Artificial Intelligence-Based Predictive Tools for Life-Threatening Diseases

Vijay Jeyakumar, Prema Sundaram, and Nithiya Ramapathiran

Abbreviations

5G	Fifth generation
AI	Artificial intelligence
AIDS	Acquired immune deficiency syndrome
ANAM-NET	Anamorphic depth embedding-based lightweight CNN
ANN	Artificial neural network
AUC	Area under the curve
BERT	Bidirectional Encoder Representations from Transformers
CI	Confidence interval
CNN	Convolutional neural network
CNN-AE	Convolutional neural network-autoencoder
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
COVNet	COVID-19 detection neural network
CRP	C-reactive protein
CT	Computerized tomography
DHIS2	District Health Information Software into a single-window

---

V. Jeyakumar (✉)
Sri Sivasubramaniya Nadar College of Engineering, Chennai, India
e-mail: vijayj@ssn.edu.in

P. Sundaram
PSNA College of Engineering and Technology, Dindigul, India
e-mail: premas@psnacet.edu.in

N. Ramapathiran
Agni College of Technology, Chennai, India
e-mail: nithiya.bme@act.edu.in

G. R. Kanagachidambaresan et al. (eds.), *System Design for Epidemics Using Machine Learning and Deep Learning*, Signals and Communication Technology, https://doi.org/10.1007/978-3-031-19752-9_8

DICOM	Digital Imaging and Communications in Medicine
DL	Deep learning
DNA	Deoxyribonucleic acid
DNN	Deep neural network
DST	Department of Science and Technology
EHR	Electronic health records
EN	Elastic net model
FDA	Flexible discriminant analysis
GLDM	Gray level dependence matrix
GLSZM	Gray level size zone matrix
GRU	Gated recurrent unit
GUIDE	Graphical User Interface Development Environment
HIV	Human immunodeficiency virus
HMC	Hamiltonian Monte Carlo
ICU	Intensive care unit
INR	International normalized ratio
IoT	Internet of Things
K-NN	K-nearest neighborhood
LASSO	Least absolute shrinkage and selection operator
LDH	Lactate dehydrogenase
LR	Linear regression
LSTM	Long short-term memory
MATLAB	Matrix Laboratory
MCHC	Mean corpuscular hemoglobin concentration
MERS-CoV	Middle East respiratory syndrome coronavirus
ML	Machine learning
MOGA	Multi-objective genetic algorithm
MRMR	Maximum relevance minimum redundancy
NA	Not applicable
NBML	National Brain Mapping Laboratory
OWL	Web Ontology Language
pH	Potential of hydrogen
PLS	Partial least square regression
RF	Random forest
RNA	Ribonucleic acid
RNN	Recurrent neural network
RT-PCR	Reverse transcriptase-polymerase chain reaction
SARS	Severe acute respiratory syndrome
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SIR	Susceptible-infected-removed
SMOTE	Synthetic Minority Oversampling Technique
SOP	Standard operating procedures
SORMAS	Surveillance Outbreak Response Management and Analysis System
SpO_2	Partial saturation of oxygen
SVM	Support vector machine

SVM-RFE	Support vector machine-recursive feature elimination
SWRL	Semantic Web Rule Language
TELINET	Telefax Library Network
VGG	Visual Geometry Group
WASH	Water, sanitation, and hygiene
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting

Origin and Background of Diseases

A medical condition that generally affects the normal structure or function of an organism and is related to specific signs and symptoms is termed a disease. This may be due to any external factor or internal body disorder resulting in pain, dysfunction, injury, or syndrome. It represents the impairment of the body's normal homeostatic balance.

The different stages of a disease are its:

- Etiology
- Pathogenesis
- Morphological changes
- Corresponding functional alterations associated with the changes

Classification

Diseases are classified into majorly four types:

1. Infectious diseases are illnesses caused by pathogens and their toxic products such as bacteria, viruses, fungi, or parasites, e.g., COVID-19.
2. Deficiency disease or malnutrition disease is caused due to deficiency, excess, or imbalance of nutrients affecting body tissues, e.g., anemia.
3. Hereditary disease or metabolic disease gives rise to alterations in the gene or abnormality in chromosomes. The two subtypes under this category are genetic and nongenetic diseases, e.g., hemophilia.
4. Diseases caused by abiotic factors (cold, high temperature, low pH) as well as nutritional imbalances are termed physiological diseases. The organs and organ systems of the body do not function properly leading to illness, e.g., diabetes.

There are other ways of disease classification.

1. Based on spread, diseases are classified as follows:
   - *Communicable disease* involves the transmission of the pathogen from an infected host through various modes of transmission, e.g., HIV infection.
   - *Noncommunicable diseases* do not spread from individual to individual, e.g., cancer.

2. Based on the time frame, diseases can be classified as follows:
    - The *acute disease* appears suddenly and persists only for a short term from a few days to weeks, e.g., common cold.
    - The *chronic disease* develops gradually and persists for long periods usually from months to years, e.g., arthritis.
3. Based on the cause of the disease, the disease can be classified as follows:
    - A *primary disease* that occurs due to some initiating cause of illness, e.g., common cold.
    - *Secondary disease* arises due to the complication caused by the primary disease or its underlying cause, e.g., rhinitis can be due to a common cold or bacterial infection.
4. Based on the body system, diseases are classified as follows:
    - *Insanity* on *mental illness* comprises a broad range of diseases which includes emotional imbalance, behavior disturbance, logical dysfunction, or cognitive impairment and can be biological or psychological, e.g., depression.
    - *Organic diseases* can be brought about by the structural or functional changes to the organ systems of the body, e.g., stroke.
5. Based on the extent, diseases can be classified as follows:
    - A *localized disease* is confined to one part of the body, e.g., eye infection.
    - *Disseminated* disease indicates the spread of disease from the site of its origin to a distant site, e.g., cancer.
    - The *systemic disease* affects the entire body leading to illness, e.g., blood pressure

The course of a disease in an individual is the time taken from the inception of the disease until its complete recovery or eventual death. The progression of a disease is a set of biological events right from its etiology (cause) to the outcome (recovery or death). The important perspectives to characterize the history of a disease are that of the family doctor who can analyze the presence of health issues in an individual and that of an epidemiologist who can discover the occurrence of new diseases and their solutions by using health records and biostatistical information [1].

Phases of Disease

There are three phases in the progression of the disease.

1. *Pre-pathogenic period* – In this period, the disease arises but the person will not present any symptoms related to the disease in their body. So, the body conditions, causative agent, and the environment play a major role.

2. *Pathogenic period* – In this period, alteration in the person's body tissue originates but the person doesn't yet find out any clinical signs of the disease. It includes two sub-phases.
   - The incubation period is a fast-evolving phase from hours to days and is the time taken for the disease-causing agent to propagate and produce symptoms after exposure to the pathogen in the host.
   - The latency period or pre-infectious period is a very-slow-evolving phase lasting from months to years and is the time taken to become infectious and capable of transmitting the disease.
3. *Clinical period* – In this period, the patient shows clinical signs and changes, and the pathology of the disease keeps continuing without treatment and can result in recovery, disability, or fatality. It includes three sub-phases.
   - Prodromal – Denotes the clinical beginning of the disease where the first symptoms appear
   - Clinical – Indicates the appearance of specific symptoms which could be identified by the physicist to choose the right treatment protocol either to treat the patient or to avoid any long-term damage
   - Resolution – Represents the terminal phase where the disease can completely resolve or progress to an advanced stage or can result in death

Transmission of Infection

This process follows the three factors mentioned below:

- An infected individual.
- A vulnerable individual.
- An effective contact between the infected and the vulnerable individual is such that the infected individual affects the vulnerable individual.

The number of effective contacts made in unit time in a given population is termed effective contact rate (β). It is nothing but the multiplication of total contact rate (γ) (whether effective or not) and the risk of infection transmission (p) and is given by the expression

$$B = \gamma^* p \tag{1}$$

Infectivity is the ability of an infectious agent to get access and amplify in the host causing an infection. Transmissibility is the chance of transmitting an infection from the infected one to the susceptible individual and can spread either by the community or local transmission. In community transmission, the epidemiological link among a group of people causing the spread of infection is not known, whereas in local transmission, the origin of infection within a specified location can be identified.

Routes of Transmission

The routes of disease transmission between different populations are essential to know about the patterns of contact and the spread of the disease based on social-economic and cultural conditions. The disease-causing agent can be transmitted in either of the following ways:

- *Horizontal disease transmission* – Infection can be acquired from the environment or unrelated people in the same generation either through direct or indirect contact.
- *Vertical disease transmission* – It is the passage of the infectious agent from the mother to the offspring which can be intracellular or extracellular.
- *Mixed-mode disease transmission* – Disease transmission can be vertical as well as horizontal depending upon the host density which can be low or high, respectively.

