International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211
Volume 11, Issue 3, March-2023, Impact Factor: 7.429, Available online at: www.ijaresm.com

# Covid-19 Future Forecasting Using Supervised Machine Learning Models

**Dr. S. Tamil Selvan[1], S.Kiruthikha[2]**

[1]Assistant Professor of Computer Science and Engineering Department , Erode Sengunthar Engineering College
[2]Computer Science and Engineering Department, Erode Sengunthar Engineering College.

----------**********----------

## ABSTRACT

The spread of COVID-19 in the whole world has put the humanity at risk. The resources of some of the largest economies are stressed out due to the large infectivity and transmissibility of this disease. The capability of ML models to forecast the number of upcoming patients affected by COVID-19 which is presently considered as a potential threat to mankind. In particular, four standard forecasting models, least absolute shrinkage and selection operator (LASSO) Support vector Machine (SVM) have been used in this study to forecast the threatening factors of COVID-19.Three types of predictions are madeby each of the models, such as the number of newly infected cases, the number of deaths, and the number of recoveries But in the cannot predict the accurate result for the patients.

To overcome the issue, Proposed method using the long short-term Integrated Average (LSTIA) predict the number of COVID-19 cases in next 30 days ahead and effect of preventive measures like social isolation and lockdown on the spread of COVID- 19.

## INTRODUCTION

### OVERVIEW OF COVID-19

COVID-19, the pandemic that is spreading worldwide, has revealed the vulnerability of human society to severe infectious diseases and the difficulty of solving this problem in a globally interconnected complex system. COVID-19 affected more than 100 countries in a span of weeks.As a result, the whole human race should not only collaborate to overcome the epidemic but also reasonably arrange to return to work and production according to the actual situation of each region and carry out geographical risk assessment.Many attempts have been conducted to find a suitable and fast way to detect infected patients in an early stage. After making chest CT scans of 21 patients infected with COVID19 in China, Guan et al found that CT scan analysis included bilateral pulmonary parenchymal ground-glass and consolidative pulmonary opacities, sometimes with a rounded morphology and a peripheral lung distribution. Consequently, COVID-19 diagnosis can be represented as an image segmentation problem to extract the main features of the disease.The disease caused by the novel coronavirus, or Coronavirus Disease 2019 (COVID-19) is quickly spreading globally. It has infected more than 1,436,000 people in more than 200 countries and territories as of April 9, 2020.

### EXPONENTIAL SMOOTHING

Exponential smoothing is a rule of thumb technique for smoothing time series data using the exponential window function. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Exponential smoothing is often used for analysis of time-series data.

Exponential smoothing is one of many window functions commonly applied to smooth data in signal processing, acting as low-pass filters to remove high-frequency noise. This method is preceded by Poisson's use of recursive exponential window functions in convolutions from the 19th century, as well as Kolmogorov and Zurbenko's use of recursive moving averages from their studies of turbulence. There is no formally correct procedure for choosing {\display style \alpha}\alpha. Sometimes the statistician's judgment is used to choose an appropriate factor. Alternatively, a statistical technique may be used to optimize the value of {\display style \alpha}\alpha

### FUTURE FORECASTING

Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends. A commonplace example might be estimation of some variable of interest at some specified future date. Prediction is a similar, but more general term. Both might refer to formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods. Usage can differ between areas of application: for example, in hydrology the terms "forecast" and "forecasting" are sometimes reserved

for estimates of values at certain specific future times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period. Risk and uncertainty are central to forecasting and prediction; In some cases the data used to predict the variable of interest is itself forecast.

Qualitative forecasting techniques are subjective, based on the opinion and judgment of consumers and experts; they are appropriate when past data are not available. They are usually applied to intermediate- or long-range decisions. Examples of qualitative forecasting methods are informed opinion and judgment, the Delphi method, market research, and simple and weighted N-Period moving averages, simple exponential smoothing, poisson process model based forecasting and multiplicative seasonal indexes. Previous research shows that different methods may lead to different level of forecasting accuracy. For example, GMDH neural network was found to have better forecasting performance than the classical forecasting algorithms such as Single Exponential Smooth, Double Exponential Smooth, and ARIMA and back-propagation neural network.

