ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 1 Issue 2, December 2021 / Page No: 188-200
Paper Id: JETA-V1I2P120 / Doi:10.56472/25832646/JETA-V1I2P120

Original Article

Predicting tomorrow's Ailments: How AI/ML Is Transforming Disease Forecasting

Mohit Surender Reddy[1], Manikanth Sarisa[2], Siddharth Konkimalla[3], Sanjay Ramdas Bauskar[4], Hemanth Kumar Gollangi[5], Eswar Prasad Galla[6], Shravan Kumar Rajaram[7]

[1]*Sr Network Engineer, Motorola Solutions, USA.*
[2]*Sr Application Developer, Bank of America, USA.*
[3]*Sr Network Development Engineer, Adobe Inc., USA.*
[4]*Senior Database Administrator, Pharmavite LLC, USA.*
[5]*Associate Consultant, KPMG, USA.*
[6]*Senior System Engineer, Infosys, India.*
[7]*Network Engineer, AT &T, USA.*

Abstract: *The most common of these technologies include artificial intelligence (AI) as well as machine learning (ML), both of which are revolutionizing the healthcare system, especially in disease prediction. Given the emerging data produced from systems in healthcare, EHRs, social media, environment, and genomics, AI and ML algorithms continue to find genuine applications in anticipating disease outbreaks or patients' health futures. Disease forecasting is looked at in this paper with regard to the different methods and algorithms being used in current practice, as well as the possibility that AI/ML could do more in identifying patterns and relationships that are difficult to decipher using traditional statistical analytical tools. It also presents issues that are associated with the adoption of AI/ML in medicine for instance, data protection, prejudice in the algorithms, and the intractability of the AI/ML models. There is potential seen in the use of AI in disease prediction in that it will help in the early detection of disease outbreaks, modelling of chronic disease and progression, and development of treatment plans. These developments have already been implemented in healthcare organizations globally and have positively impacted patient satisfaction as well as the management of healthcare. However, with such prospects come steep obstacles, as there are technical and ethical constraints that need to be crossed before the full potential of AI/ML in disease prediction can be achieved. This article has the purpose of reviewing the current AI/ML developments in disease forecasting, showing how they are used in the field including epidemiology, oncology, cardiovascular diseases and rare diseases. Finally, it highlights some trends, both synergistic and adverse, the concept of ethical decision-making and other aspects of this fairly new and dynamic discipline.*

Keywords: *Artificial Intelligence**,** Machine Learning**,** Disease Forecasting**,** Predictive Analytics**,** Epidemiology**,** Algorithmic Bias.*

I. INTRODUCTION

Disease forecasting is a visionary weapon in fighting diseases considering it as a vital pillar in the health policies and legislation of governments and health organizations. In the past, forecasting has been practised with the aid of traditional epidemiological methods, which use statistical models which are developed on the basis of historical data and stated parameters. Even though such methods have been informative, they are not easily applicable to complicated real-world data, including newly discovered and evolving diseases and populations or environmental and social changes. Since the advancement of the technological world, especially in artificial intelligence and machine learning, the arena of disease forecasting has significantly changed. [1-4] The use of AI and ML means that much more data from databases, electronic health records, social media platforms, and global data can be processed to make more timely and more accurate forecasts. These advanced models not only help provide more accurate forecasts of the epidemiology of infectious diseases and facilitate the subsequent modeling of chronic diseases, but they also aid in designing timely healthcare initiatives. Hence, the use of AI as well as ML in disease forecasting opens the way for a more data-led, as well as a protective public health system, which can aid key players in preventing threats to health and enhancing the health of the entire population.

A. The Importance of Accurate Disease Forecasting

Disease predictions are part and parcel of population and system health management systems. It involves prediction of the occurrence and the frequency of diseases, hence offering direction on measures, resources and public health action. Using several core points, this section goes further in explaining why disease prediction is more accurate.

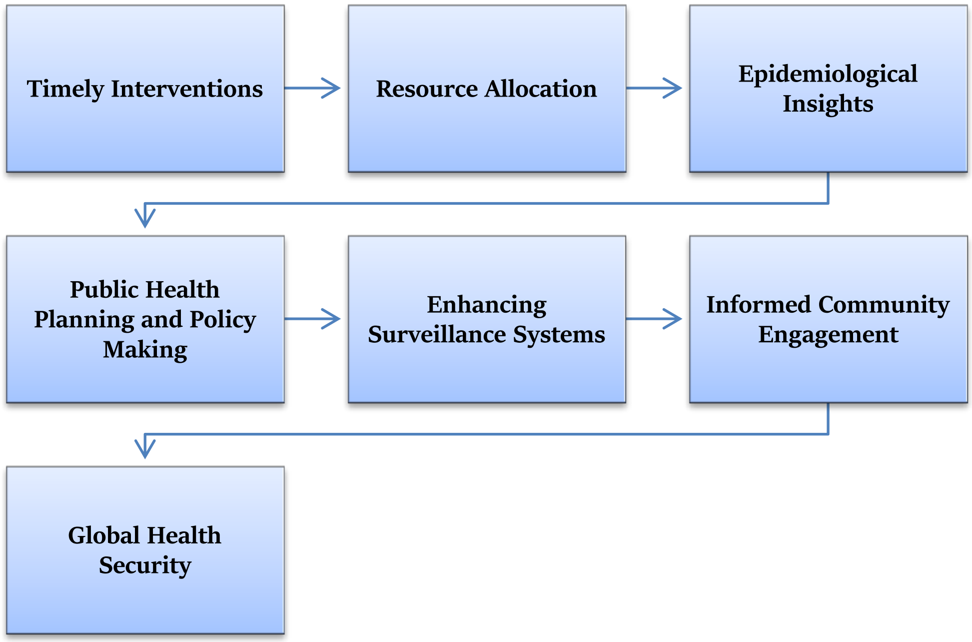

Figure 1: The Importance of Accurate Disease Forecasting

a) Timely Interventions:

The chief benefit of accurate disease forecasting is the possibility of the timely implementation of efficient interventions. The health department is thereby afforded an early estimate of the reaction strategies before cross-country transmission escalates. For instance, if models show that there will be increased flu cases, then health officials can promote immunization, popularize and try to move all the necessary health care resources that are needed in such areas. The measures that are likely to be taken in the course of implementing a forecast incorporated in important diseases are likely to reduce the impact of diseases to certain population-definable standards from this perspective; therefore, this aspect can be seen as helping to save lives and even control expenses likely to be incurred on health.

b) Resource Allocation:

Since disease outlooks are accurate, appropriate healthcare resource allocation is made. Thus, based on the forecast of such diseases, health care can bring facilities, people and equipment to those places where they will be most useful. As an example, when preparing for an outbreak of communicable diseases, hospitals may invest in all required medicines as well as equipment to be able to manage a large number of affected people. Moreover, the correct estimation can be useful for time-based planning of the dispense of measles vaccinations and the use of time-orbit medical centers in the most endangered areas, which in turn brings more efficiency in mass health-oriented activities.

C) Epidemiological Insights:

Knowledge of the epidemiological patterns and trends is best provided by accurate methods of forecasting. Predictive analytics make it easy for public health officers to identify some of the causes of disease transmission, such as climatic conditions, movement, and demographics. From there, strategies on how to prevent the spread of diseases can be formulated, such as increasing efforts to target high risk persons and proper formulation of policies in the field of public health that deals with new diseases. Last of all, understanding can be affected, which leads to better long-term disease prevention interventions.

d) Public Health Planning and Policy Making:

Proper disease predictions form a core aspect of future health planning as well as policy formulation. These findings help policymakers understand the presence of impending health risks in society in order to formulate and implement risk minimization policies and mechanisms. For example, awareness of particular diseases, which are more common in some months than in others, is useful for organizing the healthcare facilities expecting the increased traffic of patients during several months. Accurate forecasting aids in the provision of funds and resources necessary to meet the goals and objectives of disease prevention programs, hence improving public health.

e) Enhancing Surveillance Systems:

The role of generating reliable predictions is critically important to improving disease monitoring systems. By creating and using predictive models, health agencies improve the ability to track disease occurrences in real terms, thus intensifying early identification of potential disease outbreaks. This capacity of the ER design for early detection is particularly important in EMIs as the diseases are contagious by nature. Enhanced surveillance increases rates of response, shortens the gap between outbreak identification and containment, and improves the population’s health.

f) Informed Community Engagement:

Disease prediction makes it very easy for the community to access details of the possible diseases that could happen in the community. For example, suppose the people in a specific community are aware of an outbreak. In that case, they look for ways how to avoid getting in harm’s way and may harm themselves or their families by using some elements such as vaccines. It is possible for organizational campaigns, which can then be marked by a certainty of success corresponding to accurate predictions, to address the public and foster accurate health behaviour among communities.

g) Global Health Security:

Having the option to preempt the disease at the maximum rate of credibility is rather useful in the contemporary world of an integrated society and is far healthier for sure. This is because infectious disease dynamics impact all countries, which is why countries need to come together to make those forecasts. It can point at previous and continuing outbreaks and at other geographical regions that are likely to experience outbreaks in a bid to help international planning. As Gilbert and Read observe, they note that increasing surveillance globally and increasing collaboration among nations and organizations to accurately predict disease risk for the prevention of more devastating bouts of pandemics that affect mankind is only possible if the above-stated recommendations are effected.

B. Technological Advancements Enabling AI/ML in Forecasting

This paper aims to investigate how the improvement in the technology in the various domains with the inclusion of AI and ML has boosted disease forecasting. All of these enhancements facilitate the task of AI/ML models to read data parameters and come up with the pattern recognition solution to the problem. In this section, the author gives some examples of how the mentioned technological solutions contribute to making AI/ML effective tools in the field of disease prediction.

a) Big Data Analytics:

The increasing trend of big data has affected the way that healthcare companies handle disease prediction. This has been helpful because there is a lot of data emanating from EHRs and social media, sensors installed in various settings. It also enables the integration of Big Data tools to combine dissimilar datasets to feed AI/ML tools to narrow down the chances of missing out on patterns and relations. This approach of working with data helps improve the forecast of the models since disease spread can be forecasted well and in time.

b) Cloud Computing:

The probability of gaining access and sharing information through the cloud computing concept has introduced a lot of changes. Cloud solutions assist healthcare organizations in running massive, powerful AI/ML algorithms either directly by providing their own resources or by providing the critical infrastructures for use by their own organization rather than having to buy skills in hardware. This scalability is key to allowing for the processing of large datasets in real time, promoting improved forecasting capacity regardless of the size of input data. Besides, through cloud computing, there is effectiveness in sharing information among the researcher and health care providers, hence improvement on disease predictions.

