Introducing Self-Sustainable Cloud Platform for Data Management and Extraction of Actionable Knowledge for Smart Healthcare Industry: A COVID-19 Case Study

Anuradha Thakare[a] , Mohit Burkule[b], Shravani Nimbolkar[c], Marzieh Khakifirooz[d]

[a, b, c] Department of Computer Engineering, Pimpri Chinchwad College of Engineering Pune, India
[d] Department of Industrial Engineering Tecnológico de Monterrey Santa Fe, CDMX, Mexico
[a]anuradha.thakare@pccoepune.org, [b]mohitburkule5@gmail.com, [c]nimbolkarshravani@gmail.com, [d]mkhakifirooz@tec.mx

Keywords: COVID-19 pandemic, Cloud platform, Self-sustainable systems, medical data, Artificial Intelligence, Smart City.

Abstract

The novel COVID-19 is a highly contagious disease. Data scientists worldwide are attempting to respond to the pandemic by building Artificial Intelligence (AI) solutions like forecasting pandemic growth, speculating possible mutations, identifying the symptoms caused, and many more. The models require vast quantities of data to make predictions. For a newly identified virus, it may take many months or sometimes years to collect related data and prepare it for data analysis purposes, which can further delay the process of making AI solutions. Hence, there is a need for a pipeline system which can facilitate a quick transmission of medical data from healthcare providers to data scientists. This paper proposes a cloud computing platform that allows smart cities to respond to the pandemic faster, with the collaboration of public health centers and data scientists. The platform provides a structured way of identifying and utilizing collaboration opportunities between health centers and the data science community, generating actionable knowledge. The system consists of two parts: 1) The software on the hospital's side, allowing real-time data gathering and automated uploading to cloud servers, 2) The cloud system to facilitate data storing along with model building and deploying. Customers can use the deployed models on a prepaid basis, the money collected will be divided among data scientists and data providers. This unique feature ensures the healthy participation of data providers in the process of making healthcare solutions for the smart cities. A sponsor can also sponsor a project. Hence, the system will sustain on its own with the involvement of stakeholders.

1 Introduction

In recent few decades, researchers have attempted to include solutions in the healthcare domain by incorporating technologies such as cloud computing, the Internet of Things, & Artificial intelligence [1]. With AI, computers can work on complex medical problems such as de-testing chronic diseases like brain tumours, cancers, Alzheimer etc. Mishra SG et al. provides a detailed account of the importance of Artificial intelligence in healthcare [2]. Complex Neural networks and dashboards can be made for various applications such as screening and diagnosis, simulation and modelling[3]. There are AI solutions to various problems like medical diagnosis [4], breast cancer detection [5], Radiology [6, 7], diabetic retinopathy screening [8] to Drug Discovery [9, 10, 11]. Rajinder Sandhu et al. [12] has designed a health monitoring system that provides feedback on an individual's health. [13] used AI in rapid screening of COVID 19 patients using Xray Images. An AI-based start-up, Bluedot, had predicted the early signs of the COVID-19 pandemic [14] with textual data scans from various sources. Having a continuous stream of data is a fundamental requirement for any making accurate prediction systems. There is currently a disconnect between what kind of medical datasets are available, what dataset a data scientist needs, and what datasets a data scientist can find, trust, and use [15]. Before choosing the problem statement, data scientists have to look for the availability of relevant medical data; many times, by directly contacting hospitals, causing delays in developing AI models. This paper proposes an efficient cloud platform for collecting, storing, and sharing medical data to address all these limitations. By ensuring data availability, it leverages the development of smart AI solutions. Our proposed system allows 1) Collecting close to real-time medical data from authentic sources - hospitals, institutes and renowned organizations, 2) Making the reliable medical data readily accessible publicly on a single cloud platform, thereby allowing data scientists to focus only on building and increasing the performance of AI models, 3) Making it easier to deploy the AI models, 4) Returning the credits to data providers. Data scientists can also request any specific data. We also analyse the benefits of our system, especially during a pandemic. The remainder of the paper is laid out as follows; section 2 presents the related research and its evaluation in section 3. In section 4, we discuss the proposed system, followed by its advantages in section 5. Implementation plan and the challenges are discussed in section 6; the role of the proposed system during the pandemic in discussed in detail in section 7. Followed by Conclusion and Future Direction in section 8.

