The Research Monograph Series in Computing, Electrical & Communication Networks
Vol. 1 (02), February 2024, pp. 9–26

2 REVIEW OF LITERATURE

The literature survey pertaining to the study is discussed in two sections. First section explains about the contagious diseases and pandemic studies, and the second section explains the studies related to prediction models.

An exhaustive survey has been done to study, understand, analyze and propose the early warning system/procedure for dengue and COVID-19 using machine learning and the fine tuning of this work has been done using deep learning algorithms.

In this section, the articles related to the proposed model have been reviewed. Hybrid Machine Learning approach is applied for predicted the number of infections and deaths using Multi Layered Perceptron and Imperialist Competitive Calculation and it was based on the data set of the country Hungary (Cao et al., 2020).

To design, analyse, evaluate and understand the impact of the pandemic or spreading of infectious diseases without any specific availability of antiviral doses and vaccination there is a huge requirement of predictive mathematical models which plays a vital role (CIESIN 2018, Cascella et al., 2019, Centers for Disease Control and Prevention 2019, Castillo-Chavez et al., 2015, Cauchemez et al., 2009). The impact of isolation strategy was studied by applying stochastic model of SIQR and simulated the dynamics of the virus (Ceylan 2020). The evolution of COVID-19 has been analysed by computing the fractional order dynamical system on a assumed population (Chakraborty et al., 2014) and the computation system is based on the discretisation of the domain and the short memory principle. In the work, the population is divided into subgroups like susceptible, exposed, infected, deceased and the behaviour of these groups are analysed over a period of time (Chan et al., 2020). The authors have also worked on fractional calculus to study the impact of Covid-19 and applied factional natural decomposition technique to understand the dynamic behaviour of the virus. The proposed work involves highly complicated mathematical computations when compared with machine learning approaches.

Non pharmaceutical intervention approach has been applied to reduce the spread of the disease, a collection of population is divided into six groups on which the mathematical model is built by coupled differential equations (Chen et al., 2020) and the principal objective of this work is

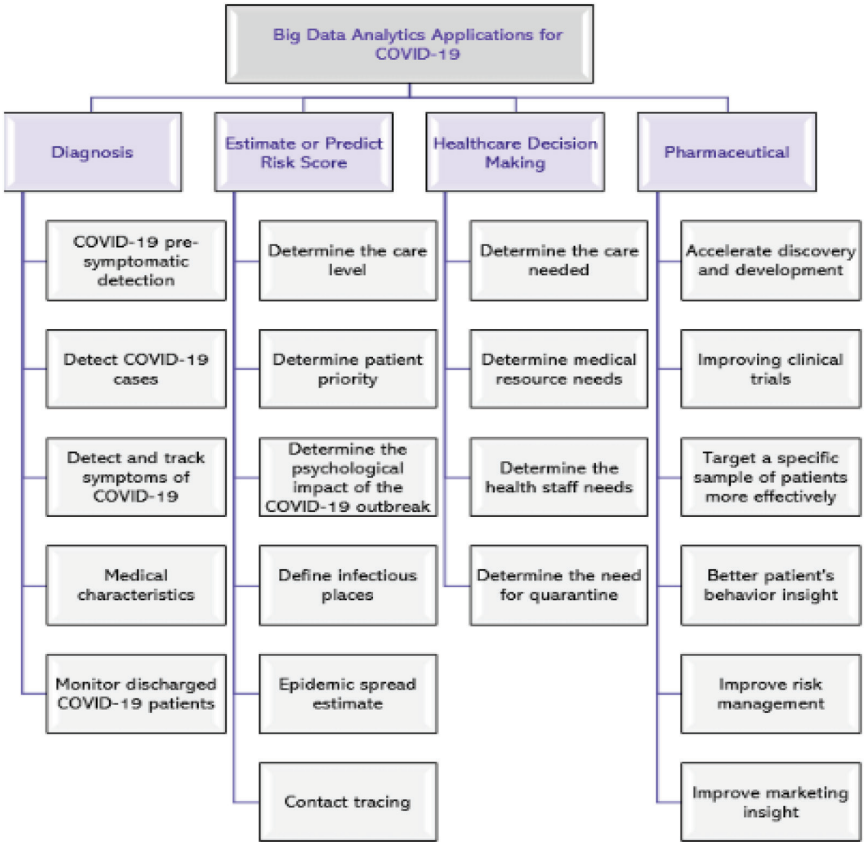

Figure 2.1. Potential application areas of big data analytics for COVID-19.

only to reduce the impact of the virus, whereas the research article assists in predicting the spread of the virus.

Researchers also focused on investing the about the possibility of reducing the spread of the virus without lockdown (Chen 2015) in which quarantine plays a vital role and the methodology applied is differential equations. Several deep learning approaches also has been proposed in order to forecast and predict the impact / spread of the virus across different global regions. Studies also discusses about the containment of the virus (Chen et al., 2017).