Mode of disease transmission can be in the following ways:

1. *Airborne infection* – Refers to the transmission of very small dry and wet aerosols of size less than 5 μm that spread through evaporated droplets which can remain living in the environment for long periods and can affect the upper and lower respiratory tract, e.g., tuberculosis.
2. *Droplet infection* – Transmission of a respiratory droplet is the most common route for respiratory infections produced by coughing, sneezing, or talking by infected hosts, wherein small and wet particles of size larger than 5 μm remain in the environment for a limited period over minimum distances. It can be transmitted either directly to the susceptible mucosal surface or indirectly through contact with the contaminated area, e.g., coronavirus.
3. *Direct physical contact* – Direct contact happens through skin-to-skin touch, kissing, and sexual contact and can lead to a variety of infectious and contagious diseases, e.g., syphilis.
4. *Indirect physical contact* – Also referred to as vehicle-borne transmission characterizes the spread of infection through the contamination of insensate objects, which can act as vehicles transmitting the infection either through a vector or an intermediate host, e.g., malaria.
5. *Transmission by other organisms* – Transmission is usually through a vector that may be of mechanical or biological origin and is not responsible to cause disease on its own but can spread infection by transmitting pathogens from one person to another, e.g., rabies.
6. *Fecal-oral transmission* – This mode of transmission poses a serious health risk for people where the fecal particles pollute food and water so that the pathogen transmits from one person to the mouth of another due to poor sanitation and personal hygiene, e.g., cholera.

Morbidity and Mortality Rates

The two important and commonly used terms in epidemiology are morbidity and mortality which are denoted as a proportion or rate and represent the frequency and severity of a specific disease condition [2].

- Morbidity is defined as a state of having a particular illness or complication which is an unnatural consequence of a disease or treatment. It can be acute or chronic and can be presented either as the incidence or prevalence of a disease, e.g., heart disease, stroke, diabetes, obesity, and infection.
- The morbidity rate represents the rate of disease occurrence in a population and evaluates the overall health of a population and their health care requirements:

$$\text{Morbidity rate}^{=} = \frac{\left(\text{Number of cases of disease or injury or disability}\right)}{\left(\text{Total population}\right)} \tag{2}$$

- Mortality indicates the number of deaths caused due to a specific illness or disease condition, e.g., in 2020, COVID-19 will be a remarkable cause of mortality.
- Mortality or death rate measures the count of deaths in a specific population scaled to the total population per unit of time.

Increased human and animal interaction due to expanded global trade events has resulted in the transmission of zoonotic pathogens and facilitated the continuous transmission of infections among mankind. The progressive population escalated the exposure and spread of novel illnesses to the increased risk for outbreaks, epidemics, and pandemics [3].

Let us review the significant terms for disease occurrence and transmission in various geographical regions [4].

- An endemic refers to the predictable rate of a disease condition that is present consistently among the population, e.g., malaria.
- An outbreak is a sudden rise in the count of people presenting a disease. It may occur in one place or extend widely and can last for days to years. It may be new to the community or has been absent for a long period, e.g., influenza.
- An epidemic spreads quickly to a large population over an extended geographic area and is considered an outbreak, e.g., polio.
- A pandemic spread across countries and continents affecting mankind globally and is an epidemic, e.g., COVID-19.

The mechanism of cross-species pathogen transfer relies on different phases.

- Phase 1 – The infectious agent infects animals particularly.
- Phase 2 – The pathogen progresses and enters the human system, but without human-to-human transference.
- Phase 3 – A few cycles of secondary infection occur between humans.

- Phase 4 – Infection persists in animals although long periods of secondary infection exist among humans but without the presence of animal hosts.
- Phase 5 – Infection persists in humans exclusively.

Pathogens of Bacteria and Virus

The risk of pathogen transmission, in turn, depends upon the animal host carrying the infectious agent, the type of cross-species interaction, and its probability. The most notable past pandemics that have distressed mankind in world history include plague, cholera, influenza, smallpox, HIV and AIDS, and coronavirus disease [5] which are shown in Table 1.

Plague or Black Death

Gram-negative rod-shaped facultative anaerobic bacteria *Yersinia pestis* causes plague which infects rodents and mammals through fleas. Fleas feed on infected rodents, receiving the bacteria which then multiplies in the flea's gut. It regurgitates the blood to new rodent hosts and spreads the infection in three different routes.

1. Bubonic – The most common form which is fatal due to an infected flea with flu-related symptoms (50–90%).
2. Septicemic – The rare form affecting 10–25% of the host resulting in progressive infection of the bloodstream.
3. Pneumonic – It occurs when the bacteria lodges in the lungs and is a rapidly fatal condition when untreated.

The three pandemic plague episodes are:

1. The initial pandemic which occurred between 541 and 543 is the plague of Justinian which originated in Egypt and passed across Rome causing a high mortality rate of 100 million people.
2. In the early nineteenth century, the second plague pandemic wave called Black Death started in Europe and eventually caused the death of 200 million people.
3. During the nineteenth century, the third plague pandemic existed in China and was endemic throughout the world in later years [6]. It is considered as a re-emerging infection by WHO affecting different parts of the world in a variety of periods and is proved to be highly fatal without immediate treatment and can be a major potential public threat.

Table 1 Life-threatening diseases and their origin

Years	Epidemics/ pandemics	Pathogens	Vectors	Death toll	Location
541–543	Plague of Justinian	*Yersinia pestis*	Fleas associated with wild rodents	15–100 million	Europe and West Asia
1347–1351	Black Death	*Yersinia pestis*	Fleas associated with wild rodents	75–200 million	Europe, Asia, and North Africa
1817–1824	First cholera pandemic	*Vibrio cholerae*	Contaminated water	1–2 million	Asia, Africa
1827–1835	Second cholera pandemic	*Vibrio cholerae*	Contaminated water	Less than a million	Europe, Asia, America
1839–1856	Third cholera pandemic	*Vibrio cholerae*	Contaminated water	1–2 million	Asia, Europe, Africa, and North America
1863–1875	Fourth cholera pandemic	*Vibrio cholerae*	Contaminated water	600,000	Russia, Europe, Africa, and North America
1881–1886	Fifth cholera pandemic	*Vibrio cholerae*	Contaminated water	Less than a million	Worldwide
1885–ongoing	Third plague	*Yersinia pestis*	Fleas associated with wild rodents	12–15 million	Worldwide
1889–1893	Russian flu	Influenza A/H3N8	Avian	1 million	Worldwide
1899–1923	Sixth cholera pandemic	*Vibrio cholerae*	Contaminated water	Less than a million	Europe, Asia, and North Africa
1918–1919	Spanish flu	Influenza A/H1N1	Avian	17–100 million	Worldwide
1957–1959	Asian flu	Influenza A/H2N2	Avian	1–4 million	Worldwide
1961–ongoing	Seventh cholera pandemic	*Vibrio cholerae*	Contaminated water	3–5 million	Worldwide
1968–1970	Hong Kong flu	Influenza A/H3N2	Avian	1–4 million	Worldwide
2002–2003	Severe acute respiratory syndrome	SARS-CoV	Bats, palm civets	Less than a million	China, Hong Kong, Canada
2009–2010	Swine flu	Influenza A/H1N1	Pigs	Less than 3 lakhs	Worldwide
2015–ongoing	Middle East respiratory syndrome	MERS-CoV	Bats, dromedary camels	Less than a lakh	Worldwide
2019–ongoing	Coronavirus disease	SARS-CoV-2	Bats, pangolins	5.7–23.8 million (as of February 2022)	Worldwide

Cholera

A Gram-negative comma-shaped facultative anaerobic bacteria *Vibrio cholerae* acutely infects the gastrointestinal tract and causes the most fatal disease called cholera. The bacteria reside in the small intestine producing cholera toxin leading to rapid dehydration resulting in hypovolemic shock and fatality. It involves the water-borne transmission of the infection which can be mild or asymptomatic [7].

Cholera remained endemic in Asia up to 1817 and then became a pandemic spreading and infecting different parts of the world leading to seven cholera pandemics due to increased globalization and foreign trade resulting in a fatal condition. Cholera cannot be completely eradicated, and its persistence is due to poor living conditions and proper sanitation and inefficient sewage systems. Also, changes in the ecosystem and climate favored genetic exchange protocols in *Vibrio cholerae* making them more virulent causing mild to life-threatening infections leading to a potential threat for mankind.

