Machine learning in drug supply chain management during disease outbreaks: a systematic review

Gunadi Emmanuel[1,2], Arief Ramadhan[3], Muhammad Zarlis[4], Edi Abdurachman[5], Agung Trisetyarso[5]
[1]Computer Science Doctoral Program, Bina Nusantara University, Jakarta, Indonesia
[2]Faculty of Science and Technology, Universitas Katolik Musi Charitas, Palembang, Indonesia
[3]School of Computing, Telkom University, Bandung, Indonesia
[4]Department of Information System Management, Bina Nusantara University, Jakarta, Indonesia
[5]Department of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Article Info

Article history:

Received Mar 1, 2023
Revised Apr 13, 2023
Accepted Apr 15, 2023

Keywords:

Disease outbreaks
Drug supply chain
Endemic
Machine learning
Prediction

ABSTRACT

The drug supply chain is inherently complex. The challenge is not only the number of stakeholders and the supply chain from producers to users but also production and demand gaps. Downstream, drug demand is related to the type of disease outbreak. This study identifies the correlation between drug supply chain management and the use of predictive parameters in research on the spread of disease, especially with machine learning methods in the last five years. Using the Publish or Perish 8 application, there are 71 articles that meet the inclusion criteria and keyword search requirements according to Kitchenham's systematic review methodology. The findings can be grouped into three broad groupings of disease outbreaks, each of which uses machine learning algorithms to predict the spread of disease outbreaks. The use of parameters for prediction with machine learning has a correlation with drug supply management in the coronavirus disease case. The area of drug supply risk management has not been heavily involved in the prediction of disease outbreaks.

Corresponding Author:

Gunadi Emmanuel
Computer Science Doctoral Program, Bina Nusantara University
Jakarta, Indonesia
Email: gunadi.emmanuel@binus.ac.id

1. INTRODUCTION

The coronavirus disease (COVID-19) pandemic event sparked research on forecasting the spread of disease outbreaks to a greater extent than in the past. The danger of COVID-19 urges the fulfillment of the need for COVID-19 drugs. The challenge for drug manufacturers and supply chain management is to meet the large and sudden demand for drugs [1]. The endpoint of drug supply chain management is on the side of the pharmacy user or patient. According to food and drug administration (FDA) [2], a long drug supply chain will complicate drug supply management and even become one of the triggers for drug shortages. Singh *et al.* [3] revealed that the pharmaceutical supply chain has a complexity that involves various stakeholders ranging from manufacturers, distributors, customers, information service providers, and the government as a regulatory agency. In addition, supply chains face environmental uncertainty, capacity planning, and inventory management [4]. Responding to the complexity of the drug supply chain, Lozano-Diez *et al.* [5] suggest taking advantage of technological developments.

The use of technology is not only in production planning, distribution processes, and storage. There must be integral management from upstream to downstream, namely primary manufacturing, secondary manufacturing, distribution centers, wholesalers, and retailers/hospitals [6], [7]. In this industrial era 4.0, all factors in supply chain activities are interrelated and should be integrated [6]. Downstream supply chain activities identifying and fulfilling user demands are directly related to activities in the distribution and manufacturing department [8]. Meanwhile, Karmaker and Ahmed [1] mentioned five indicators of drug supply resilience in drug supply activities: supply chain risk orientation, visibility, flexibility, supply agility, and collaboration. Indicators of drug supply resilience require a significant catalyst, namely information sharing. By sharing information, predictions based on field data become a source of drug demand forecasting and integration of supply processes. The downstream drug demand prediction determines all drug supply policies from upstream to downstream [9].

Predicting downstream drug demand to support drug supply management is a challenging task. There are many parameters and uncertainties from medical and non-medical variables. Identification of disease characteristics and patient care needs have dynamics of rapid change. Stakeholders involved in the drug supply chain must consider ambiguity and uncertainty from the user demand side [10], [11]. The COVID-19 pandemic is an example of the complexity of the uncertainty faced in the supply of medicines. Learning from the COVID-19 case reminds the correlation between the accuracy of predictions of epidemic cases in determining interventions, control strategies, and integrated decision-making [12]. Manufacturers and distributors expect downstream information to assist decision-makers in supply chain and patient healthcare decisions [13]. Thus, early detection of outbreaks helps stakeholders to plan drug production and distribution [14].

The development of machine learning technology in demand forecasting makes it possible to predict the spread of disease outbreaks by considering the external variables of drug distribution and availability in pharmaceutical units. The encounter of various data provides jobs for machine learning algorithms to uncover hidden patterns and the necessary knowledge from extensively stored information [15]. Furthermore, Muñoz *et al.* [13] stated that artificial neural networks in various supply chains use different variables to extract knowledge. In supply chain forecasting, the development of neural network techniques tends to be selected when anticipating market changes.

When neural networks can predict new cases or disease progression Sutariya *et al.* [16], Wieczorek *et al.* [17], Merkuryeva *et al.* [18] said that the discussion of drug demand prediction models continues to evolve to improve accuracy. Improving prediction accuracy will help drug supply planning and distribution management to prevent drug shortages. He added that predictions would be more accurate if they were closer to the drug demand point [18]. Thus, the method for developing accurate predictions of the spread of epidemics does not ignore that drug supply management uses predictions as the basis for planning. A primary drug supply management interest is the planning and distribution process of drug supply to prevent drug shortages. The research challenge is approaching the prediction of drug supply amid uncertainty about the spread of disease using machine learning methods to determine the performance of drug supply management. To the best of our knowledge, drug demand prediction activities with machine learning have not been an in-depth discussion in research related to drug supply management. The literature study aims to analyze the correlation between machine learning prediction parameters and drug supply chain management in disease transmission research over the last five years.

Next, the paper outlines in part 2 a brief theory of machine learning, drug supply chains, and disease outbreaks. Section 3 presents the method used in this literature study. Section 4 presents the results and discussion of the literature study. Finally, in Section 5, the conclusions answer the research questions.

2. THEORETICAL BACKGROUND

2.1. Overview of machine learning

Machine learning is a science that mimics the way humans learn to process data and use algorithms. Machines have adaptive mechanisms that allow learning from experience, examples, and analogies. Learning capabilities can improve as system performance improves over time. Learning mechanisms form the basis for adaptive systems [19].

The machine learning process occurs from self-learning activity without guidance. The existence of machine learning supports the field of data development science. Machine learning uses statistics and trained algorithms to make classifications and predictions [15]. Machine learning collects and stores enormous amounts of information and is constantly growing. That information exists across most application domains, such as business, the academic community, medicine, and the public. The machine learning activity extracts the collected data into knowledge to support decision-making [20]. There are three areas of machine learning: supervised, unsupervised, and reinforcement learning.

2.1.1. Supervised learning

The focus of supervised learning is on accurate predictions. Machine learning works so well that previously unseen data reveals new information. Supervised learning begins by labeling the data collected. Learning outcomes produce a certain level of performance accuracy. Supervised learning methods are usually used for classification, regression, object detection, image captions, and semantic segmentation. A simple example, this method can be applied to determine whether the male face is different from the female face. By labeling face examples and gender labels, the machine learns to obey the rules for distinguishing male and female faces [21], [22].

2.1.2. Unsupervised learning

Unsupervised learning classifies the collected data based on comparing the features and characteristics of the unlabeled data. The data classification aims to learn a reasonable and concise description of data. Meanwhile, learning techniques commonly used are clustering and data projection/preprocessing [21], [22]. The ultimate achievement of unsupervised learning focuses on data collection in subgroups, patterns, or clusters.

2.1.3. Reinforcement learning

The learning process allows feedback action to the algorithm. This method uses a dataset with a reward or punishment system. Reinforcement learning is a form of learning to map from situations to actions that maximize scalar rewards or reinforcement signals. Learning from the experience of certain circumstances, the learning agent takes appropriate steps to achieve long-term goals. Reinforcing targeted actions aim at long-term rewards. Trial seeking and delayed reward are two of the most important distinguishing features of reinforcement learning [21], [22].

2.2. Drug supply chain

The drug supply chain is a complex process with specialized operations. Strict restrictions and arrangements are made on supply and delivery channels to customers to prevent drug supply errors [18]. Many agents build supply and demand interactions. They participate in the supply process (upstream) and the patient care process (downstream). Franco and Alfonso-Lizarazo [23] stated that the pharmaceutical supply chain has the following two main elements: supply consists of i) drug manufacturers and distributors; demand consists of ii) pharmacies that order drugs from suppliers, maintain drug safety, manage drug supplies, and distributing drugs to hospitals; iii) hospitals that provide drugs to patients and order them from pharmacies or distributors; and iv) patients requiring treatment and medication Figure 1.

