Quest Journals
Journal of Software Engineering and Simulation
Volume 9 ~ Issue 4 (2023) pp: 08-14
ISSN(Online) :2321-3795 ISSN (Print):2321-3809
www.questjournals.org

Research Paper

Covid-19 Detection and Pandemic Prevention System Using Data Science

Abhimanyu Singh Negi
Dept. Of Computer Science And Engineering,
Galgotias College of Engineering And Technology,
Greater Noida, India

Prakhar Gupta
Dept. Of Computer Science And Engineering,
Galgotias College of Engineering And Technology,
Greater Noida, India

RashiSrivastava
Dept. Of Computer Science And Engineering,
Galgotias College of Engineering And Technology,
Greater Noida, India

Arun Kumar Rana
Dept. Of Computer Science And Engineering,
Galgotias College of Engineering And Technology,
GreaterNoida,

Abstract**—***The Covid-19 pandemic has caused a major disruption to the global economy and has put a strain on healthcare systems around the world. To be able to mitigate the spread of the virus, it is essential to develop an effective detection and prevention system. This system should be able to detect the presence of the virus in a population and provide early warning to the public health authorities. Data science, machine learning, and Internet of Things (IoT) technologies can be used to develop such a system. Data science can be used to analyse large amounts of data to identify patterns and trends in the spread of the virus. Machine learning algorithms can be used to identify potential hot spots and predict the spread of the virus. IoT technologies can be used to monitor the health of individuals and provide real-time data to the public health authorities.The system uses I.O.T devices to monitor the environment and detect any changes in temperature, humidity, and air quality that may indicate the presence of the virus This system can be used to provide early warning to the public health authorities and help them take necessary preventive measures to be able to contain viral infection and spread.*

Keywords**—***Internet of Things, Machine Learning Algorithms, Improving Security etc.*

R*eceived 25 Mar., 2023;* ***R****evised 05 Apr., 2023;* ***A****ccepted 07 Apr., 2023 © The author(s) 2023.* ***P****ublished with open access at* ***www.questjournals.org***

I. INTRODUCTION

The 2019 coronavirus disease (COVID-19) has had such a profound effect on the entire planet and slowed down normal human activities in such an unprecedented fashion that it will leave an enduring mark on history.humankind. distinct international locations have adopted severe measures to construct resilience in opposition to this existence-threatening sickness. Nonetheless, the pandemic's notably contagious nature has put the conventional medical models of care to the test. Hence, machine learning (ML) and artificial intelligence (AI) present new channels for efficient treatment throughout this pandemic. Designing effective diagnosis procedures and creating disease spread projections are all tasks that AI and ML can help with. **[1]**

These applications rely heavily on real-time patient monitoring and efficient information coordination, where the Internet of Things (IoT) plays a crucial role. Applications like automated drug distribution, answering patient questions, and tracing the origins of disease can all benefit from IoT. The potential use of AI, ML, and IoT technologies in the fight against the COVID-19 pandemic has been thoroughly examined in this research. A thorough overview of the enabling tools and methodologies is given together with an explanation of the current and future applications of AI, ML, and IoT. There is also a critical examination of the dangers and restrictions of the technologies.

Future research may consider National level integration of the project and keeping record for future pandemics and widespread diseases by modifying the model. The COVID-19 pandemic is currently affecting countries around the world, making the need for affordable solutions to tackle the pandemic urgent. The Internet of Things (IoT) is a network of mechanical and digitally connected objects that may transmit data over a predetermined network. The Internet of Things (IoT) is a cutting-edge technology that makes it possible for gadgets to link to networks in hospitals and other key locations to improve the fight against the COVID-19 epidemic. In order to battle the COVID-19 epidemic, this study analyses and highlights the uses of IoT by providing a perspective roadmap.The use of Internet of Things (IoT) will assist patients, doctors, healthcare professionals, and hospital management systems in recognising the signs of an infectious disease and managing COVID-19 infections around the world. **[2][3]**

