1 Conventional AI/ML for COVID-19

Covid-19 cloud data collection and using it to track the course of the disease and its associated developments using artificial intelligence and other analytical tools have proven to be successful in various regions. The data associated with Covid-19 is,

- Basic Reproduction number i.e. R0 of the coronavirus.
- Daily tests conducted.
- Positivity rate in various age, gender, and ethnic groups (epidemiological profile).
- The number of active cases in a region.
- The mortality rate due to COVID-19 in various age, gender, and ethnic groups (epidemiological profile).
- Unconfirmed death rate (suspected Covid-19 mortalities).
- Rate of recoveries.
- The number of vaccines in development and the results of various trial phases.
- Data of patients complaining of post-infection morbidities.
- Genetic profile of various viral strains.

This data is collected through mass testing, recording the data of patients coming into the clinical settings presenting the associated signs and symptoms, and the use of screening tests at various checkpoints.

Once the data has been collected, it can be processed through multiple analytical tools, which use artificial intelligence and machine learning to make the most out of this data for disease tracking, surveillance, and proper management of the limited resources to counter this pandemic. Deep learning of the data and applying artificial intelligence may help the concerned authorities in various policy-making decisions keeping in view the variables.

This data processing has applications in real-time scenarios, which include,

- Infection identification by recording the reporting of the concerned signs and symptoms.
- Tracking diagnosis by accurate detection rate in various testing procedures.
- Contact tracing of the new positive individuals for suspected infections in their contacts.
- Prediction of the number of cases by applying different scenarios.
- Predicting the mortality rate.
- Tracking the progress of various vaccines in development by checking their success in trial subjects and measuring the antibody count in subjects.
- Containing the spread of the disease by imposing the lockdown measures in the areas and regions most affected.

Covid-19 has proven to be the test case for using these advanced analytical tools in checking its spread. This will help in containing the future epidemics and pandemics, which are very likely to occur due to hazardous animal farming practices and the emergence of new strains of highly contagious viruses. Flow chart of the CAIML approach and its applications is represented in Figure 1.

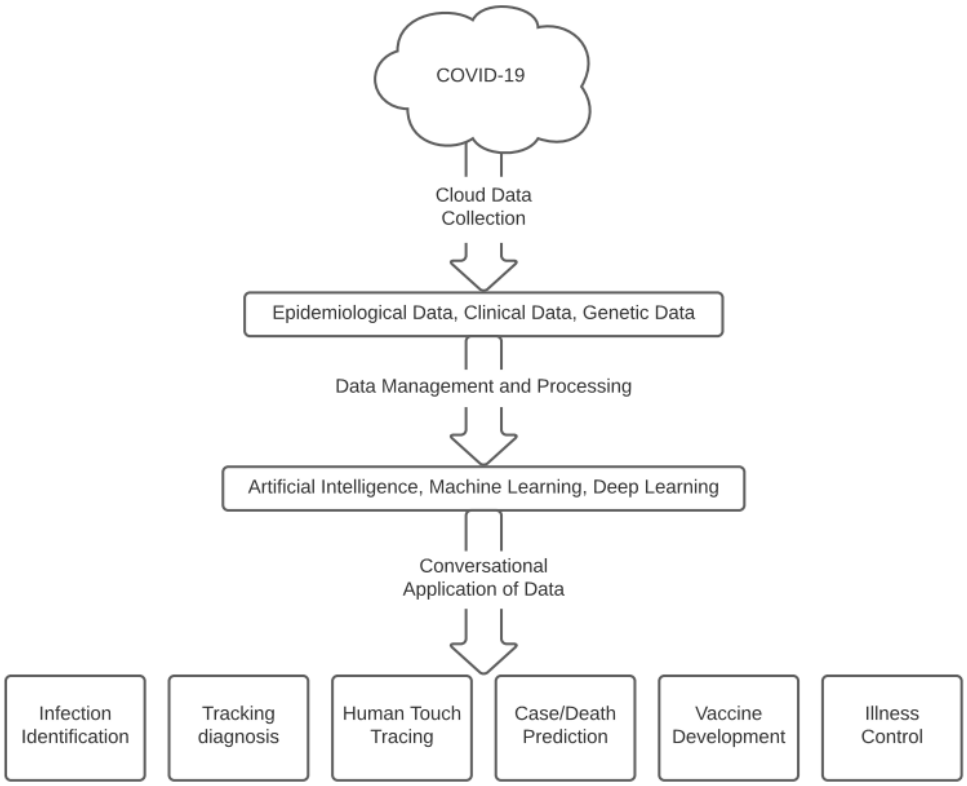

Figure 1. CAIML Covid-19 Applications

1.1 Taxonomy of COVID-19 Open Source Dataset

COVID-19 data set consists of multiple subsets of different kinds of data which are as follows

Medical Imaging

- CT Scan
- MRI Images

A comprehensive amount of data is obtained through the diagnostic imaging tool with scans of suspected patients which can then be used to differentiate and magnify the specific COVID-19 diagnostic points.