Influenza or Spanish Flu

Influenza viruses are enveloped RNA viruses (single, negative sense) belonging to the *Orthomyxoviridae* family coding for ten structural and nonstructural proteins causing severe illness and death worldwide. Influenza virus consists of four strains: A, B, C and D. Strains A and B types cause outbreaks in tropical areas and seasonal epidemics in temperate areas. They show asymptomatic or mild or classical illness and cause approximately 5 lakh deaths. They may also show severe illness and pneumonia along with secondary bacterial infection with *Streptococcus* and *Staphylococcus* species and *Haemophilus influenzae*. Strain A is endemic in humans, birds, and pigs [8].

Reassortments and genetic rearrangements lead to the antigenic shift resulting in the origination of new subtypes of virus influencing spontaneous human transmission and may result in a pandemic. The hit of a pandemic was majorly due to transmission and pathogenicity of the different types of strains and the vulnerability of the people which varied among different age groups and their earlier exposure to the virus. The constant modification and interchange of genes between different species of the influenza virus along with cross-species transmission is a critical challenge for the origin of new strains of avian virus causing low pathogenic and highly pathogenic avian influenza virus strains.

Smallpox

It was an acute, contagious, and deadly infectious disease caused by two strains of variola virus, namely, *Variola major* and *Variola minor*, belonging to the *Orthopoxvirus* family. Initial symptoms of infection were common to all viral diseases and the smallpox virus attacks skin cells leading to the formation of characteristic pimples which then spreads throughout the body ranging from ordinary, modified, malignant, or hemorrhagic forms. Pieces of evidence showed that the disease occurred in outbreaks and there were 500 million deaths in the last hundred years of its existence and it was completely eradicated globally in 1980.

HIV or AIDS

Human immunodeficiency virus falling under retrovirus family causing acquired immune deficiency syndrome (AIDS) is considered a global epidemic infecting 37.9 million people including adults and children less than 15 years of age as of 2018. The most affected region is Sub-Saharan Africa. The pandemic is heterogeneous within regions, with some regions more affected than others. After the initial infection, the person may experience a short period of influenza-like symptoms followed by a long incubation period with the progression of infection, and the immune system gets weakened, thereby elevating the risk of other common infections, inviting opportunistic infections and tumors associated with severe weight loss. This disease has a large impact on society as well as the economy.

Coronavirus

The novel coronavirus termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belongs to the *Coronaviridae* family infecting animals and humans. They have enveloped RNA virus (+ sense, single-stranded) and include four genera – Alpha, Beta, Gamma, and Delta variants. The virus was identified from an outbreak in China in Wuhan city causing the coronavirus or COVID-19 pandemic. The virus is of zoonotic origin from bats and related mammals. Human coronavirus can lodge in the upper and lower respiratory tract causing severe respiratory system diseases and sometimes to a lower extent can also cause gastroenteritis, although it is rarely fatal. It spread globally to more than 390 million people causing 5.72 million fatalities making COVID-19 one of the deadliest diseases in world history. The symptoms range from mild to deadly conditions including fever, dry cough, and tiredness and are transmitted by small airborne particles enclosing the virus. Virus-infected people can transmit the disease for 10 days even if symptoms do not develop. So, vaccination, social distancing, wearing of face mask, and quarantining are the recommended preventive measures to combat the coronavirus [9].

The common cold is caused by the Alpha-coronavirus genus, whereas the Beta-coronavirus genus causes severe respiratory tract infection. In addition, Beta-coronavirus includes three pathogenic strains of virus, namely:

1. Severe acute respiratory syndrome coronavirus (SARS-CoV)
2. Middle East respiratory syndrome coronavirus (MERS-CoV)
3. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative factor of COVID-19 inducing dreadful pneumonia

SARS-CoV originated in China and the natural reservoirs were bats and intermediary hosts were palm civets before human infection. The transmission was nosocomial and caused influenza-related symptoms. The major cause of fatality was respiratory failure and patients infected also showed pneumonia and watery diarrhea-causing serious threats. MERS-CoV showed its first emergence in Saudi Arabia. Bats were the prospective animal reservoirs and Arabian camels were the intermediary hosts infecting humans showing clinical illness from moderate to drastic pulmonary disease.

SARS-CoV-2 pandemic causing COVID-19 involving bats as an animal reservoir and pangolins as animal host before virus transmission to humans. COVID infection is asymptomatic and can show a range of mild symptoms to common complications affecting the major organs like the heart, lungs, and liver and can result in life-threatening conditions.

Disease Management System

The spillover of zoonotic organisms and the transmission dynamics of infection to mankind are remarkably escalating due to various factors such as globalization, trade of animals, animal-based food products, land use and urbanization, modifications in the living habits of the pathogen, and increased animal-human interaction [10].

This may pose new threats and challenges to public health, and there will be recurrence of epidemic and pandemic diseases. To control the spread of various diseases and to have a check over the increase, numerous protocols have been devised for disease management.

- Water, sanitation, and hygiene (WASH) programs to ensure safe drinking water access, effective sanitation methods, and hygiene practices.
- Vector control program targeting immature or adult stages of vector to control the spread of infection.
- Global surveillance program for quick detection of pathogen spillover to avoid cross-species transmission.

- Global virome project began to encounter viral transmission and to develop pathogenicity markers for life-threatening viruses.
- Next-generation sequencing of viral genome analysis to enhance novel virus detection and characterization of the virome in the respiratory system.
- Vaccine development and their clinical trials.

In the pandemics, healthcare companies have initiated the use of big data analytics and prediction tools to better understand the pathogen and cross-species infection transmission. Thus, healthcare settings expanded the use of predictive algorithms and models to gain insight into the risks and outcomes of disease and the potential impact of the pathogen on humans.

Role of Expert Systems

The expert system is one of the major applications in predicting the pandemic and epidemic illness**.** An expert system predicts the thoughts and behaviors of an individual or organization having experience in a specific sector. It contains the prior knowledge and also a set of procedural guidelines as a base for each unique case. The major qualities of an expert system are a representation of the user data, inference, and detailed explanations which are all terms that come to mind when thinking about the user interface, etc. The advantages of an expert system are higher dependency, fewer errors, lower costs, a smart database, and lower risk. Even though there are many benefits, certain limitations lack common sense without artificial intelligence (AI) [11].

The ultimate focus of an expert system-based medical decision tool is to protect, cure, and identify diseases. The evidence obtained through such a system can be taken as a second opinion which could support the doctors and medical specialists in clinical decision-making. Among the existing algorithms and predictive techniques, AI is the most promising one, where the infectious disease dataset will be collected in various forms and given as input. Based on the inputs, AI can predict disease like how mankind thinks and acts. However, in a few scenarios, unpredicted parameters are evolved during the transmission of infectious diseases such as influenza, severe acute respiratory syndrome (SARS), etc. Hence, it is difficult for AI to predict results in the early stage, especially in cancer and lymphoma diagnosis. In such conditions, AI encompasses machine learning as a subdomain in the medical expert system for the early prediction of infectious disease using Bayesian distribution, artificial neural network (ANN), and so on. From Fig. 1, the expert system containing mathematics modeling, machine learning, and AI can be used to forecast numerous types of diseases. To reduce the spread of contagious diseases, standard operating procedures (SOP), self-testing, and quarantine are not sufficient. Rather, mass screening through a noncontact ecosystem would be the best alternative by adopting AI.

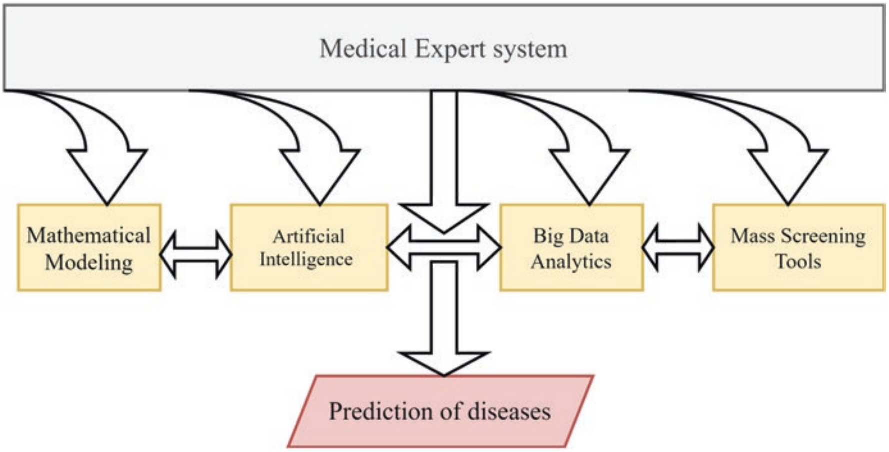

Fig. 1 Medical expert system

Mathematical Model for Predictive Modeling

Epidemic and pandemic diseases spread easily. Their outburst in the environment is due to the emergence of mass level interaction of pathogens and host bacteria in the atmosphere. To predict the pandemic disease, different types of modeling are available. Some of the specific components of spread can be predicted by mathematical models like compartment-based susceptible, infectious, and/or recovered (SIR) model, contact network model, disease control model, and many more. Mathematical modeling will follow particular rules depending on the behavior of diseases; it will also help us for the prediction of disease easily.