II. LITERATURESURVEY

Wang P.W., Horby F.G., Hayden G.F. discussed The COVID-19 coronavirus outbreak, which started in Wuhan, China, appears to be uncontrollable. As of March 1, 2020, the virus had already infected more than 558,502 individuals worldwide and caused at least 25,251 fatalities **[1].**

Unknown-cause pneumonia outbreak connected to the Huanan Seafood Market began in Wuhan, Hubei province, China. The genus of the virus beta coronavirus and has connections to the viruses that cause severe acute respiratory syndrome (SARS) and Middle Eastern Respiratory Disease (MERS) (SARS) **[3].**

Although numerous instances of pneumonia with an unclear origin were found much earlier (Dec. 8, 2019), the WHO was first notified on December 31 **[5].**

The unchecked spread of the disease was facilitated by the tardy declaration of an epidemic and the neglect to promptly notify foreign authorities. The world is currently paying close attention to this outbreak. China extended a lockdown of the centre city Wuhan to cover 20 cities and 56 million people in order to slow the spread of the virus. The viability of this attempt was first questioned by experts, who also cautioned that a repeat of the SARS pandemic posed a threat to the nation **[9].**

The pandemic is currently under control in the majority of China, but criticism of the deployment of what some have called "draconian" measures to stop its spread persists. The World Health Organization (WHO) has designated the 2019 new coronavirus outbreak as COVID-19. This unprecedented outbreak has put several governments throughout the world in a vulnerable position. The COVID-19 outbreaks' effects, which were previously limited to China's inhabitants, are becoming a source of serious concern for almost all nations in the world**. [8],[13],[15]**

The majority of these nations are currently in partial or total lockdown due to a lack of resources to combat the COVID-19 epidemic and concern over overstretched healthcare systems. More than 3 million laboratory-confirmed coronavirus illnesses had been reported as of 30 April 2020, an alarming increase in the number of cases worldwide. Since the COVID-19 outbreak, there have been several false reports, inaccurate information, and unwarranted concerns about coronavirus that have only made matters worse. In response to such actions, we give a thorough overview of all the key elements related to the COVID-19 pandemic, drawing on numerous trustworthy sources. This observation emphasises the COVID-19 outbreak's impact on the global economy in addition to the immediate health consequences. As things come to a close, we learn that the COVID-19 outbreak can be impacted less by technology, including the internet of things (IoT), unmanned aerial vehicles (UAVs), blockchain, artificial intelligence (AI), and 5G, among others**.[7]**

III. FRAMEWORK

Since research will include the healthcare records and medical jargon, proper consultation with medical professionals will be needed in each step with more of the test-based approach. Incorporating Machine Learning and I.O.T makes the model more of the experience based. Hence the database will be made for increasing the efficiency of the system**.[6]**

3.1Study design and Settings

The system will be developed in the phases with each phase incorporating some features and testing with real time scenarios. The Study Design of the project will be experiment based as the work in this domain is limited with analysis of results of each phase. Study done during this project has been extensively focused on Identifying the diseases more accurately from the symptoms and making the machine learning model more accurate with more emphasis on sample collection and increasing the diversity of data.

The system will be developed in the phases with each phase incorporating some features and testing with real time scenarios. The Study Design of the project will be experiment based as the work in this domain is limited with analysis of results of each phase. Study done during this project has been extensively focused on Identifying the diseases more accurately from the symptoms and making the machine learning model more accurate with more emphasis on sample collection and increasing the diversity of data.

Fig 3.1 Flowchart of Framework

3.2 Variables

Some Key variables that will be incorporated during the project are-
1. Age
2. O2 Levels
3. Medical Records
4. Temperature
5. Chronic Diseases.

More variables and parameters can be added in the upcoming phases of development and research**.[10]**

3.3Controls

Control Incorporated during this project are controlling the error to accuracy ratio, controlling the machine learning algorithm and removing the instances of false positive and false negative by ensuring the proper verification of the test results.