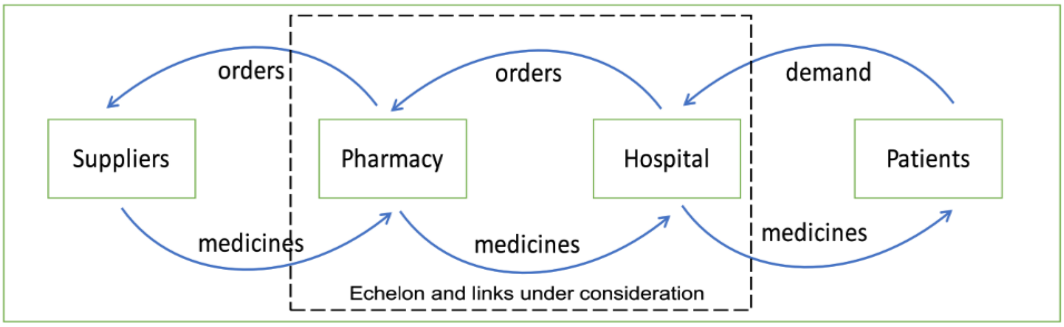

Figure 1. Interaction in the pharmaceutical supply chain [23]

Figure 1 shows the echelon and link under consideration in pharmacy and hospital. Decisions about ordering drugs from manufacturers and distributors are in these two units, as are decisions about managing drug storage and distribution. Lozano-Diez *et al.* [5] show in Figure 2, the five main areas of risk of failure of integrated operations in the case of drug supply in the public sector. Dispensing pharmacy is the meeting point of drug production and drug demand. All five areas have the same potential to contribute to failure in the drug supply chain process. Interrelated relations between areas will affect the resilience of supply. The small gap between demand and supply determines the resilience of the drug supply chain.

Dispensing position Figure 2 is in the middle, allowing you to know the gap between demand and supply from producers. The challenge for dispensing is to convey demand predictions as accurately as possible to producers. Dispensing's success in sharing information with distribution units will also provide readiness for drug supply disruptions. Predictive information helps supply chain operations to respond to unexpected risks. Zahiri *et al.* [6] stated that the problem must be local and not stop production at the primary producer level. Whereas in the area of demand, the participation of many agents and regulators is an effort to maintain drug supply in joint action [5], [6].

Figure 2. Drug supply process [5]

Previously, Anaya-Arenas *et al.* [24] said that management needs to have a level of preparedness and mitigation, response, and recovery in an emergency. A resilient supply chain is able to identify risks, ameliorate impacts, and return quickly from human or natural events [25]. A risk manager can run a resilient supply chain by designing a pharmaceutical supply chain network and proactively conducting scenario analysis to support the development of risk response contingency plans and reactively analyzing impact and implementing recovery [26]. Management needs to design a strategy so that drug supply can quickly return to its normal position or to a more optimal position. In addition, drug supply management must recognize threats, possible losses, and similar incidents that will recur [27]. Management needs information to make decisions amidst epistemic and stochastic uncertainties [28].

2.3. Disease outbreaks

A disease becomes an epidemic when an increase in the number of cases of a disease in a local community, a given geographic area, or a season is high enough to cause an outbreak. Outbreaks occur when an infectious agent spreads directly between people, through exposure to animal reservoirs, insect or animal vectors, or other environmental sources. In addition, human behavior almost always contributes to the spread. Early detection of the spread of epidemics and reporting to the health office is important to prevent the spread of epidemics and minimize negative impacts on public health, society, and the economy [29].

To conclude that there is an outbreak in an area, cases occurring in that population require an epidemiological investigation (example: rabies, plague, and polio). Diseases that are unprecedented, irregular, and rare in a region also require investigation. An increase in the number of cases of the disease may occur in a wider geographic area. Epidemic cases can occur due to [30]:

– Increase in the number or virulence of a new agent,
– Introduction of new agents into new situations,
– High mode of transmission,
– Increased susceptibility of the human body's response,
– Increased host exposure or new entry points.

Systematic investigation and risk assessment of acute public health events are necessary to prevent a wider spread of the disease. Assessment involves measuring the likelihood of disease occurrence and its eventual impact on human society. For infectious diseases, the risk will relate to the size, characteristics, and living conditions of the person migrating. Migratory human behavior poses serious epidemic risks and health impacts. The risk depends on the previous immune status of the population and the variable threat at the destination [31]. For this reason, it is important to consider resources (human, laboratory, logistics, and materials) to anticipate an outbreak. Without adequate resources, the health system is weakened and the poor are increasingly at risk of outbreaks [32].

3. METHOD

The literature review methodology proposed by Kitchenham [33] and Kitchenham *et al.* [34] became the reference for this research methodology. The research stages include planning, execution, and reporting. Figure 3 shows the process of activities in the literature review.

3.1. Planning the review

The initial study identified the need for a review study obtained from research articles that attempted to increase the accuracy of drug demand predictions based on historical data and various methods to produce high accuracy. In addition, there are indications of drug shortages (large and small scale) in many

hospital/pharmacy units. Pharmaceuticals have yet to solve this problem. The research objective was to analyze the correlation between machine learning predictive parameters and drug supply chain management in disease transmission research. This study arranges research questions to achieve research intently, as follows:
– What disease outbreaks have become research studies in the last five years?
– What prediction methods are used in research to deal with disease outbreaks?
– What parameters are considered in the use of prediction methods?
– What is the influence (correlation) of the research prediction parameters with the field of drug supply management?

This literature study covers more than one field of science (multidisciplinary). The study belongs to the field of computer science because of the machine learning methods and algorithms for prediction. In terms of a subject area, this study discusses diseases and drugs, which are part of health science. Finally, the study talks about the distribution of drugs to consumers, which is in the field of supply chain management. Therefore, the literature sources in this research are related to computer science, information systems, and pharmacy. This research uses the Publish or Perish 8 application from the digital library: Scopus and the National Library of Medicine (PubMed) to ensure the articles used were from accredited sources. The article search categories were from journals and conferences published in 2018 to 2022.

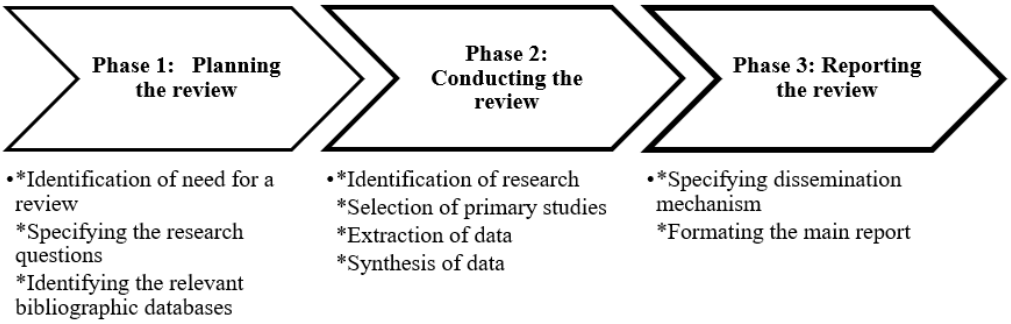

Figure 3. Phases of systematic literature review activities

3.2. Conducting the review

With the help of the Publish or Perish 8 application, search for articles that will become material for literature review using search terms in the title/keyword. The library search string used is ("machine learning") AND ("predict" OR "prediction") AND ("disease outbreaks" OR "endemic"). In the selection process, articles are subject to the following inclusion dan exclusion criteria Table 1.

Table 1. Inclusion and exclusion criteria

Inclusion	Exclusion
Only English	The research subject is not related by the research questions
Only Journal and Conference article	Any publication before the year 2018
Equivalent journal qualifications Q1, Q2, Q3, Q4	Papers with non-academic databases
Date of publication in between 2018-2022	The type of study article is not systematic literature review
Papers based on quantitative or qualitative analysis or a mix both	Studies with non-full text articles
	Duplicate paper of the same study

The selected articles have relevance to answering research questions so that the article criteria address at least one disease outbreak, refer to machine learning methods in the process of solving them, and see the prediction problem as a need in the drug supply chain as much as possible. Assessment of the quality of research articles uses the following questions so that the study articles have library materials with complete data answering research questions.
– Was there a clearly stated research objective?
– Is the article well designed to achieve its purpose?
– Were the variables considered in the study identified?
– Is the discipline clearly defined?
– Were data collection methods adequately explained?

The article management process uses the Mendeley application to facilitate the conducting and data extraction. After reading the abstract and full article content, Table 2 shows the process of selecting articles by searching relevant bibliographic databases, identifying suitable studies, and extracting the results. Searches with keywords in the Scopus and PubMed libraries found 214 study articles. In the conducting process, the study removes articles that do not meet the inclusion and exclusion criteria Table 1. The most common cases of duplication of simultaneous publications are on Scopus and PubMed. At this selection stage, there were 96 articles that had abstract content in accordance with the purpose of this literature review. After reading the contents of the articles in more detail, there are 71 selected articles.

Table 2. The outcome process of selecting, conducting, and studying relevance

Sources	Found	Candidate (Title+abstract)	Selected
Scopus	140	65	50
PubMed	74	31	21

3.3. Reporting the review

Researchers used a qualitative study approach. Concern over the issues of drug shortages, drug supply, and the development of drug demand predictions encourage researchers to approach each study material with a line of argument synthesis. The main issues in the research questions are identified and tabulated in a table and descriptions. Finally, this multidisciplinary study selected publications in Scopus-indexed international journals.

4. RESULTS AND DISCUSSION

This section reports the results of a systematic literature review study. The first part will briefly look at the general description of the selected articles. The general description of the selected articles provides a viewpoint of the feasibility of the articles for review. Furthermore, the discovery of outbreaks over the past five years, methods of predicting outbreaks, and prediction parameters are the subject of discussion on how predictive research results support drug supply management.

4.1. Overview of selected articles

There are 71 selected articles from the Scopus and PubMed digital library databases. The percentage of the number of articles from the Scopus is greater than that of the PubMed because this literature review study focuses on machine learning Figure 4. During the conducting process, there were many copies of published articles on Scopus and PubMed. The possibility occurs because the topic is multidisciplinary. Any research result working on machine learning methods will publish an article in the science field. The possibility of the same research in medical journals occurs when the research discussion is in the medical journal. Researchers must read all the article to find out the research discussion area.