Compartment Model

Since 1921, the compartment model has been used to detect infectious diseases like influenza, Ebola, and end-level kidney disease. In this model, different compartments are combined and the comparative study on the different compartments from the predicted results can be easily done. SIR is the most important model used to predict measles and Nipah viruses accurately compared to other mathematical models.

SIR Model

The most widely used compartment model for infectious disease is the susceptible-infected-removed model [12]. Since individuals fall into any one of these categories, the SIR model is a basic predictive tool, primarily functioning on individual

Table 2 SIR model-based prediction tools

Predictive tools in a compartment model	Diseases predicted	Prediction accuracy (%)
SIR model (pandemic disease)	Measles, Nipah virus infection, Asian influenza, and so on	93
SIS model (disease not having permanent immunity like a seasonal disease)	Viral fever, cold, cough	96
Vertical transmission (where children got infected from parents)	AIDS and hepatitis	93.2
Vector transmission (not spread by human to a human directly)	Malaria, swine flu	94

concern reports. This model represents the probability of the disease spreading using the data of affected S(x), prone to get infected I(x), and recovered R(x) individuals. The SIR model has been tested with the COVID-19 dataset to predict the test results [13]. The SIR model starts to work when there is a combination of the number of people who got infected and people having the possibility of infection and runs until there is a change in this combination. The result is obtained from the count of the patient cured and the death rate of patients affected by this pandemic disease and is statistically represented. If the epidemic disease is spread worldwide in one shot or even if the mortality rate is very high, the SIR model can be recommended as a predictive tool. Thus, the model can estimate the rate of recovered individuals, thereby determining the rate of infectious disease. Table 2 shows the predictive disease accuracy done on various predictive tools in the SIR model.

Limitations:

- The pathogen's characteristics change due to age, environmental conditions, and genetic and environmental factors.
- It does not consider the latent period followed by exposure.
- It severely assumes the population with no migration and the number of deaths and birth apart from epidemics.
- Quantification of uncertainty cannot be expected in this model.

Other Mathematical Models

As shown in Table 3, mathematical modeling can be used as a predictive tool in a variety of ways [14].

The Implication of Machine Learning Algorithms

In recent years, machine learning algorithms have been utilized as a predictor of infectious disease during pandemics and epidemics. Machine learning is the best use of AI in which all the input is given to the computer, which then learns and

Table 3 Mathematical modeling used for predictions

S. no	Types	Methodology/algorithm	Applications	Accuracy (%)
1	Mathematical tools for biochemical biomarkers	Receiver operator curve analysis, logical regression analysis	Ebola, AIDS, influenza	88
2	Dynamic modeling	Discrete-time epidemic modeling	Measles	95
3	Epidemic chain models	Chain model	Chicken pox, measles, mumps	93.2
4	Reversible jump Markov chain Monte Carlo methods	Bayesian network	*Staphylococcus*	94.3

begins to react based on the data and is also capable of predicting future disease progression. Unsupervised and supervised algorithms are two types of machine learning algorithms [15].

Supervised Algorithm

A supervised algorithm predicts an undefined disease based on the input information that the expert trains over the output. Examples of supervised algorithms include Bayesian, ANN, support vector machine (SVM), K-nearest neighbor, convolutional neural network (CNN), multimodal logistic regression, and linear regression (LR).

The Bayesian Networks

The Bayesian network model is a fundamental and one of the most effective machine learning algorithms with 97% accuracy for predicting and interpreting diseases [16]. It represents the numerous connections between nodes, where each node has its own set of variables. Both statistical data and simply accessible algorithm frameworks, such as STAN, Web Ontology Language (OWL), Semantic Web Rule Language (SWRL), naive Bayes, and the random forest method are used in Bayesian analysis. For convenience, graphical data can be depicted to deliver the right prediction for various forms of terrible diseases as shown in Fig. 2 [17].

The Bayesian network is mostly used to analyze unpredictably large amounts of data as well as to develop forecasting algorithms for specific diseases. The STAN Bayesian network has been built using the available medical dataset and the knowledge of experts. HMC sampling is the primary tool used in STAN to diagnose epidemic and pandemic diseases quickly and accurately. When compared to other Bayesian networks, the result obtained with this probabilistic programming approach is accurate [18]. The expert must anticipate the disease using two web

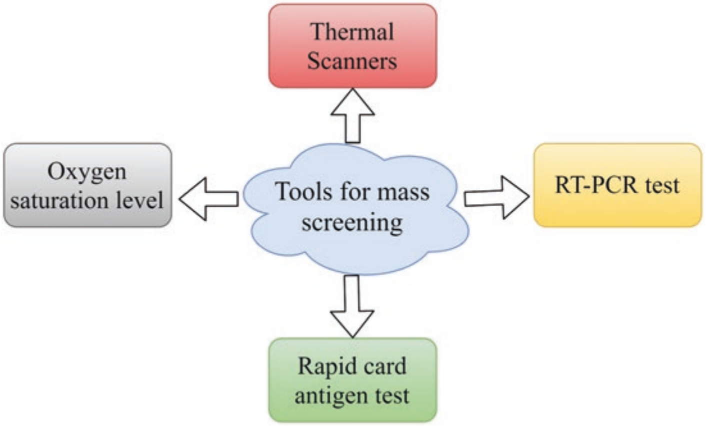

Fig. 2 Diagnostic tools for mass screening

languages, OWL and SWRL, based on medical facts [19]. Both web languages are knowledge-based prediction tools, and they can be combined with a Bayesian network to provide an outcome.

Other Machine Learning Methods

There are various other machine learning methods for disease prediction. They are as follows:

(i) Artificial Neural Network (ANN)

An ANN based on the Internet of Things (IoT) was created to forecast diseases like typhoid and severe acute respiratory syndrome (SARS). The development of an ANN network was based on the user's perception to improve the quality of life for those suffering from life-threatening diseases [19]. It is the algorithm that will revoke the working of the human brain as a network to find a solution from it. It was built using ten types of networks and different hidden layers for the early diagnosis of Middle East respiratory syndrome (MERS).

(ii) Support Vector Machine (SVM)

SVM works on the available input data and can classify the disease conditions into two categories of infected and not infected groups. There are two optimizations in SVM, namely, linear and quadratic, based on the division of data into groups. If the data are separated equally, it will be classified under linear optimization otherwise quadratic optimization [20]. In quadratic optimization, kernel values are used to obtain different results. SVM classifier can be used as a predictive tool so that it can identify the pandemic disease-related microRNA [21] and DNA variation of the specified disease. Therefore, these are the two important vital parameters needed to identify or classify at the molecular level for the prediction of the spread of disease.

(iii) K-Nearest Neighbor (KNN)

The KNN classification model can be used even when the dataset is unknown. It is completely based on the K values of the nearest neighbor data values, which are determined using the Euclidean distance. Once the dataset is available, the "K" similar neighbor in the unknown dataset can be identified and classified based on the correlation of "K" values. The K-nearest neighbor algorithm has been used to predict diseases such as Parkinson's and acute leukemia [22].

(iv) Convolutional Neural Network (CNN)

Convolutional neural networks (CNN) are automatically made up of numerous levels such as pooling, convolution, and fully connected layers, and with these features, disease prediction can be extracted from a back-projection algorithm. The input data for the network includes a dataset of symptoms. In comparison, the results showed that CNN is 85% more accurate than KNN because other factors, such as time and memory, were required in the K-nearest neighbor method.

(v) Multimodal Logistic Regression

Once trained, this network is the best method in biological sciences as the multimodal LR mostly uses the dependent variables, namely, regression and classification variables [23]. From each modal, common features can be chosen for multiple variables. Multimodal or multi-task logistic regression finds application in the study of Alzheimer's disease and cardiac arrest [24].