3.4Study Method

The Study Methods include reading previous research works related to this field and learning about the symptoms of diseases and how it is spread over a region. Along with learning about various technologies and methods of data collection that can be used in the project**.[9],[14]**

3.5 Data Collection

Data Collection will be done via device that will have sensors to measure O2 levels and Body temperature as well as including the assessment of the patient with past medical records also in consideration the collected Data will then be analyzed to make a conclusion which is initial phase will first be verified by humans to check the accuracy.

3.6Ethical Clearance

Ethical clearance is the key in this project as the project works on collecting the sensitive healthcare information of the public which can be a major concern for the security and privacy. This can be solved by getting the clearance from the Ministry of Ayush and ensuring the implementation of proper data security protocols.

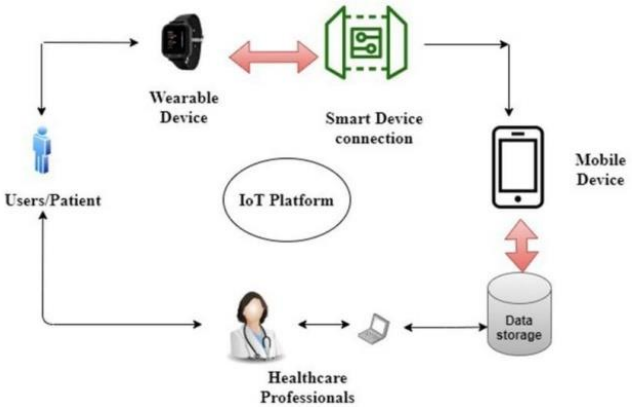

Fig 3.2 Proposed Framework

3.7Algorithm:

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)Input: CSV File Data
Output: Possible Positive results

The data clustering algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is employed in unsupervised learning. Its foundation is the concept of density-reachability and density-connectivity, which help identify clusters of points in a dataset that are separated from each other by a certain distance. It is an iterative algorithm that starts with an arbitrary starting point and then looks for points within a given radius that is density-connected. If a point is density-connected to a cluster of points, then it is added to the cluster. The algorithm continues to look for points that are density-connected to the cluster until all points are considered. It is well-suited for datasets with varying densities and can identify outliers in datasets**.[12]**

DBSCAN and k-means are two popular clustering algorithms. DBSCAN is a density-based clustering algorithm that is especially useful for finding clusters of varying density in a dataset. It uses a core distance metric to identify clusters in the data. In contrast, k-means is a centroid-based clustering algorithm that finds clusters based on their proximity to a predetermined number of centroids. K-means is better suited for finding clusters of uniform density. Both algorithms are useful for different types of data and applications, so it is important to choose the right algorithm for your use case**.[4]**

DBSCAN is distinct from other clustering algorithms in that it does not require the user to specify the number of clusters to generate. It also allows for clusters of varying densities and can identify noise and outliers. In contrast, other clustering algorithms such as K-means and hierarchical clustering require the user to specify the number of clusters they want to generate and are unable to identify noise and outliers. Additionally, these algorithms cannot handle clusters of varying densities**.[11],[16],[17]**

Steps:

3.7.1. Data is first being collected from sensors in the basic health kit.

3.7.2. That data will then be stored in the database via cloud and I.O.T.

3.7.3. A.I algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noise) will go through the data and model then will give result.

3.7.4. The result will then be compared with previous medical records to check for false alarms.

3.7.5. If positive the findings will be shared with the Govt and concerned medical authorities to give medical attention ASAP.