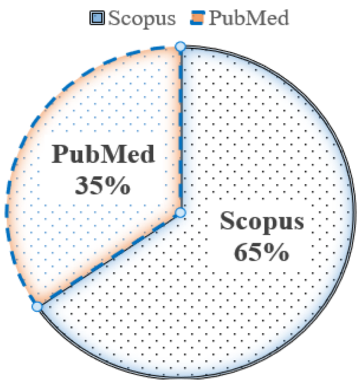

Figure 4. Comparison of articles found in the Scopus and PubMed databases by keyword

There are 24 author countries gathered in this study, but the number of author countries with more than one (1) article is 12 countries Figure 5. The composition of the author's country representation can also represent various parts of the world. The results of the selection of articles are representative enough to provide a view that the topic of using learning machines in the midst of an outbreak is a topic that is currently receiving the attention of researchers. China, England, and America still dominate research in this field. The number of developing countries that have conducted this research indicates that they are facing problems that are not much different.

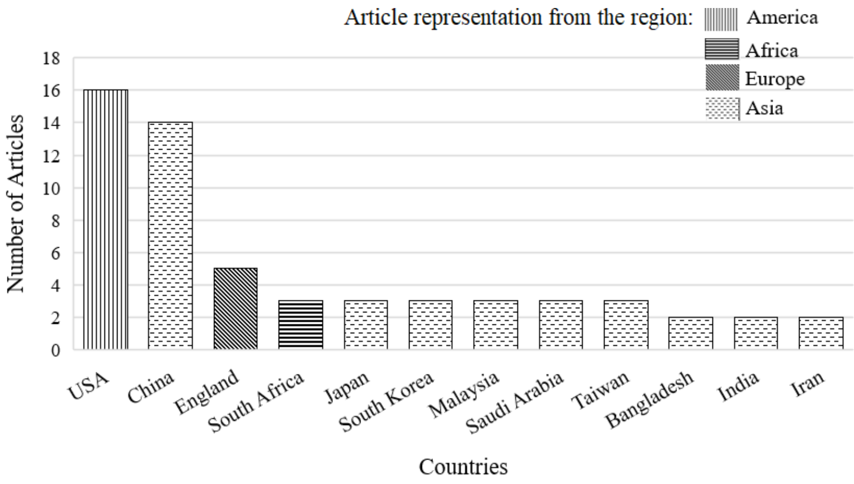

Figure 5. Composition of research countries with more than one (1) article

4.2. Disease outbreaks that have been a research study in the last five years

Of the 32 types of disease detected from the study, an outbreak study with more than two articles only had eight (8) types of disease. Figure 6 shows the eight types of epidemic disease that can be categorized into three disease categories, namely: i) epidemics with disease vectors in the form of mosquitoes (dengue fever, malaria, chikungunya, and zika); ii) outbreaks caused by viruses that attack the respiratory tract (influenza and COVID-19); and iii) outbreaks are caused by bacteria, viruses, and environmental contamination (foodborne food and diarrhea). Each category is marked with the same pattern in Figure 6.

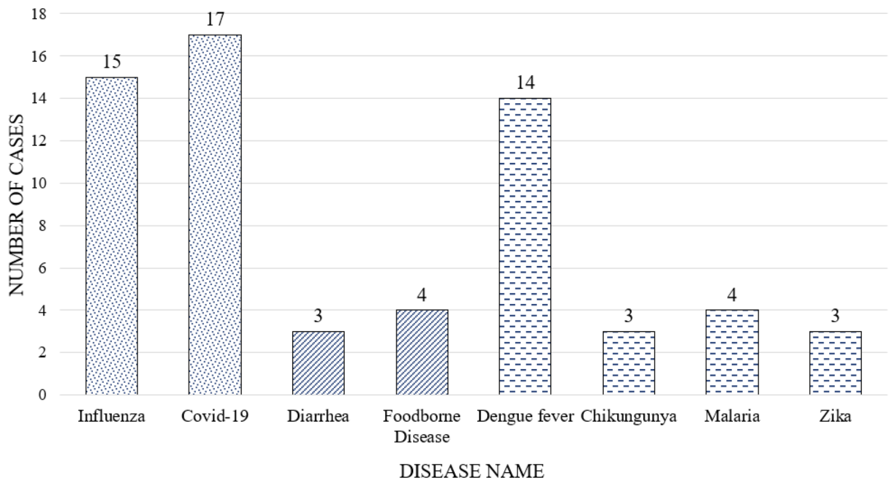

Figure 6. Number and distribution of diseases in case studies

The three categories of disease outbreaks identified can occur in one country, several adjacent countries, and a wider area. Each country or continent region has different demographic, socio-cultural, economic community characteristics that are of great concern to researchers. The severity of the outbreak is also a concern for researchers, thus encouraging them to participate in preventing or inhibiting the rate of spread of outbreaks with research on predicting the spread of outbreaks.

4.3. Predictive methods used in research to address disease outbreaks

In this literature study, the trend method approach is taken from the methods discussed in more than two (2) articles that conduct studies of certain disease outbreaks. Table 3 (see in Appendix) shows the eight outbreaks identified using machine learning algorithms according to the cases each one faces [35]–[91]. An interesting phenomenon is when more than one outbreak uses the same machine-learning method. There are indications that the use of machine learning is by the objectives to be achieved by research, but also (needs to be investigated more deeply) parameters that are not much different are used as parameters in learning.

Based on the three outbreak categories in this study, the commonly used machine learning techniques are as follows: i) epidemics with disease vectors in the form of mosquitoes (dengue fever, malaria, chikungunya, and zika) using decision trees, regression, support vector regression (SVM), boosting algorithm, neural network; ii) outbreaks caused by viruses that attack the respiratory tract (Influenza and COVID-19) using naïve Bayes, SVM, boosting algorithm, decision tree learning, k-nearest neighbors (KNN), regression, neural network; iii) outbreaks are caused by bacteria, viruses, and environmental contamination using Bayesian Regression models, linear regression models, convolutional neural networks (CNNs), long short-term memory networks (LSTMs) and SVMs. Different approaches can complement the limitations of previous research or improve what is still lacking, such as increasing accuracy in supervised learning.

In the case of technique selection, Benedum *et al.* [76] developed machine learning (ML), regression, and time series models to predict dengue rates over weekly time periods. The data for the forecast uses weekly dengue fever outbreak monitoring data and available weather data. Meanwhile Jain *et al*. [74] considers the use of parameters of meteorological data, clinical, disease surveillance lag variables, socioeconomic, and spatial dependence coding on dengue transmission. In one-month intervals, they developed generalized additive models (GAMs) to synchronize the relationship between predictors and clinical data of dengue hemorrhagic fever (DHF). Both of these studies sufficiently show that the use of techniques is always related to the parameters. Comparison between ML techniques to obtain high accuracy needs to consider the level of parameter development.

An interesting study was conducted by Raizada *et al*. [75], when they combined machine learning algorithms for prediction studies of mosquito vector-borne disease outbreaks for all three diseases (chikungunya, dengue, and malaria). They built the CNN-multimodal disease outbreak prediction (CNN-MDOP) algorithm. Future studies can increase the accuracy by 88% by increasing the accuracy of demographic data, meteorological data, and disease outbreak prediction. Saha *et al.* [86] found a challenge to elucidate the etiology of meningitis in Bangladesh. At that time, they conducted validation and application studies of unbiased metagenomic next-generation sequencing (mNGS) to elucidate the etiology of meningitis in Bangladesh. They used cerebrospinal fluid (CSF) specimens from patients admitted to the largest children's hospital. In cases of disease of unknown cause (idiopathic), mNGS identified a potential bacterial or viral etiology in 40% of supervised machine learning algorithms. In this case, the discovery of disease-causing knowledge is useful for determining the production and distribution of drugs for healing.

4.4. The parameters to be considered in the use of the prediction method

With the formation of three categories of disease outbreaks, the characteristics of the disease are found in clinical data, patient history, and the spread of disease that is influenced by natural factors. The parameters used for the eight types of outbreaks are also not much different. Parameters are determined for training because they are involved in determining whether the disease can be dangerous or not.

Parameter areas that are widely used are related to disease symptoms, disease history, distribution patterns, climate-meteorological aspects, demographic information about the outbreak environment, mobility behavior of sick people, and real-time data about the rate of disease spread. It needs an in-depth discussion, prediction of the spread of disease is not only done for health prediction or knowledge. Accurate predictions will correlate with drug supply and availability in pharmaceutical units.

Complete parameters in the disease spread prediction model are not guaranteed to improve accuracy. Environmental aspects, the level of socio-economic awareness of the community, the characteristics of the causes of outbreaks, and geographical conditions have dynamic features that will affect the prediction results. For Santosh [56], continuous and unprecedented factors must receive attention in designing complex predictive models. He proposes not to rely solely on random parameter-driven stochastic or discrete indicators. Epidemiological characteristics (of the disease) can form the basis of initial consideration for predictive models. According to Pourghasemi *et al*. [55], the selection of predictor variables was based on a literature review and expert knowledge related to disease outbreaks. These considerations are relevant when finding several conditions that differ from one area to another.