2 Related Research

2.1 Data-driven systems and their role in pandemic

Data-driven systems are specialized software or applications that facilitate data acquisition, data handling/ maintenance, and presentation. Data is the fuel for data-driven systems. These systems have transformed telemedicine, remote health monitoring systems, wearables, etc., especially during COVID-19. It is the volume and rate at which data is generated which makes things complex [16]. Generally, healthcare datasets collected in clinical and non-clinical segments are majorly divided into seven types-Biomedical images, Electronic health records, Social Media, Clinical text, Genomic data, Biomedical signals and Sensing data [17]. By collecting such a variety of data, an AI-driven algorithm identified the outbreak before public health officials[18]. These data-driven systems should exploit the data to gain more knowledge about the virus and its behavior. A common way to disseminate data about infections such as COVID-19 is through data visualizations and simulated disease models [19]. E.g., The John Hopkins COVID-19 dashboard [20], real-time data visualization and forecasting tool by the Institute for Health Metrics and Evaluation [21]. Google launched a platform for COVID 19 data where all data collected from different resources has made available [22].

2.2 Cloud computing and other services

Cloud computing has valuable applications in healthcare and its support services due to its capability of handling varying demands [12]. Obtaining and analyzing healthcare data is a critical factor in getting ahead of and slowing the spread of disease, which is the case with the current COVID-19 pandemic[23]. Cloud-based applications can address existing health care issues, as discussed by Griebel et al. [24], Harsha et al. [25], and Kuo [26]. Despite the expected benefits, however, the rate of adoption and successful use of cloud computing in the healthcare sector remains considerably low [27, 28]. The highest adoption of cloud computing is into the banking industry at 59%. In contrast, the healthcare sector has been reported the least with 39% adoption [29]. Omar Ali et al. mentioned numerous cloud computing issues in healthcare, such as technological issues, security and privacy issues, etc. [30]. Rolim et al. [31] proposed a cloud system capable of automatically compiling patient data. The system discussed by Venkatesh et al. [32] can eradicate the errors which often occur while manually collecting the data, increasing the quality of data. Rao et al. [33] put forth 'Dhatri', which is a pervasive cloud initiative. The use of cloud computing has helped physicians to access the medical data of patients whenever required. As stated by Adamu et. al.[34], PaaS(Platform as a Service) provides an environment as a service for developers to code and develop software application platforms suitable for appropriate SaaS(Software as a Service) subscribers. PaaS services can be open-source or commercial and support programming languages like python and R, libraries, tools and frameworks. Data scientists choose a suitable platform depending on the factors like data size, availability of tools and computational capability. Google Colaboratory, IBM Cloud Dashboard, Kaggle Notebook, Microsoft Azure ML, MatrixDS, Amazon Sagemaker are some services that allow model building and data visualization. A data repository is an online service that holds data in an organized manner for research use. Usually, there are some restrictions on who can submit data to the data repository. However, many open data repositories allow anyone to submit data. University-owned data repositories usually require a university/academia login to upload or access data. In some cases, data scientists are even required to integrate data from different sources to form a new dataset. If the data is not available anywhere, then they have to painstakingly collect the data on their own using proper medical equipment.

3 Evaluation of Related Research

There are many limitations of current online and offline data searching methods. While searching for data online, there can be multiple sites showing similar datasets from different sources. Also, there may be many versions of a similar dataset. There may also exist bogus or phishing websites that may claim to have a particular dataset but redirect to some malicious page. For large datasets, few data hosting sites provide APIs to download data to the training environment directly. In the absence of direct APIs, data scientists will have to download it from the hosting website locally and then upload it to the training environment causing inconvenience and unnecessary delay. When searching for medical datasets offline, the process of contacting a data-providing organization and getting the required data from them consumes much time. Also, data might get corrupt, or modified which poses a question on its authenticity. Such manual hunting of the data is a very inefficient & tedious task that impedes the development of solutions. The use of medical data, which includes any personal information, may be prohibited by licenses or terms & conditions [35]. Data scientists might realize the need for additional features that are required during model building, which can improve the model performance. However, there is usually no way to communicate these feature requirements to the data provider. Moreover, deploying the model for public use requires its hosting on some servers. This process demands some knowledge of website building and hosting– an unsaid prerequisite for being a good data scientist.