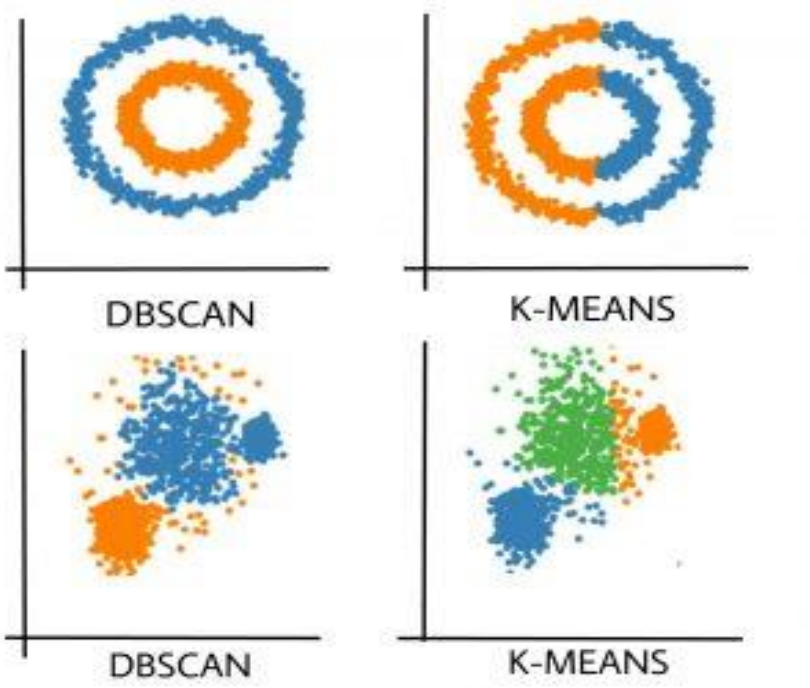

Fig 3.3 DBSCAN Vs K-Means

3.8 System Design

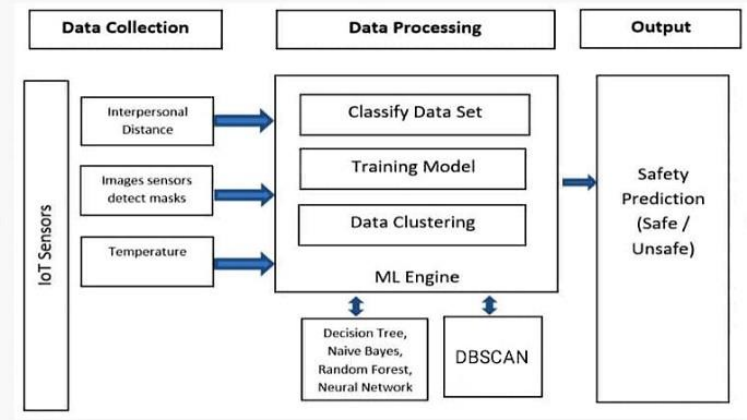

Fig. 3.4 Functional Specification of System

3.8. FUNCTIONAL SPECIFICATION OF SYSTEM
The following sub-systems make up the Covid-19 detection and pandemic prevention system:

3.8.1. Clients and web browser instant messaging system. The Directory Server is used to authenticate clients and store their identities. When the chatter starts the next talking session via SMTP, the online messages are transmitted to the messaging server. Suspicious phrases extracted from instant chats sent between chatters using predefined database criteria.

3.8.2. Active Chat Monitoring and Suspicious Chat Detection systems use OBIE and categorization rules to monitor the system. Using ontology, data pre-processing extracts domain and context (OBIE). CBA is used to build meaningful association rules.

3.8.3. Information database (Message DB, code word DB, Ontology DB, short word) The design of the Active Chat Monitoring and Suspicious Chat Detection system monitoring system in Instant Messaging System (IMS) is shown in Figure 1, along with the interrelationships amongst diverse subsystems so as to locate suspicious phrases based at the context of the on the spot message. Fig. 3.5 Flowchart IV endA network-received sickness called COVID-19 exhibits symptoms similar to those of the flu and bacterial pneumonia.