From Table 4 (see in Appendix) the parameter column shows several indications of data collection from government health institutions and other trusted health organizations/associations. Recent developments and disease outbreak predictions take data from web searches Google Trends and social media such as

Twitter and Weibo. Dissemination of information from patients, families, and people affected by the outbreak on the internet media can be a source of information in predicting disease outbreaks. Identification of "sick posts" by social media users reporting symptoms and diagnoses for either themselves or others regarding the COVID-19 outbreak can be classified by machine learning [53]. The prediction results using the social networking site (SNS) model in tracking disease outbreaks can detect COVID-19 accurately, as an initial stage, before official information from the outbreak information center [54]. It is similar to taking data from the main topic of each news article. Subsequent research has led to the identification of news articles and related events that refer to the symptoms of the disease in question [35], [36].

4.5. Correlation of research prediction parameters with drug supply management area

Each outbreak has its own characteristics according to the disease vector and its spread. These characteristic differences affect the determination of parameter weights which can improve prediction accuracy. The discussion of determining close predictive parameters on outbreak characteristics requires further research. However, the third column of Table 4 shows the relationship between the parameters used in the drug supply management area. Based on drug supply chain theory and the five integrated operational failure risk areas in the case of public sector drug supply (Figures 1 and 2), we identify the parameters used in the prediction of the eight outbreaks into supply chain management areas: planning, production, distribution, storage, and risk of supply chain.

Table 5 shows the correlation of the predicted spread of eight outbreaks using predictive parameters that relate to at least two areas of drug supply management. In influenza outbreaks, predictive parameters showed correlations in determining the amount and type of drug production and distribution, management of drug storage, and planning of type and distribution pattern. However, in this literature study, articles predicting the spread of influenza outbreaks have not found parameters that discuss the risk of drug supply.

In the case of COVID-19, Table 5 shows that the researchers developed the most extensive parameters. The spread of COVID-19 and the high death toll have prompted a large-scale study. The correlation between the parameters used and the area of drug supply management indicates a large study. There is a sizable difference in research between COVID-19 and other outbreak cases. In cases of others outbreaks, there are not many prediction parameters that consider the risk areas of drug supply.

Table 5. Parameter used to eight types of diseases

Diseases	Planning	Production	Distribution	Storage	Risk of Supply Chain
Influenza	✔	✔	✔	✔	0
COVID-19	✔	✔	✔	✔	✔
Diarrhea	✔	0	0	0	✔
Foodborne Disease	✔	0	✔	✔	0
Dengue fever	✔	✔	✔	✔	0
Chikungunya	0	✔	✔	✔	0
Malaria	✔	✔	✔	✔	0
Zika	✔	✔	✔	✔	0

✔ =parameters are in the areas of drug supply management
0=parameters are not in the areas of drug supply management

4.6. Practical implications and limitations

The correlation of research prediction parameters to drug supply management has practical implications for predictions by machine learning and drug supply management. In terms of developing machine learning, increasing the accuracy of predicting drug demand is developing by taking data from social media and real-time media for monitoring the spread of outbreaks [35], [36], [43], [45], [46], [54], [57], [90]. The development of drug demand prediction parameters supports the development of drug distribution chain models. Predicting the spread of outbreaks in an area is a consideration for the resilience of the drug supply chain. Parameters of the outbreak spread can be an early indication of the need for a drug supply model on a national and regional scale.

The spread of the COVID-19 outbreaks and the dangers to human health have triggered the use of deep learning machine technology by no longer relying on historical disease and patient data. The basic prediction parameters from disease characteristic data, patient distribution data, data from trusted government/private health institutions, data related to climatological-meteorological conditions, and socio-demographics of the developing community to traffic data related to epidemic diseases. For drug manufacturers, data on developments to predict the spread of disease form the basis for planning drug production and distribution. Predicting drug demand by observing the real-time conditions where an outbreak occurs or is likely to expand will add value to the accuracy of the prediction. In more depth, future research can examine the predictive parameters of types of diseases with high mortality values in a region.

The role of machine learning in predicting the spread of disease outbreaks triggers the integrated drug supply management process. We can take the example of demographic parameters in predictions. Mobility and population density will affect the forecasting of disease outbreaks [56], [72]. From the prediction results with demographic parameters, drug supply management cannot ignore the risk factors of the supply chain. In the case of parameters, changes in social and economic characteristics for types of foodborne diseases and diarrhea can be significant in predictions. Dynamic changes in the field can affect prediction accuracy and drug supply management. Practical implications, changes in the direction of research from the smooth distribution of drugs and predictions of drug demand need to find common ground towards integrated drug supply information management.

This study attempts to place predictive parameters as a factor integrating drug demand prediction research in drug supply management. The limitation of this study is that it does not provide specific recommendations or solutions to improve drug supply chain management or predict disease outbreaks due to the extensive discussion on disease outbreaks in general. In future research, predictive parameter research using the machine learning method of a disease outbreak in a region can be one of the factors in developing drug supply management.

5. CONCLUSION

This literature study found three categories of epidemic causes with eight types of diseases, namely i) mosquitoes as disease vectors (dengue fever, malaria, chikungunya, and zika); ii) viruses that attack the respiratory tract (influenza and COVID-19); and iii) bacteria, viruses, and environmental pollution (foodborne food and diarrhea). The character of the vector carrying the disease also determines the characteristics of the spread of the disease and the prediction method used. The use of machine learning methods to predict the spread of disease during epidemics is still within the technical scope of using combination/hybrid/ensemble methods. No single approach method has high accuracy. Disease characteristics and external (non-clinical) factors determine the prediction method used. In addition to clinical and non-clinical characteristic parameters, public data via social media tends to attract researchers' interest.

This study also reveals that predictive results can support drug planning, production, distribution, and storage. Meanwhile, there are not many studies using predictive results to assess risk factors for drug supply. This research contribution identifies a strong correlation between the learning parameters of the COVID-19 outbreak in machine learning and the field of drug supply management.

This literature review can be a starting point for further studies focusing on the latest methods in prediction for various disease categories that will assist future research. Comparative research on prediction methods and parameter determination tests in the risk area of drug supply to what extent provides increased prediction accuracy. Likewise, the effectiveness and efficiency of the prediction approach will be the attractiveness of its application to various types of diseases. The achievement of effectiveness and efficiency has led to research into integrated drug supply systems and deep-learning hybrid methods.

APPENDIX

Table 3. Methods of approaches to eight types of diseases

Epidemic of a disease (number of case articles)	Selected articles	Approach method*
Influenza (15)	[35]–[49]	– Convolutional neural network (CNN), long short-term memory network (LSTM), least absolute shrinkage and selection operator (LASSO) – Bayesian, ProMED, the program for monitoring emerging diseases, and HealthMap – The dynamic network marker/biomarker (DNM/DNB) method – The cross industry standard process for data mining (CRISP–DM), Boolean weighting, term frequency weighting (TF), inverse document frequency weighting (IDF), and TF-IDF, FastText, random forest (RF), naïve Bayes, support vector regression (SVM), C 4.5 decision tree (DT), k-nearest neighbors (KNN), and AdaBoost, linear regression model – Auto-regressive integrated moving average (ARIMA), RF, SVM, and extreme gradient boosting (XGBoost) – Feedforward neural network (FFNN), beta regression, box Jenkins model, multivariate adaptive regression splines, variable selection, Bayesian model averaging – Improved particle swarm optimization algorithm to optimize the parameters of support vector regression (IPSO-SVR) – Linear autoregression (AR), linear network autoregression, random forest, training and hyperparameter selection – LSTM, the Pearson correlation coefficient (PCC)

Table 3. Methods of approaches to eight types of diseases (*continue*)