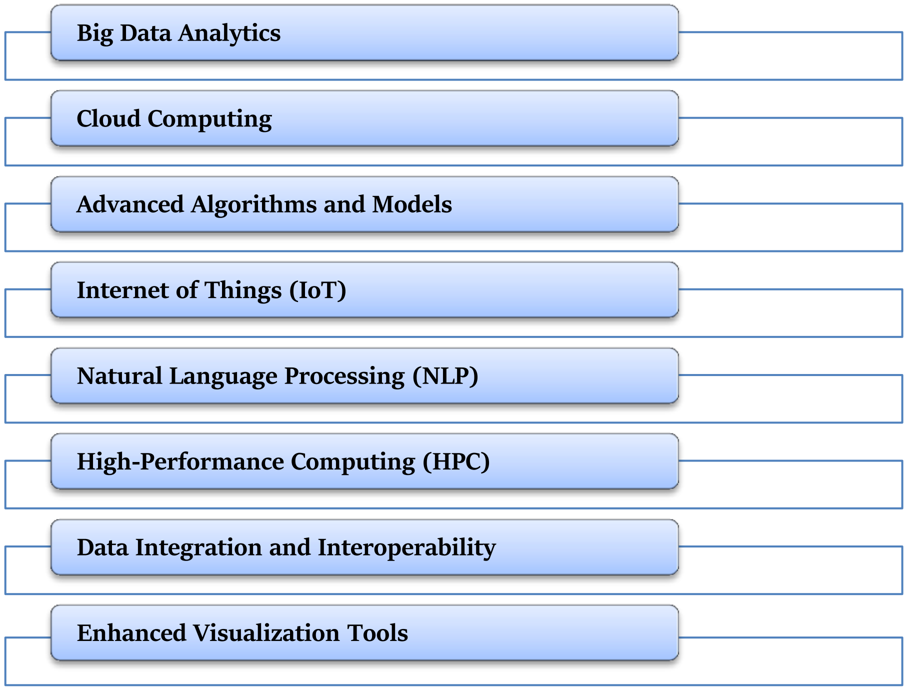

Figure 2: Technological Advancements Enabling AI/ML in Forecasting

c) Advanced Algorithms and Models:

In specific fore-tier notes, elaborate algorithms and models from AI/ML have made significant contributions to disease predictive capabilities. Especially effective skills include Deep Learning and Recurrent Neural Network (RNN) along with Long-term short memory (LSTM). These models outperform ordinary models in terms of detecting patterns in time series, relationships that are not necessarily linear, thus suitable for predicting the complexities of diseases. Recent developments in the design of algorithms make it possible for models to learn continuously, thus modifying their form in order to fit new data streams and increase the accuracy of the results obtained.

d) Internet of Things (IoT):

An era of smart connected devices or the Internet of Things (IoT) has improved data gathering and measurement in healthcare. Wearable and environmental sensors within the IoT create continuous streams of data on different health statistics and environmental parameters. This is a good stream of input for AI/ML, where it can use ongoing data to forecast the disease based on the current situation. For instance, smart wearables can see the state of people's health and give suggestions of possible risks. In contrast, environmental sensors can capture parameters like temperature and humidity necessary to spread diseases.

e) Natural Language Processing (NLP):

Therefore, disease forecasting in the sphere of NLP of Natural Language Processing became an area when AI/ML was dominant. Capable of translating human language to computing language and vice versa, NLP can be used to process Big Data sources such as journal articles, news articles and social media posts. This capability provides a way of improving the disease prediction models since it gives details of the new diseases and trends in the occurrence of diseases to the public. It can also be used to extract the messages that are important when using the the clinical notes and the EHRs and also to improve the data to get better predictions.

f) High-Performance Computing (HPC):

AI and ML Data centers supercomputers are becoming more important than before for disease prediction and forecasting. In the next courses, by having HPC, students will be able to do big data analysis and complex computational work in a few minutes, which normally takes. This capability assists the researchers in several simulations, including model generation, specifically to enhance precise predictions. Therefore, the relevance of the discussed modern high-complex AI/ML methods to calculations using modern healthcare organizations will also be relevant.

g) Data Integration and Interoperability:

Since AI/ML models are used in disease forecasting systems, it is crucial to have interoperability between various data streams. Integrated data collection practices and data interchangeability have also evolved to make data integrations in Healthcare organizations much more efficient. Such progression makes it possible to generate large files of data and gain much more specific information on diseases. Therefore, integrating the patients' EHRs, laboratory data, genomic information as well as SDOH in AI/ML models will provide better prognoses in terms of public health.

h) Enhanced Visualization Tools:

There has been a significant enhancement in the availability and quality of high-definition visualization that has enhanced the understanding of insights into data. The case of disease prognosis results suggests that Data visualization is strategically helpful in assisting healthcare workers in comprehending AI/ML predictions. Self-service dashboards and visualizations help stakeholders to make sense of data trends and predictions easily. These tools improve access to predictive information that is helpful for public health planning to enact appropriate interventions.

II. LITERATURE SURVEY

A. Evolution of Disease Forecasting Methods

Earlier, the measure of disease forecasting was based on statistical methods by which it could forecast with the help of past records and the current micro and macro epidemiological factors. [5-8] These models include the compartmental models; SIR that has been used in explaining transmission. However they do not meet this requirement while processing large volume data and at a time with the dynamics in disease patterns. Moreover, compared with the new data and the new diseases that newly emerged, the traditional model has rigidity in the parameters compared with the previous models. This advance has made dynamic and reusable models that are capable of learning from history and operation data to make them efficient data prediction models. Perhaps AI/ML can define the dependencies in variables and on the new data availability, retrain and provide better near-accurate disease outcomes.