Textual

- Clinical Records.
- Patient history.
- Research articles.
- Patient interactions on social media posts.

This enormous data set can be used to analyze the varying trends of disease course and finding the rare signs and symptoms and post-infection morbidities.

Speech

- Cough
- Breathing

This can be recorded in audio data-set and vast pool of this audio data set may be helpful in recognizing specific patterns easily identified through artificial intelligence technologies.

This data set has multiple uses and real-time application in the case of pandemic as mass testing makes it difficult for individual attention on each case the artificial intelligence can be used to:

- Make diagnosis through image learning in case of radiological diagnostic testing
- Proper segmentation of images in various stages of diseases
- Natural language processing on scholarly articles to streamline the new advancements in diagnosing and treatment capabilities.
- Sentimental analysis on social media posts to see varying trends in mental health issues and checking the following of SOPs in place.
- Audio analysis of breathing and cough to make it easy for diagnosing where advanced diagnostic procedures are scarce.
- Devising the non-pharmaceutical interventions to limit the use of already dire resources and counter the lags in supply chain.

The tools which can be used for these purposes include artificial intelligence in the form of machine leaning, statistical analysis of the data and its representation in comprehensive graphical forms. Big data sources for development of these applications should be open-source and services like GitHub can be used to develop softwares and making sure their availability to responsible parties for countering pandemic.

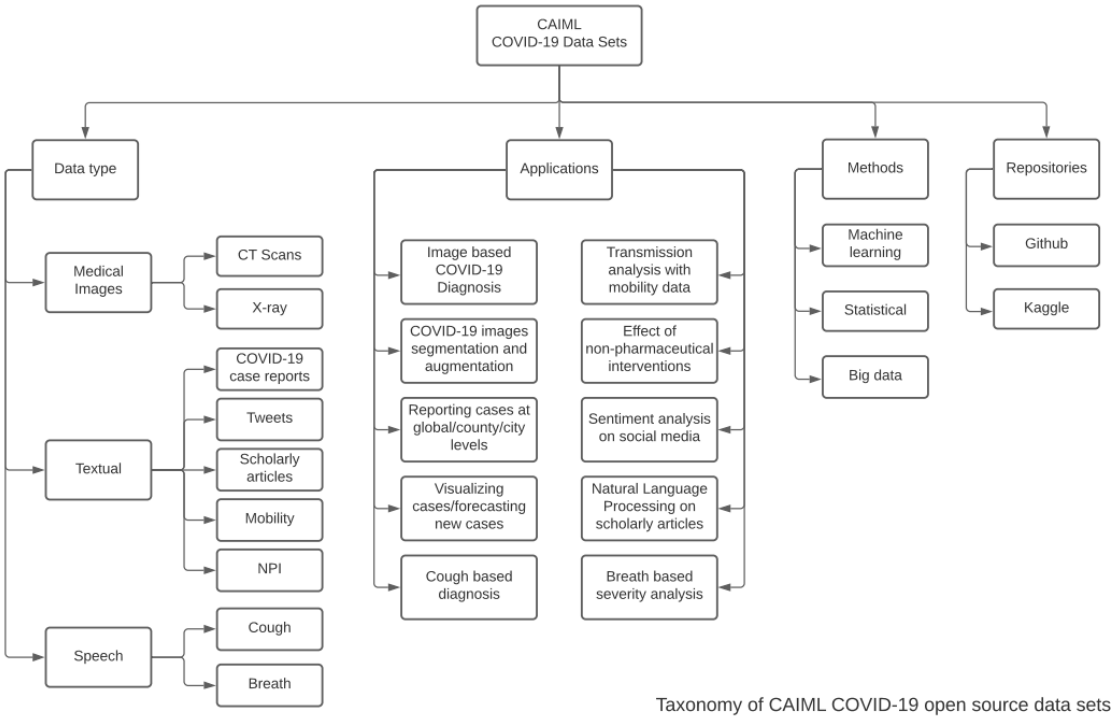

Figure 2. Taxonomy of CAIML COVID-19 open source dataset

Based on the flow chart of the CAIML approach, its applications and taxonomy, the procedures to use CAIML for COVID-19 purposes has been discussed in detail below.