Epidemic of a disease (number of case articles)	Selected articles	Approach method*
Influenza (15)	[35]–[49]	– The augmented regression with general online data (ARGO) model, the net model, the autoregressive model – RF and SVM, multi-layer perceptron (MLP) models, logistic regressor (LR) models, medical statistics – The classification and regression tree-decision tree (CART-DT) algorithm, SVM and backward propagation (BP) neural network, combining CART, XGBoost, and LightGBM – LASSO regression, analysis retrospective
COVID-19 (17)	[41], [44], [50]–[64]	– Distributed machine learning and blockchain, neural network – DT, extra trees, KNN, multilayer perceptron, SVM, RF – Curve-fitting models, agent-based models, susceptible-exposed-infectious-removed or susceptible, infected, and recovered (SEIR/SIR) models – Spatial modeling, risk mapping, and change detection of COVID-19, RF, machine learning technique (MLT), LASSO algorithm – The time series forecasting method, ML approach, WEKA – Mathematical approaches based on susceptible, infected, and recovered (SIR) cases and susceptible, exposed, infected, quarantined, and recovered (SEIQR) cases models, prophet algorithm, Broyden–Fletcher–Goldfarb–Shanno (BFGS), conjugate gradients (CG), limited memory bound constrained BFGS (L-BFGS-B), and Nelder–Mead – A geographic information system (GIS)-based machine learning algorithm (MLA), SVM, – ARIMA, FFNN, MLP, and LSTM – Genomic features, analysis of variance (ANOVA), LASSO, ridge regression shrinkage – A hybrid architecture including CNN, LSTM, convolutional autoencoder (CAE) – LSTM, CNN, decision tree – Logistic regression (LR), SVM, XGBoost, MLP – Natural language processing (NLP), TF-IDF, Glove, and FastText, SVM, XgBoost, LR, and RF, SVR, Prophet, and deep learning models like LSTM – RF, FFNN, Transfer: TrAdaBoost algorithm, ML algorithms – RF and SVM models, deep neural network (DNN) in the form of MLP models. Logistic regressor (LR), medical statistics – ML, linear regression model – SVM, KNN, dan naïve Bayes (NB), N-gram; TF-IDF
Diarrhea (3)	[65]–[67]	– CNNs, LSTM, a traditional ML, SVM – SVMs, LSTM, CNNs, relevance estimation and value calibration (REVAC)
Foodborne Disease (4)	[68]–[71]	– LASSO, XGBoosting – XGBoost, multigraph structured LSTM network – The language model BERTweet, a variant of bidirectional encoder representations from transformers (BERT)-Twitter – Bayesian regression model, linear regression model, ElasticNet regression model, support vector machine regression (SVR) model, gradient boosting regression (GBR) model
Dengue fever (14)	[41], [72]–[84]	– Generalized additive models (GAM); quasi-poisson regression; distributed lag non-linear models (DLNM) – Predictive flushing-mosquito model (PLUM), univariate flagging algorithm (UFA), statistic model – LSTM, transfer learning (TL) – Decision trees (CART), MLP, SVM (linear, polynomial, radial basis function (RBF)), and Bayes network (TAN) – The ARIMA, RF, and ANN models – DT, DNN, and the logistic regression – Bayes network (BN), SVM, RBF tree, decision table and naive Bayes – ML, regression, and time series models to forecast – Local-in-time SVM classifiers, SVMs, a binary classifier – SVM, leave-time-out cross validation (LTO-CV) – RF, FFNN, TrAdaBoost algorithm, ML algorithms – Statistical analysis, CNN multimodal disease outbreak prediction (CNN-MDOP) algorithm
Chikungunya (3)	[75], [85], [86]	– Statistical analysis, CNN-MDOP – Latent class cluster analysis (LCCA), Newton-Raphson algorithms – Unbiased Ribonucleic acid (RNA) metagenomic next-generation sequencing (mNGS), the Idseq bioinformatics pipeline and ML
Malaria (4)	[75], [87]–[89]	– RF – The k-means clustering, XGBoost, – Statistical analysis CNN-MDOP – Lag-regression analysis
Zika (3)	[41], [90], [91]	– Backward propagation neural network (BPNN), gradient boosting machine (GBM), RF – SVM, semi-supervised learning (SSL), and deep neural network – RF, FFNN, TrAdaBoost algorithm, ML algorithms

*The approach method used generally does not only take one method of approach, but a combination/hybrid/assemble approach.

Table 4. Parameter used to eight types of diseases and correlation in the drug supply chain

Epidemic of a disease (number of case articles)	Parameters used	Parameter correlation in the drug supply chain area
Influenza (15)	– Symptoms of disease – Measures to control epidemic spread, host and environmental susceptibility, pathogen transmission, population density, and health care capacity – Historical disease information in the clinic and new discriminant input patterns – Sentiment analysis and keyword occurrences related to disease – Graph of disease visit surveillance data both from the health department and patient medical records – Records of historical disease symptoms, Google search data, meteorological data - Atmospheric data – Historical/non-real time data from the Epidemic Center and real-time Twitter data – Availability of Regional Datasets on the spread of disease and social media – Word embedding options regarding outbreak, tokenization, and time series conversion data – Google search activity, real-time weather and local information, flu-related Twitter micro-blogs, electronic health record data, and disease activity synchronicity history across regions – Data on symptoms of similar diseases – Data on medical records and data on the spread of epidemics from government health departments – The daily number of influenza incidence and meteorological data: temperature, precipitation, wind speed, relative humidity, air pressure, daily radiation – Keyword search logs – Web search queries	– Dynamics of information related to the development of up-to-date and reliable outbreak information data: the current epidemiological, demographic and meteorological situation, and the environment (consideration of the amount and type of drug production and drug distribution) – Update historical data comparing with historical drug supply (drug storage and distribution considerations) – Consideration of the type and quantity of drug supply following the epidemic pattern (consideration of drug production and supply) – External: Twitter, online news articles, and Google Flu Trends (GFT) up-to-date and reliable data: current epidemiological, demographic and environmental situation. Early indication of drug demand dynamics in a particular area. External information accompanies existing data in the health department and hospital/pharmacy (consideration of planning types of drugs and supplies) – External: updating the data base on disease symptoms and pathology to determine the type of drug and the need for similar drugs (consideration of the production of the type of drug needed)
COVID-19 (17)	– Patient data records – Symptoms of disease – Characteristics of the disease and level of care of the patient, including: i) hospital setting/capacity; ii) test capacity/rate (daily); iii) demographics; iv) population density; v) vulnerable people; and vi) income versus commodities (poverty) – 16 variables – active case data, died, and correlations between them. Transmission patterns and real demographic, climatological, and distribution cluster situations – Limited to data on confirmed cases, recovered cases and death cases – 16 effective factors: minimum temperature of coldest month, maximum temperature of warmest month, precipitation in wettest month, precipitation of driest month, distance from roads, distance from mosques, distance from hospitals, distance from fuel stations, human footprint, density of cities, distance from bus stations, distance from banks, distance from bakeries, distance from attraction sites, distance from automated teller machines (ATMs) and density of villages – Data on variations, trends and predictions from 2 health institutions – Disease characteristics (Virus) – Data on confirmed cases by time – Genome database and reservoir host disease characteristics – Embedding hybrid words related to epidemic cases – Comparison of patterns/symptoms of similar diseases – Data on medical records and data on the spread of epidemics – Disease data structure	– Management of utilization of patient data records locally with privacy protection (Planning - hospital management considerations) – Preparedness for disaster mitigation of hospitals and pharmaceutical facilities, including the capacity of the facilities and the specific types of drugs needed (consideration of the risk of drug supply) – Updates on disease symptoms and pathology databases to determine drug types (consideration of drug supply and storage planning) – Data acquisition from several health institutions made awareness of the need for various stakeholders to build transparency of information on the dynamics of public health (consideration of planning and development of health units and risk management) – External: Twitter, Weibo (similar to Twitter); update data base of disease symptoms and pathology to determine the type of drug. Indication of the subsequent distribution of the drug in a specific area (drug distribution considerations – drug supply readiness) – Demand for similar drugs, up-to-date historical data for subsequent drug supply data, and development of vaccines and new drugs (consideration of proper quality drug production and prediction of the amount of supply to areas recorded by data)
Diarrhea (3)	– Climate variables in different geographic areas: rainfall, humidity, evaporation, and temperature climate variables – Climate change data: rainfall, humidity, evaporation, and temperature – Regional-geographical data on disease spread, climate change	– Up-to-date and reliable data: current epidemiological, demographic, and environmental situation (considerations of drug supply planning and risk management)

Table 4. Parameter used to eight types of diseases and correlation in the drug supply chain *(continue)*

Epidemic of a disease (number of case articles)	Parameters used	Parameter correlation in the drug supply chain area
Diarrhea (3)	– Climate variables in different geographic areas: rainfall, humidity, evaporation, and temperature climate variables – Climate change data: rainfall, humidity, evaporation, and temperature – Regional-geographical data on disease spread, climate change	– Up-to-date and reliable data: current epidemiological, demographic, and environmental situation (considerations of drug supply planning and risk management)
Foodborne Disease (4)	– Information on environmental factors and foodborne diseases – Information from case reporting systems, outbreak monitoring systems, molecular tracking networks – Non-real time data center of the outbreak and real-time Twitter – Information on foodborne disease risk factors caused by the spread of water pollution, sick case data on Weibo, geographic location data of the water pollution plants	– Updating of the disease symptom and pathology data base for determining the type of drug (consideration of drug production and distribution) – External: Twitter and Weibo, symptom and pathology data base updates for drug determination in certain areas (drug distribution considerations – drug supply readiness)
Dengue fever (14)	– Local meteorological data, clinical data, disease surveillance lag variables, socioeconomic data and spatial dependence coding data on DHF transmission – Data on the spread of dengue fever, population density, entomology and environment – Data on Human dengue cases per month and data on basic demographic characteristics (e.g., sex, age, nationality, and residential address) and time of occurrence related to the disease (e.g., date of onset, diagnosis, and death) – Meteorological data: 15 meteorological variables (i.e., extreme wind speed, maximum wind speed, average wind speed, minimum pressure, maximum pressure, average pressure, average water pressure, minimum air temperature, maximum air temperature, temperature average air, highest average daily temperature, lower average daily temperature, average daily precipitation, number of days with precipitation, and average relative humidity maximum pressure, average pressure, pressure average water, minimum air temperature, maximum air temperature, average daily highest temperature, average daily precipitation, number of days with precipitation, average relative humidity, geographical location – Baseline patient data, current data of rainfall, sea surface temperature, and temperature of outbreak sites – Clinical symptoms of dengue, a real-time surveillance system, climate variables such as temperature, wind speed, humidity and rainfall – Data on historic dengue cases, satellite-derived estimates for vegetation, precipitation, and air temperature, as well as population counts, income inequality, and education – Information on laboratory test results, four main variables [age, fever, white blood cell (WBC) count, and platelet count], compatible with clinical and epidemiological knowledge - vector population – Data on symptoms of similar diseases – Demographical data: positive cases; meteorological data: temperature, humidity, rainfall	– Provision of drugs based on Medical Record data (planning considerations in pharmacy units) – Consideration of the type and quantity of drug supply following the epidemic pattern (consideration of drug production) – Update on disease symptoms and pathology data base for determining drug types (consideration of drug production and supply in pharmaceutical units) – Update historical data on the spread of epidemics with historical data on drug supply, documentation of pharmaceutical unit warehouse data and information on the need for similar drugs (consideration of drug storage) – Up-to-date and reliable data: current epidemiological, demographic, and environmental situation (considerations of drug planning and distribution)
Chikungunya (3)	– Demographic data: positive cases; meteorological data: temperature, humidity, rainfall – Data on clinical symptoms and variations in phenotypic response – Symptoms of disease and patient test results	– Comparison of recent historical data on the spread of epidemics with historical drug supply (consideration of stock storage and distribution of drugs to certain areas) – Update data base on symptoms of disease and pathology for determining the type of drug (drug production considerations)
Malaria (4)	– Real-time information on transmission dynamics, parasite movements and reservoirs of infection – Data on malaria incidence and climate variability (rainfall, temperature and surface radiation) – Demographical data: positive cases; meteorological data: temperature, humidity, rainfall – Data on disease outbreak characteristics, tropical climate variability and associated sea surface temperatures over the Pacific and Indian Oceans	– Consideration of the type and quantity of drug supply following the epidemic pattern (consideration of drug production) – Up-to-date and reliable data: current epidemiological, demographic, and environmental situation (considerations of drug planning and distribution) – Comparison of recent historical data on the spread of epidemics with historical drug supply (consideration of stock storage and distribution of drugs to certain areas)