2.1 CONTAGIOUS DISEASES AND PANDEMIC STUDIES

Recent years have seen at least six large-scale outbreaks — Hantavirus pulmonary syndrome, severe acute respiratory syndrome, H5N1

influenza, and Ebola (Gostin et al., 2016). Infact, H1N1 can be considered as the first pandemic of this twenty first century. Nearly eighteen thousand people lost their lives due to this Pandemic (Rewar et al., 2015).

Ebola has a different mortality rate, where in ten days from the day of infection most of them reach the critical terminal stage as per WHO data of 2003. Nearly six hundred infections with a mortality rate of thirty percent were observed during H7N9 (Su and He, 2015).

Acute Hemorrhagic Conjunctivitis (AHC) occurred in 1981 has lower severity, this was also declared as Pandemic because of its explosive nature (Donaldson et al., 2009).

Maurice in 2016 stated that as per the world band statistics, the impact of Ebola on economy was huge with around two billion dollars of financial loss and nearly eleven thousand lives lost. Zika is another virus that is spread of thirty four nations as per Troncoso, 2016 analysis. All these Pandemics remember and recollect the damage caused by Spanish flue in 1918 (Lin et al., 2016).

The first Coronavirus case was reported in 2019 December at City of Wuhan. In short time, it has become a Pandemic that world ever witnessed since Spanish flu. It has changed its ugly face into a major or severe Pandemic that has created human economic impact globally and loss of numerous valuable lives. Near about 106.1 crores of infected cases reported globally and nearly twenty three lakh people were put down by this Pandemic in USA, Brazil, United Kingdom, Russia and India. There was drastic change in scenario in India alone, where in 2021 there were more than one crore cases and the number of people lost their lives counted to approximately more than one and half lakh (Focà et al., 2020).

Animals and humans connectivity enhances the risk of epidemic or pandemic. Often animals are spreaders to humans, when it comes to most of the viruses. Bats, cows and other pet animals spread virus to humans. Human to human spread then becomes more fast and easy. This infectious disease spread data is available with scientists from last few decades, the computerization of this data, analysis and inferences are drawn for the benefit of the future research. But this data is seldom useful in predicting the outburst of future epidemic/pandemic. This data may not be always useful to prevent the spread of Pandemic or in designing effective containment measures. The reason is most obvious, the nature of the Pandemic is explosive and the impact is immeasurable in the initial phase (Jia et al., 2019, Joachims 2002).

When it comes to Coronavirus (CoV), it is a positive stranded RNA. This is a respiratory virus. The name is so given due to crown shaped tips on its surface. When severity is high, Coronavirus infection can become more intense and transforms into pneumonia. Patients with the history of

acute respiratory diseases are prone to the intense level of infection, with the mild infection itself the cases moves from mild to intense phase in no time. The cause of death in most of the cases is multiple organ failure (Chen et al., 2011, Kaggle 2020).

Bio-scientists and researchers across the globe did extensive research and the process is ongoing as far as the remedy is concerned. The fusion of virus cells with lung cells namely Alivio shall be prevented by the ACE2. This enzyme contains a transmembrane that prevents this fusion (Zhonghua Jie He et al., 2020).

A fifty four year old Korean living in Wuhan entered in Korea on 20th January, 2020. First hand of symptoms are chillness and cramps observed after two days of entry. He was isolated in Myongji hospital and tested positive for coronavirus infection on twenty sixth January, 2020. The confirmation was done by swab test. By the time, unfortunately the man transmitted virus to his friend in a restaurant. This incident happened on the first day of symptoms to this old man. His friend became a carrier and transmitted it to his family members. A person in a church also got infected and all of them tested positive in the first week of February, 2020 (Khajanchi Subhas and Sarkar Kankan 2006). These are the first set of infection spread outside the geography of China due to coronavirus [Jaegyunlim 1].

According to Cortegiani et al. (2020), no specific dosage of apt medicine is prescribed clinically yet for Coronavirus treatment. Some nations preferred using Chloroquine, which is effective in few cases but has an advantage of less cost (Khozeimeh et al., 2021). Preclinical studies have proven the effectiveness of this drug but long-term results are due. Tropical conditions, patient medical history and other prevalent conditions are hampering the usage of drug to severe to risk cases. In china also such clinical trials are in progress, high quality statistical inferences and data is most sought out across the globe in the treatment of COVID-19.

In 2015 Carlos Castillo suggested learning outcomes of Ebola Pandemic (Kishor Sonti and Sundari, 2021). These may useful in mitigating the further Pandemic eruptions. The preventive cost always more than management cost in the case of such Pandemic diseases. For example, Vaccination drive costs huge amount of money than offering Covid management services to the public.