1.2 Disease Detection

Covid-19 is a pandemic that was identified in Wuhan, China in late 2019. 17 years ago, similar coronavirus was detected in the form of SARS that was thought of being caused by unknown contagious agent. In lieu of these pandemic occurring after every 10 to 20 years, advancement in the health care system made a new paradigm of using Artificial Intelligence technique (AI) by implementing several Machine Learning tools (ML) for detecting, analyzing and decision making process. AI tools are used not only to detect the Covid breakouts but also to predict their spread across the world. However, AI–driven technology needs a sufficient amount of training data with a set of annotated data in ML to train the classifiers using supervised learning method (SLM).

1.2.1 Using Active Learning

For detecting COVID-19, Conventional AI/ ML needs to deploy "active learning based cross population training/ testing models" employing multi-model data [1].

AL tool is a self-learning mechanism also known as Incremental Learning. IL assists the learning modal to adapt the upcoming data without being fail to recall the previous limited knowledge. Meanwhile passing through the learning phase, the variations in data with the passage of time can be evaluated using Anomaly Detection (AD) methods. Figure 2 shows the process of AL mechanism below.

In AL-driven tool, AD helps in detecting rare events that convey the suspicious observations that are significantly differing from the majority of data or normal data for the specific event.

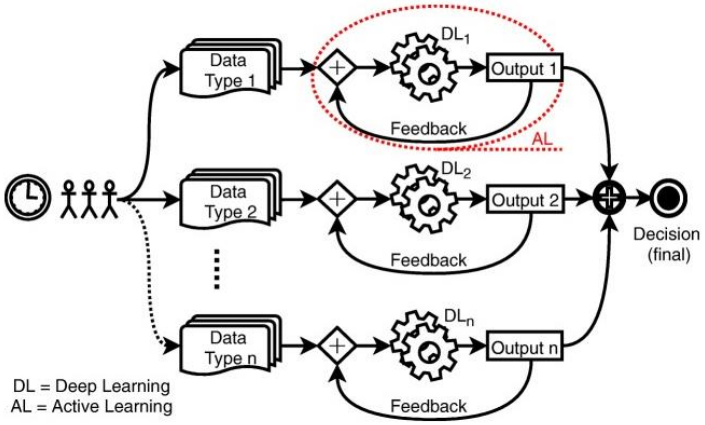

Figure 3. Active Learning modal scheme [1]

On collecting the data with various attributes including gender, age, location, patient ID, this raw and unstructured form is then passed through pre-processing step. Then features are extracted for converting them into the probabilistic value. These features are then supplied to ML algorithms. Figure 2 below represents the flow of all the process starting from data collection till ML classification for detection of COVI19.

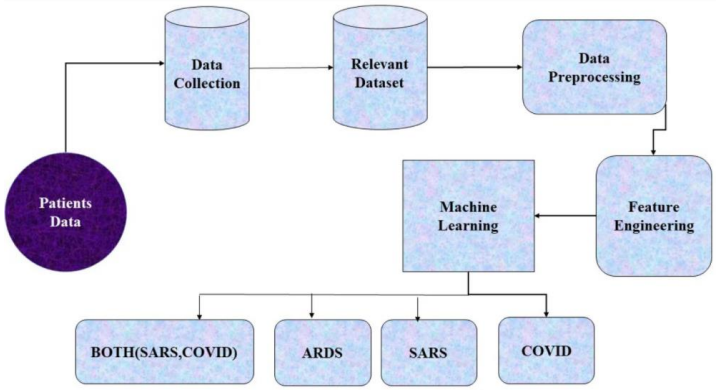

Figure 4. Flow of various steps involved in detecting Disease[2]

1.2.2 Using Logistic Regression

This algorithm calculates the numerical variables on the basis of their association with label where, various features are selected and give in as input. The algorithm then figures out the class probability. These probabilities are calculated using the following Eq (1):

$$P(y = k|x) = \frac{exp^{\varphi^{T}\theta_k}}{1 + \sum_{k=1}^{3} exp^{\varphi^{T}\theta_k}} \forall\, k = 1,2,3$$

$$P(y = 0|x) = \frac{exp^{\varphi^{T}\theta_k}}{1 + \sum_{k=1}^{3} exp^{\varphi^{T}\theta_k}} \forall\, k = 1,2,3$$