Table 4. Parameter used to eight types of diseases and correlation in the drug supply chain *(continue)*

Epidemic of a disease (number of case articles)	Parameters used	Parameter correlation in the drug supply chain area
Zika (3)	– Meteorological and environmental covariates, demographic distribution of mosquitoes, socioeconomic factors - human movement and urbanization – Data provided by MedIsys. Information on article title, description, publication date, original URL, language code, category indicating disease name, latitude and longitude are displayed as initial data for prediction – Comparison of symptoms of similar diseases	– Consideration of the type and quantity of drug supply following the epidemic pattern (consideration of drug production) – External: the news article data. update data base on symptoms of disease and pathology for determining the type of drug (consideration of drug planning and distribution) – Need for similar drugs (consideration of drug production and storage)

REFERENCES

[1] C. L. Karmaker and T. Ahmed, "Modeling performance indicators of resilient pharmaceutical supply chain," *Modern Supply Chain Research and Applications*, vol. 2, no. 3, pp. 179–205, 2020, doi: 10.1108/MSCRA-04-2020-0006.

[2] FDA, "Drug shortages: Root causes and potential solutions - A report by the drug shortages task force," US Food and Drug Administration (FDA), 2019.

[3] R. K. Singh, R. Kumar, and P. Kumar, "Strategic issues in pharmaceutical supply chains: a review," *International Journal of Pharmaceutical and Healthcare Marketing*, vol. 10, no. 3, pp. 234–257, Sep. 2016, doi: 10.1108/IJPHM-10-2015-0050.

[4] H. T. Pham *et al.*, "The impact of global risks on supply chain performance. An empirical study on construction sector in the COVID-19 pandemic," *International Journal of Quality and Reliability Management*, 2022, doi: 10.1108/IJQRM-03-2022-0084/FULL/XML.

[5] J. Lozano-Diez, J. Marmolejo-Saucedo, and R. Rodriguez-Aguilar, "Designing a resilient supply chain: An approach to reduce drug shortages in epidemic outbreaks," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, no. 21, May 2020, doi: 10.4108/eai.13-7-2018.164260.

[6] B. Zahiri, J. Zhuang, and M. Mohammadi, "Toward an integrated sustainable-resilient supply chain: A pharmaceutical case study," *Transportation Research Part E: Logistics and Transportation Review*, vol. 103, no. 2017, pp. 109–142, Jul. 2017, doi: 10.1016/j.tre.2017.04.009.

[7] N. Shah, "Pharmaceutical supply chains: Key issues and strategies for optimisation," *Computers and Chemical Engineering*, vol. 28, no. 6–7, pp. 929–941, 2004, doi: 10.1016/j.compchemeng.2003.09.022.

[8] M. S. Amalnick, N. Habibifar, M. Hamid, and M. Bastan, "An intelligent algorithm for final product demand forecasting in pharmaceutical units," *International Journal of System Assurance Engineering and Management*, vol. 11, no. 2, pp. 481–493, Apr. 2020, doi: 10.1007/s13198-019-00879-6.

[9] R. Siddiqui, M. Azmat, S. Ahmed, and S. Kummer, "A hybrid demand forecasting model for greater forecasting accuracy: the case of the pharmaceutical industry," *Supply Chain Forum*, vol. 23, no. 2, pp. 124–134, 2022, doi: 10.1080/16258312.2021.1967081.

[10] G. Candan, M. F. Taskin, and H. R. Yazgan, "Demand forecasting in pharmaceutical industry using neuro-fuzzy approach," *Journal of Military and Information Science*, vol. 2, no. 2, pp. 41–49, Mar. 2014, doi: 10.17858/jmisci.06816.

[11] E. Hofmann and E. Rutschmann, "Big data analytics and demand forecasting in supply chains: A conceptual analysis," *The International Journal of Logistics Management*, vol. 29, no. 2, pp. 739–766, May 2018, doi: 10.1108/IJLM-04-2017-0088.

[12] P. Manliura Datilo, Z. Ismail, and J. Dare, "A review of epidemic forecasting using artificial neural networks," *International Journal of Epidemiologic Research*, vol. 6, no. 3, pp. 132–143, 2019, doi: 10.15171/ijer.2019.24.

[13] E. G. Muñoz, N. S. Cossío, S. del M. R. Cedeño, S. E. L. Ricardo, Y. C. Hernández, and E. O. Crespo, "Application of neural networks in predicting the level of integration in supply chains," *Journal of Industrial Engineering and Management*, vol. 13, no. 1, pp. 120–132, Feb. 2020, doi: 10.3926/jiem.3051.

[14] N. Ben Yahia, M. D. Kandara, and N. B. BenSaoud, "Integrating models and fusing data in a deep ensemble learning method for predicting epidemic diseases outbreak," *Big Data Research*, vol. 27, Feb. 2022, doi: 10.1016/j.bdr.2021.100286.

[15] B. Abdualgalil and S. Abraham, "Efficient machine learning algorithms for knowledge discovery in big data: A literature review," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 3880–3889, 2020.

[16] V. Sutariya, A. Groshev, P. Sadana, D. Bhatia, and Y. Pathak, "Artificial neural network in drug delivery and pharmaceutical research," *The Open Bioinformatics Journal*, vol. 7, no. 1, pp. 49–62, Dec. 2013, doi: 10.2174/1875036201307010049.

[17] J. Wieczorek Michałand Siłka, D. Połap, M. Woźniak, and R. Damaševičius, "Real-time neural network based predictor for cov19 virus spread," *PLOS ONE*, vol. 15, no. 12, 2020, doi: 10.1371/journal.pone.0243189.

[18] G. Merkuryeva, A. Valberga, and A. Smirnov, "Demand forecasting in pharmaceutical supply chains: A case study," *Procedia Computer Science*, vol. 149, pp. 3–10, 2019, doi: 10.1016/j.procs.2019.01.100.

[19] M. Negnevitsky, *Artificial intelligence: A Guide to intelligent systems*, 2nd ed., vol. 2. Harlow Essex: Pearson Education Limited, 2002.

[20] C. Vercellis, *Business intelligence: Data mining and optimization for decision making*, 1st ed. West Sussex: John Wiley & Sons Ltd., 2009.

[21] R. S. Sutton, "Introduction: The challenge of reinforcement learning," in *Machine Learning*, vol. 8, no. 3–4, Boston: Kluwer Academic Publisher, 1992, pp. 225–227, doi: 10.1007/BF00992695.

[22] D. Barber, *Bayesian reasoning and machine learning*. Cambridge University Press, 2012, doi: 10.1017/CBO9780511804779.026.

[23] C. Franco and E. Alfonso-Lizarazo, "Optimization under uncertainty of the pharmaceutical supply chain in hospitals," *Computers & Chemical Engineering*, vol. 135, Apr. 2020, doi: 10.1016/j.compchemeng.2019.106689.

[24] A. M. Anaya-Arenas, J. Renaud, and A. Ruiz, "Relief distribution networks: a systematic review," *Annals of Operations Research*, vol. 223, no. 1, pp. 53–79, 2014, doi: 10.1007/s10479-014-1581-y.

[25] C. S. Singh, G. Soni, and G. K. Badhotiya, "Performance indicators for supply chain resilience: review and conceptual framework," *Journal of Industrial Engineering International*, vol. 15, no. S1, pp. 105–117, Dec. 2019, doi: 10.1007/s40092-019-00322-2.