(vi) Linear Regression

Usually, independent variables can work on linear regression. If there are two variables present in the same application, then linear regression can be used so that a comparative study could be made between them [25]. Depending on the type of application employed, the algorithm can differentiate between linear and logistic regression. In Table 4, the various machine learning tools used for the prediction of the pandemic disease with their accuracy level are depicted [26].

Unsupervised Algorithm

The unsupervised algorithm is used to analyze and aggregate unlabeled symptom datasets of diseases. Using this algorithm, the identification of hidden patterns and data grouping can be accomplished without the assistance of a human being. The two techniques involved in unsupervised algorithm are as follows:

Clustering

It is the process of combining similar datasets into one cluster and dissimilar datasets into another cluster. Depending on the common characteristics, clustering can be done easily for the prediction of the disease, e.g., chronic diseases like diabetes.

Table 4 Machine learning tools used for the prediction of disease

S. no	Machine learning tools	Techniques	Predicted disease	Percentage of prediction
1	ANN	Three-layered network with 1000 hidden layers, under the curve (AUC) specificity, and predictivity are determined	Breast cancer, Naples plague, SAARC, H1N1 Flu	P curve value almost accurate to expert result AUC = 0.939 (P < 0.001)
2	SVM	Select the set of variables and classify the individual into two categories. C-I (infected) vs C-II (not infected/maybe infected). The classification was done on common variables for pandemic diseases	Respiratory syndrome, MERS, Ebola, flu	93.34%
3	K-nearest neighbors	Alignment-free method for the prediction of new virus by a comparative study of the sequence of molecules	SARS, H1N1, MERS, West Nile fever	94.30%
4	CNN	Classification using 16 layers, visual group network, mobile net, U-Net on imaging	Malaria, smallpox, SAARC, influenza H1N2	93%
5	Multimodal LR	The classifier network will classify the dataset into four different layers of classes so that prediction can occur	SARS, ARDS, malaria, influenza	96.2%
7	Linear regression	Tenfold cross-validation techniques using Synthetic Minority Oversampling Technique (SMOTE) data	Malaria, influenza, heart disease	Accuracy = 95.2% AUC = 99%

Association

It is one of the unsupervised algorithms which can predict disease from the specified symptoms by the technique of association, e.g., genetic disease.

Deep Learning Model for Big Medical Data Analytics

According to the World Health Organization, 12.8 million people died worldwide because of past pandemic diseases. This pandemic has generated a large amount of data, which is either structured or unstructured; the main job of Big Medical Data Analytics is to work on this large dataset and improve the efficiency of data processing to provide useful information to the public to prevent epidemic diseases such as influenza, flu, and malaria [27].

Big data processing includes data consolidation and fine-tuning and removing unwanted data to generate the required data in various forms. Multi-objective genetic algorithm (MOGA) is the algorithm used to generate big data in various efficient forms. This allows technicians to speed up data processing for early prediction of pandemic disease but is the most time-consuming step [28].

The Internet of Things (IoT) and big data analytics are playing an important role in the prediction against pandemic disease to collect enormous amounts of data, to view disease information, dissipate pandemic disease risk, and take precautions against disease prevention. For processing all this data, a neural network model can be employed which provides 98% accuracy when compared to other machine learning models [29].

Another type of big data analytics is the gene prediction algorithm, where most of the predictive analysis for the epidemic disease is dependent on the type of gene prediction. Particularly, the genes that are coded like protein-coding genes will be collected from the infected person and studied for the identification of a specific gene responsible for a disease using an National Brain Mapping Laboratory (NBML) algorithm. According to the findings, big data analytics can predict disease with an accuracy of 96.05% [30].

Big data is simply a massive database gathered from the public in various scenarios. The main challenge in big data analytics is maintaining the privacy of public data. Digital Imaging and Communications in Medicine (DICOM), electronic health records (EHR), or IoT-based tools were deployed to store the data in the appropriate location for future analysis and diagnostic purposes. STATA and MATLAB R2021 are the tools used to maintain the privacy of the big data that are used in big data analytics [29].

Tools Used for Mass Screening

According to the modeling study shown in Fig. 2, mass screening techniques utilized can reduce the spread of pandemic disease by 60% and the mortality rate can be reduced from 0.66% to 0.19%. As a result, most experts believe that mass screening could reduce the spread of infectious diseases [31].

Predictive Tool for Mass screening

(a) Lyfas COVID Score is a DST-funded Android application developed by Bangalore-based Acculi Labs. It is a mobile phone application in which the individual places his/her index finger on the camera, and it analyzes 95 biomarkers, namely, capillary pulse, blood volume variations, etc., using different processors and sensors to track the small changes in the physiological and path-

Table 5 Mass screening software for prediction

S. no	Name of the software	Developer	Year	Technology
1	DHIS2 (District Health Information Software)	University of Oslo	2020	National information system platform for data integration around 70 countries. Case-based surveillance to track the situation easily
2	SORMAS	Helmholtz Centre for Infection Research	2020	Mobile e-health systems are mainly for disease control systems
3	Go. Data	WHO and GOAN	2020	It is the data collection tool for Ebola, MERS, and COVID-19. From the pathological data collected by this app, contact list tracking can be done

ological conditions of the body. It can be used as a predictive app for mass screening.

(b) Graphical User Interface Development Environment (GUIDE) is a platform that combines many predictive software tools, such as District Health Information Software into a single-window (DHIS2), Surveillance Outbreak Response Management and Analysis System (SORMAS), and GO. The various tools introduced by the vendors are shown in Table 5.

Thus, these expert systems for predicting pandemic and epidemic diseases help in the management and prevention of COVID-19 as follows.

COVID-19 Prediction Tools

In 2019 SARS-CoV-2 is impacting global health and becoming a life-threatening disease. Many countries impose curfews and lockdown based on the impact of the COVID-19 virus and its variants. These restrictions by central and state public administrations affect the daily routines of individuals, professionals, and both private and public organizations. It creates a major impact on the global economy, thereby affecting all the sectors like education, tourism, manufacturing, automobile, etc. Many COVID-19 medication solutions like vaccines, drugs, self-testing home kits, diagnosis, and treatment procedures are being proposed by healthcare researchers and professionals in various instances. As the SARS-CoV-2 virus is mutating into different characteristics and spreading as an airborne disease, the rate of morbidity and mortality across the countries has become more critical to estimate [32]. The statistical data about the variants of coronavirus is very rare and almost not available for prediction and estimation. The evolving field of radiomics provides knowledge extraction from the clinical images that are acquired from different sources along with biological results.

AI plays a major role in the prediction and diagnosis of COVID-19 infections using various sources of datasets like CT X-ray images, reverse transcriptase-polymerase chain reaction (RT-PCR) examination results, the blood sample features, and the symptoms identified from the individuals. Many machine learning (ML) algorithms and CNN-based frameworks have been developed and published by researchers since January 2020. Apart from COVID-19-related predictions, recent technologies like the Internet of Things (IoT), fifth generation (5G) communication, and blockchain technologies are also adopted for the effective diagnosis and treatment of coronavirus-affected patients. AI-based noncontact measurement techniques of physiological parameters of an individual or group have become another research avenue during this pandemic situation [33]. Both ML and deep learning-based algorithms are applied to predict the spread of the disease, disease classification (pneumonia, non-pneumonia, COVID-19, and healthy), diagnosis, mortality risk, ICU admission, survival rate, and discharge period as illustrated in Fig. 3.

Machine Learning Algorithm-Based Predictive Tools

COVID-19 disease management is being carried out by many professionals to find out the mortality rate, ICU admission, prediction of disease, disease classification, and survival rate of the severely affected persons. To accomplish these tasks, LR, Extreme Gradient Boosting (XGBoost), SVM, neural networks (NN), random forest (RF), recurrent neural network (RNN), long short-term memory (LSTM), and least absolute shrinkage and selection operator (LASSO) have been considered either individually or as a combined model with others to arrive at the best results [34]. The performance measures like sensitivity, specificity, F1-score, accuracy, and AUC are incorporated in the proposed system to justify their outcomes.

Li et al. [35] collected a total of 413 patient data such as age, CT scan results, cough and fever signs, and blood sample results. These significant features were considered to distinguish influenza and COVID symptoms using the XGBoost model. They achieved a sensitivity and specificity of 92.5% and 97.9%, respectively. A similar dataset was collected [36] from the University Medical Centre Ljubljana to predict COVID-19 diagnosis using random forest (RF), deep neural network (DNN), and XGBoost. These data were collected from 5333 patients. They concluded that the MCHC, eosinophil ratio, prothrombin, INR, prothrombin percentage, and creatinine are the important factors that differentiate COVID infections from other bacterial infections.