**SUPERVISED MACHINE LEARNING**

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an The parallel task in human and animal psychology is often referred to as concept learning.

Typically, the input object is transformed into a feature vector, which contains a number of features that are descriptive of the object. The number of features should not be too large, because of the curse of dimensionality; but should contain enough information to accurately predict the output. Complete the design. Run the learning algorithm on the gathered training set. Some supervised learning algorithms require the user to determine certain control parameters. These parameters may be adjusted by optimizing performance on a subset (called a validation set) of the training set, or via cross-validation. Evaluate the accuracy of the learned function. After parameter adjustment and learning, the performance of the resulting function should be measured on a test set that is separate from the project.

## SYSTEM ANALYSIS

**EXISTING SYSTEM:**

COVID 19 is currently considered a potential threat to humanity. In four standard prediction models, such as linear regression (left to right), at least complete summary and select operator, Support Vector Machine (SVM), have been used to predict COVID-19 threatening factors in this study. Predictions are made on each of the models, such as the number of new infections, the number of deaths, and the number of recurrences over the next 10 days. For the effects of the study it demonstrates a promising mechanism for the use of these methods in the current context of COVID 19 infection. Predictions are made on each of the models, such as the number of new infections, the number of deaths, and the number of recurrences over the next 10 days. For the effects of the study it demonstrates a promising mechanism for the use of these methods in the current context of COVID 19 infection.

COVID-19 does not seem to affect children severely; many pediatrics wards have been focused more on the emergency of COVID-19-related issues. For this reason, attention on many other acute and chronic diseases, especially those rarer, may be lacking. This scarcity of interest may cause, particularly in childhood, severe problems, or even death.

**DRAWBACKS OF EXISTING SYSTEM:**

- COVID-19 problem cannot predict the exact result from the patients.
- Difficult to Monitor Performance - It is not easy for managers to monitor their staffs' progress and performance without them being in the same office space.
- This is especially escalated if the job role requires a lot of "background duties" that can't be monitored on a work's system.
- Financial burden on the world, Morbidity and mortality Social and mental distance between people.

**PROPOSED SYSTEM:**

Machine learning methods proved to be effective for prediction due to automatically extracting relevant features from the training samples, feeding the activation from the previous time step as input for the current time step and networks self-connections. According to the results of the model analysis, we believe that the emergency intervention measures adopted in the early stage of the epidemic, such as blocking, restricting the flow of people, and increasing the support, had a crucial restraining effect on the original spread of the epidemic.

It is a very effective prevention and treatment method to continue to increase investment in various medical resources to ensure that suspected patients can be diagnosed and treated in a timely manner. The epidemic trends long short-

term Integrated Average (LSTIA) of were first fitted and analyzed in order to prove the validity of the existing mathematical models.The results were then used to fit and analyze the situation of COVID-19. The prediction results of three different mathematical models are different for different parameters and in different regions. The prediction obtained by the proposed method of various components (number of positive cases recovered number of cases, etc.) will be accurate within a certain range and will be a beneficial tool for administrators and health officials.

**ADVANTAGES OF PROPOSED SYSTEM:**

- Lifestyle modifications
- Health awareness
- Importance of health
- Advantages of real-time and fast, which can predict the incidence trend of infectious diseases as early as possible, and are suitable for data analysis of a large number of people.
- The sensitivity, spatial resolution and accuracy of its prediction result is improved

**FEASIBILITY STUDY**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- BEHAVIORAL FEASIBILITY

**ECONOMICAL FEASIBILITY:**

The economic feasibility study (EFS) should demonstrate the net benefit of the proposed application in light of the benefits and costs to the agency, other state agencies and the general public as a whole. The agency must submit its EFS and request for approval to the Office of Financial Management (OFM) prior to accepting or disbursing electronic funds/benefits. Approval from OFM is required for pilot and permanent applications, and both Internet and retail applications.