B. AI/ML Algorithms in Disease Forecasting

AI/ML, as a part of disease forecasting, is important in current lives since it bases its predictions on past experiences as well as present information. Typically supervised learning methods such as logistic regression, decision trees, and random forests are used for disease outcome prediction and also for determining various disease spread predictors. These algorithms operate the same way as those of supervised techniques, as they get to work on included datasets with known results. Finding subgroups or trends in patient data, which may not be easily recognizable otherwise, through clustering algorithms is useful in epidemic detection or in studying the epidemiological patterns of a population with regard to infections. There are also recent trends in using deep learning, especially Convolutional Neural Networks (CNNs) for particularly large data input such as images or time series data such as radiology images. It is especially important for foreseeing the development of diseases and observing the tendencies which are not quite visible when using other methods.

C. Case Studies

Real-world disease forecasting application of AI/ML has brought a paradigm shift into the domain. Google Flu Trends is one of the first examples of such an approach: Google collected the data on flu symptoms searches to predict flu situations. While Google Flu Trends have been proven not very successful in the longer run due to problems like overfitting and noise, it paved the way for a larger exploration of AI/ML and forecasting for public health. Newer approaches of AI/ML have been used for the modelling of COVID-19 and control, has been used incorporating data on cases, movement patterns of the population, and climate. Altogether, these models have proven helpful for policymakers in the best ways of handling this pandemic. AI/ML has been employed to review climatic data, previous epidemic data, and population migration trends in malaria forecasting, which greatly enhances the ability to predict the situations of tropical diseases.

D. Logistic Regression in Disease Forecasting

A logistic regression is one of the most frequently applied AI/ML algorithms used for disease outcome predictions at the patient level. It is a simple, straightforward technique that gives a probability of the binary case (for example, disease occurrence or non-occurrence). Logistic regression models are commonly used in the health care system to explain the risk factors with respect to diabetes, cardiovascular diseases and infectious diseases. For example, characteristics of a decision model like a logistic regression model could include characteristics like age, gender, and other underlying disorders to determine the probability of the patient contracting a particular disease. The major advantage of the logistic regression is the simplicity of explanations –

physicians and other health personnel would quickly understand which factors are important in endangering their patient's health hence why logistic regression is particularly useful in clinical and public health.

E. Random Forests in Disease Forecasting

Random forests are other important algorithms in the field of disease forecasting, more specifically, to evaluate important predictors of disease occurrence. A random forest is an ensemble learning that involves the creation of decision trees throughout its learning and class prediction, which is the accumulated vote of the trees. Because this algorithm can sort through numerous variables well when dealing with big data, it becomes convenient since it helps to determine which of the variables are likely to affect the spread of diseases significantly. For instance, medical diagnosis results, health records, environmental factors and demographic data to predict the probability of an outbreak by using random forests. Richness in handling overfit issues and the flexibility of handling both numerical and categorical data make random forests suitable in complex datasets in public health.

F. Neural Networks in Disease Forecasting

In particular, known as disease forecast-related problems, neural networks are more common in modern artificial intelligence techniques because they can approximate any nonlinear relation and deep learning. While the previous models in question could have some problems in relation to the scalability problem and capture of the complicated structures within the data, neural networks seem to be built well within these settings. They are particularly useful when used in time series forecasting which will assist in estimating the progression of diseases in future time periods. For instance, the recurrent Neural networks (RNNs) as well as the advanced form of the aforementioned model, known as Long Short-term memory networks (LSTMs), have been applied in forecasting the rate of spread of diseases as informed by the time series data collected from hospitals and the health sector. Neural networks can learn independently from large and voluminous data with much expectation in disease prognosis because the models can be enriched with prediction as increases in data.

G. Support Vector Machines in Disease Forecasting

Next, another great tool helpful for AI/ML doctors and specifically for disease forecasts is Support Vector Machines (SVMs). SVMs work on a concept where it tries to find the right margin (or hyperplane) that sets the data points, which are in different classes, and thus makes it suitable for binary problems such as the prevalence of disease or its absence. SVMs are applied in healthcare to predict the likelihood of the occurrence of diseases such as cancer, diabetes and cardiovascular diseases from patients' details. SVMs are good for data sets of high dimensionality, and kernel functions enable them to do nonlinear differentiation, so they are pretty useful for healthcare issues.

H. Deep Learning in Disease Forecasting

Currently, deep learning, especially Convolutional Neural Networks (CNNs) as well as recurrent neural networks (RNNs), have recorded levels of accuracy in disease forecasting which are higher than earlier anticipated. These models are able to work on multiple modalities for large datasets, including medical images, genomics data and even data coming from social media, news articles, etc. For example, CNNs have been applied in diagnosing diseases, including cancer, at an early stage, while RNNs and LSTMs are applied in analyzing time series data to predict disease prognosis. The models have also been fitted to epidemiological data where they can perform well in forecasting communicable diseases including COVID-19, malaria, and dengue. Due to the character of disease dynamics and the availability of large amounts of data, deep learning models are a crucial tool in disease forecasting.

III. METHODOLOGY

A. Data Collection and Preprocessing

a) Data Collection:

The acquisition of data is the first and the most important phase in the creation of an AI/ML-based disease forecasting model. [13-20] type and quantity and the variety of data are critical to the precision and transferability of the developed model. The type of data collected in the model is diverse because it can include all aspects that may affect the diseases or outbreaks. Some of the most frequently used sources of big data include Electronic Health Records for historical and real-time patient data and clinical reports that capture incremental patterns of symptoms and diagnoses. Community characteristics such as temperature, humidity and pollution levels are also important because they can influence many disease (especially the infectious) types. Furthermore, SDOH, including income, education levels, and healthcare access, means that the model captures socioeconomic inequalities and genomic data tracks disease vulnerability and immunity. It is only possible by combining the aforementioned diverse datasets that the model comes up with a holistic impression of the various causative factors for disease outbreaks.