[26] J.-M. Lawrence, N. U. Ibne Hossain, R. Jaradat, and M. Hamilton, "Leveraging a Bayesian network approach to model and analyze supplier vulnerability to severe weather risk: A case study of the U.S. pharmaceutical supply chain following Hurricane Maria," *International Journal of Disaster Risk Reduction*, vol. 49, Oct. 2020, doi: 10.1016/j.ijdrr.2020.101607.
[27] S. Alqudah, H. Shrouf, T. Suifan, and S. Alhyari, "A moderated mediation model of lean, agile, resilient, and green paradigms in the supply chain," *International Journal of Supply Chain Management*, vol. 9, no. 4, pp. 158–172, 2020.
[28] D. E. C. Yu, L. F. Razon, and R. R. Tan, "Can global pharmaceutical supply chains scale up sustainably for the COVID-19 crisis?," *Resources, Conservation and Recycling*, vol. 159, Aug. 2020, doi: 10.1016/j.resconrec.2020.104868.
[29] Manitoba Public Health, "Epidemiological investigation of outbreaks," *Communicable Disease Control Unit*, pp. 1–5, 2018.
[30] M. Porta, *A dictionary of epidemiology*, 5th ed. New York: Oxford University Press, 2008.
[31] WHO, "Outbreak surveillance and response in humanitarian emergencies - WHO guidelines for EWARN implementation," *World Health Organization*. WHO, pp. 1–68, 2012.
[32] C. F. Houlihan and J. A. G. Whitworth, "Outbreak science: Recent progress in the detection and response to outbreaks of infectious diseases," *Clinical Medicine, Journal of the Royal College of Physicians of London*, vol. 19, no. 2, pp. 140–144, 2019, doi: 10.7861/CLINMEDICINE.19-2-140.
[33] B. Kitchenham, "Procedures for performing systematic reviews-joint technical report," Keele University Technical Report TR/SE-0401, UK, 204AD, Jul. 2004.
[34] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.
[35] P. P. Schneider *et al.*, "Using web search queries to monitor influenza-like illness: an exploratory retrospective analysis, Netherlands, 2017/18 influenza season," *Eurosurveillance*, vol. 25, no. 21, pp. 1–10, May 2020, doi: 10.2807/1560-7917.ES.2020.25.21.1900221.
[36] O. Șerban, N. Thapen, B. Maginnis, C. Hankin, and V. Foot, "Real-time processing of social media with SENTINEL: A syndromic surveillance system incorporating deep learning for health classification," *Information Processing & Management*, vol. 56, no. 3, pp. 1166–1184, May 2019, doi: 10.1016/j.ipm.2018.04.011.
[37] H. Ge *et al.*, "How to determine the early warning threshold value of meteorological factors on influenza through big data analysis and machine learning," *Computational and Mathematical Methods in Medicine*, vol. 2020, pp. 1–13, Dec. 2020, doi: 10.1155/2020/8845459.
[38] H.-Y. Cheng *et al.*, "Applying machine learning models with an ensemble approach for accurate real-time influenza forecasting in Taiwan: Development and validation study," *Journal of Medical Internet Research*, vol. 22, no. 8, Aug. 2020, doi: 10.2196/15394.
[39] F. Wang *et al.*, "Regional level influenza study with geo-tagged Twitter data," *Journal of Medical Systems*, vol. 40, no. 8, Aug. 2016, doi: 10.1007/s10916-016-0545-y.
[40] A. N. Desai *et al.*, "Real-time epidemic forecasting: challenges and opportunities," *Health Security*, vol. 17, no. 4, pp. 268–275, 2019, doi: 10.1089/hs.2019.0022.
[41] K. Roster, C. Connaughton, and F. A. Rodrigues, "Forecasting new diseases in low-data settings using transfer learning," *Chaos, Solitons & Fractals*, vol. 161, Aug. 2022, doi: 10.1016/j.chaos.2022.112306.
[42] P. Chen, E. Chen, L. Chen, X. J. Zhou, and R. Liu, "Detecting early-warning signals of influenza outbreak based on dynamic network marker," *Journal of Cellular and Molecular Medicine*, vol. 23, no. 1, pp. 395–404, 2019, doi: 10.1111/jcmm.13943.
[43] A. Alessa and M. Faezipour, "Preliminary flu outbreak prediction using twitter posts classification and linear regression with historical centers for disease control and prevention reports: Prediction framework study," *JMIR Public Health and Surveillance*, vol. 5, no. 2, Jun. 2019, doi: 10.2196/12383.
[44] J. A. Bishop *et al.*, "Improving patient flow during infectious disease outbreaks using machine learning for real-time prediction of patient readiness for discharge," *PLOS ONE*, vol. 16, no. 11, Nov. 2021, doi: 10.1371/journal.pone.0260476.
[45] B. Jang, I. Kim, and J. W. Kim, "Effective training data extraction method to improve influenza outbreak prediction from online news articles: Deep learning model study," *JMIR Medical Informatics*, vol. 9, no. 5, 2021, doi: 10.2196/23305.
[46] I. Alsmadi, Z. Almubaid, and H. Al-Mubaid, "Disease outbreak detection using search keywords patterns," in *EPiC Series in Computing*, 2020, vol. 69, pp. 362–351, doi: 10.29007/z8tp.
[47] E. L. Aiken, A. T. Nguyen, C. Viboud, and M. Santillana, "Toward the use of neural networks for influenza prediction at multiple spatial resolutions," *Science Advances*, vol. 7, no. 25, Jun. 2021, doi: 10.1126/sciadv.abb1237.
[48] C. Poirier *et al.*, "Influenza forecasting for French regions combining EHR, web and climatic data sources with a machine learning ensemble approach," *PLOS ONE*, vol. 16, no. 5, May 2021, doi: 10.1371/journal.pone.0250890.
[49] M. Soliman, V. Lyubchich, and Y. R. Gel, "Complementing the power of deep learning with statistical model fusion: Probabilistic forecasting of influenza in Dallas County, Texas, USA," *Epidemics*, vol. 28, Sep. 2019, doi: 10.1016/j.epidem.2019.05.004.
[50] I.-H. Kao and J.-W. Perng, "Early prediction of coronavirus disease epidemic severity in the contiguous United States based on deep learning," *Results in Physics*, vol. 25, Jun. 2021, doi: 10.1016/j.rinp.2021.104287.
[51] Nikhil, A. Saini, S. Panday, and N. Gupta, "Polynomial based linear regression model to predict COVID-19 cases," in *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, Aug. 2021, pp. 66–69, doi: 10.1109/RTEICT52294.2021.9574032.
[52] H. R. Pourghasemi *et al.*, "Assessment of the outbreak risk, mapping and infection behavior of COVID-19: Application of the autoregressive integrated-moving average (ARIMA) and polynomial models," *PLOS ONE*, vol. 15, no. 7, Jul. 2020, doi: 10.1371/journal.pone.0236238.
[53] C. Shen, A. Chen, C. Luo, J. Zhang, B. Feng, and W. Liao, "Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in Mainland China: Observational infoveillance study," *Journal of Medical Internet Research*, vol. 22, no. 5, May 2020, doi: 10.2196/19421.
[54] Y. Didi, A. Walha, M. Ben Halima, and A. Wali, "COVID-19 outbreak forecasting based on vaccine rates and tweets classification," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–16, Oct. 2022, doi: 10.1155/2022/4535541.
[55] H. R. Pourghasemi *et al.*, "Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020)," *International Journal of Infectious Diseases*, vol. 98, pp. 90–108, Sep. 2020, doi: 10.1016/j.ijid.2020.06.058.
[56] K. C. Santosh, "COVID-19 prediction models and unexploited data," *Journal of Medical Systems*, vol. 44, no. 9, Sep. 2020, doi: 10.1007/s10916-020-01645-z.
[57] S. S. Aljameel *et al.*, "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent covid-19 outbreaks in Saudi Arabia," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, 2021, doi: 10.3390/ijerph18010218.