**TECHNICAL FEASIBILITY:**

Technical feasibility is the process of proving that the concept is technically possible. The objective of the technical feasibility step is to confirm that the product will perform and to verify that there are no production barriers. Technical Information: The technical feasibility step generates knowledge about the product or process's design, performance, production requirements, and preliminary production costs.

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**BEHAVIORAL FEASIBILITY:**

Behavioral feasibility pertains, in part, to the degree to which public attitudes and beliefs about restoration of wildlife species are shaped by the abilities of local residents and other stakeholders to identify and build on restoration-related opportunities and to overcome or mitigate potential restoration-related problems.

For this reason, the proportion of residents who hold either positive or negative attitudes about restoration should never be used as the sole social input and treated as a surrogate referendum about restoring a species. Rather, the ways in which community members think about possible impacts in the context of their own community, and the ways in which this context affects attitudes, must be understood.

## SOFTWARE DESCRIPTION

**FRONT END: JAVA**

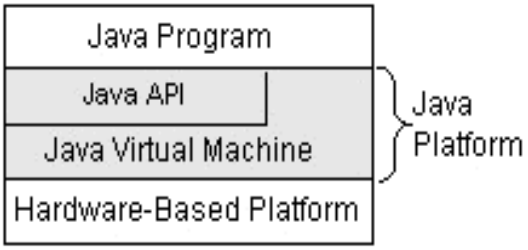
The software requirement specification is created at the end of the analysis task. The function and performance allocated to software as part of system engineering are developed by establishing a complete information report as functional representation, a representation of system behavior, an indication of performance requirements and design constraints, appropriate validation criteria.

**FEATURES OF JAVA**

Java platform has two components:

- The *Java Virtual Machine* (Java VM)
- The *Java Application Programming Interface* (Java API)

The Java API is a large collection of ready-made software components that provide many useful capabilities, such as graphical user interface (GUI) widgets. The Java API is grouped into libraries (*packages*) of related components.
The following figure depicts a Java program, such as an application or applet, that's running on the Java platform. As the figure shows, the Java API and Virtual Machine insulates the Java program from hardware dependencies.




As a platform-independent environment, Java can be a bit slower than native code. However, smart compilers, well-tuned interpreters, and just-in-time byte code compilers can bring Java's performance close to that of native code without threatening portability.

**SOCKET OVERVIEW:**
A network socket is a lot like an electrical socket. Various plugs around the network have a standard way of delivering their payload. Anything that understands the standard protocol can “plug in” to the socket and communicate.

Internet protocol (IP) is a low-level routing protocol that breaks data into small packets and sends them to an address across a network, which does not guarantee to deliver said packets to the destination.

Transmission Control Protocol (TCP) is a higher-level protocol that manages to reliably transmit data. A third protocol, User DatagramProtocol (UDP), sits next to TCP and can be used directly to support fast, connectionless, unreliable transport of packets.

**CLIENT/SERVER:**
A server is anything that has some resource that can be shared. There are compute servers, which provide computing power; print servers, which manage a collection of printers; disk servers, which provide networked disk space; and web servers, which store web pages. A client is simply any other entity that wants to gain access to a particular server.

A server process is said to “listen” to a port until a client connects to it. A server is allowed to accept multiple clients connected to the same port number, although each session is unique. To manage multiple client connections, a server process must be multithreaded or have some other means of multiplexing the simultaneous I/O.

**RESERVED SOCKETS:**
Once connected, a higher-level protocol ensues, which is dependent on which port user are using. TCP/IP reserves the lower, 1,024 ports for specific protocols. Port number 21 is for FTP, 23 is for Telnet, 25 is for e-mail, 79 is for finger, 80 is for HTTP, 119 is for Netnews-and the list goes on. It is up to each protocol to determine how a client should interact with the port.

**JAVA AND THE NET:**
Java supports TCP/IP both by extending the already established stream I/O interface. Java supports both the TCP and UDP protocol families. TCP is used for reliable stream-based I/O across the network. UDP supports a simpler, hence faster, point-to-point datagram-oriented model.