In a recent research Wu (2020) and his team of colleagues analyzed the risk factors of respiratory syndromes. As per their mention, the deaths due to COVID-19 still have a factor of uncertainty. As per their data, a retrospective study of a group of two hundred and one patients infected with COVID-19 was admitted into Jinyintan Hospital in Wuhan between

25th December 2019 and 26th January, 2020. The observed data of final follow-up was mentioned as 13th February, 2020 (Lanata et al., 2015). The epidemiological study, clinical and other laboratories details confirmed the following results after extensive analysis. 52.4% of patients died out of 201. Most of them have comorbid issues (Lagier et al., 2020). The key inference drawn out of this extensive clinical observation is patients with higher age are at more risk. Patients with less immune system are also affected more than healthy ones that happens naturally.

The initial studies pertinent to COVID-19 Pandemic naturally took place in China as it was the epicenter. Yang X and his team observed the patients affected with COVID-19 in Wuhan and made key observations in 2020 (Yang et al., 2020). SARS-CoV-2 pneumonia observed in patients admitted into Jin Yin-tan hospital of Wuhan, China. Fifty two seriously infected patients of COVID-19 were observed. This observation took place during December 2019 to 26th January, 2020. Factors such as age, medical data, comorbid issues, and clinical values are observed. A comparative study was carried out between the recovered and deceased. Acute respiratory distress syndrome (ARDS) was observed with patients' requirement to artificial breathing i.e. in general terms Ventilator support. Twenty eight day mortality study is another key outcome (Lee et al., 2004).

The observations include patients with median age of 59.7. 61.5% of the patients died at twenty eighth day. 67% of patients observed to have ARDS and the death was miserable. 15% of patients suffered with kidney malfunction and rest has issues such as Cardiac arrest. 5% of the patients required long term hospitalization. The key observation of this study was COVID-19 has considerable mortality rate. The interpretations include (Lee et al., 2004), survival time is one to two weeks after onset of symptoms. Patients more than sixty five years and ARDS are more prone to the risk of death. COVID-19 is going to create an impact on health infrastructure of society.

Yang et al. (2020) suggested that SARS CoV 2 is the cause of the new infection named COVID-19 (Yang et al., 2020). The initial study reveals that around 18K people died due to COVID-19 in China. Most of the ill-fated belongs to comorbidity. They also stated that this is the third intense outbreak of Coronavirus in last two decades. SARS has created impact in 2002, 2003 and it is now reappeared in most intense form as COVID-19. The set of symptoms included cold to sever illness and death. It is firmly believed that this strain emerged out of wild life market of Hubei Province of Wuhan city and transmitted to humans. Bats, Pangolins and reptiles are stated as the primary reason of transmission. The group also suggested severe isolation procedures, home isolation to cut the chain of transmission (Lee 2019). This is in resemblance with history witnessed

most rude attack on humans, the Spanish flu. Closing national borders, restricted air travel, strict measures of mobility reduction are suggested. Research is also suggested for finding appropriate vaccine. Antiviral drugs are to be made ready to contain the spread. This study also observed the global impact on economy and health care systems. They also inferred that his sudden change will have psychological impact on citizens across the globe and may induce racial attacks on Asians.

In the same year, Yao et al. (2020) proposed the use of hydroxychloroquine in the treatment of COVID-19. The background of this study is same, the origin of COVID-19 leading to Wuhan again. Chloroquine was used earlier in treating SARS family of viruses. It is stated that in the initial days (Lei et al., 2018, 2020), Chloroquine was the natural choice of medicos in treating this virus. This drug infact was used to treat malaria and other mosquito bite related infections. This was also observed to be a safe medication for autoimmune situations. They stated in their research that Hydroxychloroquine may be more useful than Chloroquine. The reason stated is the immuno-modulatory property of the later drug (Li et al., 2008, Li 2011). This new suggestion also fruitful in tackling the cytokine effect, which is observed in the critical illness conditions of the patients. As per their report, substantial clinical evidence is still to obtain as a support of their study. Extensive clinical studies suggested that Hydroxychloroquine was found more effective than Chloroquine in treating new strain of SARS CoV 2.

Yin-Wong et al. (1998) stated that SARS and its family of coronaviruses are the prime reason for Pneumonia (Yin and Wunderink 2018, Letko et al., 2020). They reported that human coronavirus is considered to be less harmful as per earlier studies. But the outbreak of Middle East Respiratory Syndrome (MERS) and SARS suggested that the impact of human coronavirus is an important pathogen in respiratory diseases. This review paper focus upon the epidemiology studies, pathology and other clinical observations related to human coronavirus.

Zhao et al. (2020) did a profiling study ACE in 2020. According to this study, WHO reported nearly seventy six thousand coronavirus infections by the third week of February, resulting nearly two thousand five hundred mortality cases. The global presence of this disease is also increasing exponentially. ACE2 RNA profiling is observed in lungs of humans (Li and Xia, 2020). This study throws light on the therapeutic strategy. This investigation provided the needed biological background for further studies.