Many articles related to COVID-19 are published based on machine learning and deep learning frameworks. Some of them are listed in Table 6.

Few researchers utilized LR to differentiate COVID infections from other respiratory disorders. Most of the datasets, considered as inputs to the expert systems, are patient-related information such as age; medical history; smoking habits; self-estimated results like cough, fever, sore throat, smell, and taste sensations; and complete blood count components [45, 48, 49]. Apart from this, LR has been adopted by

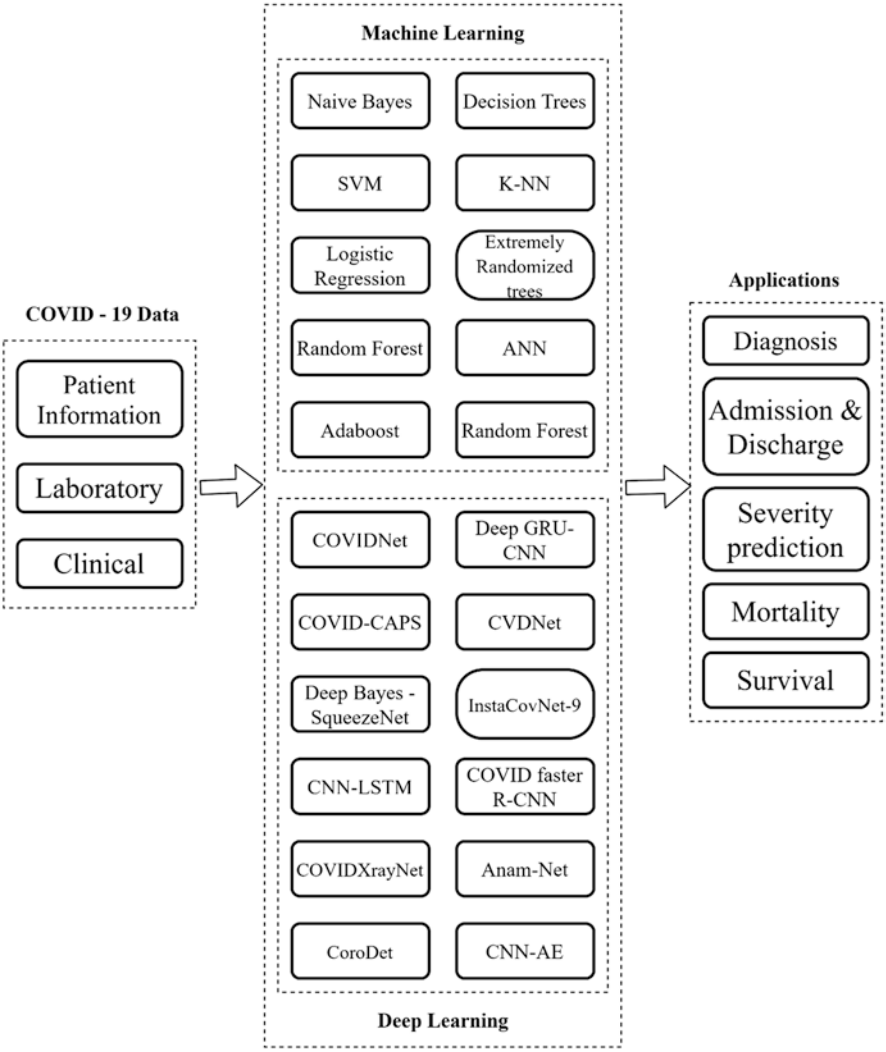

Fig. 3 Machine learning and deep learning approaches used for COVID-19 applications

many for the early prediction of COVID patients' mortality rate. Hu et al. [44] considered more than ten algorithms to find the mortality rate based on the data derived from 183 patients. Finally, they concluded that LR, partial least square (PLS) regression, elastic net, RF, and FDA yielded the best results by comparing the percentages of AUC. An interesting fact is that most of the non-survivors belong to the male gender and are older. From the results obtained by [50, 51], the proposed predictive tools suggest a few important factors that cause severe illness to coronavirus-infected patients. They are elevated heart rate, chronic obstructive pulmonary disease (COPD), comorbidities, respiration rate, CRP, partial saturation of oxygen (SpO_2), and LDH.

Table 6 ML and DL in COVID-19 diagnosis

Source	Sample size	Features	Algorithms/methodology	Performance
Chest CT image [37]	Training/ validation: 106	Shape features, gray level size zone matrix (GLSZM), gray level dependence matrix (GLDM)	Feature selection: Maximum relevance minimum redundancy (MRMR)	AUC: 0.95 ± 0.029
	Testing: 46			Accuracy = 88% ± 4.6 (95% CI: 0.88–0.89)
				Sensitivity = 88% ± 6.6 (95% CI = 0.87–0.9)
			Classification: XGBoost	Specificity = 89% ± 7 (95% CI = 0.87–0.9)
Chest CT images [38]	Testing: 322	Visual features	COVID-19 detection neural network (COVNet)	AUC: 0.96
	Training: 3000			Sensitivity: 90% [95% CI: 83%, 94%]
				Specificity = 96% [95% CI: 93%, 98%]
Routine blood tests [39]	600 samples	Leukocytes, platelets, monocytes, eosinophils, and patient age quantile	DNN, CNN, RNN, LSTM	Accuracy: 92.5%
				Precision: 86.96%
				Sensitivity: 100%
				Specificity:85%
				F1-score: 92.20%
Clinical and demographic data [40]	628 samples	Noninvasive features: S_PO_2, age, and cardiovascular disorders	Support vector machine-recursive feature elimination (SVM-RFE)	AUC: 0.92 ± 0.03
				Sensitivity: 0.81 ± 0.08
				Specificity: 0.91 ± 0.05
Chest X-ray images [41]	Pneumonia infected: 1628	NA	Ensemble deep convolutional neural network model	Precision: 98.33%
				Recall: 98.33%
	COVID-19-positive patients: 1628			Accuracy: 99.71%
	Healthy patients: 2345			
Chest X-ray images [42]	Normal: 1583	NA	Deep Bayes-SqueezeNet	Correctness: 98.26%
	Pneumonia: 4290 COVID-19: 76			Recall: 98.25%
				Accuracy: 98.26%
				F1-score:98.25%

Source	Sample size	Features	Algorithms/methodology	Performance
Routine blood tests [43]	235 samples	Lymphocytes, leukocytes, and eosinophils	SVM	AUC: 85%
				Sensitivity: 68%
				Specificity: 85%
Patient data [44]	183 samples	Demographic, clinical, and laboratory data	Linear regression (LR), partial least square (PLS) regression, elastic net model (EN), flexible discriminant analysis (FDA)	AUC: 89.5%
				Sensitivity: 89.2%
				Specificity: 68.7%
Patient information [45]	43,752 samples	Age, gender, prior medical conditions, smoking habits, fever, sore throat, cough, shortness of breath, loss of taste or smell	Linear regression (LR)	AUC: 73.7%
CT and patient data [46]	413 samples	Age, CT scan result, temperature, lymphocyte, fever, coughing	XGBoost	Sensitivity: 92.5%
				Specificity: 97.9%
Patient information [47]	3524 samples	Sex, age	LR	AUC: 83%

Multiple algorithms are also considered for COVID data analytics by the researchers for the early prediction of disease diagnosis and the survival rate of the patients [38, 52]. Many methods have been deployed with a set of COVID data to predict the results. Each of their results ends up with certain predictions, cannot be taken as a universal solution for the same cause.

Deep Learning-Based Predictive Tools

Generally, the predictions can be performed through statistical approach and medical data analytics. As the number of COVID morbidity, diseased, recovered, and death rates is increasing day by day across the countries, the traditional methods such as mathematical modeling and statistical-based approaches are not encouraging to deal in a huge volume. Hence, the recent approach like deep neural network-based models is being imposed by many researchers. Many deeper neural network models with optimizers are adopted for COVID applications.

Karthik, R et al. [34] proposed a pixel-based upsampling module that encompasses an encoder to segment the chest CT images. This model provides a dice score of 85.43% and a recall of 88.10% for a dataset containing 50 samples of 3D CT scans chosen selectively from 1110 studies. A fuzzy rank-based algorithm along with the CNN model has been used to classify COVID and non-COVID cases from the CT scan images [53]. To rank the decision scores, three neural network models, namely, Visual Geometry Group (VGG), Wide ResNet, and Inception, were utilized. Optimized data augmentation and hyperparameter tuning techniques associated with CovidXrayNet yield 95.82% accuracy with only 30 epochs of training in finding the COVID disease classification [54].