**INETADDRESS:**
The InetAddress class is used to encapsulate both the numerical IP address and the domain name for that address. User interacts with this class by using the name of an IP host, which is more convenient and understandable than its IP address. The InetAddress class hides the number inside. As of Java 2, version 1.4, InetAddress can handle both IPv4 and IPv6 addresses.

**TCP/IP CLIENT SOCKETS:**

TCP/IP sockets are used to implement reliable, bidirectional, persistent, point-to-point and stream-based connections between hosts on the Internet. A socket can be used to connect Java's I/O system to other programs that may reside either on the local machine or on any other machine on the Internet.

There are two kinds of TCP sockets in Java. One is for servers, and the other is for clients. The Server Socket class is designed to be a "listener," which waits for clients to connect before doing anything. The Socket class is designed to connect to server sockets and initiate protocol exchanges.

The creation of a Socket object implicitly establishes a connection between the client and server. There are no methods or constructors that explicitly expose the details of establishing that connection. Here are two constructors used to create client sockets

Socket (String host Name, intport) - Creates a socket connecting the local host to the named host and port; can throw an Unknown Host Exception or anIO Exception.

Socket (InetAddressipAddress, intport) - Creates a socket using a pre existing Inet Address object and a port; can throw an IOException.

A socket can be examined at any time for the address and port information associated with it, by use of the following methods:

- InetAddressgetInetAddress ( ) - Returns the InetAddress associated with the Socket object.
- IntgetPort ( ) - Returns the remote port to which this Socket object is connected.
- IntgetLocalPort ( ) - Returns the local port to which this Socket object is connected.

Once the Socket object has been created, it can also be examined to gain access to the input and output streams associated with it. Each of these methods can throw an IO Exception if the sockets have been invalidated by a loss of connection on the Net.

Input Streamget Input Stream ( ) - Returns the InputStream associated with the invoking socket.

Output Streamget Output Stream ( ) - Returns the OutputStream associated with the invoking socket.

**URL:**

The Web is a loose collection of higher-level protocols and file formats, all unified in a web browser. One of the most important aspects of the Web is that Tim Berners-Lee devised a saleable way to locate all of the resources of the Net. The Uniform Resource Locator (URL) is used to name anything and everything reliably.

The URLprovides a reasonably intelligible form to uniquely identify or address information on the Internet. URLs are ubiquitous; every browser uses them to identify information on the Web.

## PROJECT DESCRIPTION

**PROBLEM DEFINITION:**

COVID-19 mainly spreads through the air when people are near each other long enough,[a] primarily via small droplets or aerosols, as an infected person breathes, coughs, sneezes, sings, or speaks. Transmission via fomites (contaminated surfaces) has not been conclusively demonstrated. It can spread as early as two days before infected persons show symptoms (presymptomatic), and from asymptomatic (no symptoms) individuals. People remain infectious for up to ten days in moderate cases, and two weeks in severe cases. The standard diagnosis method is by real-time reverse transcription polymerase chain reaction (rRT-PCR) from a nasopharyngeal swab. Preventive measures include social distancing, quarantining, ventilation of indoor spaces, covering coughs and sneezes, hand washing, and keeping unwashed hands away from the face. The use of face masks or coverings has been recommended in public settings to minimize the risk of transmissions. There are no proven vaccines or specific treatments for COVID-19 yet, though several are in development. Management involves the treatment of symptoms, supportive care, isolation, and experimental measures.

**OVERVIEW OF THE PROJECT:**

Due to the growing magnitude of number of cases and its subsequent stress on the administration and health professionals, some prediction methods would be required to predict the number of cases in future.The prediction of various parameters (number of positive cases, number of recovered cases, etc.) obtained by the proposed method is

accurate within a certain range and will be a beneficial tool for administrators and health officials.A better understanding of such opposition ahead of a COVID-19 vaccine is therefore critical for scientists, public health practitioners, and governments.Advantages of real-time and fast, which can predict the incidence trend of infectious diseases as early as possible, and are suitable for data analysis of a large number of people.

