Potential Variables in Pharmaceutical Drug Prediction Research with Machine Learning Approach: A Literature Review

Gunadi Emmanuel[1], Yulyani Arifin[2], Ilvico Sonata[3], Muhammad Zarlis[4]
Computer Science Department-Binus Graduate Program-Doctor of Computer Science,
Bina Nusantara University, Jakarta, Indonesia[1, 2, 3]
Information System Management Department-Binus Graduate Program-Master of Information System Management,
Bina Nusantara University, Jakarta, Indonesia[4]

***Abstract*—As a downstream component of the drug supply chain, pharmaceutical installations often face uncertainty in drug demand. Predicting pharmaceutical drugs using a machine learning approach enables the development of new variables that can enhance the performance of medicine prediction. Amidst limited data and a choice of prediction algorithms, the accuracy of variable selection is significant for drug prediction performance. This study remaps the scope of variables from previous studies related to drug demand prediction and machine learning performance to develop further significant variables. Investigating research literature on significant variables in drug demand prediction with machine learning models published in 2020-2024. The systematic literature methodology uses the Kitchenham method. Mapping problems, discussion areas, and data availability result in ten categories of issue areas, each with its respective data needs and algorithm choices. A qualitative exploration of issue areas identifies potential variables for pharmaceutical drug prediction, including drug consumption, epidemiology, drug management, supply chain-patient domicile, and pharmacotherapy. Mapping potential variables facilitates the availability and integration of data relevant to local or regional characteristics, enabling further research on the characteristics of data and algorithm choices.**

Keywords—Drug demand; machine learning; pharmaceutical installations; prediction; potential variables

I. Introduction

Drug availability is a crucial issue during the COVID-19 Pandemic. Hospital management and health services need to build an optimal drug inventory control system in the future. Amid the uncertainty of drug demand, the planning and procurement process becomes a determining stage in pharmaceutical logistics operational management [1]–[3]. The drug planning and procurement process in hospital pharmacy installations is related to the drug supply chain from pharmaceutical wholesalers (distributors) [4], [5].

Both manufacturers and distributors require high downstream demand prediction results to form the basis for drug supply planning [6]. Accurate downstream planning supports the analysis of supply chain visibility, lead time, inventory status, and supplier performance. Pharmaceutical logistics can calculate drug availability variables, lead time, and drug costs from suppliers [7]. These results will improve internal pharmaceutical supply chain performance and optimize inventory management and distribution activities [8]. The accuracy of demand predictions generated by algorithms anticipates elements of uncertainty or fluctuations in drug demand [7], [9]. In the supply chain, drug demand uncertainty is a category of external uncertainty and risk. These external risks are identified from events that deviate from the general pattern and may change the prediction performance [9], [10].

In the pharmaceutical industry, stakeholders must identify external variables of drugs in the demand prediction process. Prediction results do not only anticipate events that disrupt drug distribution and provision planning downstream in the supply chain. Furthermore, the results of medicine demand prediction maintain the balance of supply against drug demand and mitigate against events that deviate from the general pattern in determining the amount and pattern of consumption, epidemiology, and available budget [11], [12]. Stakeholders can leverage machine learning models for prediction, saving time and costs, and allowing for a more focused investment in clinical trials [13]. Machine learning models can capture variables of changing market conditions and consumer demand. However, machine learning requires the availability of good, reliable, and real-time features and data to achieve high performance [14].

Beyond drug variables, several variables in drug prediction may contribute to changes in drug demand, including strong seasonal patterns, uncertain demand, and a lack of historical data [15]. Each variable involved in the prediction may have different data characteristics. Prediction variables can exhibit non-linear correlations with several parameters of their dependent variables, such as historical sales data in conjunction with demographic and epidemiological data (including disease type, disease progression, or patient distribution level) [16].

The greater the data variation, the more the machine learning model obtains aspects of demand variability. The model can predict consumption frequency, consumption amounts over time, and consumption patterns based on geographic areas [17]. Increasing data dimensions, computational complexity, storage requirements, the ambiguous or noisy nature of data, and device capacity also demand higher model performance [18]. Among the many features used, machine learning will identify several significant ones. Dimensionality reduction can impact the performance, complexity, time, and space requirements in machine learning. Similarly, some irrelevant or redundant features can significantly reduce classifier performance [19].

The selection of variables from epidemiological data of dengue fever patients considers variables that affect the development of dengue fever. In their research, Syahrir [20] built a model based on the development variables of each stage of the patient's disease. The model known as susceptible-expose-infectious-removal (SEIR) is used to predict drug needs in the face of dengue fever outbreaks in Indonesia. Based on data on the number of hospitalized patients with dengue fever from 2014 to 2016, the prediction model employed a Moving Average (MA), regression, and Exponential Smoothing (ES). In subsequent studies, they developed Single Exponential Smoothing (SES), Double Exponential Smoothing (DES), Linear Regression (LR), Quadratic regression, Bayesian regression, the winters method, Autoregressive Integrated Moving Average (ARIMA), and the SEIR model for prediction [21]. They predicted the number of patients and drug demand based on time variables while ignoring other epidemiological variables, such as climate and environmental health.

Meanwhile, Ho [22] employed machine learning algorithms, including the Decision Tree (DT), Deep Neural Network (DNN), and Logistic Regression, to identify cases of dengue fever. Using the algorithm, they identified four primary variables: age, body temperature, white blood cell count, and platelet count. Then, based on the doctor's recommendations and epidemiological characteristics, they added Hemoglobin (Hb) and gender parameters. The two additional parameters did not improve the prediction performance. With four main variables, the performance of the DT, DNN, and logistic regression models provided a sensitivity of 90%. Other research on dengue cases utilised climate variables [23]–[25], dengue case data, population demographics, entomology, and environmental factors [26], [27]. The use or addition of specific variables is not necessarily significant to the performance of the machine learning model until high prediction performance is validated.

In the development of the Earth Cognitive System for COVID-19 (ECO4CO) information system for handling the COVID-19 Pandemic, the display of notification services, alerts, and scheduled communications, as well as geographic localization for end-users, supports logistics management of the Pandemic's drug supply [28]. In general, the pharmaceutical industry can further optimize drug planning and procurement. Specifically, ECO4CO helps aggregate data in information systems to help pharmaceutical installations predict drug needs [29]. Integrated data will facilitate the acquisition of significant variables for prediction. Government policy encouragement, especially in developing countries, can accelerate the development of integrated health information systems, particularly when potential predictive variables are sourced from multiple government and private institutions.

Health disruption situations, such as the COVID-19 Pandemic, require efforts to identify variables that are not yet clearly significant (referred to as hidden variables) in building drug prediction models. Machine learning can help identify hidden variables and pinpoint significant factors in drug demand, even outside of historical data on drug demand or sales [9]. According to Koala [30], significant variables for predicting drug demand can vary from one case or place to another. If the prediction refers to changes in consumer attitudes, the time series approach will influence the determination of significant variables. Machine learning can perform a clustering analysis approach to identify customer demand behavior. In addition, regression analysis produces continuous-valued functions for products that cannot be stored for long in storage [31]. These approaches direct significant variables for predicting drug demand.