Zhou et al. (2020) published this article in Lancet. This is a retrospective cohort study. This study is special as the age group of adults and early adulthood are taken for analysis. Patients from famous

hospitals and pulmonology clinics are considered for this study. The sample data includes their demographic, clinical and psychological observations. Electronic Medical Records (EMRs) are compared between the people survived and not survived (Lim et al., 2020). Regression studies carried out to assess the risk factors associated with death rate. The inferences drawn out are interesting. Potential risk factors are once again old age, but the novelty of this study is early prognosis of the diseases. Isolation measures, alternative medicine treatments are the possibilities, if the early diagnosis is done in the patients is one of the outcome of this study.

Influenza was and is the main cause of deaths across the nation from few decades (Soliman et al., 2019). To support the work in this research, we found an interesting work, the team used deep learning algorithms in the prediction of seasonal spread of influenza in Dallas of United States of America (Lin et al., 2020).

The team made use of feed forward neural networks blended with statistical models. Regression, Average and absolute shrinkage methods with selected operators were used. Bayesian model was also used in the forecast of this influenza virus. In their work, seventy five hidden nodes and two hidden computation layers were used, the deep learning mechanism is arrived using cross validation method. The mean square error obtained is 0.005 (Min Chen 2017). When analyzed the factors, age, geographical background, diabetic history, blood pressure values, Cardio diseases history became crucial in the mortality rate (Masuda and Holme 2013, Lina et al., 2020). Many such methods are presented in the Section 2.2 (Masuda and Holme 2013, Lupia et al., 2020).

2.2 PREDICTION MODELS

The literature suggests deep learning algorithms yields better results. Various surveys provide information related to prediction models. Developed countries already are making use of technological advancements. Developing countries such as India are marching towards enjoying the right fruits of technology. Information technology revolution has helped to major extent. Number of cases reported, days to day analysis, data sets are supplied by various agencies. Most of the data is empirical (Loeffelholz and Tang 2020).

This literature survey is exhaustive and concentric on medical data analysis models and their efficiency. So in advance to the actual discussion of specific prediction models presentation, a detailed mention about necessity, classification and efficacy related factors are presented (Mills et al., 2015, Yin-Wong et al., 1998).

Some of the parameters such as health history, population, climate, financial commitments and academic background play vital role in containing and spreading the diseases of Pandemic category. Mathematical models, epidemiology studies proven that ambient temperature plays a greater part in spreading the disease. Further research on this aspect has divergent opinion (Bhandari et al., 2009, 2010, Bhardwaj et al., 2017, Benvenuto et al., 2020, Perotte et al., 2015).

COVID-19 predicting models rallied around population size/age, temperature and seasonal variations. Various deep learning algorithms such as Convolutional neural network (CNN), Recurrent neural networks (RNN), Bidirectional recurrent neural networks (BRNN), Long short-term memory (LSTM) and Bidirectional Long Short-Term Memory Networks (BLSTM) (Liu et al., 2020, Nemati et al., 2016, Murugan and Ramachandran 2012).

Body area networks, health gadgets generate huge amount of data every day. This data has to be analysed properly, processed efficiently and should be interpreted logically. This is a big exercise of decision making (Biswas et al., 2020, Lupia et al., 2020). Minimal cost shall be incurred from patients in this process of data handling. Physicians work should be minimized with the tech-assistance from these modern day equipment and algorithms. This is in fact the immediate need of Health care industry with the onset of a Pandemic (Togacar et al., 2020, Wolpert 1992).

Medical, Image processing, Climate prediction, Business management, Natural language processing, Speech recognition and Automobile industry are few applications of deep learning algorithms (Buhmann 2003, Lowen et al., 2007, Mills et al., 2015).

Deep learning has inherent scalability, whereas machine learning offers classification and prediction but not necessarily scalable. In fact, the basics or fundamentals of deep learning derived from the human neural systems. The imitation is successful to the major extent but the journey is still on. The processing units are classified broadly into input, output and hidden layers. At each layer, the efficiency is reciprocated to weights (Watkins 2020). The weights update leads to increased efficacy of the overall system. Each time the weight is update, the iteration count increases. This total process depends on the activation function of the system (Burges et al., 2007, Yin-Wong et al., 1998).

Generally, Linear, sigmoid and rectangular functions are used as activations functions. The basic steps of building such artificial systems are training, predicting and validating. This training is also otherwise known as updating. Expert systems and other decision making models also closely follows the natural human cognitive system (Wu et al., 2018, Mohammad et al., 2019, Peng et al., 2020, Zamir et al., 2021).

Neural networks are the epicentres of deep learning systems/ algorithms. Feed forward, recurrent, self-organizing and modular neural networks are available prominent classifications (Cao et al., 2020). Discussions related to these networks are taken care in the Chapter 3. But for time being the simplest version shall be termed as; feed forward has no loopback, recurrent has loop back.