A dynamic CNN model constructed from the modification of MobileNet and ResNet finds COVID-19 features from both chest CT and X-ray images. It has provided 99.6% test accuracy [55]. Anamorphic depth embedding-based lightweight CNN (Anam-Net) is a lightweight CNN [56] to find the anomalies from COVID-19 chest CT images. It extracts very few parameters to attain its task. This model has been implemented in Raspberry Pi, and other mobile-based Android applications to make it a point of care solution. Many COVID research articles explore the implications of deeper neural networks (DNN) on 3D chest images. To name a few, 3D DNNs are COVID-Nets, CNNBERT, LuMIRa, Telefax Library Network (TELINET), etc.

Summary

In this chapter, a complete overview of pandemic and epidemic diseases, disease transmission ways, and various pathogens evolved across the globe at different instances was discussed. The pandemic data can be analyzed through some statistical tools or mathematical models to predict the spread of disease and mortality and morbidity rates. During the mid-nineteenth century, mathematical models were imposed to know the severity of infectious diseases. During the interpretation by mathematical modeling, data inconsistency, nonavailability of global data, and lack of clinical and nonclinical data dropped the efficiency of the predicted results. Due to its simplicity, understanding of the data had been done through various assumptions. Constructing the mathematical model and execution time to arrive at the prediction results were not cost-effective. Data preprocessing and the removal of outliers are time-consuming processes. To overcome these limitations, machine learning-based algorithms have been adopted for various findings like severely infected person recovery rate, ICU admission, discharge rate, demographic widespread rate, etc. Such results helped both private and public sectors to take necessary steps before announcing night curfews and lockdowns across the country or part of it. Nowadays, the predictive peaks in data visualization alert the government to plan for healthcare establishments, procurement of drugs, and recurring and nonrecurring medical facilities.

The analytics on COVID data supports the governments to know the number of citizens who have taken up first, second, and booster shots (if required). It also helped the pharmaceutical companies to know their vaccine efficacy during drug trials. The major factor to be considered while imparting machine learning-based predictive tools is data inconsistency due to variants of pathogens. During this COVID pandemic, almost 42 crore people were infected and around 58 lakhs were deceased. This is the highest pandemic infection rate after the French flu. This data is publicly available and announced by various governments daily. Apart from that, many have not been revealed to the outside world. Still, data collection is a crucial part of medical data analytics and predictions. To handle billions and billions of CT chest data, patient information, and blood-related features, in which machine learning algorithms were not capable of handling in a large volume, many researchers are proposing deeper neural networks to handle high-dimensional data. Still, a benchmark network to handle heterogeneous medical data is a golden opportunity for data analysts to explore. However, during this pandemic situation, the public is being advised to follow all safety procedures and protocols like social distancing, sanitizing the hands often, wearing masks, and vaccinating to help everyone to "stay safe."

Acknowledgments The authors would like to sincerely thank reviewers for their thoughtful comments and efforts toward improving this chapter. Also, they would like to extend their gratitude to the editors for considering this chapter and including it in the book.

References

1. M. Porta (ed.), *Natural history of disease. A Dictionary of Epidemiology*, 5th edn. (Oxford University Press, 2014), pp. 193–194
2. L. Shaw-Taylor, An introduction to the history of infectious diseases, epidemics and the early phases of the long-run decline in mortality. Econ. Hist. Rev. **73**(3), E1–E19 (2020)

3. J.F. Lindahl, D. Grace, The consequences of human actions on risks for infectious diseases: A review. Infect. Ecol. Epidemiol. **5**, 30048 (2015). https://doi.org/10.3402/iee.v5.30048
4. D. Grennan, What is a pandemic? JAMA **321**(9), 910 (2019). https://doi.org/10.1001/jama.2019.0700
5. J. Piret, G. Boivin, Pandemics throughout history. Front. Microbiol. **11**(631736), 1–16 (2021). https://doi.org/10.3389/fmicb.2020.631736
6. I. Ansari, G. Grier, M. Byers, Deliberate release: plague – A review. J. Biosaf. Biosecur. **2**, 10–22 (2020). https://doi.org/10.1016/j.jobb.2020.02.001
7. F.R. Chowdhury, Z. Nur, N. Hassan, L. von Seidlein, S. Dunachie, Pandemics, pathogenicity and changing molecular epidemiology of cholera in the era of global warming. Ann. Clin. Microbiol. Antimicrob. **16**, 10 (2017). https://doi.org/10.1186/s12941-017-0185-1
8. J.H. Beigel, F.G. Hayden, Influenza therapeutics in clinical practice-challenges and recent advances. Cold Spring Harb. Perspect. Med. **a038463** (2020). https://doi.org/10.1101/cshperspect.a038463
9. M.A. AlBalwi, A. Khan, M. AlDrees, U. Gk, B. Manie, Y. Arabi, et al., Evolving sequence mutations in the middle east respiratory syndrome coronavirus (MERS-CoV). J. Infect. Public Health **13**, 1544–1550 (2020). https://doi.org/10.1016/j.jiph.2020.06.030
10. A. El-Sayed, M. Kamel, Climatic changes and their role in emergence and re-emergence of diseases. Environ. Sci. Pollut. Res. **27**, 22336–22352 (2002). https://doi.org/10.1007/s11356-020-08896-w
11. M.A. Kadhim, M.A. Alam, H. Kaur, Design and implementation of fuzzy expert system of back pain diagnosis, international journal of innovative technology & creative. Engineering **1**(9), 16–22 (2011)
12. T. Di Noia, V.C. Ostuni, F. Pesce, G. Binetti, D. Naso, F.P. Schena, E. Di Sciascio, An end stage kidney disease predictor based on an artificial neural networks ensemble. Expert Syst. Appl. **40**(11), 4438–4445 (2013)
13. J.C. Boyd, Mathematical tools for demonstrating the clinical usefulness of biochemical markers. Scand. J. Clin. Lab. Invest. **57**(sup227), 46–63 (1997)
14. B.F. Finkenstädt, B.T. Grenfell, Time series modeling of childhood diseases: A dynamical systems approach. J. R. Stat. Soc.: Ser. C: Appl. Stat. **49**(2), 187–205 (2000)
15. M.I. Jordan, T.M. Mitchell, Machine learning: Trends, perspectives, and prospects. Science **349**(6245), 255–260 (2015)
16. H.A. Karaboga, A. Gunel, S.V. Korkut, I. Demir, R. Celik, Bayesian network as a decision tool for predicting ALS disease. Brain Sci. **11**(2), 150 (2021)
17. D. Akila, D. Balaganesh, Semantic web-based critical healthcare system using Bayesian networks. Mater. Today Proc. (2021). https://doi.org/10.1016/j.matpr.2021.01.870
18. V. Jackins, S. Vimal, M. Kaliappan, M.Y. Lee, AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. J. Supercomput. **77**(5), 5198–5219 (2021)
19. K. Ganji, S. Parimi, ANN model for users' perception on IOT based smart healthcare monitoring devices and its impact with the effect of COVID 19. J. Sci. Technol. Policy Manag. **13**(1), 6–21 (2021)
20. W. Yu, T. Liu, R. Valdez, M. Gwinn, M.J. Khoury, Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes. BMC Med. Inform. Decis. Mak. **10**(1), 1–7 (2010)
21. Q. Jiang, G. Wang, S. Jin, Y. Li, Y. Wang, Predicting human microRNA-disease associations based on support vector machines. Int. J. Data Min. Bioinform. **8**(3), 282–293 (2013)
22. N.Z. Supardi, M.Y. Mashor, N.H. Harun, F.A. Bakri, R. Hassan, Classification of blasts in acute leukemia blood samples using k-nearest neighbor, in *2012 IEEE 8th International Colloquium on Signal Processing and Its Applications* (2012), pp. 461–465.
23. L. Prompers, N. Schaper, J. Apelqvist, M. Edmonds, E. Jude, D. Mauricio, L. Uccioli, V. Urbancic, K. Bakker, P. Holstein, A. Jirkovska, Prediction of outcome in individuals with diabetic foot ulcers: Focus on the differences between individuals with and without peripheral arterial disease. The EURODIALE study. Diabetologia **51**(5), 747–755 (2008)