Researchers generally use correlation testing between predictive features to identify significant variables in machine learning models. In the case of the COVID-19 Pandemic, many studies have used new variables from the field. The use of new variables highlights the limited scope of previous similar research, particularly in identifying potentially significant variables in predicting drug demand in pharmaceuticals using a machine learning approach. Mapping previous research areas can contribute to the study of developing machine learning models for predicting drug demand in the pharmaceutical industry, utilizing potentially significant variables. Therefore, this study analyzes research related to pharmaceutical drug prediction using a Machine Learning approach. The study redefines the scope and area of variables in drug demand prediction research.

Mapping the use of variables in previous research suggests the direction for developing a machine learning approach to predict pharmaceutical drug demand. The implications of this study can encourage the government to develop policies related to the use of drug supply calculation methods and significant prediction variables. At the same time, accelerating the integration of national health data into the National Health Network increases the availability of data and the variability of drug prediction variables [32]. To sharpen the direction of research on potential variables in drug prediction with this machine learning approach, researchers formulated research objectives (RO): RO1. Identifying areas of drug demand, research issues, and data availability for prediction from previous studies. RO2. Examining the performance of using machine learning algorithms for drug demand prediction. RO3. Categorizing potential variables to be developed in predicting drug demand.

Next, we discuss literature studies related to drug demand design in pharmacy and machine learning methods commonly used to predict drug demand. Next, we will describe the systematic literature review method. Next, the results and discussions related to research issue areas and finding variables. The final section is the conclusion and the direction of further studies related to drug demand variables.

II. Related Studies

Generally, drug demand calculations in pharmacies utilize internal inventory data. Drug demand calculations are central to drug management and the pharmaceutical supply chain. Drug supply concerns often center on drug stock and demand. In a study of drug planning in several pharmaceutical institutions, Burinskiene [33] utilized historical drug sales data from multiple pharmacies within a chain to compare prediction performance results at each pharmacy location. The prediction target was to improve accuracy at each pharmacy point. The computational process took into account seasonal data trends, outlier handling, and the handling of missing data. However, differences in pharmacy locations with their variable trends were not a consideration in the computation. In the field, medication supply planning considers many variables and data sources that can support data analysis and prediction. These data include variables such as historical demand, patient demographics, clinical information from Electronic Health Records (EHRs), epidemiology, supply chain, and external elements such as market dynamics and regulations [6].

Information on the extent of drug demand and the variables that influence changes in drug demand are external factors that have not received much attention. Drug demand prediction has a close causal correlation with drug consumption patterns [30]. Correlating external factors with increasing demand prediction accuracy requires careful analysis of market trends, customer behavior, suppliers, and technology [31]. According to Koala [30] factors that can influence drug demand prediction include socioeconomic factors such as gender, age, patient origin, annual income, and social class. Second, patient health factors include comorbidities, insurance, and lifestyle (smoking, alcohol, and obesity). The third factor includes institutional facilities and healthcare personnel. However, Koala does not provide a basis for external variables that remain attached to the dynamics of drug demand in a time series sequence.

Similarly, the behavior of drug supply chain actors focuses on predicting the public's need or demand for drugs to support health services, particularly timely access to drugs in the right amount, type, and quality [34], [35]. Supply chain management strives to ensure that provision and availability match demand [36]. Therefore, analyzing demand prediction variables is necessary to anticipate changes in drug demand [37].

Periods of increased risk of disease spread, such as dengue fever, can change drug demand during specific periods [34]. Likewise, in disaster areas, factors such as limited data on the number of disaster victims, distribution, and the increasing number of victims, and substitution of certain types of drugs (which cannot be replaced by others) can add variables to the prediction of drug needs [20]. In dealing with these issues, the machine learning approach can handle the problems of interdependent variables, complexity of causal factors, and non-linear relationship patterns [36]. Additionally, machine learning can analyze unstructured data sets and identify significant factors that impact supply chain performance [15]. In the case of predicting drug sales volume using sales promotion stimuli, Al-Gunaid [34] compared the performance of five algorithms (LR, Random Forest (RF), Neural Network, Support Vector Regression (SVR), and the Levenberg-Marquardt algorithm). During the analysis, previously unseen factors may emerge due to the increasing correlation between features. Marketing promotions have a high correlation coefficient because they are closely related to variables that influence consumption behavior. Consequently, sales predictions increase. However, the promotion variable as a planned stimulus for supply chain actors is only a small part of the other variables that influence the performance of drug demand prediction.

Dynamic market changes reveal hidden factors that necessitate methods to capture these factors, thereby enhancing data-driven predictions [38], [39]. Machine learning learns from all available data, including environmental parameters, to make informed decisions. Each time new historical data becomes available, the learning process continues to occur, enabling the making of new decisions [40]. Machine learning cannot solve all problems, but it can improve accuracy in certain cases and datasets [41]. Several machine learning algorithms are available for demand prediction based on historical data, including random forests, gradient boosting, and neural networks. Note that environmental changes can affect the performance of prediction algorithms over time [6]. A non-negligible variable is that drug evaluation often results in switching to different drugs to enhance pharmacological effects or mitigate side effects [34].

III. METHODOLOGY

Machine learning has become a widely used method for predicting demand for goods, including in the pharmaceutical sector, specifically for drug demand [30], [42]–[45]. This study examines the mapping of the scope and area of research variables for predicting drug demand in machine learning models. For this reason, this research uses a qualitative literature study method. A systematic review allows researchers to analyze the phenomenon of research on the use of drug demand prediction variables based on machine learning models. This literature study method adapts the steps of the Kitchenham literature review method, except that the data synthesis section is replaced by an analysis of the results and concludes with a report on the study results [46]. Fig. 1 shows the steps of this systematic review research.

1) Research identification: Uncertainty in drug demand necessitates the use of an accurate machine-learning approach with variables that are truly significant for predicting performance. Therefore, the investigation requires as many original articles from previous studies as possible. The original articles targeted are primary research related to machine learning models and drug prediction. The identification of potential variables in previous studies is based on the study's problem area, the context of the problem, data availability, and the performance of the machine learning algorithm used. Meanwhile, the scope of the research is multidisciplinary, covering drug supply chain information systems, pharmacy, and informatics. Therefore, the discussion of the research follows the questions:

- RQ1. What are the areas of drug demand research issues and data available for prediction from previous research?
- RQ2. How is the machine learning algorithm performed for drug demand prediction?
- RQ3. What are the potential variable findings from the selected articles identified in the categories that can build drug predictions with machine learning models?

Furthermore, this research requires quality primary (original) research articles. For this reason, article exploration is conducted in digital libraries, including Scopus, PubMed, and Google Scholar. Researchers use the Publish and Perish 8 Application to search for research literature sources. We also use Mendeley Desktop for documentation and article processing. The selection of search keywords refers to the research questions. Article searches on the application combine keywords in three searches by compiling search keywords for each search, namely: “drug demand AND prediction method AND machine learning”; “prediction AND drug AND pharmaceutical AND demand”; and “medicine OR pharmaceutical OR drug demand AND prediction method AND machine learning”.