Kohonen algorithms rally around unsupervised learning, the updating is based on Euclidean distance (Chung 2015). Fragmentation is the central principle in the case of modular networks. Complex networks are decomposed or divided in to smaller autonomous units. Each unit provides an output, final output is the summation of fragments (Cortegiani et al., 2020, Pinter et al., 2020).

Sometimes, deep learning networks are also implemented with sparse encoders. There is a further classification of discriminative and generative, when it comes to Modified Long Short Term Memory networks (Cossock and Zhang 2006, Rahangdale and Raut 2019, Pyle et al., 2017, Won et al., 2020).

In the case of epidemic diseases, generally discriminative approach is followed. The problem formulation is based on prediction. Hidden layer role is significant in such problems. Training in such networks is broadly categorized into supervised, unsupervised and partially supervised (semi-supervised) (Prajapat et al., 2020). Semi-supervised models are generally used in the prediction of epidemics (https://github.com/CSSEGISandData, Liu et al., 2020, Lippi et al., 2020).

There is another classification of networks, namely, memory based and model based. Performance metrics in memory based networks are correlation. In model based the metrics re related to Bayesian models and Semantics (Davies 2013, Rajashree et al., 2017, Phan 2020, Mills et al., 2015, Lu et al., 2020, Lowen et al., 2007).

An interesting report says that 50% of the American population spend their maximum earning in the treatment of diseases. Therefore, health insurance is compulsory in this region. Almost 18% of USA GDP is on healthcare spending, particularly in the treatment of chronic diseases. In Asian region, for example, in China, the death rate due to chronic disease mounts to 86%. To reach out actual or near accuracy, the structured and unstructured data is often combined (Davies 2013, Ling et al., 2020, Wu et al., 2020). In this complex scenario, large geographical area, different climatic conditions, divergent lifestyles of people compel the data scientists to merge the structured and unstructured data to reach holistic view (Ray and Reich 2018, Peng et al., 2020, Murugan et al., 2019, Min Chen et al., 2017). Hybrid models yielded better results as per the observations from the literature (Li et al., 2008, Rochwerg et al., 2017).

The risk prediction is usually carried out by the combination of one or more neural networks models. For example, CNN and multimodal disease risk prediction are combined in a work (De Wit et al., 2020, Rose-John 2018). The observations arrived at reveal the fact that the accuracy is obtained by the combination of structured and text data. The convergence point is arrived with such combination in the case of CNN models.

Noise models, dependent systems models, statistical models are often combined to reach a prediction model. This sometimes also poses challenges in comprising certain performance metrics. But advantages overweigh the cons, while making these hybrid models (Barrelet et al., 2009, 2011).

Contagious diseases prediction models are also categorized into agent based and regression based models. Further to mention, regression based time series, compartmental models and auto regressive models can be considered (Bennardo et al., 2020, Roosa et al., 2020). These models generally do the predictions of point, associative and full scale.

Most important discussion related to data mining algorithms is about the source or origin of the data. This origin includes, historical data, Wikipedia, Climatological tabulations and web based information. In general, the clinical data is classified into medical images, clinical notes, vitals, demography related and laboratory observations. The diagnosis follows into prognosis, auxiliary method and early warning signals (Wu et al., 2018). Data collection from various sources helps in accurate prediction (Cauchemez et al., 2009, Cao et al., 2020, Cascella et al., 2019, Wu et al., 2018).

Health monitoring systems such as Electrocardiogram (ECG), Electroencephalogram (EEG) and Magnetic Resonance Imaging (MRI) generates good amount of medical data of a patient. Sometimes these data are unstructured, in the post-operative conditions, intensive care units also generate the data that is crucial in arriving at the normalcy conditions are restored in a patient or not. Further diagnosis and treatment depends on such information. Data analytics, deep learning algorithms support to arrive at near accurate inferences.

Above all, today health gadgets, Internet of Medical Things (IoMT) revolution is generating huge amount of health data of individuals and cohorts (Centers for Disease Control and Prevention 2019). In clinical decision making and diagnosis, preventive or proactive approach is possible with the careful analysis of this data. In times of epidemic or a Pandemic such data analysis, prediction towards prevention or containment of disease becomes crucial (Ceylan 2020).

Data analytics related to medical data involves many factors. Each factor has got various reasons and sources, in the specific context of a

Pandemic. Selection of factors or features is one of the crucial stages of machine learning. This total framework is generally known as 'Learning to Rank" (Rigutini et al., 2008, Chakraborty et al., 2014).

The feature significance can be well understood with the application of supervised machine learning methods to learning to rank (Chan et al., 2020). These are trained using point wise and pair wise variants of input and output (Chen 2015, Chung et al., 2015, Chen et al., 2011, 2017, 2020, Chin et al., 2020, Zhonghua Jie He et al., 2020).