1.2.3 Using Support Vector Machine (SVM)

It is a supervised ML technique that classifies the text into various useful categories. It collects 'n' no. of features with the labels. Here data points of the training set are taken as; $(y_k, x_k)_1^n$ where n is the considered number of features [2]. These are then taken in the form of table where they are proceeded to be given as input to the SVM for classifier using Eq. 2.

$$y(x) = sign\left[\sum_{k=1}^{n} \alpha_k y_k \varphi(x, x_k) + b\right]$$

$$\varphi(x, x_k) = \begin{cases} x_k^t x = Linear\ SVM \\ (x_k^t x + 1)d: PolynomialSVM\ in\ degreed \\ \exp\left(-\|x - x_k\|_2^2 / \sigma^2\right): RBFSVM \end{cases}$$

Where αk = + real value

B= real constant

Then the hyperplane is made to perform the classification being categorized into three classes including (COVID-19, SARS, ARDS).

1.2.4 Using Decision Tree

It is an alternative approach for classification, partitioning the text input into regions. It then classifies each region independently. This algorithm divides the space recursively as regards to the available inputs which are then classified at bottom of tree. The leaf node categorizes the text into classes. For creating a decision tree, a function known as "splitting criterion" is required [3]. It defines the method to maximize the performance by splitting the data. "Information gain ratio" is used that is taken to be equal to information gain to intrinsic info and is represented in Eq. 3 below.

$$IGP(EX, \alpha) = {}^{IG}/_{IV}$$

Where IG= information gain, IV = intrinsic information.

1.2.5 Using Random Forest

Random Forest is collaborative ML algorithm relying on data collected through Decision Trees which are trained on "mutually independent subsets of the original data" with the purpose of gaining classifier with "lower variance and bias" [4]. These trained independent data-points of Decision Trees are gained from original dataset by:

- sampling with instance replacement
- picking up the random subset of features

Random Forest is considered to be the class of "probability scoring classifiers" (PSC) where a probability score is allocated by instance modal to each class. The abstention is executed based on two thresholds;

$\alpha, \beta \in [0,1]$

where '1' indicates the positive class and '0' indicates the negative class.

Hence, instance modal is categorized:

positive if "*score* (1) > α and *score* (1) > *score* (0)", and

negative if "*score* (0) >β and *score* (0) > *score* (1)", other than that, modal is not selected.

In regards to these models, the performance is assessed on the basis two class instances only.

1.3 Disease Diagnose

Application of ML algorithm for diagnosing the disease automatically has become popular in medical field. Deep learning (AI research Field) is enabled to gain the desired results irrespective of the need of feature extraction. It has been working successfully for diagnosing the problems including arrhythmia diagnose, skin cancer, breast cancer and pneumonia detection from chest X-ray images. COVID-19 pandemic's sudden upsurge has demanded the need for proficiency in this field with the development in the automated diagnose systems based on AI/ ML algorithms.

1.3.1 Using Conventional Neural Network (CNN)

Deep Neural network is the increase in size of network with number of layers therefore named after convolution. CNN technique has convolution layer to extract the important features from input while reducing the structure size for computational performance, and connecting with the layer known as neural network. CNN modal is built by joining each layer. Thereafter, its core factors are tuned to fulfill the specific task including item recognition and classification [5].

Considering the example for X-ray image data-points, number of filters are increased consisting of convolutional layers and pooling layers. These layers are taken as CNN layers with several sizes and filter values. Convolution operation is defined as [6]:

$$(X * K)(i.j) = \sum_{m}\sum_{n} k(m,n)X(i-m, j-n)$$

1.4 Epidemic Forecasting

AI algorithm based methodologies are used in order to predict the real time confirm cases of COVID-19 to overcome restrictions in the application of epidemiological approach, and aiding in public health and policy making decisions.

1.4.1 Modified Auto-encoder for Modeling Time Series

MAE is used to predict the number of growing new confirmed cases of COVID-19 [7]. In classical auto encoder, no. of nodes in layers decreases gradually from input to the latent layer. While in MAE the layers are considered as 8, 32, 4 and 1 for input node layer, first, second and output layer respectively. Here, 8 days' segment of time series as sample data and 128 as training samples are viewed. On the first day of segment a random element Z is selected and choose 7 successive days as the other days to form a segment of time series. Consider I be index of segment j_i, be column index of matrix Z. The i^{th} time series is represented by $Z_{ji}, Z_{ji+1}, Z_{ji+2}, Z_{ji+3} \ldots\ldots Z_{ji+7}$, where data is normalized to $x_{j_{i+k}} = \frac{Z_{ji+k}}{S}$, k= 1,2,…7 taking $S = \frac{1}{8}\sum_{k=0}^{7} Z_{ji+k}$.