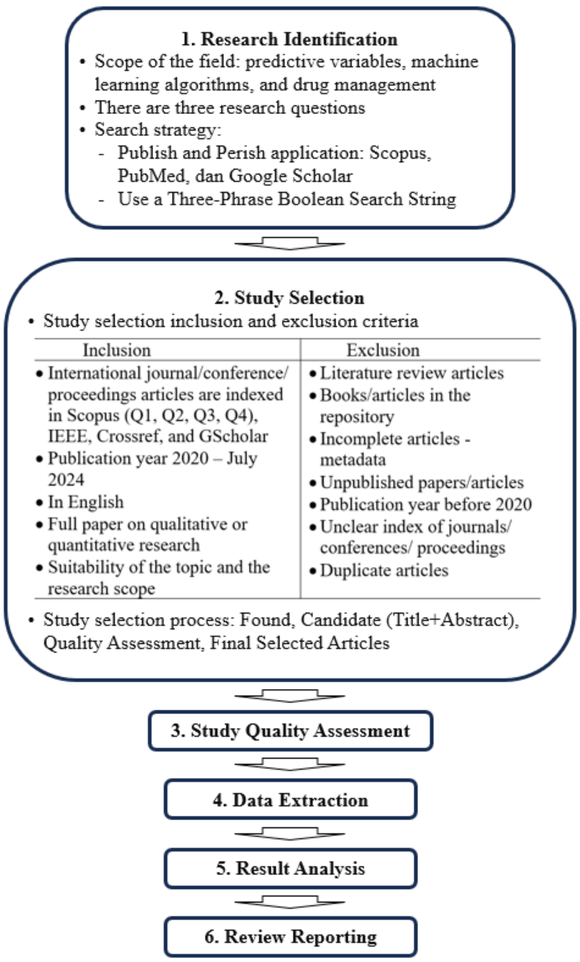

Fig. 1. Research stages - adapted from kitchenham [47], [46].

2) Study selection: After obtaining the articles, the following article selection process involves reading the title, keywords, and abstract to filter out articles that do not match the topic. The method of skimming the articles ensures that the discussion of the article aligns with the research criteria. Furthermore, the fulfillment of the inclusion and exclusion criteria helps in obtaining articles that align with the research objectives. The study requires that selected articles analyzed have a similar research scope, research design, and type of research. The similarity of selected articles suggests that the research articles are closely related to this study. The article has complete data regarding the information needed in this study. The article includes objective and valid measurement techniques for the research model. Differences in methods in presenting research results will follow the objectives of each study. In machine learning research, it is possible to use the same data sources, allowing this literature study to incorporate research literature sources that utilize the same data. Research questions help the stages of analysis of articles according to research topics.

3) Research quality assessment: The researcher evaluates whether the journal's quartile level can ensure a minimum level of study quality. However, the researcher does not group the quality of the articles. The researcher limits observations to variables used to reduce bias in the analysis.

4) Data extraction: The search results for selected articles can be presented in tables or graphs containing the researchers' names, the year of research, problems, data availability, research methods/designs, research variables, and research results. Selected articles can answer research questions that contain the name and year of research, the algorithm used, and the variables used in machine learning. Likewise, the order of appearance follows the discussion of the findings.

5) Result analysis: The analysis of the selected article review results is qualitative in nature. The findings are presented in tables and figures. The analysis of the results, including the research issue area, data availability, and machine learning model performance, leads to the mapping of potential variables to predict drug demand. This presentation is a form of descriptive research synthesis.

6) Review reporting: The final activity of this research is to communicate the results and contributions of the research effectively in international scientific journals. The reporting of this literature study is directed primarily at journals in the areas of computer science or information systems. This research topic is very close to the selection of variables used in machine learning studies.

IV. Results and Discussion

The first part of this session presents the results of the article selection process, quality article publishers, and the number of article topics that tend to increase. Next, we will present the results in the order of the research questions before discussing them. Mapping the problem areas of the selected articles and the variables used to predict drug demand helps organize and systematize the discussion into several subsections.

A. Selected Articles

The use of the Publish and Perish Application makes it easier to search for selected articles and confirms that these articles meet the qualifications as study material. The process of searching for articles with three keyword phrases obtained 886 articles. Selection according to the research stages follows the keyword phrase groups. At the quality assessment stage, the researcher eliminated duplicate articles. Ultimately, the article selection process yielded 23 selected articles (Table I).

TABLE I. STUDY ARTICLE SELECTION PROCESS

Boolean search string	Found	Candidate *Title+Abstract*	Quality assessment	Final selected
1) "Drug Demand AND Prediction Method AND Machine Learning".	277	64	13	11
2) "Prediction AND Drug AND Pharmaceutical AND Demand"	279	45	15	4
3) "Medicine OR Pharmaceutical OR Drug Demand AND Prediction Method AND Machine Learning"	330	39	12	8
Total	**886**	**148**	**40**	**23**

Determining the scope of the study limits the number of articles that can be found before they are analyzed. Fig. 2 presents the mapping of twenty-one selected articles from the library database. The composition of the articles shows that the selected articles have met international standards and are worthy of analysis. The publisher is also reputable for publishing high-quality papers.

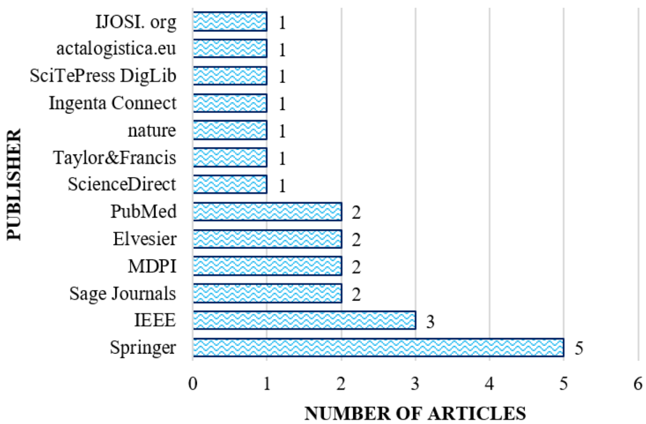

Fig. 2. Composition of selected article publications and the library database.

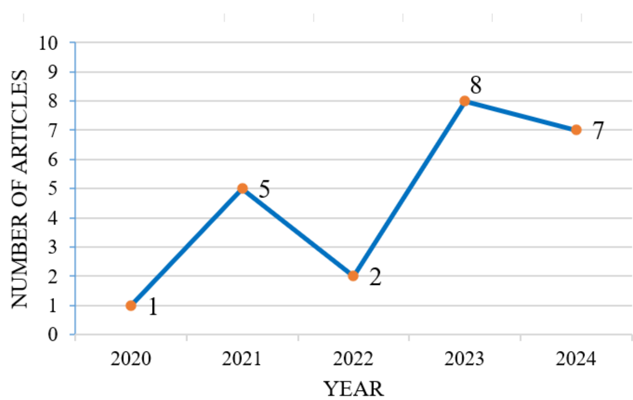

Fig. 3. Number of selected article publications 2020 - July 2024.

This study examines articles published between 2020 and July 2024 to demonstrate the novelty of the research and the article. During that period, the findings of the research topics on drug demand prediction and machine learning were quite encouraging, as the number of related research studies reached more than 50% compared to the previous year, as shown in Fig 3. This finding is interesting because there has been research interest in drug availability in pharmaceutical installations over the past two years.