**MODULES DESCRIPTION**

In this project, contains the following modules:

- DATA
- ESTIMATION PROCESS
- DATA-DRIVEN METHODS TO PREDICT COVID-19
- DATA PRE PROCESSING
- PREDICTION OF ACCURACY
- CLASSIFICATION

**DATA:**

The data information includes the cumulative confirmed cases, the cumulative number of deaths, newly confirmed cases, and the cumulative number of cured cases provinces. We also used the data on the recent diagnoses in South Korea, Iran, and Italy, it includes the data, and here, the data comes from official notifications from various countries.All data are from the dailycase report and the update frequency of data is one day.

**ESTIMATION PROCESS:**

In different control stages, the Basic reproduction number changes greatly and it affects the intensity of control directly. In addition, the incubation period of the virus affects the speed of transmission directly. These two parameters need to be estimated. Current literature shows that the uncontrolled Basic reproduction. Therefore, we chose the valuation range in the corresponding range. For the controlled Basic reproduction number, the range of valuation was selected in the range of [0, 1.5].

**DATA-DRIVEN METHODS TO PREDICT COVID-19:**

The data has been used (when the first case of COVID-19 was reported in India) with 80% data is used for training and rest 20% for forecasting and validation purposes. The resulting plot showing the total number of confirmed cases, the observed data is the data used for training purposes, official data (green line) indicates the official data available and forecasted data indicates the forecast of a total number of confirmed cases. From this graph, it is observed that the forecasted number of total confirmed positive cases closely matches with the available official data.

**DATA PRE PROCESSING:**

Data Preprocessing is a technique that is used to convert the raw data into a clean data set. The dataset is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues.