Improper diagnosis and follow-up treatment is possible, when doctors follow the empirical information based on previous clinical history. This is more evident in the case of Psychological treatments. Attention deficit hyperactivity disorder (ADHD), Depression, Anxiety disorders and Schizophrenia requires proper clinical data (Cortegiani et al., 2020, Rothe et al., 2020, Sahoo et al., 2021).

Big data techniques are useful in the precision based public health models. The success of these models often contributes to the overall health efficiency component in the society. This success also will be reflexive in the proper tracking of the air, water, insect and seasonal diseases (Cossock and Zhang 2006, Shaman and Kandula 2015, Russell et al., 2020, Sahoo et al., 2021).

As stated by Stoddard et.al, many vector based pathogens contribute to the transmission dynamics of an epidemic or pandemic, this is crucial in terms of mobility of humans. Behaviorism plays a major role, mob-behavior is uncontrollable at times (Shaman and Karspeck 2012), long periods of isolation or quarantine results in such strange behavior patterns of humans. This also contributes to the rise in cases (Covid https://github.com/ CSSEGISandData).

Swarm Intelligence based prediction models also paved the way for the novel destinations in computational research. Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) by Kennedy and Eberhart in 1995 derived inspiration from the nature (Shoeibi et al., 2020). Kishor Sonti and Sundari (2020) stated in a book chapter on "Artificial Swarm Intelligence" and its potential in the prediction models. Nature versus Nurture dichotomy was discussed with relevant examples. The discussion also extended to Prediction based models in energy sector as main and health care systems as sub-category (Yang et al., 2020, Zhou et al., 2020).

Dunitz et al. (2015) suggested the prediction model for hyperlactatemia. Sepsis is a state where harmful microorganisms' presence shall be seen in the blood. EMR records of Boston's Beth Israel Deaconess Medical Centre (BIDMC) gave the clue for the analysis and characterization of Sepsis. Different risk categories are identified and at every stage of processing the medical nature of the data is taken into cognizance. The estimates

predicted the possible cardiac arrest, and limb malfunctioning. Over one hundred and forty six patients were taken for this study (Russell et al., 2020). Therefore, designing a prediction model involves robust data acquisition, classification and prognosis based on models.

In Lanata et al. (2015), and team suggested a novel system known as PSYCHE. Communication technologies are assisting the physicians in the treatment of the mental disorders. Proper diagnosis and treatment of mental disorders is possible only with the availability of the accurate data. Wearable electronics has provided the necessary boost to this research (Shrestha and Mahmood 2019). Flexible electronic systems and gadgets evolution has greater influence on the generation of huge medical data (Yu et al., 2019).

Embedded sensors, textile sensors, piezoelectric sensors are making novel inroads in the development of flexible medical electronic systems. Emotional intelligence is the discussion of newer generation. T-shirts displaying the emotions of individuals are designed and available in the commercial market. This is a resultant of the fusion of flexible electronics, artificial intelligence and Internet of Medical Things.

Various researchers worked on learning to rank method. Few of them are presented in this section. Rahangdale and Raut (2019) presented the topic of regularization for feature selection in learning to rank. Learning to rank is a novel topic that attracted the attention of various researchers. This is very closely related to deep learning algorithms. The regularization is achieved by the updating of weights in the neural networks (Shoeibi et al., 2020, Zamir et al., 2021). This is done by constant training of hidden layers. The outcome of the neural network depends on the rate of the training. This regularization is achieved in the recent past with the evolution of learning in rank methods.

Li (2011) presented a paper on "Learning to rank". In this paper, a fundamental treatment has been given to understand the basic problems related to learning to rank. Single vector machine techniques are also discussed.

Crammer and Singer (2002) had written an interesting article on ranking. Rank prediction rule was discussed that assigns a ranch which closest to the true rank. Cossock and Zhang (2006) proposed regression based subset ranking. Discounted cumulated gain was discussed in this work (Shaman et al., 2013). This criterion measured the quality of the entities. This work also necessitates the significance of computationally efficient approaches.

In Li et al. (2008) presented a work on gradient boosting based learning to rank. The realization of the boosting tree algorithm is also discussed in their work. Herbrich (2000) has done a work on ordinal

regression is another step forward towards the use of regression techniques in classifiers (Soliman et al., 2019).

Data retrieval is most tricky. In Joachims (2002) presented a work on the optimization of the search engines. This is based on click through data. A controlled experiment was conducted and the theoretical information was compared. Several training examples were considered in the experimental investigations related to the retrieval of data. Support vector machine approach is also discussed in this work. In Burges et al. (2007) proposed a technique of learning to rank using multivariate cost functions.

This knowledge of learning to rank was extended to the neural based systems in Rigutini et al. (2008) and colleagues. Preference function based approach is used and training was carried out in arriving at the desired result. All this knowledge base is used by Wu K and friends in 2020 in the modeling of COVID-19. This outbreak influence in China and globally was studied. The quantitative documentation of four phases of this outbreak has given good directionality for the studies across the western part of the globe (El Bcheraoui et al., 2020, Singh et al., 2021, Zhao et al., 2020).