B. Research Issue Areas and Data Availability (RQ1)

The answer to the first question (RQ1) of this study is completed by presenting the research issue areas and data availability from selected articles. Issues and data availability represent the perspectives of previous researchers on drug prediction to maintain drug availability in pharmacies. The issues raised by each researcher can show the perspective used to solve the problem. This study identified ten research issue areas, as discussed in Table II. Several research issue areas have relationships with others (combination of areas). Subudhi's research issue area [48] has a combination of issue areas, including Epidemiology, Patient Medical Record Data, and Pharmacy Drug Management. We learn to observe the correlation of the problem context and the availability of data that can be used to solve it. The combination of research issue areas raises discussions about data availability and the central point of the problem to be solved using machine learning.

Six selected articles discussing the research issue areas of drug supply chain and usage/sales history in Table II. Data availability in these areas includes discussions of time series, historical market consumption data, and inventory data. Real-time and integrated data are research challenges in the supply chain and usage/sales history area. Interestingly, the availability of data that affects the dynamics of drug demand has not been developed, particularly about disease factors (epidemiology) that determine the magnitude of drug demand. In addition to standing alone, epidemiological variables are also related to the research issue areas of Geography, Medical Records, Pharmaceutical Drug Management, and Disasters. In these areas, epidemiological variables are also considered to influence drug demand. In these areas, epidemiological variables are also considered to influence the prediction of drug demand.

The correlation between research issue areas requires an analysis of the significance of the predictive variables. The same variable that is not significant for drug prediction can change significantly, and vice versa. The combination of ten research issue areas illustrates the complexity of drug management. The more combinations of issue areas, the higher the data dimension, thus requiring more careful data handling for machine learning.

TABLE II. ISSUE AREAS, PROBLEMS, AND DATA AVAILABILITY

Research Issue Area	Problem Context	Data Availability	Ref.
Medical data record - pharmaceutical drug management	– The development of the number, diversity, and sources of data requires an appropriate prediction model to improve accuracy. – Inappropriate determination of significant variables in production, consumption, and storage stock predictions affects prediction accuracy.	– The electronic logistic management information system (eLMIS); for a five-year (2015–2019); available monthly or quarterly. Scope of research: Regional - Local Area. – Kerman Hospital data consisting of 283 features, covering more than 9351 different drugs and 121,690 patients (three years). Scope of research: Pharmacy.	[17], [49]
Epidemiology	– The level of emergency and spread of the disease affect the allocation of drugs. – The different and complex handling, distribution, storage, and administration methods for each drug make it a challenge to predict drug availability.	– Three major pharmaceutical companies producing insulin in the US. Scope of research: National companies. – Internal hospital data from Fast Healthcare Interoperability Resources (FHIR). Data collected from 2017 and 2021. Scope of research: Blood bank.	[55], [56]
Epidemiology - disaster	– The uncertainty of the impact of a disaster requires the prediction of drug supplies with a fluctuating cases number.	– Data on the number of COVID-19 infections in Shanghai. Period 20 January 2020 - 24 September 2022; Emergency Unit Scope of research: Hospital	[57]
Epidemiology - medical data record - pharmaceutical drug management	– Prediction models with the complexity of patient medical conditions, disease characteristics, and drug availability require more than one data source. The difficulty of designing a drug demand prediction model, especially for emergency units.	– Medical Data Record from the Mass General Brigham (MGB) Healthcare database. Emergency units from 14 hospitals. Scope of research: Local – Regional Hospital	[48]
Epidemiology - geographic	– For high-spread diseases such as COVID-19, epidemiological analysis becomes an important variable in predicting drug needs related to geographic distribution (from district level to national level).	– Data from: Lazio region health system (open datasets), Italian Medicine Agency (AIFA) (medicines and equipment), Italian National Civil Protection dataset (patient care/treatment - regional level), National Agency of Regional Health Services (AGENAS) (medical facilities). Scope of research: regional health services	[28]
History of usage/sales data	– Design of accurate drug demand prediction model for the needs of the drug supply pattern. – There is a demand for drugs/drug groups following seasonal trends, a prediction model is needed that is suitable for time series. – The need for a prediction approach method for complex pharmaceutical sales data sets.	– Monthly Sales Dataset from Kaggle. Transaction data of 57 drugs. – Weekly sales data for pharmaceutical products from the Kaggle. Transactions of 57 drugs with 600,000 transactions (2014–2019). Scope of research: algorithm comparison, Pharmacy, improving accuracy with different algorithmic approaches. – Historical sales data and market trends. A dataset of 600,000 transaction records from Kaggle (2014–2019). Scope of research: public data – eLMIS public health – Rwanda: to perform demand prediction of medicine. – Data from the analytical company AMS (Market Data in the Pharmaceutical Industry) in the Middle East and North Africa region, especially in Iraq. Scope of research: Time series dynamic	[35], [58]–[62]
History of usage/sales data - geographic	– A model approach to drug demand that does not only occur in one geographical point but also with the nearest drug demand point. – Non-communicable diseases have characteristics that require optimization of the management of essential drugs for non-communicable diseases (consumption or distribution).	– The National Health Service (NHS) in England Data. Prescription data in each practice August 2010 - December 2019. Scope of research: Adjacent pharmacy. – The electronic logistic management information system (eLMIS) data (2015–2019). Scope of research: Regional.	[63], [64]
Pharmaceutical drug management	– Enterprise planning (ERP) systems require an accurate demand forecasting model.	– Actual six different drug classes sales data (from 6 pharmacies in Pakistan). Data was collected from sources including injections, tablets, capsules, and suspensions of various potencies. Scope of research: Pharmaceuticals.	[65]
Drugs - patient behavior (self-medication)	– Patients have difficulty finding medicine for their illnesses independently (self-medication).	– The UCI Machine Learning Repository data - Drug Review with the subject Health and Medicine 6 features. Scope of research: Public data	[66]
Drug supply chain	– Ignoring that each region and demographic has its own drug needs – Manufacturers need accurate prediction of drug demand from downstream. On the other hand, pharmacies do not have visibility of drug supply – Data limitations: the real-time and patient prescription data for demand prediction – Data limitations, many hidden factors, changing market situations	– Sales data from PharmaGuide covering 22 pharmacies in Ontario, Canada (2014–2021) and historical drug shortage data from the Drug Shortages Canada website. Scope of research: Pharmacies – Simulation for application with sales data collected during 2014–2019. Scope of research: Pharmacies – Data from electronic data interchange (EDI). Weekly data on distributor-level sales volumes and inventory levels for 133 unique national drug codes (NDCs) from July 2007–August	[9], [50], [67]–[70]

Research Issue Area	Problem Context	Data Availability	Ref.
	– Variables determining demand prediction in the drug supply chain – Method of grouping and classifying products at the downstream level to improve the accuracy of drug demand prediction.	2017. Scope of research: Distributors – Database from an integrated services company, Coopservice Group, Italy. Scope of research: National companies – Data available for 519 samples containing historical information for 43 product types. Scope of research: Pharmacies	

Data availability relies on pharmaceutical logistics data sources, patient medical records, health services, independent health service units or hospitals, drug manufacturers and distributors, as well as pharmacies or hospital pharmacies. Complete and quality data helps find correlations between data variables. Complete and high-quality data help find correlations between data variables. The available dataset builds the initial assumption of the prediction variables. Nabizadeh [49] attempted to gather data sources from all departments but found the effectiveness of prediction to be limited when considering only one department. In contrast, Pall [50] used drug sales data from 22 pharmacies and historical drug shortage report data from Health Canada to obtain prediction results that help estimate drug demand. From this phenomenon, the correlation between the research issue area and the variables used determines the effectiveness of the available dataset.