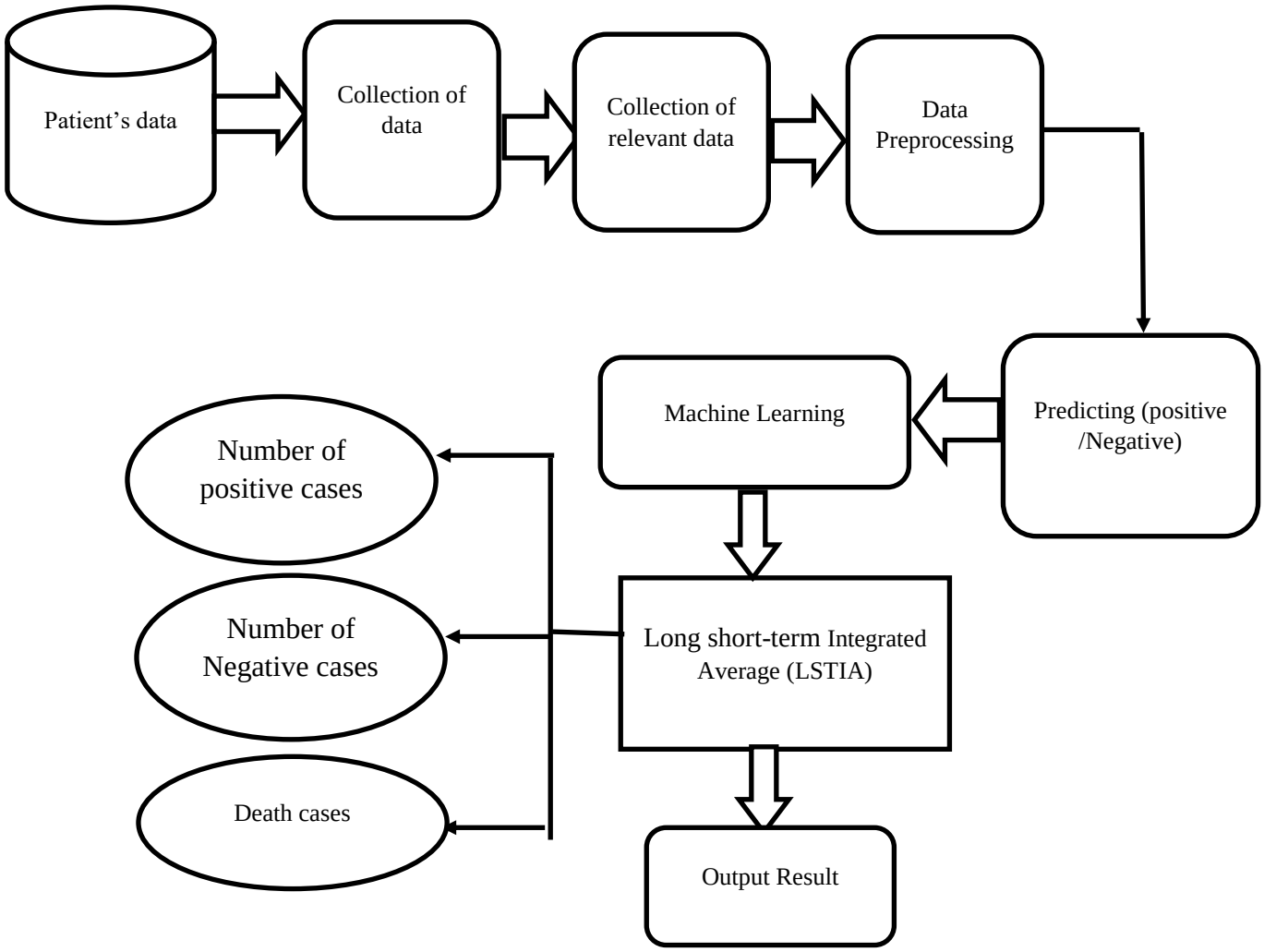
**PREDICTION OF ACCURACY:**

This technique is suitable to use predictive neural networks or characteristic data as such infection event or non-event binomial effects. The prediction accuracy of various measurements can be used for different purposes. They include the rate at which normal (non-predicted prediction correctly predicts sensitivity (non-infectious disease), accuracy (predicted percentage of predicted trend), positive predictive value, negative predictive value (correctly predicted infection rate is)), the ratio is Expected predictions are a measure of the likelihood that the increase in the entire process exceeds the accuracy of the individual).

**CLASSIFICATION**

The classification technique predicts the target class for each data set point. With the help of the classification approach, a risk factor can be associated with patients by analyzing their patterns of diseases.

**SYSTEM FLOW DIAGRAM**




**DATABASE DESIGN**

Database design is the process of producing a detailed data model of a database. This logical data model contains all the needed logical and physical design choices and physical storage parameters needed to generate a design in a Data Definition Language, which can then be used to create a database. A fully attributed data model contains detailed attributes for each entity.

The term database design can be used to describe many different parts of the design of an overall database system. Principally, and most correctly, it can be thought of as the logical design of the base data structures used to store the data.

In the relational model these are the tables and views. In an object database the entities and relationships map directly to object classes and named relationships. However, the term database design could also be used to apply to the overall process of designing, not just the base data structures, but also the forms and queries used as part of the overall database application within the database management system (DBMS). The most important consideration in designing the database is how information will be used. The main objectives of designing a database are:

**DATA INTEGRATION**In a database, information from several file are coordinated, accessed and operated upon as through it is in a single file. Logically, the information are centralized, physically, the data may be located on different devices, connected through data communication facilities.

**DATA INTEGRITY**

Data integrity means storing all data in one place only and how each application to access it. This approach results in more consistent information, one update being sufficient to achieve a new record status for all applications which use it.

**CONCEPTUAL DESIGN**

The next step is to form a concise description of the data requirements using a high level data model. This description would be independent of storage requirements. This step involves identifying entities involves in the system, and the relationship between the different entities.

**INPUT DESIGN**

The goal of designing input data is to make data entry as easy, logical and error free from errors as possible. In entering data, operators need to know the following: The allocated space for each field. Field sequence, which much match that in the source document. The format in which data fields are entered for example, filling out the date field is required through the edited format mm/dd/yy. When we approach input data design, we design the source document. Let us elaborate on each step.