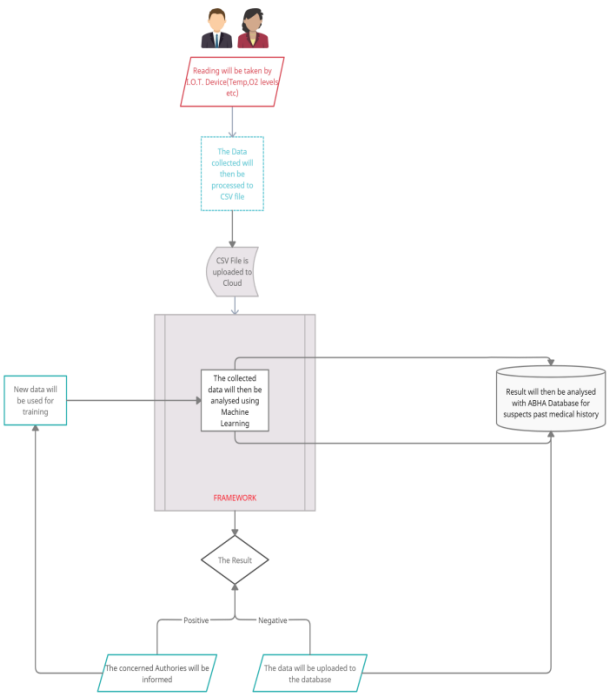

Fig. 3.5Flowchart

IV. CONCLUSION

It has had a devastating impact on the economy, social life, education, and technology throughout the entire planet. To lower the risk of COVID-19, many health organisations agree that it's crucial to keep your distance and use a mask. Another one of COVID-19's warning signs and symptoms is fever. In this research, we have put forth a novel method that makes use of ML algorithms to track the COVID-19 risk in public areas by measuring temperature, physical separation, and the use of masks.By including more risk parameters in the system, our model is also capable of making judgements to calculate risk using ML algorithms. In this paper, a novel strategy has been proposed. for identifying safe and risky regions while monitoring crowds utilising IoT technologies and machine learning techniques. IoT sensors gather information for the feature vector from both interior and outdoor locations. For classification, grouping, and prediction, ML techniques are used. With the same training and testing datasets, the decision tree algorithm achieves 94.50% accuracy, the DBSCAN classifier achieves 99.37% accuracy, and the random forest algorithm achieves 97.32% accuracy. The K-means algorithm is used for clustering, and compared to other algorithms, it performs better for two clusters than expectation maximisation (EM).The results of time performances for training and testing the model are likewise quite encouraging. In contrast to the neural network approach, which needs 0.89 s to generate the model, locally weighted learning only needs 0.01 s. In general, testing models takes less time than training them. For 800 crowd data points, locally weighed learning requires 1.72 seconds, DBSCAN classifier requires 0.07 seconds, support vector machine requires 0.04 seconds, and the other algorithms require only 0.01 seconds.

By assessing risks using datasets, we were able to attain excellent validation rates for our model. Using the DBSCAN classifier, we were able to attain a 99.37% classification accuracy. In our model, we employed a fixed dataset. But when they're inside, individuals frequently move at a specific pace. This study's drawback is connected to the precise collection and usage of a significant amount of data from moving objects. In our future work, we intend to incorporate additional ML methods, such as deep learning strategies, to enhance our model.

V. LIMITATION

•The model cannot be 100% accurate so it will never replace the medical test.
•Users must use provided chat application.

VI. FUTURE SCOPE

An continuous pandemic known as COVID-19 has put the health of people all around the world in grave peril recently. In this work, a strategy for predicting the risk of COVID-19 using DL has been developed. The framework analyses the actual dataset of daily data and applies deep learning algorithms to generate forecasts for the days ahead. This analysis establishes the top-tier activation function for M-LSTM, particularly a set of deep reinforcement learning guidelines to enhance the outcomes of predictions. The suggested approach was compared to popular current algorithms like LR and LSTM. The results of this study demonstrate how effectively the DL approach can forecast COVID-19 cases in the future.Typically, it can be deduced that the model's predictions are in line with the virus's fame; this can aid in understanding and reducing the virus' propagation.

In order to address the COVID-19 situation, it may be very helpful to act promptly and make informed judgements. We recommend employing a semi-supervised hybrid scheme to identify COVID-19 and social media sites in the near future in order to stop further spread. Also, a dashboard displaying the anticipated results will be made available using Google statistics Cloud. Provide a tool that can identify questionable words based on the surrounding sentences. The items listed below can be used for it.