The ten research area issues are divided into two categories. First, the main category encompasses areas such as pharmaceutical drug management, drug supply chain, and epidemiology. Second, additional categories occur due to research needs involving variables and data in geographic areas, medical records, and patient behaviour. Likewise, epidemiology can also develop in areas such as climatology and disaster management. Historical consumption and sales data can also involve areas such as supply chain management or pharmaceutical drug management. According to Mbonyinshuti [17], historical data on essential drug consumption extracted from eLMIS is an area of pharmaceutical drug management. In addition to historical data, patient origin data from medical records can help determine the geographic location of patients who require essential drugs. Finally, we understand that many studies predict the spread of COVID-19 to help distribute drugs more precisely [28], [51]–[54].

In the Epidemiology - Disaster, the hospital emergency unit is a frontline unit that requires drug resources according to the cases that occur. There is a correlation between disease prediction and the prediction of the type and amount of drugs required to treat the disease. Correlation can be observed in the analysis of patient data, diseases, and the medications they consume. Patient medical record data is a significant source for predicting patient drug needs [48]. The network of emergency units between many hospitals can produce data on the spread of infectious diseases as a significant variable. Collecting prescription data from several health service points provides information for analyzing drug demand in a particular distribution area [49], [64], [68]. With this analysis, distributors can formulate a drug distribution chain strategy. Drug demand information (from prescriptions) in the blockchain network can be used to identify the source of drug demand [71].

C. *Machine Learning Algorithms for Drug Demand Prediction (RQ2)*

The literature of this study, as shown in Fig. 4, indicates that ARIMA, LR, RF, eXtreme Gradient Boosting (XGBoost), K-Nearest Neighbours, Support Vector Machine (SVM), ANN, multi-layer perceptrons (MLP), and LSTM algorithms are employed in more than two research studies. These algorithms are widely recognized for their performance. The selection of prediction algorithms is closely related to the performance and working method of the algorithm.

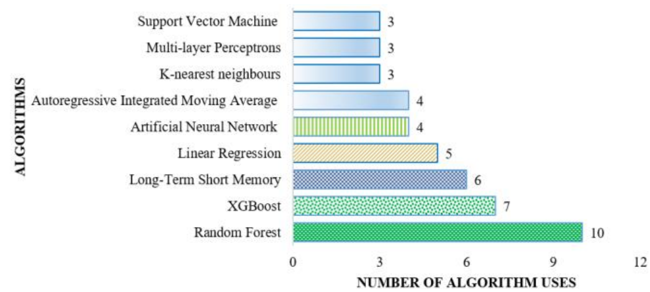

Fig. 4. The most widely used algorithm sequence in machine learning studies.

Additionally, the selection of machine learning algorithms is correlated with data availability factors and significant variables in prediction models. The choice of deep learning algorithms, such as ANN, MLP, and LSTM, has not dominated shallow machine learning algorithms. The complexity of the prediction research issue area is evident in its limited scope. To answer RQ2, Table III presents the performance of machine learning algorithm implementations for drug prediction based on selected articles.

Comparing multiple machine learning algorithms to improve prediction accuracy is a limited research area. The development of methods to handle cases of research area combinations presents an opportunity to enhance machine learning algorithms. The Machine Learning approach can simultaneously combine learning algorithms for big data and several causal variables with non-linear relationships. Machine Learning approaches can simultaneously combine learning algorithms for big data and several causal variables with non-linear relationships [36]. The study by Subudhi [48] took place in the Intensive Care Unit (ICU). The epidemiology, medical data record, and pharmaceutical drug management areas encouraged the use of ensemble models to handle cases of multi-source data. The average F1 score prediction results were higher than 0.83. Likewise, in the study by Zhang and Li [57], the disaster and epidemiology areas face challenges in processing disaster data and handling diseases that arise from disasters. The combination method produces R2= 0.9996 for the BILSTM-GASVR combination prediction model.

TABLE III. MEASUREMENT OF MACHINE LEARNING ALGORITHM PERFORMANCE IN SELECTED ARTICLES

Research issue area	Algorithm	Measurement	Ref.
Medical data record - pharmaceutical drug management	Linear regression, RF, ANN	RF (accuracy 76%)	[17]
	RF, neural network (NN)	RF model accuracy 83.3% (for a month) and 81.3% (for a season - two years data)	[49]
Epidemiology	RF, multiple regression, ANNs	ANN (the correlation coefficient value between output and response value is 93,1%)	[55]
	XGBoost, RF, dummy classifier, and Optuna	XGBoost (Accuracy 40,4%) (Specificity 98,75%)	[56]
Epidemiology - disaster	Bilateral long-short-term memory and genetic algorithm support vector regression (BILSTM-GASVR), bilateral long-short-term memory (BILSTM), genetic algorithm support vector regression (GASVR), informer	BILSTM-GASVR combined prediction model (R^2=99,96%)	[57]
Epidemiology - medical data record - pharmaceutical drug management	AdaBoost, bagging, gradient boosting, RF, XGBoost, ExtraTrees, logistic regression, decision tree, linear discriminant analysis, quadratic discriminant analysis, MLP Classifier, perceptron, passive AggressiveClassifier, linear SVC, Gaussian NB, K-neighbors, naive Bayes, SVM, NN	ICU admission: ensemble and tree-based models had significantly higher F1 scores than all other model types. Mortality: all ensemble-based models had average F1 scores higher than 0.83.	[48]
Epidemiology - geographic	Linear regression, K-nearest neighbours (KNN), recurrent neural networks (RNN)	The medicinal purchases prediction for the regions of Lazio and Piemonte has yielded a low RMSE 16%	[28]
History of usage/sales data	Linear regression, RF, SVM, XGBoost	XGBoost (accuracy: > 95% every drug sale)	[35]
	Probabilistic neural network, generalized regression neural network, radial basis function neural network (RBF-NN), long-term short memory neural network, stacked LSTM	Shallow neural networks. RMSE: Probabilistic Neural Network (6,21), Generalized Regression Neural Network (6,01), RBF-NN (6,59)	[58]
	ARIMA, LSTM	ARIMA (short term prediction), LSTM (long term and complex patterns)	[62]
	Model Bayesian, regresi proses gaussian (GPR): exponential squared, Revised Mat´ern, Rational Quadratic kernels	Kernel Ensemble MSE ($9,09 \times 10^{-21}$), MAE ($7,48 \times 10^{-11}$), RMSE ($9,53 \times 10^{-11}$), Score R^2 (1,0)	[60]
	ARIMA, LSTM	The best LSTM accuracy RMSE 2.0 (training) and 2.043 (testing); R^2=0.952 (training) and 0.912 (testing).	[61]
	Multi-layer perceptrons (MLP), convolutional neural network (CNN), LSTM	The best sales LSTM performance with RMSE=1.28(k) and MEA=0.85(k); for USD Price RMSE=0.75, and MEA=0.44.	[59]
History of usage/sales data - geographic	ARIMA, XGBoost, gate recurrent unit (GRU), graph convolutional network (GCN), spatiotemporal graph convolutional network (STGCN), attention-based spatial-temporal graph convolutional network (ASTGCN)	ASTGCN (MAE 0,0957, RMSE 0,1241 – short period), STGCN (MAE 0,1139 RMSE 0,1465 – long period)	[64]
	RF	Accuracy: R-square 0.78 (training) and 0.71 (testing)	[63]
Pharmaceutical drug management	Autoregressive integrated moving average (ARIMA), Holts Winter, ARHOW	ARHOW (RMSE 18; MAPE 1,99)	[65]
Drugs - patient behavior (self-medication)	RF, naïve Bayes, long-short term memory	RF (Accuracy 83%)	[66]
Drug supply chain	Facebook prophet, LSTM neural networks, and XGBoost	XGBoost (has the lowest MAPE value of 5 out of 8 drug sales)	[69]
	XGBoost	able to address 59% of major drug shortages	[50]
	XGBoost	has the lowest MSE of 6 out of 8 types of drugs	[68]
	LR, moving average (MA), SVR, RF, basic exponential smoothing (ES), RNN	RNN (NMSE 2.69±0.37; NMAE 0.30±0.06)	[9]
	LR, RF, KNN, SVM, MLP	SVM (MSE 2,02 Medicine; 1,89 month)	[70]
	K-Means, Elbow method, Classification And Regression Trees (CART), MLP- neural network, RBF-NN, polynomial (or group method of data handling (GMDH) neural network	MLP-NN (cluster 1 is MLP2 and MLP4, cluster 2 is MLP4 and GMDH10, cluster 3 is MLP4, cluster 4 is MLP2, for all samples is MLP2.	[67]