4 Proposed Cloud platform for autonomic real-time collection, analysis and sharing of medical data

4.1 Background

The first step for any data science project is the data acquisition and integration from multiple sources. Next, data is pre-processed, cleaned and filtered. Finally, a model architecture is chosen and trained until satisfactory performance is obtained and then it is deployed. We are proposing a cloud platform paired with a local subsystem that will make this whole process efficient and speedy, especially during a pandemic. We do this by providing a pipeline-like system for a pseudo-real-time data flow and a common platform for developing and deploying the AI models.

4.2 Proposed System

The cloud platform can be visualized in two main parts as shown in Figure 1. The cloud platform and a local subsystem. The latter assists the former in automating the data upload. The two main subparts are discussed below.

a) Cloud platform (a website)

This cloud platform will link data providers to various stakeholders such as data scientists, clients, and sponsors.

The website will serve the following purposes

i. Platform to display and provide access and find to the medical data.

ii. Platform to deploy the AI models

iii. Platform for the customers to use the deployed ML models via API.

b) Local sub-system (Configuration Software)

The local sub-system is software installed and configured on the machine available on the data provider side. It serves the most crucial aspect, i.e. data acquisition and pre-processing. It ensures that data is clean, consistent, and accurate, moreover, it should not contain any individual's personal information to preserve one's privacy. In this way, this software will automate the data cleaning, and will upload the data on the website. The local sub-system and the website will interact with each other in a pipeline-like manner, providing a continuous flow of medical data from data providers to the data consumers. This reduces the efforts for searching medical data. The system serves four main stakeholders: 1) Hospitals: Hospitals will register on the website as data providers, after which they will be committed to provide the requested medical data generated at the hospital. 2) Data scientists: The Data scientists will register as data consumers. They can request desired medical data as per their problem statement. They will get access to plenty of medical data through the website to build various machine learning models. They can deploy these models using the API provided on the website. 3) The sponsors: These are the people who will register on the platform to fund various initiatives and bring up different problem statements.

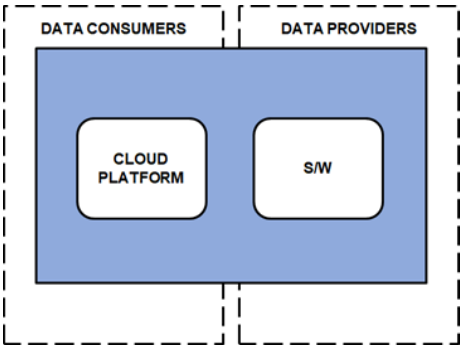

Figure 1: System overview of Self-sustainable Cloud Platform

They may hire data scientists to work on a wide range of medical problems, especially during a pandemic.

4) The customers: They can use the models developed & deployed by the data scientists via API on the pre-paid basis.