- Detecting possible suspects of COVID-19.
- Detecting pandemic before happening.
- Making it available to everyone.
- Using that data for sending it to database.

REFERENCES

[1]. Wang P.W., Horby F.G., Hayden G.F. GaoA novel coronavirus outbreak of global health concern.Lancet. 2020;395(10223):470–473. [PMC free article] [PubMed] [Google Scholar]
[2]. S. Tuli, S. Tuli, G. Wander, P. Wander, S.S. Gill, S. Dustdar, R. Sakellariou, O. Rana, Next generation technologies for smart healthcare: challenges, vision, model, trends and future directions, Internet Technol. Lett.e145.
[3]. AtulSharma,a, SwapnilTiwari,bManasKantiDeb,b and Jean Louis Marty Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies.
[4]. DBSCAN++: Towards fast and scalable density clustering,Jennifer Jang 1 Heinrich Jiang.
[5]. R.K.Mojjada,A.Yadav, A.V. Prabhu and Y.Natarajan Machine learning models for covid-19 future forecasting,2020.
[6]. SM Abu Adnan Abir, S.N Islam, A. Anwar, A.N. Mahmood, A.M.T Oo Building resilience againstCOVID-19pandemic using artificial intelligence, machine learning, and IoT: A survey of recent progressIoT 1 (2), 506-528, 2020.

[7]. The COVID-19 pandemic preparedness ... or lack thereof: from China to ItalySimone Villa, Andrea Lombardi, 3,DavideMangioni, 4 Giorgio Bozzi, 3 Alessandra Bandera, 2 3Andrea Gori, 1and Mario C. Raviglione.
[8]. Y. Zhang, B. Jiang, J. Yuan, Y. Tao the impact of social distancing and epicenter lockdown on the COVID-19 epidemic in mainland China: a data-driven SEIQR model study medRxiv (2020)
[9]. Next generation technologies for smart healthcare: challenges, vision, model, trends and future directions ShreshthTuli, ShikharTuli, Gurleen Wander, Praneet Wander, Sukhpal Singh Gill, SchahramDustdar, RizosSakellariou, Omer Rana.
[10]. NishthaManchanda, Nursing Officer (BL Kapoor Max) reg. Delhi Nursing Council DNC. (Covid 19 and how to detect it in preliminary stages)
[11]. Chakraborty, Sanjay. (2011). Analysis and Study of Incremental DBSCAN Clustering Algorithm.International Journal of Enterprise Computing and Business Systems. 1.
[12]. Techniques to Enhance the Performance of DBSCAN Clustering Algorithm in Data Mining Authors: ShyamLal, Dr.Vaishali Singh
[13]. Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death) Cases based on Machine Learning Analysis Mohammad Khubeb Siddiqui1, Ruben Morales-Menendez1, Pradeep Kumar Gupta2, Hafiz M.N. Iqbal1, Fida Hussain1, Khudeja Khatoon3 and Sultan Ahmad.
[14]. Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review Author links open overlay panelNorahAlballa, Isra Al-Turaiki.
[15]. Response to "Body temperature correlates with mortality in COVID-19 patients". Anne M. Drewry, Richard Hotchkiss & Erik Kulstad.
[16]. DBSCAN Clustering Algorithm Based on Big Data Is Applied in Network Information Security Detection Yan Zhang Department of Information Engineering, ShiJiaZhuang University of Applied Technology, Shijiazhuang 050081, Hebei, China.
[17]. A Technical Survey on DBSCAN Clustering Algorithm NidhiSuthar (Department of Compute Engineering, HashmukhGoswami collage of Engineering, Vahelal, Gujarat.nidhisuthar2610@gmail.com), Prof. IndrjeetRajput (Head of Department of Compute Engineering,HashmukhGoswami collage of Engineering, Vahelal, Gujarat.), Prof.Vinitkumar Gupta (Department of Compute Engineering,HashmukhGoswami collage of Engineering, Vahelal, Gujarat)