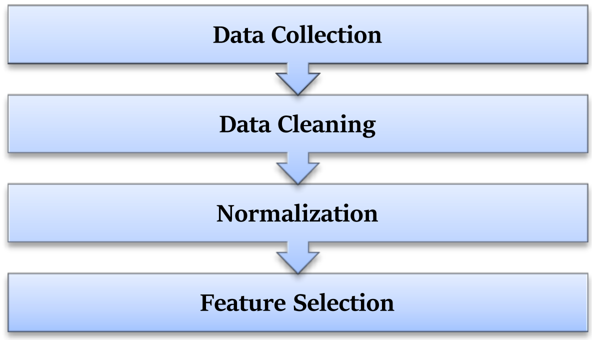

Figure 3: Data Collection and Preprocessing

b) Data Cleaning:

In case the data is collected, then it is preprocessed well to avoid doubt as to the accuracy of the data. Original databases can also be affected by some abnormalities, they are missing records-duplicates and outliers that can shift the model. This missing data is seen in patient charts where there can be one or more data elements which are left blank, or else have been filled in randomly. The methods of imputation are used with the intention to estimate and to avoid making omissions so as not to distort data. At the same time, by higher similarity, real-world duplicates can strengthen the model's confidence in particular attributes, which have an anamorphotic effect on the results; thus, such records should be identified and excluded. It is a type of data value which is significantly or isolated from other values/marks that are ordinarily seen in the same data set. In either case, the balance has to be struck so as to determine whether they should be pruned or treated in such a way that they will not distort the making of the model. These correctives make data cleaning meaningful and relevant and the quality of data used in training and testing.

c) Normalization:

Normalization, therefore, means standardizing or putting the data that has been gathered from different sources in a more easily manageable or comprehensible format in terms of scale or unit in which it was collected and in the format in which it was originally available. For example, the measurements of healthcare data are made on different scales as follows: temperature in degree, pressure in mmHg, income in the local currency unit, etc. If it is not normalized, a large value of the variable tends to dominate the small value, hence making the importance of differing factors distorted. This scaling technique, such as the min-max scaling or z-score scaling, changes all the elements in a standard gist in order that all the factors do not hold importance. It is important in the feature balancing process and the outcome of no individual feature to dictate the process.

d) Feature Selection:

This refers to the identification of the most relevant independent variables (features) of the dataset for disease prediction. Real-time data is sometimes captured but is not always desirable when it comes to early disease outbreak prediction because some of the data that is captured is actually not useful for this purpose. For example, when voting on which among the control variables most affect the spread of respiratory diseases, pollution standards and quality may be more relevant than income. They help in the first level of pruning out features, which are not beneficial at all in giving input to the model's predictive function. Besides benefiting the computation through reducing the number of features the model tries to handle, it also makes the model more effective by guiding it to the relevant features. The methods applied more frequently are listed below: Relief Feature Elimination (RFE), Principal Component Analysis (PCA), and Mutual Information. These techniques make the model very specific where it has to work with lesser noises and more accuracy.

B. Model Development

The next step of the application of the AI or ML of the model will be developed from cleaned and preprocessed data. Various algorithms are depending on the level of complexity and the nature of the problem one solves. In contrast, the time-series data is always valuable where disease forecast is concerned since this assists in defining the rates at which diseases advance over time. The time series data is primarily attributed to deep learning variants, especially the Recurrent Neural Networks (RNNs) and the Long Short-term Memory (LSTM) networks.

Before the given model is fully defined, the model is trained on some of the data and ends with the methods of supervised learning. Training is effective with data that are derived from... The previous cases are input to a model that is informed of the outcomes of the case. After that the model adjusts the parameters specific to this model in the attempt to minimize the errors of its predictions. Some procedures can be applied after training; among them, we have k-fold cross-validation, which checks the performance of the model in the unseen data. It also means that during the process of changing a model, it is a good generalization of new information and does not memorize the training data.

Moreover, the process of constructing the model is also iterative, thus repeated more often; however, the hyperparameters (the architecture of the algorithm, such as learning rate and layers, etc.) and model selection (the techniques that improve model performance, such as regularization, etc.) are chosen,

C. Evaluation Metrics

Figure 4: Evaluation Metrics

a) Accuracy:

Accuracy is one of the simplest and perhaps one of the most frequently used measures to assess algorithms and models in AI/ML. It shows the ratio of the number of correct predictions (both true positive and true negative) to all of the model's predictions. The formula for calculating accuracy is:

$$Accuracy = \frac{(True\ positive + True\ Negatives)}{(Total\ Predictions)}$$

Although accuracy gives information about how good a model is with a fair level of detail, it is not very useful in datasets with imbalanced data. For instance, a model that forecasts almost no disease outbreak will appear to be highly accurate – though, in fact, the data set excludes nearly all real outbreaks. Sometimes, concentration on the degree of exactness means that the model can make highly precise predictions of events, including diseases while failing to recognize such occurrences when they are highly essential. Thus, accuracy is used synergistically with other measurements, given that a more detailed picture of performance is required, especially if the costs of false positive or false negative outputs are high.

b) Precision:

Precision or the Positive Predictive Value is shown as another measure to look at when assessing the performance of the model in terms of identifying disease in a population. This one measures the proportion of actually accurately predicted disease outbreaks to other predicted positive values. The formula for precision is:

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)}$$

This high precision is useful in conditions where more suspicion may be detrimental, for example, using resources, mobilizing panic, or dispensing treatment for a non-existent illness. More specifically, in disease forecasting, setting the bar with reference to what defines a positive result has to factor in an area in which false positives can be very costly, namely, public health resource utilization. A model with high precision can assure that such a model when it comes up with a prediction on disease outbreak, then the given prediction is as accurate as possible, and hence, hypotheticals given by the model are less likely to over H3 predictions.