Each algorithm has its own variation in prediction accuracy (see Table III). Using public data, Ankita et al. [62] applied different algorithms according to the prediction target, following both short-term and long-term seasonal trends. In other words, the complex pattern of historical drug consumption data affects accuracy. Using the same data source, Mbonyinshuti et al. [17] employed different significant variables for essential drug target prediction, as the predicted target variables are more specific to non-communicable diseases [63]. Another perspective on the same data source can provide new knowledge. Every change in data can change the pattern and prediction results. We need an analysis study to explore each algorithm in solving time series problems [58]. Analysis of low machine learning performance requires further research because generalization requires carefulness in approaching the data source [17], [50], [56].

The actual value factor also opens up the possibility of developing significant variables in predicting drug demand. In the case of Engelke [56], they evaluated data from multiple systems within a hospital to form the basis for predicting platelet concentrate (PC) transfusion. The low prediction recall value (0.3648) indicates a significant false negative (FN) value.

The patient is predicted not to need a PC, but in reality, the patient needs it. So, the availability of historical health data on the patient's condition from the EHR can affect the prediction performance. In addition, Mirshekari [60] utilized the power of Bayesian and Gaussian process regression (GPR) models for time series analysis, specifically the kernel ensemble, which can process low-dimensional variables and provide complete and high-quality data. This Bayesian technique optimization produced a prediction model with high accuracy. Again, real-time data on drug demand trends drives the algorithm to deliver better accuracy. Ultimately, the significant variables from the available variable data drive the performance of the combined prediction model algorithm [57], [68].

D. The Potential Variables in Pharmaceutical Drug Prediction (RQ3)

Manufacturers and distributors require accurate product predictions in production and distribution planning to ensure patients receive products with sufficient quality and quantity. Thus, variables related to drug and medical device products also include information on the area of need and the magnitude of the decision. In addition, epidemiological variables need parallel studies with predictions of disease spread. Table IV shows that the areas of research issues and prediction variables can continue to develop through machine learning.

TABLE IV. RESEARCH ISSUE AREAS AND RESEARCH VARIABLES

Research Issue Area	**Variables**		**Ref.**
Medical data record - pharmaceutical drug management	– Patient identity and demographics – Type of disease	– Doctor's prescription – Type and amount of medication	[17], [49]
Epidemiology	– Patient demographics – Disease diagnosis – Mortality – Obesity	– Cholesterol – Medication type – Patient condition	[55], [56]
Epidemiology - disaster	– Number of cases – Type and quantity of medication	– History of medication consumption	[57]
Epidemiology - medical data record - pharmaceutical drug management	– Patient's medical condition – Patient's medical history	– Treatment history – Laboratory test results – Mortality	[48]
Epidemiology - geographic	– Number of new cases – Geographic mortality	– Number of outpatients/inpatients	[28]
History of usage/sales data	– Drug sales history data – Drug types and quantity – Date/time of sale – Therapeutic group	– Company Type – Product Source – Number of Competitors – Price	[35], [58]–[62]
History of usage/sales data - geographic	– Drug sales history data – Drug types and quantity – Name and number of prescriptions	– Consumer region – Location of hospital/health service – Time of consumption	[63], [64]
Pharmaceutical drug management	– Name, type, and number of drugs		[65]
Drugs - patient behavior (self-medication)	– Name and condition of the drug, – Contents and properties of the drug	– Drug rating – Consumption time	[66]
Drug supply chain	– Drug sales history data – Seasonal trends – Number of prescriptions and patients – Number of sales per time – Name, type, and number of drugs – Number of orders per time – Timeliness of order arrival	– Supplier and number of orders – Number of drug stock – Storage qualifications – Frequency of consumption – Patient demographics – Product price – Type of disease	[9], [50], [67]–[70]

Variables spread across ten research issue areas assume that they have been tested in their research as significant variables in machine learning algorithms. The challenge is how the model performs when these variables are absent [50]. Likewise, expert judgment needs to provide comprehensive analysis of the variables' significance [65]. In general, the availability of data on demand variables for each type of drug can reveal the seasonal trend of the drug, allowing the model to produce more accurate pattern predictions, as consumption trend data is synchronous with seasonal patterns [62]. The addition of other variables becomes significant when they support the prediction pattern, including the average number of days of drug supply per patient, the duration of drug supply, previous drug shortages, drug hierarchies, and different therapeutic groups [50].

Determining research issue areas, especially those involving combinations, helps researchers identify potential variables for predicting drug demand. As a result, machine learning utilizes richer, multi-source data for learning. Multimodal data from various hospital information systems opens the discovery of potential variables from geospatial, epidemiological, and climatological data [56], [63]. Potential variables can be identified through the analysis of disease epidemiology spread. Real-time data and trends related to geographic distribution (from local to regional and national levels) are essential data requirements for training machine learning prediction models [28]. Epidemiological variables are not limited to disease names but also include disease characteristics, disease sources, and drugs that control the disease. Variables predicting disease spread correlate with variables in drug prediction [49].

Furthermore, an analysis of variables is needed to determine which variables remain significant in the present and the future [55]. The development of complex modern pharmaceutical predictions requires careful analysis to explore significant variables [69]. Variable development requires analysis of the relationship between one variable and another. Finding the relationship between location variables, drug prices, and weather conditions can improve system performance [68]. As a development, classification methods can help drug reviews identify the relationship between disease names, drug names, and disease indications, enabling patients to find suitable drugs for their conditions [66], [67], [72].

Table IV also presents the findings of the same variables from several other studies. The need for the same variables arises. Fig. 5 shows the frequency of variables used in selected articles.