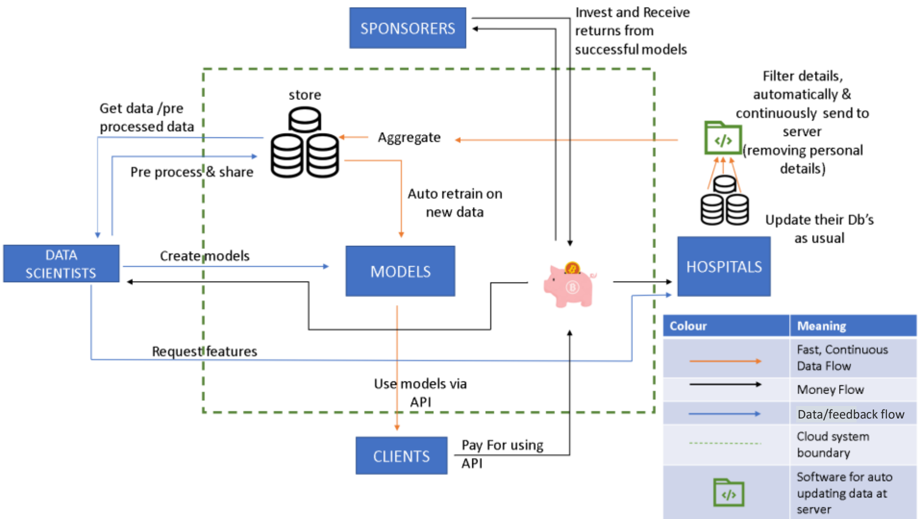

Figure 2: Working of Self-sustainable Cloud Platform

As illustrated in Figure 2, the proposed system serves different stakeholders in many ways and expedites the development of the AI solutions. Data scientists can freely use this medical data available on website to train their models. However, only a small fraction of data will be visible to them. They can build various machine learning models on the website and train them using this medical data. Occasionally, data scientists realize that certain new features may be helpful to enhance the model performance. They can communicate these requirements to the data providers by posting their requests for one or more features. The data requests can be served by data providers such as hospitals or medical institutes. Consequently, data scientists will be free to choose any medical problem statement for their research without worrying about data availability. Data Scientists can also collaborate and help medical communities more efficiently. In this way, the system bridges the communication gap between the data providers & data consumers. Since medical data will be continuously provided with regular updation, the data scientist can use it to plot and create real-time dashboards. Data scientists can also deploy their models free of cost on the website to showcase their work publicly. However, to use the deployed model, the customers will have to pay; part of those funds will return to data providers in exchange for medical data and also to the data scientist who created the model. Sponsors can identify a specific medical problem and hire skilled data scientists to collaboratively work on it while encouraging data providers to give related data.

5 Advantages of the proposed system

i. Single Platform - from raw data to insights: Data Scientist can focus on getting better models instead of struggling to acquire the correct data. ii. Better Collaboration: The collaboration of data providers with data scientists will increase, and data requirements can be easily conveyed. iii. Returning credits to data providers: Data providers will get the true value for their data; they will be paid as per its utility. They will get money for providing good quality and accurate medical data. In this way, they can be given the proper credits that they deserve for providing medical data. They can also be rewarded based on how well they serve a request. This can further be examined using the feedback given by the data consumers based on crucial aspects such as the quality & quantity of the data. The proposed system can also contribute well in times of epidemic or pandemic, discussed in section 7.

6 Proposed Implementation plan and challenges

This section elaborates on the tools and technologies required for implementing the proposed system and the challenges associated with it. The Configuration Software can track changes in a local database with tools like Liquibase, Db deploy, Ruckus, dbsource or database SQL triggers that can notify about a new record or a new attributes added. After the changes are committed to the local database, they are captured and sent to the de-identification module which will remove any personally identifiable information in the data. This process, as discussed in the literature [36,37,38,39], can be automated. After de-identification process, the data is periodically sent to the cloud; this can be implemented using any networking software. The cloud can store this data using Amazon S3 (a scalable storage infrastructure). The environment for model building, training, and deployment can be provided using Amazon Sagemaker. The API services can be implemented using Sagemaker Neo. All the monetary transactions over the cloud platform can be carried out using cryptocurrencies such as Bitcoin, which Amazon AWS supports. Hence, the premise of technical feasibility is clearly defined with the tools and services mentioned above, which guarantees its implementation. Even after a successful implementation, the proposed system can face several challenges. A few of them are listed below based on their severity along with their possible solutions: 1) preserving patient’s privacy (High severity): As stated by Jose et al. [40], proper security measures should be followed while handling healthcare data electronically.