24. D. Zhang, D. Shen, Alzheimer's disease neuroimaging initiative. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage **59**(2), 895–907 (2012)
25. X.D. Zhang, Machine learning, in *A matrix algebra approach to artificial intelligence*, (Springer, Singapore, 2020), pp. 223–440
26. P.S. Kohli, S. Arora, Application of machine learning in disease prediction, in *2018 4th International conference on computing communication and automation* (*ICCCA*) (IEEE, 2018), pp. 1–4.
27. K. Indhumathi, K.S. Kumar, A review on prediction of seasonal diseases based on climate change using big data. Mater. Today Proc. **37**, 2648–2652 (2021)
28. A. Kumar, T.V. Kumar, Multi-objective big data view materialization using MOGA. Int. J. Appl. Metaheuristic Comput. **13**(1), 1–28 (2022)
29. R. Chauhan, H. Kaur, V. Chang, An optimized integrated framework of big data analytics managing security and privacy in healthcare data. Wirel. Pers. Commun. **117**(1), 87–108 (2021)
30. O.A. Sarumi, C.K. Leung, Adaptive machine learning algorithm and analytics of big genomic data for gene prediction, in *Tracking and Preventing Diseases with Artificial Intelligence*, (Springer, Cham, 2022), pp. 103–123
31. N. Johanna, H. Citrawijaya, G. Wangge, Mass screening vs lockdown vs combination of both to control COVID-19: A systematic review. J. Public Health Res. **9**(4), 523–531 (2020)
32. P. Schwab, A.D. Schütte, B. Dietz, S. Bauer, Predcovid-19: A systematic study of clinical predictive models for coronavirus disease 2019. arXiv preprint arXiv:2005.08302 (2020)
33. J. Vijay, K. Nirmala, S.G. Sarate, Chapter 5: Non-contact measurement system for COVID-19 vital signs to aid mass screening – An alternate approach, in *Cyber-Physical Systems*, ed. by R. C. Poonia, B. Agarwal, S. Kumar, M. S. Khan, G. Marques, J. Nayak, (Academic Press, 2022), pp. 75–92
34. R.M. Karthik, M. Hariharan, D. Won, Contour-enhanced attention CNN for CT-based COVID-19 segmentation. Pattern Recogn. **125**, 108538 (2022)
35. L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, J. Xia, Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy. Radiology **296**(2), E65–E71 (2020). https://doi.org/10.1148/radiol.2020200905
36. M. Kukar, G. Guncar, T. Vovko, S. Podnar, P. Cernelc, M. Brvar, M. Zalaznik, M. Notar, S. Moskon, M. Notar, COVID-19 diagnosis by routine blood tests using machine learning. Sci. Rep. **11**, 10738 (2020)
37. I. Shiri, M. Sorouri, P. Geramifar, M. Nazari, M. Abdollahi, Y. Salimi, B. Khosravi, D. Askari, L. Aghaghazvini, G. Hajianfar, A. Kasaeian, H. Abdollahi, H. Arabi, A. Rahmim, A.R. Radmard, H. Zaidi, Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients. Comput. Biol. Med. **132**, 104304 (2021)
38. S. Li, Y. Lin, T. Zhu, M. Fan, S. Xu, W. Qiu, C. Chen, L. Li, Y. Wang, J. Yan, et al., Development and external evaluation of prediction models for mortality of covid-19 patients using the machine learning method. Neural Comput. Applic., 1–10 (2020)
39. S.B. Rikan, A.S. Azar, A. Ghafari, J.B. Mohasefi, H. Pirnejad, COVID-19 diagnosis from routine blood tests using artificial intelligence techniques. Biomed. Signal Process. Control **72**, 103263 (2021). https://doi.org/10.1016/j.bspc.2021.103263
40. M. Mahdavi, H. Choubdar, E. Zabeh, M. Rieder, S. Safavi-Naeini, Z. Jobbagy, A. Ghorbani, A. Abedini, A. Kiani, V. Khanlarzadeh, R. Lashgari, E. Kamrani, A machine learning-based exploration of COVID-19 mortality risk. PLoS One **16**(7), e0252384 (2021). https://doi.org/10.1371/journal.pone.0252384
41. P. Kedia, R.K. Anjum, CoVNet-19: A deep learning model for the detection and analysis of COVID-19 patients. Appl. Soft Comput. **104**, 107184 (2021)

42. F. Ucar, D. Korkmaz, COVIDiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease (COVID-19) from X-ray images. Med. Hypotheses **140**(2020), 109761 (2019)
43. A. F. De Moraes Batista, J.L. Miraglia, T.H.R. Donato, A.D.P. Chiavegatto Filho, COVID-19 diagnosis prediction in emergency care patients: A machine learning approach.medRx (2020). https://doi.org/10.1101/2020.04.04.2005209
44. C. Hu, Z. Liu, Y. Jiang, O. Shi, X. Zhang, K. Xu, et al., Early prediction of mortality risk among patients with severe COVID-19, using machine learning. Int. J. Epidemiol. **49**(6), 1918–1929 (2020)
45. S. Shoer, T. Karady, A. Keshet, S. Shilo, H. Rossman, A. Gavrieli, et al., A prediction model to prioritize individuals for sars-cov-2 test built from national symptom surveys. Med **2**(2), 196–208 (2020). https://doi.org/10.1016/j.medj.2020.10.002
46. W.T. Li, J. Ma, N. Shende, G. Castaneda, J. Chakladar, J.C. Tsai, L. Apostol, C.O. Honda, J. Xu, L.M. Wong, T. Zhang, A. Lee, A. Gnanasekar, T.K. Honda, S.Z. Kuo, M.A. Yu, E.Y. Chang, M.R. Rajasekaran, W.M. Ongkeko, Using machine learning of clinical data to diagnose covid-19: A systematic review and meta-analysis. BMC Med. Inform. Decis. Mak. **20**, 247 (2020). https://doi.org/10.1186/s12911-020-01266-z
47. A.K. Das, S. Mishra, S.S. Gopalan, Predicting covid-19 community mortality risk using machine learning and development of an online prognostic tool. PeerJ **8**, e10083 (2020)
48. R.P. Joshi, V. Pejaver, N.E. Hammarlund, H. Sung, S.K. Lee, A. Furmanchuk, H.-Y. Lee, G. Scott, S. Gombar, N. Shah, S. Shen, A. Nassiri, D. Schneider, F.S. Ahmad, D. Liebovitz, A. Kho, S. Mooney, B.A. Pinsky, N. Banaei, A predictive tool for identification of sars-cov-2 PCR-negative emergency department patients using routine test results. J. Clin. Virol. **129**, 104502 (2020). https://doi.org/10.1016/j.jcv.2020.104502
49. M. Tordjman, A. Mekki, R.D. Mali, I. Saab, G. Chassagnon, E. Guillo, R. Burns, D. Eshagh, S. Beaune, G. Madelin, et al., Pre-test probability for sars-cov-2-related infection score: The Paris score. PLoS One **15**(12), e0243342 (2020). https://doi.org/10.1371/journal.pone.0243342
50. Z. Zhao, A. Chen, W. Hou, J.M. Graham, H. Li, P.S. Richman, H.C. Thode, A.J. Singer, T.Q. Duong, Prediction model and risk scores of ICU admission and mortality in covid-19. PLoS One **15**, e0236618 (2020)
51. H. Huang, S. Cai, Y. Li, Y. Li, Y. Fan, L. Li, C. Lei, X. Tang, F. Hu, F. Li, X. Deng, Prognostic factors for covid-19 pneumonia progression to severe symptoms based on earlier clinical features: A retrospective analysis. Front. Med. **7**, 643 (2020). https://doi.org/10.3389/fmed.2020.557453
52. M. Nemati, J. Ansary, N. Nemati, Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data. Patterns **1**, 100074 (2020)
53. R. Kundu, H. Basak, P.K. Singh, et al., Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans. Sci. Rep. **11**, 14133 (2021). https://doi.org/10.1038/s41598-021-93658-y
54. M.A. Maram, J.P. Monshi, V. Chung, F.M. Monshi, CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. Comput. Biol. Med. **133**, 104375 (2021)
55. G. Jia, H.-K. Lam, X. Yujia, Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method. Comput. Biol. Med. **134**, 104425 (2021)
56. N. Paluru, A. Dayal, H.B. Jenssen, T. Sakinis, L.R. Cenkeramaddi, J. Prakash, P.K. Yalavarthy, Anam-net: Anamorphic depth embedding-based lightweight CNN for segmentation of anomalies in COVID-19 chest CT images. IEEE Trans. Neural Netw. Learn. Syst. **32**(3), 932–946 (2021). https://doi.org/10.1109/TNNLS.2021.3054746