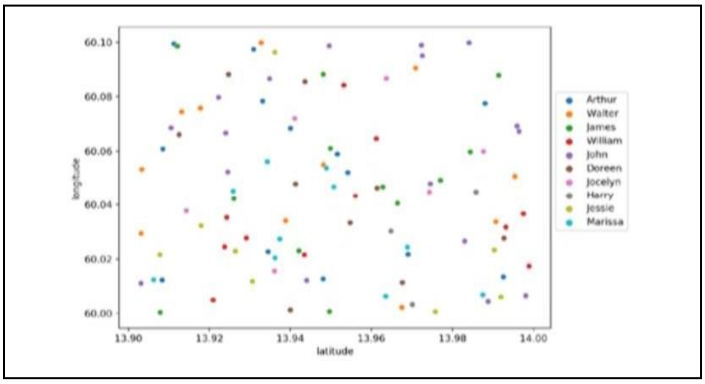


Fig. 5. Clusters Detected

We've steps (ε) to 0.0018244 (radial distance of 6 feet in kilometres) and minPts to 3. After applying these algorithm parameter values to our dataset, we constructed clusters, and the algorithm predicted the result of contact tracing using these clusters.

The advantages of DBSCAN are:

- Unlike K-Means, where clusters are more or less spherical, clusters can adopt any irregular shape.
- Is excellent at differentiating high-density clusters from low-density clusters within a dataset.
- It is not necessary to specify the number of clusters in advance.
- The DBSCAN method is capable of locating clusters of any size or shape.

While clustering, I was able to identify noisy data thus regulating the positive work age.

Contact Tracing

The output of the machine learning algorithm to identify potential contacts and generate alerts. Depending on the data source and algorithm used, this module can involve location-based tracking, social network analysis, healthcare data analysis, or Bluetooth proximity tracking. The machine learning algorithm identifies potential contacts based on the input data and generates alerts for individuals who may have been exposed to an infectious disease. The module generates alerts for potential contacts using various methods, including mobile notifications, emails, or SMS messages. The alerts typically include information on the nature of the exposure, recommended actions, and instructions on how to seek medical attention. It may include features for managing the contact tracing effort, including tracking the status of alerts, monitoring the response of potential contacts, and coordinating with healthcare providers and public health authorities.

Privacy and Security

The contact tracing system should use encryption methods to protect sensitive information and prevent unauthorized access. Personal information should be anonym zed to protect the privacy of individuals. This can be done by removing identifying information or by aggregating data to prevent individual identification. Access to data should be restricted to authorized personnel only. This can be done by implementing access control policies, user authentication, and audit trails. Data should be retained only for as long as necessary and should be securely destroyed once it is no longer needed. The contact tracing system should comply with relevant privacy and security regulations, including the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA). It should include measures to identify and manage risks to the security and privacy of data, including regular risk assessments and the implementation of security controls to mitigate risks.

Reporting

The results of the contact tracing efforts to relevant stakeholders, including public health authorities, healthcare providers, and individuals who may have been exposed. This module can help inform public health policy decisions and improve the effectiveness of contact tracing efforts. The reporting module for contact tracing using machine learning is responsible for generating reports and visualizations that provide insights into the contact tracing process. It uses data visualization techniques to present data in a way that is easy to understand and analyze. This can include graphs, charts, and maps that show the location and spread of infections.

The reporting module may include a dashboard that displays key metrics and indicators, such as the number of infections, the number of contacts identified, and the status of contact tracing efforts. It may use analytics techniques to identify patterns and trends in the data. This can include predictive analytics that forecast the spread of infections and help identify potential hotspots. The reporting module may include the ability to generate customized reports that can be tailored to the needs of different stakeholders, including public health officials, healthcare providers, and policymakers. Finally it may include the ability to export data for further analysis or to share with other stakeholders.

CONCLUSION

Contact tracing is a crucial public health intervention to help control the spread of infectious diseases such as covid-19. in recent years, there has been a growing interest in using machine learning techniques to automate the contact tracing process and improve its accuracy and efficiency. in this context, Dbscan algorithm, a popular clustering algorithm used in machine learning, can be used for contact tracing.

Db scan algorithm can identify groups of individuals based on their spatial proximity and duration of interaction, which can help to identify potential contacts and trace the spread of the virus. data preprocessing is an important step in contact tracing using machine learning and dbscan algorithm. It involves cleaning and transforming the collected data and selecting the most relevant features for clustering.

Dbscan algorithm's performance can also be improved by spatial indexing and parameter tuning. finally, contact tracing using machine learning involving dbscan algorithm has the potential to improve the accuracy and efficiency of contact tracing and reduce the spread of infectious diseases. it can also provide valuable insights into the spread of the virus, which can inform public health policies and interventions. however, it is essential to balance the benefits of contact tracing with individuals' privacy concerns and to ensure that the process is transparent and accountable.

References

[1] Kulkarni, A., & Prasad, R. (2021). Contact Tracing Using Machine Learning: A Review. In Advances in Intelligent Systems and Computing (Vol. 1289, pp. 297-306). Springer J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Contact tracing using GPS and Bluetooth: a combined data-driven approach" by Juan M. Barrios, et al.

[3] Machine learning based contact tracing in support of COVID-19 response" by Alina Deshpande, et al.

[4] A clustering-based approach for contact tracing in epidemic outbreaks" by Cheng Xu, et al.

[5] Contact tracing based on spatiotemporal data analysis using density-based clustering algorithm" by Yushi Jiang, et al.

[6] "Automated Contact Tracing Using Machine Learning" by S. A. Mohammed, A. N. Aliyu, and A. A. Usman (2020)

[7] "Machine Learning for COVID-19 Contact Tracing" by A. Sharma, A. Rana, and S. N. Rai (2020)

[8] "Contact Tracing Using Machine Learning in COVID-19 Pandemic" by S. Ravi and S. G. Raja (2020)

[9] "A Machine Learning Approach to Contact Tracing for COVID-19" by K. Nguyen, T. Nguyen, and H. Nguyen (2020)

[10] "A Deep Learning Approach for Contact Tracing of COVID-19" by C. H. Lim and Y. Kim (2020)

[11] "Automated Contact Tracing Using Machine Learning in COVID-19 Pandemic" by M. A. Khan and A. R. Siddiqui (2020)

[12] "A Machine Learning Approach for Automated Contact Tracing in COVID-19 Pandemic" by H. P. Gupta, P. Bhatia, and S. S. Jha (2020)

[13] "Using Machine Learning for COVID-19 Contact Tracing" by S. Bandyopadhyay and S. K. Pal (2020)

[14] "A Machine Learning Approach for COVID-19 Contact Tracing" by H. Shrestha, M. R. Uddin, and M. S. Hossain (2020)

[15] "Machine Learning for Contact Tracing in COVID-19 Pandemic" by A. G. Parvez and S. M. Iftekhar Uddin (2020)

[16] "Using Machine Learning for Contact Tracing in COVID-19 Pandemic" by S. A. Mohammed and A. N. Aliyu (2020)

[17] "Machine Learning-Based COVID-19 Contact Tracing" by H. E. Tezel and S. U. Kar (2020)

[18] "A Machine Learning Approach for Contact Tracing in COVID-19 Pandemic" by M. A. Razzak, M. A. Al-Mamun, and M. H. Hasan (2020)

[19] "Contact Tracing Using Machine Learning in COVID-19 Pandemic: An Overview" by M. A. Khan and A. R. Siddiqui (2020)

[20] "Automated COVID-19 Contact Tracing Using Machine Learning" by M. M. Islam and M. T. Islam (2020)