The software installed and configured on the data provider side would automatically filter any sensitive information before uploading data to servers, preventing back-tracing a patient using his information. 2) Self-sustainability(Low severity): The system will only need hosting fees, part of which can be collected from clients that use the API to access the deployed AI models. 3) Dataset Exploitation by data scientists (Medium severity): Returning the credits to the data providers is one of the primary aims of the system. If all the medical data is shown to the data scientists, then they might download all the data and train their models on their local machines and then sell these models independently without giving any credit to the data providers, thereby bypassing our system altogether. To avoid this, we propose a restriction wherein the data scientist can only see some part of the entire dataset. However, the model can be trained on the entire data and be deployed and used with API calls only via the proposed cloud platform. Models can also be downloaded if the data scientist pays some fee upfront. In this way, the data scientist will be obliged to give due credits to the data providers. 4) Ensuring Quality Data: (High severity) As studied by Adir Even et al. [41], the Utility of a particular data can help us better assess the quality of data. If a certain data is of better quality, then it would be used by many data scientists. Since the quality data trains a quality AI model, thereby attracting many clients for it. Hence, the quality of the data will be good if it has trained a model which is used by a more significant number of clients; thus, the usability of the data can be assessed, and the incentives to data providers can be decided.

7 Role of the proposed system during pandemic/epidemic

Occurrence of a disease or virus that affects many people simultaneously in a particular locality or community is called epidemic. It becomes a pandemic when it crosses international boundaries. If an epidemic is detected at an early stage, measures can be taken, and a pandemic can be avoided. Hospitals keep a record of symptoms and diseases for all their patients. The aggregate number of symptoms is usually constant or shows seasonal variability. However, a sudden growth in one or more symptoms over a region may indicate an outbreak. Thus, if data from all such hospitals belonging to different geographical locations is collected periodically and analyzed, an epidemic can be detected earlier. This is possible with our proposed self-sustainable cloud platform, which also allows advanced models to be trained on this collected data. According to WHO, there are six stages in a pandemic[42]. The proposed cloud system can be used in each phase of the pandemic. During the initial phases (1 to 3), the probability of a pandemic is low, the virus circulating between animals' groups only. Animal to human infection might be observed, causing small clusters of diseases in humans. An abnormal rise in fever (or other symptoms) in certain regions can be easily spotted with our system in place for the data is continuously collected from the hospitals and analyzed. In the 4th phase, human to human transmission is observed, and the virus can sustain community-level outbreaks. Our system helps in accurately mapping and monitoring the spread of the virus. Infected regions can be isolated (with systems described by Poh-Chin Lai[43] relying on real-time data). In the 5th phase, the virus causes community-level outbreaks in at least two countries. The pandemic is confirmed in the 6th stage. In both of these phases, advanced AI models for the diagnosis and prognosis of the disease could be built and deployed directly through our system. Data scientists can collaborate and create better solutions while working closely with the hospitals or medical research institutes. Neumann et al.[44] reviewed the H1N1 emergence in 2009 and its potential to spread again if not properly controlled. Our system can also assist in the drug trial process. The results from vaccine trials could be shared and analyzed. Sets of patients on whom the vaccine may be effective can be identified. Vaccine distribution can also be mapped. Similarly, the seasonal return of the virus and some mutations causing changes in symptoms or mortality rate can be mapped effortlessly with the collected data. Additionally, our system will also provide the necessary real-time data for all other AI applications mentioned by Raju Vaishya et al.[45]

8 Conclusion and Future Research Direction

The availability of plenty of quality data is the primary need to develop AI solutions. This paper discusses a self-sustaining cloud platform, which expedites the process of data acquisition, collaboration and development of AI-based solutions using accurate medical data, while simultaneously fulfilling the needs of all of its stakeholders. It facilitates better communication between the data providers and data consumers, ensuring a continuous flow of quality data. The proposed self-sustainable cloud platform is extremely helpful while facing global pandemics, making smart cities responsive and resilient to such sudden outbreaks. By seeking government support, additional incentives can be given to hospitals to promote medical data sharing. There is a massive scope of integrating this cloud platform with IoT to build a advanced data-sharing systems. From collecting data using various equipment to transmitting them to the cloud, everything can be automated. A similar system can be used for finance or business data to build good prediction models and gauge the impact of the pandemic on the business and economy at large.