Applications, trends and associated issues related big data; deep learning was discussed by Yu et al. (2019). Precision medicine is the most sought out in the present times.

Deep learning techniques driven health data analysis and research enabled the design and implementation of the precision medicine (Enemark 2009). Accurate data prediction, comprehensive data analytics and exhaustive literature survey only yields the desired set of results. Powerful techniques such as big data show the way in this direction, also discussed a holistic view of studies related to deep learning approaches (Soliman et al., 2019, Zhenfeng et al., 2020).

The work carried out by Yang et al. (2015) and friends was the baseline for many recent researchers. They presented this work in 2015 in a journal of computational biology. The fast growth in inference techniques is due to the recent developments in the mathematical modeling. All these trends assisted in the prediction of influenza related epidemics. Even though influenza is a seasonal phenomenon observed in most of the tropical and sub-tropical regions, onset of epidemic rules out all the general cases. In this work, authors discussed about Kalman filter along with susceptible infected recovered model. The work successfully predicted atleast three weeks before epidemic predictions (European Medicines Agency 2020, Stoddard et al., 2009).

Almost a half- decade before Yang et al. (2014) proposed about the modeling, filtering and retrospective forecasting of influenza transformed

epidemic. Markov chains, Kalman filters and Maximum likelihood estimations are used in this work (Togacar et al., 2020). Ensemble filters are also used to predict the peaking of the epidemic with good accuracy.

Spatial heterogeneity factor was emphasized by Yang et al. (2016) in the work on predicting local outbreaks. This work is carried by considering the neighborhoods of city of New York in 2016. Development of appropriate network models, data acquisition and assimilation are used to forecast or predict the influenza in the city (Stoddard et al., 2009).

Lancet published paper on acute illness caused in COVID-19 patients of Wuhan was an observational study. This work was carried out Yang et al. (2020). This comparative study provided primary outcome as twenty eight days cushion period for ADRS affected patients. Second outcome is the clinical data inference suggested that patients with acute respiratory diseases are having the major risk towards death. Seven hundred and ten patients are considered as the sample size for this experiment (Trappenberg 2019). Around 70% of the patients found to require the ventilator support, around 50% of the patients died in the first twenty eight days (Velásquez and Lara 2020). Fever, chillness and muscle pains are the common symptoms observed in the first week of onset of symptoms in the patients. The critical examination of the clinical data resulted in the following interpretations. First, the severity of the disease takes the healthy patients within a week to the intense care units of the hospitals, the risk of ICU admission is more in the patients with age equal and above to sixty five; medical care management becomes cumbersome for the governments and local bodies within months from the eruption of pandemic in a region or nation.

Zhenfeng Lei et al. (2020) suggested the need for frequent use of deep learning algorithms in medical data examination. The knowledge dissemination pertinent to the health care data results in developing the timely preventive mechanisms (Wang et al., 2020, Zuk et al., 2009). This reflects in reaching out to herd immunity soon as well as to develop needed health infrastructure to mitigate the challenges comes with the different phases or waves of a typical Pandemic like COVID-19. This work also suggested the use of Knowledge graphs (KG). Mixed use of KGs and conventional techniques yielded good results.

Zhou et al. (2020) studied influence on COVID-19 in adult hospitalized patients. The potential factors leading to increased mortality rate are studied. This is a retrospective cohort study based on assessing the risk factors associated with the intensity of the disease (Yin and Wunderink 2018, Zamir et al., 2021, Zhonghua Jie He et al., 2020).

Zuk et al. (2009), Watkins (2020) studied the humidity in relation with transmission capacity of the influenza virus. This is a probabilistic approach, a sample ensemble is considered and the studies are carried

(Wang et al., 2020). The performance metrics are identified and measured using the experimental data (Won et al., 2020). It was inferred that climatic conditions play major role in the transmissibility of the Pandemic.

This exhaustive literature survey enabled this research into a potential pathfinder for analyzing different prediction models and in arriving at the model discussed in Chapter 3.

2.3 RESEARCH GAP

The studies pertaining to the contagious diseases transmission, analysis and prediction are vast and ever updating. The information available in the past few decades unveils the fact that there is a predominant research gap exists. Research is much sought out in the area of better prediction models and an attempt has been made in this research to add a meaningful yet effective addition to address this gap.

Kohenen algorithms, regression analysis, SVM classifiers are used in past in estimation and prediction of contagious diseases. But more accurate results shall be arrived using latest computational advancements such as machine learning and deep learning. In fact, computational intelligence is driving the global scientific and technological developments, therefore this approach needs much attention and it has been chosen to address this research gaps.

2.4 PROBLEM FORMULATION

After extensive survey, analysis and understanding the research problem has been formulated. The accurate estimation and prediction of contagious diseases is very important to effectively mitigate the associated challenges in economy, health and in restoring public life. Table 2.1 and the literature survey in this chapter provided an idea and gives necessary thrust to formulate a research problem in arriving at better prediction models.