Fig. 5 shows that drug sales/consumption (quantity, type, time) (11) and drug data (name, condition, quantity, expired date, and price) (12) are the most widely used variables in drug prediction. Together with the doctor's prescription parameters (name and drug quantity) (2) and drug reviews (content and properties of the drug) (1), these variables are the most dominant among the other variables. The epidemiological variable group uses the most variables, namely the number of disease cases (3). The variables of health service location - distributor (5) and patient domicile (3) represent discussions related to the drug supply chain. Meanwhile, the prediction pattern using the susceptible-expose-infectious-removal (SEIR) method uses many patient identity and condition variables.

The various research variables found can be grouped into five categories: epidemiology, drug consumption, drug storage and ordering (drug management in pharmacies), supply chain and patient domicile (drug distribution - area - health service distribution), and patient identity and condition (pharmacotherapy). The third research question (RQ3) is answered. These five categories represent potential variables for predicting drug demand, as identified in selected articles. This study recommends the following five categories for selecting potential variables for drug prediction using machine learning. The five categories demonstrate that obtaining significant variables by measuring the proximity between variables using machine learning algorithms can overlook field conditions, such as the spread of disease and the distribution of drugs from a health service. However, the five categories of potential variables are identified as significant variables in machine learning models for drug demand prediction.

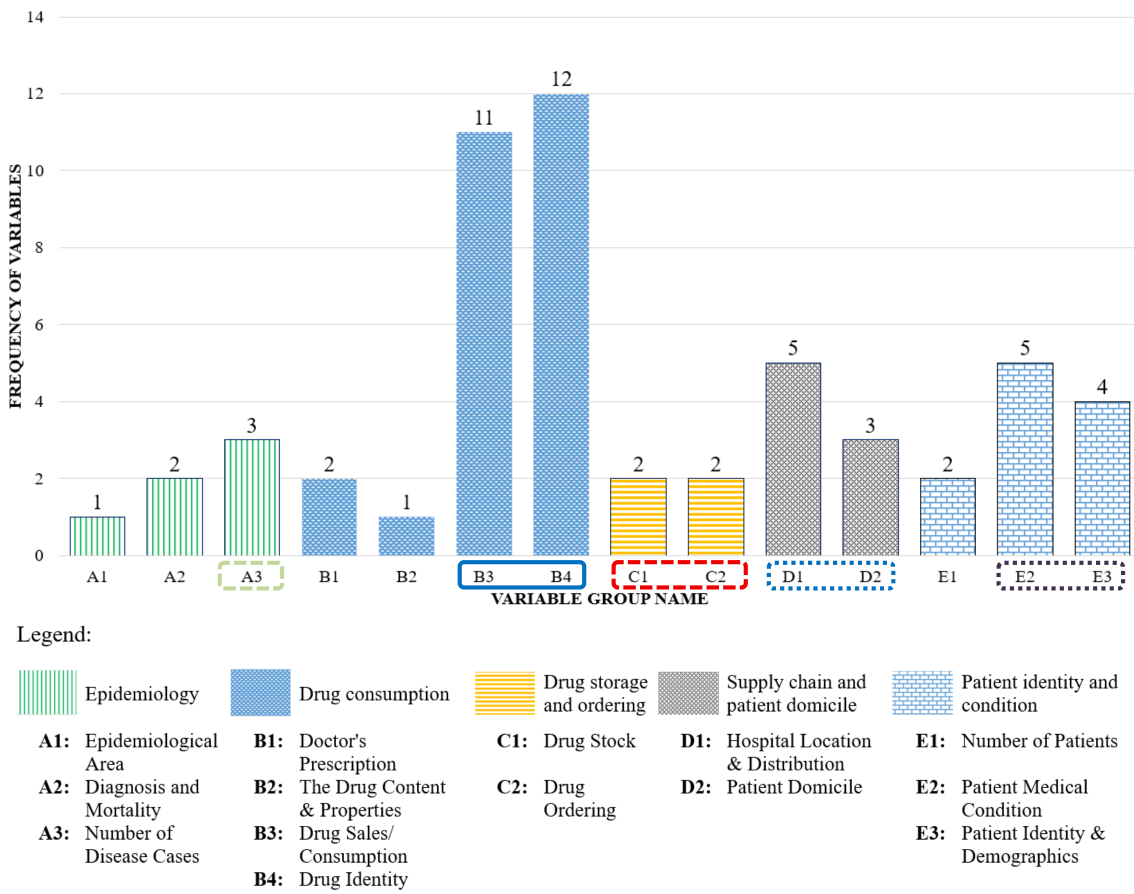

Fig. 5. The Frequency of variable use in research.

E. Integrated Datasets, Algorithms, and Potential Variables in Pharmaceutical Drug Prediction

The study begins by exploring potential variables by identifying issue areas. Classifying issue areas facilitates the study in sorting out the problem context and data availability in building machine learning models for drug demand prediction. Of the ten issue areas, the drug supply chain and drug consumption are closely related to the problem of predicting drug demand. This closeness arises because machine learning models aggregate historical data on drug consumption for prediction purposes. Increasing accuracy is the target of drug demand prediction research. Therefore, model development relies on algorithms that can enhance prediction accuracy. However, several studies have encountered obstacles due to limited data or restricted access to data (see Table II). The use of public data is one solution in building machine learning models for prediction [35], [58], [62], [69]. Indirectly, the development of drug demand prediction models ignores uncertain environmental conditions. Dynamic data changes, diverse data types, and varied data sources, which tend to increase data complexity, are not widely observed in this study. Therefore, further research is needed to understand how algorithms recognize data patterns and select significant features.

This study also examined the combination of medication consumption, disease, and treatment management. Dataset limitations often hinder the development of machine learning models. Potential variables can be a valuable option for developing machine learning algorithms, especially given the difficulty of obtaining certain types of data, which can reduce predictive performance. The results of this study, which paired the issue area with research variables (see Table IV), can be used as a consideration in developing machine learning models for regression, classification, and clustering. Selecting potential variables with close relevance can improve the performance of machine learning algorithms, particularly by reducing model complexity. Potential variables are rationally selected based on their relevance to the case and their proximity to the case, before being tested in machine learning algorithms using available data [73].

A limitation of this study is that many articles were included in the initial search because the search keywords were still too broad. The suitability of the search keywords captures more secondary research articles. Meanwhile, the discovery of primary research articles often involves fewer search keywords that match the contents of the primary research article. Selecting articles requires much careful effort to read the entire article. The number of selected articles is not yet representative of a particular continental region, thus requiring further study at the national level.

V. CONCLUSION

Mapping of research problem areas and significant variables in machine learning models for drug demand prediction produces ten research problem areas. The dynamics of data needs and algorithm choices for solving research problems are evident in the mapping. The challenge in the combination area is the increasingly complex and evolving data dimensions and algorithm choices. The selection of algorithms is still dominated by RF, XGBoost, and LSTM (for deep learning) machine learning algorithms. Efforts to improve accuracy dominate the research, providing a comparative perspective of various algorithms to improve predictions. Meanwhile, this study is in the area of significant variables that often depend on data limitations and algorithm performance.

This study identifies five potential variables in pharmaceutical drug prediction: drug consumption, drug management, epidemiology, supply chain-patient domicile, and pharmacotherapy. These five variables have not established themselves as a benchmark model or framework for designing drug demand predictions. However, mapping these variables can provide perspective for drug supply chain actors in developing machine learning models for downstream drug prediction.