Needless to say, therefore, that the input data is the lifeblood of a system and have to be analyzed and designed with at most case and consideration.

The decisions made during the input design are

- To provide cost effective method of input
- To achieve the highest possible level of accuracy
- To ensure that input is understand by the user.

Input design is a process of converting a user-oriented description of the input to the computer-based system. This design is important to avoid errors in the input process and show the correct direction to the management for getting the correct information from the computerized system.

A source document differs from a turnaround document in that the former contains data that change the status of a resource while the latter is a machine readable document. Transaction throughput is the number of error-free transactions entered during a specified time period. A document should be concise because longer documents contain more data and so take longer to enter and have a greater chance of data entry errors.

**OUTPUT DESIGN:**

Output design generally refers to the results and information that are generated by the system for many end-users; it should be understandable with the enhanced format. The Output of the software is used to make the remote installation of the new software in the system and, it is awake the immediate alert to the system that should be enhanced it as the input to the system. Output is the main reason for developing the system and the basis on which they evaluate the usefulness of the application.

Computer output is the most important direct source of information to the user output design deals with form design efficient output design should improve the interfacing with user. The term output applies to any information produced by an information system in terms of displayed. When analyst design system output, they Identify the specific output that is needed to meet the requirements of end user. Previewing the output reports by the user is extremely important because the user is the ultimate judge of the quality of the output and, in turn, the success of the system

As the outputs are the most important sources of information to the users, better design should improve the system's relationships with user and also will help in decision-making. Form design elaborates the way output is presented and the layout available for capturing information.

**REFERENCES**

[1] Alsaeedy, A. A. R., & Chong, E. (2020). Detecting Regions at Risk for Spreading COVID-19 Using Existing Cellular Wireless Network Functionalities. IEEE Open Journal of Engineering in Medicine and Biology, 1–1.
[2] Sear, R. F., Velasquez, N., Leahy, R., Restrepo, N. J., El Oud, S., Gabriel, N., … Johnson, N. F. (2020). Quantifying COVID-19 content in the online health opinion war using machine learning. IEEE Access, 1–1.
[3] Hu, S., Gao, Y., Niu, Z., Jiang, Y., Li, L., Xiao, X. … Yang, G. (2020). Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification from CT Images. IEEE Access, 1–1.
[4] Zhang, Y., Li, Y., Yang, B., Zheng, X., & Chen, M. (2020). Risk Assessment of COVID-19 Based On Multisource Data from a Geographical View. IEEE Access, 1–1.
[5] Abdel-Basset, M., Mohamed, R., Elhoseny, M., Chakrabortty, R. K., & Ryan, M. (2020). A hybrid COVID-19 detection model using an improved marine predator's algorithm and a ranking-based diversity reduction strategy. IEEE Access, 1–1.

[6] F. Petropoulos and S. Makridakis, "Forecasting the novel coronavirus covid-19," Plos one, vol. 15, no. 3, p. e0231236, 2020.
[7] G. Grasselli, A. Pesenti, and M. Cecconi, "Critical care utilization for the covid-19 outbreak in lombardy, italy: early experience and forecast during an emergency response," Jama, 2020.
[8] C. P. E. R. E. Novel et al., "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19) in china," Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi, vol. 41, no. 2, p. 145, 2020.
[9] Y. Grushka-Cockayne and V. R. R. Jose, "Combining prediction intervals in the m4 competition," International Journal of Forecasting, vol. 36, no. 1, pp. 178–185, 2020.
[10] N. C. Mediaite. Harvard professor sounds alarm on 'likely' coronavirus pandemic: 40% to 70% of world could be infected this year. Accessed on 2020.02.18. [Online]. Available: https://www.mediaite.com/news/harvardprofessor-sounds-alarm-on-likely-coronavirus-pandemic-40-to-70-ofworld-could-be-infected-this-year/