References

[1] R. Shah, A. Chirc, "IOT and AI in healthcare: A systematic literature review", *Issues in Information Systems* ,**volume 19** (Issue 3) (2018) https://iacis.org/iis/2018/3_iis_2018_33-41.pdf

[2] S. G. Mishra, A. K. Takke, S. T. Auti, S. V. Suryavanshi, M. J. Oza, "Role of artificial intelligence in health care", *Biochemistry* ,**volume 11** October 24. https://www.tsijournals.com/articles/role-of-artificial-intelligence-in-health-care.pdf

[3] S. Latif et al., "Leveraging Data Science to Combat COVID-19: A Comprehensive Review," *in IEEE Transactions on Artificial Intelligence*, **volume 1**, no. 1, pp. 85-103, Aug. 2020, doi: 10.1109/TAI.2020.3020521. doi:10.13140/RG.2.2.12685.28644/4.

[4] S. Das, S. Biswas, A. Paul, A. Dey, "Ai doctor: An intelligent approach for medical diagnosis", *Industry Interactive Innovations in Science, Engineering and Technology*, **volume 11** 173-183doi:https://doi.org/10.1007/978-981-10-3953-9 17.

[5] N. Houssami, G. Kirkpatrick-Jones, N. Noguchi, C. I. Lee, "Artificial intelligence (ai) for the early detection of breast cancer: a scoping review to assess ai's potential in breast screening practice", *Expert Review of Medical Devices*, **volume 16** (5) (2019) 351–362. doi:10.1080/17434440.2019.1610387.

[6] P. Savadjiev, C. J., D. A., et al., "Demystification of ai-driven medical image interpretation: past, present and future", *European radiology,* **volume 29** (2019) 1616-1624.https://doi.org/10.1007/s00330-018-5674-x

[7] D. A. Bluemke, "Radiology in 2018: Are you working with ai or being replaced by ai?", *Radiology* ,**volume 287** (2) (2018) 365–366.

[8] A. Grzybowski, P. Brona, G. Lim, "Artificial intelligence for diabetic retinopathy screening: a review", *Eye* **volume 34** (2020) 451-460.https://doi.org/10.1038/s41433-019-0566-0

[9] E. Smalley, "Ai-powered drug discovery captures pharma interest:, *Nat Biotechnol* ,**volume 35** (2017) 604–605. doi:https://doi.org/10.1038/nbt0717-604.

[10] M. Benhenda, "ChemGan challenge for drug discovery: can ai reproduce natural chemical diversity?" *arXiv e-prints* 2017 https://arxiv.org/abs/1708.08227

[11] N. Fleming, "How artificial intelligence is changing drug discovery", *NatureMay;* **volume 557** (7707) (2018) S55–S57. doi:10.1038/d41586-018-05267-x.

[12] R. Sandhu, H. K.Gill, S. K. Sood, "Smart monitoring and controlling of pandemic influenza a (h1n1) using social network analysis and cloud computing", *Journal of Computational Science,* **volume 12** (2016) 11–22. doi:10.1016/j.jocs.2015.11.001.

[13] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, et al., "Can ai help in screening viral and covid-19 pneumonia?", *IEEE ACCESS,* **volume 19**. doi:10.1109/ACCESS.2020.

[14] "How this Canadian start-up spotted coronavirus before everyone else knew about it", Accessed-July 4, 2020. https://www.cnbc.com/2020/03/03/bluedot-used-artificial-intelligence-to-predict-coronavirus-spread. html

[15] A. Chapman, E. Simperl, L. Koesten, G. Konstantinidis, L.-D. Iba´n˜ez, E. Kacprzak, P. Groth,"Dataset search: a survey", *The VLDB Journal* , **volume 29**, pages251–272 (2020).doi:10.1007/s00778-019-00564-x.

[16] Y. Wang, L. Kung, T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations", *Technological Forecasting and Social Change,* **volume 126** , pp. 3-13,(2018) doi:10.1016/j.techfore.2015.12.019.

[17] C. A. Alexander1, L. Wang, "Big data and data-driven healthcare systems", *Journal of Business and Management Sciences,* **volume 6** (3) (2018) 104–111. doi:10.12691/jbms-6-3-7.