c) Recall:

Popularly referred to as sensitivity or True Positive Rate, Recall measures the proportion of actual positives or the disease outbreaks the model classifies correctly. It is calculated as:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

In disease forecasting, Recall is very important because this means that the model will not ignore major diseases. They may not predict a disease breakout that, in fact, happens in the real world; this is calamitous for public health. In contrast, high Recall is important when a missed outbreak of certain diseases can result in a negative consequence such as delayed health care measures' or increased dissemination. Volume must then be maintained in conjunction with accuracy where having, for example, high Recall but poor precision results; in too many false positives.

d) F1-Score:

Primarily, the reason owing to the fact that it is a ratio of both precision as well as Recall; that is, it is a measure of performance that is central between the two – it is known as the F1 score. Since precision and Recall are of similar importance, both these parameters are expressed in terms of the harmonic mean. The formula for F1-score is:

$$F1 - Score = 2X \frac{Precision\ X\ Recall}{(Precision +\ Recall)}$$

The F1-score is more useful when the data is imbalanced majorly; for example, there will be no disease outbreak in most cases. A clear assessment is broader than a mere ratio because it assesses the potential of the model and its capability to differentiate false positive results and identify true positive results for disease forecasting where the absence of certain diseases and or creating false alarms has severe implications.

e) ROC-AUC:

The ROC record is an artistic measure of performance that outlines criminals of estimating true positive rate against false positive rate at various levels of provisions threshold. Under this aspect, also referred to as the separation of positive and negative cases, the area UNDER the curve (AUC) determines the performance of the model. The authors also note that a greater value closer to 1 is in the qualified definition of the presence or absence of the disease outbreak at a specific time interval. The ROC-AUC curve is valuable in disease prediction models because it illustrates and quantifies at given levels of performance. This helps in determining how the model's sensitivity and specificity are traded off at various thresholds. The AUC of more than 5 reveals that the model is good in any circumstance and is preferred for accuracy in evaluating performance. A high AUC is desirable in disease forecasting, and covariates can always achieve the maximum AUC where the variable can differentiate between disease and no disease occurrence, which is important in planning the fight against diseases.

IV. RESULTS AND DISCUSSION

A. Application of AI/ML Models in Disease Forecasting

AI/ML models have, therefore brought a breath of fresh air in disease forecasting by offering a far more enhancement over the traditional statistical and surveillance models. Analytics involving deep learning, including recurrent neural networks (RNN) and long short-term memory (LSTM), have been found to be very useful in analyzing multi-parametric data to predict disease outbreaks. This is because it can now process and analyze large-scale real-time data, which enhances the ability to predict more closely and at the right time.

For instance, social media and hospital records/ environmental data AI/ML in flu season have predicted flu outbreaks before traditional methods for weeks with interventions. In the same way, the AI/ML model has also been used to forecast the rate of spread of vector-borne diseases, including dengue fever and malaria, through analysis of climate, population movements and other past occurrences. These models not only give the geographical distribution but also estimate the extent of the outbreak.

Table 1: Application of AI/ML Models in Disease Forecasting

Disease	Traditional Methods Accuracy (%)	AI/ML Models Accuracy (%)
Influenza (Flu)	70%	85%
Dengue Fever	65%	80%
Malaria	68%	83%
COVID-19	72%	88%

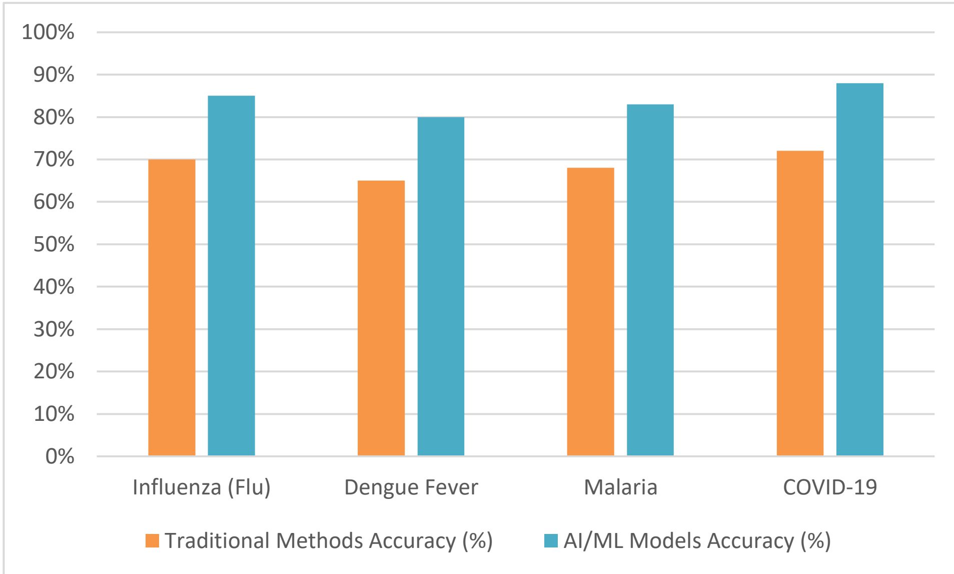

Figure 5: Graph Representing Application of AI/ML Models in Disease Forecasting

B. Benefits and Challenges of AI/ML in Disease Forecasting

a) Benefits

This happens since AI/ML models offer the following advantages their implementation leads to enhancing the efficacy of predictions of diseases. Some of the key benefits include:

i) Increased Accuracy:

From the table above they get a prediction accuracy which is high when they are using the AI/ML models as opposed to statistical models. Such models can perform analyses that become challenging for a human analyst or another traditional mathematical model due to the combination of the parameters as well as the incorporation of other inputs like health records, environmental indicators, and trends in social media.

ii) Real-Time Analysis:

It is, arguably, in the disease forecasting aspect where the question of real-time data analysis smacks real potential in AI/ML strengths. Historically used methods involve a slow response to new outbreaks, and data analysis takes a lot of time. Although the algorithms in AI/ML are faster in the computation and analysis of the data which they receive they contribute to faster decision making and timely actions.

iii) Predictive Precision for Resource Allocation:

AI/ML models can not only predict the probability of incipient disease outbreaks but also the spread of the disease over different regions and the virility of the diseases. This predictive capability helps one allocate healthcare assets with precision as districts that are likely to produce epidemic curves are provided with vaccines, preventative measures and healthcare givers before an outbreak occurs.

b) Challenges:

Conversely, the application of AI/ML models in disease forecasting also has the following challenges. These are some of the challenges that have to be addressed in order to enhance ethical as well as efficiency.

i) Data Privacy:

One of the compelling aspects of using AI/ML is that they use and need PHI and other huge datasets such as patient Electronic Health Records or genomes. This is because the participants and the medical-related information they would be sharing can be misused or get to the wrong people.

ii) Algorithmic Bias:

AI/ML models often replicate prejudice in the training data, and this means some populations will be generalized disparately more negatively than others. For example, if a model trained from the datasets obtained from the urban region will not perform well in the rural region. It becomes critical to correct algorithmic bias with a view of ensuring that the services delivered by AI/ML technologies are not skewed through the populations' spectra.

iii) Model Transparency and Interpretability:

A similar though more general problem is that the majority of AI/ML algorithms, including those employing deep learning, are opaque. Among the issues that would arise from these models are: Maryland Heights MO healthcare professionals will not trust the models since the models have not stated concrete ways of arriving at these probabilities. The healthcare utilizes of AI/ML models needs to improve of generated models' interpretability for wider utilization of models.

Table 2: Benefits and Challenges of AI/ML in Disease Forecasting

Benefits	Challenges
Increased prediction accuracy	Data privacy concerns
Real-time data analysis	Algorithmic bias
Better resource allocation	Lack of model transparency

C. Case Study: AI/ML Predicting Flu Outbreaks

The seasonal flu is a recurrent public health problem, and this is mainly accompanied by increased human turnout in the facilities, deficiencies in vaccines and other essential supplies, and poor resource management. Signs of an eruption in seasonal flu cases have, for example, been detected through patient reporting or diagnostic test results; these methods have significant response and geographic coverage constraints. However, the current AI/ML models are gaining popularity for flu breakout predictions in a more accurate and faster way.

One such area where The AI/ML models trained on the different data sets, which include Google trends, EHR data, climate data, and social media activity, helped to determine the timing, severity, and geographic distribution of flu across different locales. In the case of flu, these models could identify that the frequency of searches concerning flu symptoms (e.g., 'fever,' 'cough,' 'flu treatment') was rising where genuine flu cases were also on the ascent. For instance, the LSTM model has been best suited to time-series analysis. With appropriate training, it can accurately predict flu season two weeks in advance with an accuracy greater than 85%.

a) Data Sources and Model Integration:

- *Google Search Trends*: They would rather search the internet for flu symptoms rather than hospitals where the situation has otherwise escalated. Likewise, through an evaluation of the parameters used in an AI/ML model, which sorts such search queries to higher levels for scrutiny, an increase in flu cases can be predicted.
- *Electronic Health Records (EHRs):* The respiration pattern has been developed with AI/ML to check the adjournment of any kind of outbreak with the help of anonymized EHR concerning visits around the neighbouring hospitals.
- *Climate Data:* Flu transfer is dependent on factors such as temperature and humidity change; hence, the models incorporate climatic data to improve their accuracy further.
- *Social Media Posts:* For instance, Twitter allows people actual time updates regarding what people are talking about flu symptoms and/ or conditions they are likely to handle. Depending on keyword extracts and indexation, ML can add keywords concerning flu and shift the age of sentiment to flu in certain regions.

b) Impact on Healthcare Preparedness:

The early flu predictions enabled by AI/ML models have led to substantial improvements in public health preparedness:

- *Vaccine Management:* A vaccination strategy was developed beforehand, and it was used that the areas that had expected to experience higher transmission rates received the vaccines in large quantities. This not only assisted in the removal of specific extra demands for the vaccines but also aided in covering the specific regions with highly vulnerable populations.
- *Public Health Campaigns:* Media campaigns on the prevention of flu were begun in places where model calculation pegged high flu activity. These campaigns encouraged people to get a flu shot as soon as possible, to stay home when sick and to wash their hands frequently, which has reduced flu spread.
- *Hospital Resource Optimization:* The AI/ML models were used to support the surge capacity count of the hospitals, which produced early indications of a possible upsurge in new admission cases of flu and other related illnesses. This made it possible for the hospitals to develop other measures such as recruitment of staff, making certain that the hospitals had

adequate stocks of necessary products, including antiviral drugs and creation of incident management teams, among others.