[58] C.-S. Yu *et al.*, "A COVID-19 pandemic artificial intelligence-based system with deep learning forecasting and automatic statistical data acquisition: Development and implementation study," *Journal of Medical Internet Research*, vol. 23, no. 5, May 2021, doi: 10.2196/27806.
[59] S. Tiwari, S. Kumar, and K. Guleria, "Outbreak trends of coronavirus disease-2019 in India: A prediction," *Disaster Medicine and Public Health Preparedness*, vol. 14, no. 5, pp. e33–e38, Oct. 2020, doi: 10.1017/dmp.2020.115.
[60] Y. Yang *et al.*, "Reservoir hosts prediction for COVID-19 by hybrid transfer learning model," *Journal of Biomedical Informatics*, vol. 117, no. 286545, 2021, doi: 10.1016/j.jbi.2021.103736.
[61] I. Rahimi, A. H. Gandomi, P. G. Asteris, and F. Chen, "Analysis and prediction of COVID-19 using SIR, SEIQR, and machine learning models: Australia, Italy, and UK aases," *Information*, vol. 12, no. 3, Mar. 2021, doi: 10.3390/info12030109.
[62] S. Warnat-Herresthal *et al.*, "Swarm learning for decentralized and confidential clinical machine learning," *Nature*, vol. 594, no. 7862, pp. 265–270, Jun. 2021, doi: 10.1038/s41586-021-03583-3.
[63] A. Kunjir, D. Joshi, R. Chadha, T. Wadiwala, and V. Trikha, "A comparative study of predictive machine learning algorithms for COVID-19 trends and analysis," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2020, pp. 3407–3412, doi: 10.1109/SMC42975.2020.9282953.
[64] A. B. Gussow, N. Auslander, Y. I. Wolf, and E. V Koonin, "Prediction of the incubation period for COVID-19 and future virus disease outbreaks," *BMC Biology*, vol. 18, no. 1, Dec. 2020, doi: 10.1186/s12915-020-00919-9.
[65] T. Abdullahi and G. Nitschke, "Predicting disease outbreaks with climate data," in *2021 IEEE Congress on Evolutionary Computation (CEC)*, Jun. 2021, pp. 989–996, doi: 10.1109/CEC45853.2021.9504740.
[66] T. Abdullahi and G. Nitschke, "Disease outbreaks," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, Jul. 2021, pp. 205–206, doi: 10.1145/3449726.3459414.
[67] T. Abdullahi, G. Nitschke, and N. Sweijd, "Predicting diarrhoea outbreaks with climate change," *PLOS ONE*, vol. 17, no. 4, Apr. 2022, doi: 10.1371/journal.pone.0262008.
[68] Y. Du and Y. Guo, "Machine learning techniques and research framework in foodborne disease surveillance system," *Food Control*, vol. 131, Jan. 2022, doi: 10.1016/j.foodcont.2021.108448.
[69] S. Chen, J. Xu, L. Chen, X. Zhang, L. Zhang, and J. Li, "A regularization-based extreme gradient boosting approach in foodborne disease trend forecasting," *Studies in Health Technology and Informatics*, vol. 264, pp. 930–934, 2019, doi: 10.3233/SHTI190360.
[70] D. Tao, D. Zhang, R. Hu, E. Rundensteiner, and H. Feng, "Crowdsourcing and machine learning approaches for extracting entities indicating potential foodborne outbreaks from social media," *Scientific Reports*, vol. 11, no. 1, Nov. 2021, doi: 10.1038/s41598-021-00766-w.
[71] M. Zhang, D. Guo, J. Hu, and W. Jin, "Risk prediction and assessment of foodborne disease based on big data," in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management*, Nov. 2019, pp. 1–6, doi: 10.1145/3356998.3365776.
[72] T.-S. Ho *et al.*, "Comparing machine learning with case-control models to identify confirmed dengue cases," *PLOS Neglected Tropical Diseases*, vol. 14, no. 11, Nov. 2020, doi: 10.1371/journal.pntd.0008843.
[73] N. Ochida *et al.*, "Modeling present and future climate risk of dengue outbreak, a case study in New Caledonia," *Environmental Health*, vol. 21, no. 1, Dec. 2022, doi: 10.1186/s12940-022-00829-z.
[74] R. Jain, S. Sontisirikit, S. Iamsirithaworn, and H. Prendinger, "Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data," *BMC Infectious Diseases*, vol. 19, no. 1, Dec. 2019, doi: 10.1186/s12879-019-3874-x.
[75] S. Raizada, S. Mala, and A. Shankar, "Vector borne disease outbreak prediction by machine learning," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Oct. 2020, pp. 213–218, doi: 10.1109/ICSTCEE49637.2020.9277286.
[76] C. M. Benedum, K. M. Shea, H. E. Jenkins, L. Y. Kim, and N. Markuzon, "Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore," *PLOS Neglected Tropical Diseases*, vol. 14, no. 10, Oct. 2020, doi: 10.1371/journal.pntd.0008710.
[77] S. Anno *et al.*, "Spatiotemporal dengue fever hotspots associated with climatic factors in Taiwan including outbreak predictions based on machine-learning," *Geospatial Health*, vol. 14, no. 2, pp. 183–194, 2019, doi: 10.4081/gh.2019.771.
[78] F. Yavari Nejad and K. D. Varathan, "Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, 2021, doi: 10.1186/s12911-021-01493-y.
[79] N. A. M. Salim *et al.*, "Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques," *Scientific Reports*, vol. 11, no. 1, Jan. 2021, doi: 10.1038/s41598-020-79193-2.
[80] J. Ong *et al.*, "Mapping dengue risk in Singapore using random forest," *PLOS Neglected Tropical Diseases*, vol. 12, no. 6, Jun. 2018, doi: 10.1371/journal.pntd.0006587.
[81] C. M. Benedum, O. M. E. Seidahmed, E. A. B. Eltahir, and N. Markuzon, "Statistical modeling of the effect of rainfall flushing on dengue transmission in Singapore," *PLOS Neglected Tropical Diseases*, vol. 12, no. 12, Dec. 2018, doi: 10.1371/journal.pntd.0006935.
[82] J. Xu *et al.*, "Forecast of dengue cases in 20 Chinese cities based on the deep learning method," *International Journal of Environmental Research and Public Health*, vol. 17, no. 2, Jan. 2020, doi: 10.3390/ijerph17020453.
[83] S. F. McGough, L. Clemente, J. N. Kutz, and M. Santillana, "A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles," *Journal of The Royal Society Interface*, vol. 18, no. 179, Jun. 2021, doi: 10.1098/rsif.2020.1006.
[84] N. Zhao *et al.*, "Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia," *PLOS Neglected Tropical Diseases*, vol. 14, no. 9, Sep. 2020, doi: 10.1371/journal.pntd.0008056.
[85] O. M. Vidal *et al.*, "Chikungunya outbreak (2015) in the Colombian Caribbean: Latent classes and gender differences in virus infection," *PLOS Neglected Tropical Diseases*, vol. 14, no. 6, Jun. 2020, doi: 10.1371/journal.pntd.0008281.
[86] S. Saha *et al.*, "Unbiased metagenomic sequencing for pediatric meningitis in Bangladesh reveals neuroinvasive chikungunya virus outbreak and other unrealized pathogens," *mBio*, vol. 10, no. 6, Dec. 2019, doi: 10.1128/mBio.02877-19.
[87] E. Diez Benavente *et al.*, "A molecular barcode to inform the geographical origin and transmission dynamics of Plasmodium vivax malaria," *PLOS Genetics*, vol. 16, no. 2, Feb. 2020, doi: 10.1371/journal.pgen.1008576.
[88] P. Martineau *et al.*, "Predicting malaria outbreaks from sea surface temperature variability up to 9 months ahead in Limpopo, South Africa, using machine learning," *Frontiers in Public Health*, vol. 10, Aug. 2022, doi: 10.3389/fpubh.2022.962377.
[89] O. Nkiruka, R. Prasad, and O. Clement, "Prediction of malaria incidence using climate variability and machine learning," *Informatics in Medicine Unlocked*, vol. 22, 2021, doi: 10.1016/j.imu.2020.100508.

[90] J. Kim and I. Ahn, "Infectious disease outbreak prediction using media articles with machine learning models," *Scientific Reports*, vol. 11, no. 1, Feb. 2021, doi: 10.1038/s41598-021-83926-2.
[91] D. Jiang, M. Hao, F. Ding, J. Fu, and M. Li, "Mapping the transmission risk of Zika virus using machine learning models," *Acta Tropica*, vol. 185, pp. 391–399, 2018, doi: 10.1016/j.actatropica.2018.06.021.

BIOGRAPHIES OF AUTHORS

Gunadi Emmanuel received the B.Eng. degree in electrical engineering from the Sekolah Tinggi Nasional, Yogyakarta, Indonesia in 2001, and the master's degree in Information Technology from the Universitas Gadjah Mada, Yogyakarta, Indonesia in 2019. Everyday, he works as a lecturer in the informatics program, Faculty of Science and Technology, Musi Charitas Catholic University, Palembang, Indonesia. He is currently studying further at the Computer Science Doctoral Program at Bina Nusantara University, Jakarta, Indonesia. His current research interests include the application of artificial intelligence in the pharmaceutical field, e-learning, information technology, e-government. He can be contacted at email gunadi.emmanuel@binus.ac.id.

Arief Ramadhan works as a Lecturer and Researcher at School of Computing, Telkom University, Bandung, Indonesia. He is also a Senior Member of IEEE. His areas of research focus include information technology, gamification, metaverse, information systems, data analytics, enterprise architecture, e-business, e-government, e-learning, e-tourism, and business intelligence. Currently, he is supervising doctoral students who are interested in this field of study. He can be contacted at email arieframadhan@telkomuniversity.ac.id.

Muhammad Zarlis is a Lecturer in the Information Systems Management Study Program at Bina Nusantara University. He completed his Bachelor of Science in Physics from USU's Faculty of Mathematics & Natural Sciences (FMIPA) in 1984. He continued his Master's in Computer Science at the University of Indonesia's Postgraduate Program (UI) with a year-long Sandwich Program at the University of Maryland (UoM) USA (1990). Then, he completed his Doctoral Program (Ph.D.) study in Computer Science at Universiti Sains Malaysia (USM) Pulau Pinang Malaysia in 2002. He has been a Professor of Computer/Information Science since 2009. He is active in services and research in the field of intelligent systems, computer security, and calculations. In 2006 and 2017, he received the Satyalancana Karya Satya Award for 20 and 30 years of Honor from the President of the Republic of Indonesia. He can be contacted at email muhammad.zarlis@binus.edu.

Edi Abdurachman is working as Chair of the Transportation and Logistics Management Doctoral Program at the Trisakti Institute of Transportation and Logistics. Inauguration as Professor of Statistics in 2008 at Bina Nusantara University. He holds a Bachelor of Science degree from the Bogor Agricultural Institute, Indonesia (1979), and a Master of Science in Statistics (1981) from the same University. Later, he also obtained his second Master's Degree in Statistics from Iowa State University - USA, and his Ph.D. degree in Statistics from the same University. He is also a Thesis Advisor and Dissertation Promoter. His research interest is applied statistics for behavioral science and management. He can be contacted at email edia@binus.ac.id.

Agung Trisetyarso is the Head of Concentration in Computer Science, Department of Doctor of Computer Science at Bina Nusantara University. He completed his Bachelor of Science degree (2000) and achieved a Master of Science degree from the Department of Physics, Bandung Institute of Technology (2002). He completed his doctoral program in quantum information and computation at Keio University, Japan (2011). Before teaching at Bina Nusantara University, he was a Teaching Staff at the Informatics Department of Telkom University (2011-2015). He can be contacted at email atrisetyarso@binus.edu.