[18] E. Niiler, An AI epidemiologist sent the first warnings of Wuhan virus,2020 https://www.wired.com/story/ai-epidemiologist-wuhan-public-healthwarnings

[19] J. M. Plasek, C. Tang, Y. Zhu, Y. Huang, D. W. Bates, "Following data as it crosses borders during the covid-19 pandemic", *Journal of the American Medical Informatics Association,* **volume 27** ,pp. 1139–1141. (2020) doi:10.1093/jamia/ocaa063.

[20] E. Dong, H. Du, L. Gardner, "An interactive web-based dashboard to track covid-19 in real time", *The Lancet. Infectious diseases,* **volume 20**.,5 ,pp. 533-534 doi:10.1016/S1473-3099(20)30120-1.

[21] B. Xu, B. Gutierrez, S. Mekaru, K. Sewalk, L. Goodwin, A. Loskill, E. L. Cohn, Y. Hswen, S. C. Hill, M. M. Cobo, A. E. Zarebski, S. Li, C.-H. Wu, E. Hulland, J. D. Morgan, L. Wang, K. O'Brien, S. V. Scarpino, J. S. Brownstein, O. G. Pybus, D. M. Pigott, M. U. G. Kraemer, "Epidemiological data from the covid-19 outbreak real-time case information", *Scientific Data* , **volume 7** (106). doi:10.1038/s41597-020-0448-0.

[22] K. JH., Google cloud launch covid-19 data sharing platform. https://healthitanalytics.com/news/hca-google-cloud-launch-covid-19-data-sharing-platform

[23] P. Kumar, "Literature based study on cloud computing for health and sustainability in view of covid19", *International Journal of Innovative Technology and Research*, **volume 0**, oai:ojs.ijitr.com:article/2566 https://core.ac.uk/download/pdf/327105038.pdf

[24] L. Griebel, H.-U. Prokosch, F. Ko¨pcke, D. Toddenroth, J. Christoph, I. Leb, I. Engel, M. Sedlmayr, "A scoping review of cloud computing in healthcare", *BMC Medical Informatics and Decision Making,* **volume 15** (17). doi:10.1186/s12911-015-0145-7.

[25] H. S. Gardiyawasam, P. Vladimir, P. A. Oleshchuk, "preserving mechanisms for enforcing security and privacy requirements in e-health solutions", *International Journal of Information Management,* **volume 36** (6 part B) (2016) 1161–1173. doi:10.1016/j.ijinfomgt.2016.07.006.

[26] T.-W. Kuo, A.-C. Lee, "Assessing measurement noise effect in run-to-run process control: Extends ewma controller by kalman filter", *International Journal of Automation and Smart Technology,* **volume 1** .(2011) doi:10.5875/ausmt.v1i1.71.

[27] P. L. Bannerman, "Cloud Computing Adoption Risks: State of Play", *Asia Pacific Software Engineering Conference,* 2010. doi:10.1.1.900.5869.

[28] A. Rosenthal, P. M. Maya, H. L. Jean, S. D. Koester, P. Reynolds, "Cloud computing: A new business paradigm for biomedical information sharing", *Journal of Biomedical Informatics,* **volume 43** (2) (2010) 342–353. doi:10.1016/j.jbi.2009.08.014.

[29] "Ascending cloud: The adoption of cloud computing in five industries Executive summary, A report from The Economist Intel- ligence Unit", Published on Feb 26 2016, Accessed on July 4, 2020.https://www.scribd.com/document/497726528

[30] O. Alia, A. Shrestha, J. Soar, S. F. Wamba, "Cloud computing-enabled

healthcare opportunities, issues, and applications: A systematic review", *International Journal of Information Management*, **volume 43** (2018) 146–158. doi:10.1016/j.ijinfomgt.2018.07.009.

[31] C. Oberdan, F. Rolim, C. Luiz Koch, B. Westphall, A. Jorge Werner, G. Fracalossi, S. Salvador, "A. cloud computing solution for patient's data collection in health care institutions", *Second International Conference on eHealth, Telemedicine, and Social Medicine*, *St. Maarten 2010*, pp. 95–99. (2010) doi:10.1109/eTELEMED.2010.19.