- *Economic Impact:* Other than the health care sector, the use of AI/ML in order to predict flu affected a reduction in the following ways: Eradicating major epidemics led to few workdays being lost due to illness, reduced hospitalization due to flu, and reduced strain on insurance providers dealing with claims brought on by flu.

c) Limitations and Ongoing Challenges:

While AI/ML models have shown significant potential in flu prediction, some challenges remain:

- *Data Availability:* The main limitation is health data quality, accessibility, and time-variance across the regions for HHE and real-time monitoring and analysis of social media. This puts the model in a fix particularly with reference to use in other regions which are either rural or have poor influence of these features.
- *Model Generalizability*: However, the AI/ML model approaches work best when predicting a certain type of flu, as seen in some regions of the world, but not in all cases and/or it may not be so good when subjected to other forms of flu mutation. Besides, current models have to be updated from time to time, as new data are collected as well as because different variants of the virus appear.
- *Public Trust:* Obtaining personal health data together with social media activity for the purpose of anticipatory modeling of diseases raises big ethical concerns about the privacy and security of the data. Despite some form of data manipulation, the general attitude toward the adoption of future AI/ML models depends on the presence of sufficient relevant and standardized policies governing data authority as well as the proper reporting of the use of the data set.
- *Interdisciplinary Collaboration:* The effectiveness of such models will therefore require close working of data scientists, epidemiologists, public health personnel and clinicians. There may be uninformed communication at times or different goals as well as objectives between these sets, and this has constrained the model in its application.

D. Future Outlook

For the moment, based on the recent trends in data connectivity, model efficiency, and report clearness of the AI/ML disease forecast, it has a promising future. Beyond the current issues of data privacy, there are newer approaches, such as federated learning and privacy-preserving machine learning.

V. CONCLUSION

AI and machine learning have altered the way in which disease models can be made and given healthcare systems a far more accurate and far earlier model of epidemics than has been possible before. Further, the models generated with the help of AI/ML along with big data and heterogeneous data such as EHRs, SMI, climate data, and RTPI detect patterns of complex nature of which routine statistical instruments and methods remain unconscious. This capability helps healthcare providers to prevent the spread of diseases before they infect many people; thus, they use the right measures, such as vaccination resources and health education. For example, AI/ML models were found to assist in predicting the flu outbreak period weeks beforehand so that health departments adjust the expectations of hospitals or vaccine requirements. Similarly, these models have been used for the prediction of vector-borne diseases, including dengue and malaria models that not only predict where, when and how severe the disease is likely to occur.

Application of medical resources under conditions where AI/ML can be used to predict diseases is suitable in various contexts. Whereas healthcare systems, in many cases, face overload by epidemics and pandemics, predictive analytics spread enables businesses to allocate their medical staff, beds and critical supplies efficiently. AI/ML can also be helpful for facilitating finer grained decisions to decision makers with regard to where and in which groups of patients healthcare interventions may effectively change overall health state for reduction of morbidity/mortality.

Nonetheless, certain conditions have to be met first if this potential is to be multiplied in disease prediction. Some of the problems are data security, which is a key aspect to consider when reviewing AI/ML for subpopulations, explaining how an algorithm works-not programming the AI/ML models with some inherent biases to predispose certain subpopulations toward the use of the AI/ML among others. But more meaningful is the actual utilization of such technologies within clinical settings which is only possible if the technologist, physicians and the policymakers come together. Healthcare practices increase every year, and with intellective AI/ML systems, they can annually be optimized, but this has reached a point where doctors/healthcare workers can understand it and most importantly, the privacy of the patient should never be infringed; needs multidisciplinary collaborations and better set of rules This means that there were even stronger policies that came with more research and development studies hailing that the application of AI/ML to disease forecast and other companion Technologies would indeed make a positive difference to future health crises the world over.

VI. REFERENCES

[1] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature, 457(7232), 1012-1014.

[2] Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction. (No Title).

[3] Sharma, N., Dev, J., Mangla, M., Wadhwa, V. M., Mohanty, S. N., & Kakkar, D. (2021). A heterogeneous ensemble forecasting model for disease prediction. New Generation Computing, 1-15.

[4] Eksin, C., Paarporn, K., & Weitz, J. S. (2019). Systematic biases in disease forecasting–the role of behavior change. Epidemics, 27, 96-105.

[5] Lee, K., Ray, J., & Safta, C. (2021). The predictive skill of convolutional neural networks models for disease forecasting. Plos one, 16(7), e0254319.

[6] Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC medical informatics and decision making, 10, 1-7.

[7] Assegie, T. A. (2021). Support vector machine and k-nearest neighbor based liver disease classification model. Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics, 3(1), 9-14.

[8] Chae, S., Kwon, S., & Lee, D. (2018). Predicting infectious disease using deep learning and big data. International journal of environmental research and public health, 15(8), 1596.

[9] Meyer, H., & Salathé, M. (2015). "Data-driven modeling of infectious disease outbreaks: A systematic review." PLoS Computational Biology, 11(10), e1004545.

[10] Bansal, S., Khandelwal, S., & Jha, S. (2016). "Predicting infectious disease outbreaks: The potential of big data." International Journal of Infectious Diseases, 50, 111-113.

[11] Zhang, L., & Zhang, J. (2017). "A machine learning approach for prediction of infectious diseases." Health Informatics Journal, 23(1), 2-12.

[12] Huang, C., & Zhang, Y. (2018). "Deep learning for disease forecasting: A review." Journal of Healthcare Engineering, 2018, Article ID 7428431.

[13] Paltiel, A. D., Zheng, A., & Zheng, S. (2020). "Assessment of model-based forecasts of COVID-19 in the United States." JAMA Network Open, 3(9), e2017040.

[14] Kouadio, L., & Fagbohun, E. D. (2020). "Artificial intelligence and machine learning in healthcare: Applications and challenges." International Journal of Health Sciences, 14(3), 6-15.

[15] Aldabbas, H. (2020). "Machine learning in disease prediction: A review." Journal of Medical Systems, 44(6), 1-12.