From this, by estimating the segments and weights in MAE, the final forecasted number of the collected confirmed cases for each province/city are measured.

1.5 Sustainable Development

Sustainable development is the combination of social, economic, and environmental factors involving political goals. It is considered as the enhancement in the quality of life in a positive way in the lives of future generation. Researchers have used Artificial intelligence based methods and approaches to practically implement the sustainable development in times of COVID-19 to sustain any country's economic growth. These approaches are:

- Group Method of data handling
- Regression analysis

1.5.1 Group Method of Data Handling

GMDH is a type of neural network self-organized method that is used for pattern recognition, complex system optimization and modelling, problem prediction etc. It is also known as the "Polynomial of the Ivakhnenko equation" which can predict the value of (y_i) from the function (f) for every input vector (X). Equation below shows the input and output function of GMDH as:

$$y = f(x_{i1}, x_{i2}, x_{i3}, x_{i4} \ldots\ldots\ldots\ldots x_{im})$$

$$i = 1, 2, 3, \ldots\ldots\ldots, m$$

$$Y = a + \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} C_{ij} x_i x_j + \cdots\ldots\ldots$$

Where Y = output, m = no. of data, x = input data

The bivariate form of this model is calculated as:

$$y = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2$$

1.6 Patient Management

Considering the environment of ICU (Intensive care unit) with a lot of COVID-19 patients, some of them seems to be incubated while others need the noninvasive ventilation. On the other hand, doctors and nurses uses personal protective equipment for their safety. Patient management is possible using Artificial Intelligence (AI) and machine learning (ML) approaches which can be used to predict the patients with high health crisis risks. In addition to that, making software solutions that can automatically determine the crisis risk is also possible for personalized medicine. Moreover, using AI, it will be possible to predict the laboratory tests with high precision and accuracy [8].

In order to manage the diagnoses of COVID-19 disease from other community like pneumonia, AI deep learning is very helpful tool, while ML approach of genetic variants from asymptomatic, mild or severe COVID-19 patients is used to classify the potential patients; performing vulnerability of patients to potential COVID-19 infection that is possible using AI based DL approach [8].

AI based reasoning engines and "medical assisted decision systems (CDSS)" are helpful for managing the COVID-19 patients. The designs need highly accurate and comprehensive data points while collecting and storing the data of COVID-19 patients for which registry system is vital. Since a wide number of countries are focusing on controlling the epidemic by producing and manufacturing vaccines and drugs, therefore, a full devotion has been given to the execution of rapid and accurate registry system in regards to disease. These efforts are only possible by implementing AI based approaches that doubles up the effort and assist in managing the patients more efficiently.

Reference

[1] K. C. Santosh, "AI-Driven Tools for Coronavirus Outbreak: Need of Active Learning and Cross-Population Train/Test Models on Multitudinal/Multimodal Data," *J. Med. Syst.*, vol. 44, no. 5, p. 93, May 2020, doi: 10.1007/s10916-020-01562-1.

[2] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *Int. J. Inf. Technol.*, vol. 12, no. 3, pp. 731–739, Sep. 2020, doi: 10.1007/s41870-020-00495-9.

[3] S. H. Kassani, P. H. Kassasni, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning Based Approach," p. 19.

[4] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, "Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study," *J. Med. Syst.*, vol. 44, no. 8, p. 135, Aug. 2020, doi: 10.1007/s10916-020-01597-4.

[5] M. A. Al-antari, C.-H. Hua, and S. Lee, "Fast Deep Learning Computer-Aided Diagnosis against the Novel COVID-19 pandemic from Digital Chest X-ray Images," In Review, preprint, Jun. 2020. doi: 10.21203/rs.3.rs-36353/v1.

[6] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, Jun. 2020, doi: 10.1016/j.compbiomed.2020.103792.

[7] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, "Artificial Intelligence Forecasting of Covid-19 in China," p. 20.

[8] S. Rahmatizadeh, S. Valizadeh-Haghi, and A. Dabbagh, “The role of Artificial Intelligence in Management of Critical COVID-19 patients,” *J. Cell. Mol. Anesth.*, vol. 5, no. 1, Apr. 2020, doi: 10.22037/jcma.v5i1.29752.