[32] V. Venkatesh, A. Rai, T. A. Sykes, R. Aljafari, "Combating infant mortality in rural india: evidence from a field study of ehealth kiosk implementations", *MISQ,* **volume 40**, 2,pp. 353-380,(2016) doi:10.2307/30036540.

[33] G. S. V. Rao, K. Sundararaman, J. Parthasarathi, "Dhatri - a pervasive cloud initiative for primary healthcare services", *14th International Conference on Intelligence in Next Generation Networks*,pp. 1-6 (2010) 10.1109/ICIN.2010.5640918 .

[34] I. Adamu, K. I. Musa, S. A. wahab, S. Umar, "Cloud computing as a service era: Origins, current and future trends", *IOSR Journal of Computer Engineering (IOSR-JCE),* **volume 21** (3),pp. 34–39, (2019) doi:10.9790/0661-2103063439.

[35] P. Mork, K. P. Smith, C. Wolf, K. B. Samuel, K. Sarver, I. Vayndiner, B. T. Blaustein, "Facilitating discovery on the private web using dataset digests", *International Journal of Metadata, Semantics and Ontologie* pp. 451-455, (2018) doi:10.1504/IJMSO.2010.034042.

[36] Y. C. ZengjianLiu, X. W. BuzhouTang, Q. Chen, H. Li, Q. D. Jingfeng Wang, S. Zhu, "Automatic de-identification of electronic medical records using token-level and character-level conditional random fields", *Journal of Biomedical Informatics,* **volume 58**, S47–S52, (2015) doi:10.1016/j.jbi.2015.06.009.

[37] Z. Jiang, C. Zhao, Y. BinHe, J. Jiang, "De-identification of medical records using conditional random fields and long short-term memory networks", *Journal of Biomedical Informatics,* **volume 75**, S43–S53, (2017). doi:10.1016/j.jbi.2017.10.003.

[38] B. He, Y. Guan, J. Cheng, K. Cen, W. Hua, "Crfs based de-identification of medical records", *Journal of Biomedical Informatics,* **volume 58**, S39–S46, (2015). doi:10.1016/j.jbi.2015.08.012.

[39] A. Dehghan, A. Kovacevi, G. Karystianis, J. A. Keane, G. Nenadica, "Combining knowledge- and data-driven methods for de-identification of clinical narratives", *Journal of Biomedical Informatics*, **volume 58** , S53–S59, (2015). doi:10.1016/j.jbi.2015.06.029.

[40] J. Luis, Fernа´ndez-Alema´n, I. C. Sen˜or, P. A´ ngel Oliver Lozoya, A. Toval, "Security and privacy in electronic health records: A systematic literature review", *Journal of Biomedical Informatics,* **volume 46** (3), pp. 541–562, (2013). doi:10.1016/j.jbi.2012.12.003.

[41] A. Even, G. Shankaranarayanan, "Utility-driven assessment of data quality", *ACM SIGMIS Database:the DATABASE for Advances in Information Systems*, **volume 38,**2 pp. 75-93(2007). doi:10.1145/1240616.1240623.

[42] NIH2, Accessed on July 30 2020.https://www.ncbi.nlm.nih.gov/books/NBK143061/table/ch4.t1/?report=objectonly

[43] P.-C. Lai, C. B. Chow, H. T. Wong, K. H. Kwong, Y. W. Kwan, S. H. Liu, W. K. Tong, W. K. Cheung, W. L. Wong, "An early warning system for detecting h1n1 disease outbreak – a spatio-temporal approach", *International Journal of Geographical Information Science,* **volume 29** (7) pp. 1251–1268, (2015). doi:10.1080/13658816.2015.1030671.

[44] G. Neumann, T. Noda, Y. Kawaoka, "Emergence and pandemic potential of swine-origin h1n1 influenza virus", *Nature,* **volume 459**, pp. 931–939 (2009) doi:10.1038/nature08157.

[45] R. Vaishya, M. Javaid, I. H. Khan, A. Haleem, "Artificial intelligence (ai) applications for covid-19 pandemic", *Diabetes and Metabolic Syndrome: Clinical Research and Reviews,* **volume 14** (4), pp. 337–339, (2020). doi:10.1016/j.dsx.2020.04.012.