The effective guidelines are possible only with the proper techniques and tools. This research problem was identified and formulated in the purview of providing potential early warning system with machine learning and deep learning approach for contagious diseases and thereby extension of this work to a typical pandemic, namely, COVID-19. The core of the problem shall be defined as follows:

Identifying influencing factors, analyzing their impact on the spread of contagious diseases such as dengue and extension of this work to a pandemic scenario. Processing raw and authenticated data to estimate and predict the pattern of spread and other associated factors. Proposing an early warning system/procedure to alert the related health authorities.

Table 2.1. Summary of surveys on big data analytics in the healthcare field.

SI. No.	Source	Domain	Key Contribution
1	Abouelmehdi, K. et al., 2017	Healthcare security and privacy	Discussed healthcare data security and privacy issues, and the mechanisms and strategies available for healthcare data privacy, security, and user access.
2	Alex, C.A. et al., 2017	Heart attack prediction and prevention	Identified the uses and technologies of big data analytics in this area, as well as challenges and concerns regarding patient privacy.
3	Mehta, N. et al., 2018	General healthcare	Defined the scope of big data analytics and its applications in healthcare, and provided strategies to overcome its challenges.
4	Shahid, N. et al., 2019	Health care organizational decision-making	Identified the main characteristics and drivers of market up- take of Artificial Neural Networks (ANN) for healthcare-related regulatory decision-making.
5	Mardani, A. et al., 2019	Healthcare and medical problems	Reviewed traditional and fuzzy decision-making methods applied to nine areas of healthcare and medical problems.
6	Bahri, S. et al., 2019	Healthcare sector applications	Discussed the impact of big data on various stakeholders and the challenges.
7	Saheb, T. et al., 2019	IoT and healthcare industry	Identified research trends of the Internet of Things Big Data Analytics model (IoTBDA) in the healthcare industry, and demonstrated the influence of the IoT BDA model on the design, development, and application of IoT-based innovations in healthcare services.

(*Continued*)

Table 2.1. (*Continued*)

SI. No.	Source	Domain	Key Contribution
8	Radcliffe, K. et al., 2019	Medical decision-making	Described the current state of research related to collective intelligence.
9	Palanisamy, V. et al., 2019	Patient-centric healthcare system	Presented several analytical approaches from various stake holders perspectives and reviewed the different big data frameworks in terms of data sources, analytical capability, and application areas. Also, it discussed the impact of big data on improving the healthcare ecosystem.
10	Galetsi, P. et al., 2019	Public health and healthcare organizations	Provided a better understanding for governments and health policymakers about how developing a data-driven strategy could improve public health and the functioning of Sensors 2021, 21, 2282 4 of 27 healthcare organizations and explain the challenges associated with this improvement.
11	Agbehadji, I.E. et al., 2020	COVID-19 detection and contact tracing	Explained the potentials of nature-inspired computing (NIC) models for accurate COVID-19 detection and optimized contact tracing.
12	Wang, J. et al., 2020	COVID-19 medical images	Discussed the role of medical imaging integrated with artificial intelligence (AI) in combating COVID-19.
13	Albahri, O. et al., 2020	COVID-19 medical images detection and classification in terms of evaluation and benchmarking	Highlighted the gaps and challenges, and proposed a detailed methodology for the benchmarking and evaluation of AI techniques used in all COVID-19 medical images classification tasks.

(*Continued*)

Table 2.1. (*Continued*)

SI. No.	Source	Domain	Key Contribution
14	Vaishya, R. et al., 2020	COVID-19 pandemic	Explained the role of AI in fighting pandemics.
15	Schmidt, B.-M. et al., 2020	Data harmonization (DH) and health management decision-making	Collected definitions and concepts of DH and addressed the causal relation between DH and decision-making in health management.
16	Galetsi, P. et al., 2020	Healthcare aspects	Provided an overview of the big data analytics publication dynamics in healthcare and discussed several examples to this field.
17	Salazar-Reyna, R. et al., 2020	Healthcare engineering systems	Synthesized and analyzed publications covering data analytics, big data, data mining, and machine learning in the field of Healthcare Engineering Systems.
18	Khan, Z.F. et al., 2020	Mobile health (m-health)	Explored AI applications and big data analytics to provide insights for users to plan resource use for specific challenges in m-health, and proposed a m-health model based on AI and big data analytics.

2.5 OBJECTIVES OF THE RESEARCH WORK

- To design an Early Warning System (EWS) for prediction and/or identification of onset of epidemic disease(s) in a specific region
- To identify contributing attributes to contagious diseases and to pandemic
- To Collect and Process raw data related to contagious diseases using machine learning and deep learning approach and then extending the study, analysis to pandemic scenario.

Chapter 3 presents the work carried out using machine learning algorithms in the case of dengue and COVID-19.