Potential variables will not be significant in the model without the development of integrated data availability that is appropriate and relevant to the characteristics of the drug supply chain, endemic diseases, and public health services. Potential variables can be the basis for updating government policies related to health data integration and drug procurement guidelines in pharmacies. This research provides a basis for future research. The challenge of building a machine learning model for drug demand prediction involves working with multiple pharmacies across different geographic and demographic areas.

REFERENCES

[1] Y. A. Mashuri et al., “The response to COVID-19 among drug retail outlets in Indonesia: A cross-sectional survey of knowledge, attitudes, and practices,” Lancet Reg. Heal. - West. Pacific, vol. 22, p. 100420, 2022, doi: 10.1016/j.lanwpc.2022.100420.

[2] W. Ying, Y. Qian, and Z. Kun, “Drugs supply and pharmaceutical care management practices at a designated hospital during the COVID-19 epidemic,” Res. Soc. Adm. Pharm., vol. 17, no. 1, pp. 1978–1983, Jan. 2021, doi: 10.1016/j.sapharm.2020.04.001.

[3] Mudriyan, A. Razak, Sukri, M. A. Arifin, A. Syam, and Y. Thamrin, “Analysis of Management of Consumable Medical Materials during the Covid-19 Pandemic in Pharmaceutical Installations Health Department of Baubau City,” NeuroQuantology, vol. 20, no. 8, pp. 926–938, 2022, doi: 10.14704/nq.2022.20.8.NQ44100.

[4] H. A. Sani, A. Asriwati, M. Nyorong, M. Sari, and I. Ginting, “Analysis of Drug Availability Management at the Regional General Hospital Pharmacy Installation,” J. La Medihealtico, vol. 4, no. 2, pp. 87–102, 2023, doi: 10.37899/journallamedihealtico.v4i2.837.

[5] M. H. Ambon, “Evaluation of Drug Planning and Procurement at the Pharmacy Installation of Regional General Hospital Dr. M. Haulussy Ambon,” Medico-Legal Updat., vol. 20, no. 4, 2020, doi: 10.37506/mlu.v20i4.1981.

[6] M. Niaz and U. Nwagwu, “Managing Healthcare Product Demand Effectively in The Post-Covid-19 Environment: Navigating Demand Variability and Forecasting Complexities,” Am. J. Econ. Manag. Bus., vol. 2, no. 8, pp. 316–330, 2023, doi: https://doi.org/10.58631/ajemb.v2i8.55.

[7] C. Franco and E. Alfonso-Lizarazo, “Optimization under uncertainty of the pharmaceutical supply chain in hospitals,” Comput. Chem. Eng., vol. 135, p. 106689, 2020, doi: 10.1016/j.compchemeng.2019.106689.

[8] K. Moons, G. Waeyenbergh, and L. Pintelon, “Measuring the logistics performance of internal hospital supply chains – A literature study,” Omega (United Kingdom), vol. 82, pp. 205–217, 2019, doi: 10.1016/j.omega.2018.01.007.

[9] X. Zhu, A. Ninh, H. Zhao, and Z. Liu, "Demand Forecasting with Supply-Chain Information and Machine Learning: Evidence in the Pharmaceutical Industry," Prod. Oper. Manag., vol. 30, no. 9, pp. 3231–3252, 2021, doi: https://doi.org/10.1111/poms.13426.

[10] M. Wang and F. Jie, "Managing supply chain uncertainty and risk in the pharmaceutical industry," Heal. Serv. Manag. Res., vol. 33, no. 3, pp. 156–164, Jun. 2019, doi: 10.1177/0951484819845305.

[11] L. P. L. Wulandari, A. A. S. Sawitri, and A. Hermansyah, "The potential roles of pharmacy medication sales data to augment the syndromic surveillance system in response to COVID - 19 and preparedness for other future infectious disease outbreaks in Indonesia," Int. J. Health Plann. Manage., vol. 37, no. 1, pp. 30–39, Jan. 2022, doi: 10.1002/hpm.3356.

[12] I. Vanany, M. H. Ali, K. H. Tan, A. Kumar, and N. Siswanto, "A Supply Chain Resilience Capability Framework and Process for Mitigating the COVID-19 Pandemic Disruption," IEEE Trans. Eng. Manag., 2021, doi: 10.1109/TEM.2021.3116068.

[13] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," Int. J. Intell. Networks, vol. 3, no. February, pp. 58–73, 2022, doi: 10.1016/j.ijin.2022.05.002.

[14] Tangi Ndakondja Angula and Abraham Dongo, "Assessing the impact of artificial intelligence and machine learning on forecasting medication demand and supply in public pharmaceutical systems: A systematic review," GSC Biol. Pharm. Sci., vol. 26, no. 2, pp. 140–150, Feb. 2024, doi: 10.30574/gscbps.2024.26.2.0071.

[15] D. Ni, Z. Xiao, and M. K. Lim, "A systematic review of the research trends of machine learning in supply chain management," Int. J. Mach. Learn. Cybern., vol. 11, no. 7, pp. 1463–1482, 2020, doi: 10.1007/s13042-019-01050-0.

[16] R. Alfred and J. H. Obit, "The roles of machine learning methods in limiting the spread of deadly diseases: A systematic review," Heliyon, vol. 7, no. 6, p. e07371, 2021, doi: 10.1016/j.heliyon.2021.e07371.

[17] F. Mbonyinshuti, J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, "The prediction of essential medicines demand: A machine learning approach using consumption data in Rwanda," Processes, vol. 10, no. 1, pp. 1–13, 2022, doi: 10.3390/pr10010026.

[18] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," Appl. Intell., vol. 52, no. 4, pp. 4543–4581, Mar. 2022, doi: 10.1007/s10489-021-02550-9.

[19] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," J. Appl. Sci. Technol. Trends, vol. 1, no. 1, pp. 56–70, 2020, doi: 10.38094/jastt1224.

[20] I. Syahrir, Suparno, and I. Vanany, "Drug Supplies Planning in Hospital for Epidemic Attack using SEIR Model," J. Phys. Conf. Ser., vol. 1179, no. 1, 2019, doi: 10.1088/1742-6596/1179/1/012150.

[21] I. Syahrir, Suparno, and I. Vanany, "A Proposed Model for Drug Demand Forecasting and Ordering Inventory System for Dengue Endemic," Oper. Supply Chain Manag., vol. 15, no. 1, pp. 69–78, 2022, doi: 10.31387/oscm0480331.

[22] T. S. Ho et al., "Comparing machine learning with case-control models to identify confirmed dengue cases," PLoS Negl. Trop. Dis., vol. 14, no. 11, pp. 1–21, 2020, doi: 10.1371/journal.pntd.0008843.

[23] N. A. M. Salim et al., "Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques," Sci. Rep., vol. 11, no. 1, pp. 1–9, 2021, doi: 10.1038/s41598-020-79193-2.

[24] P. Siriyasatien, A. Phumee, P. Ongruk, K. Jampachaisri, and K. Kesorn, "Analysis of significant factors for dengue fever incidence prediction," BMC Bioinformatics, vol. 17, no. 1, pp. 1–9, 2016, doi: 10.1186/s12859-016-1034-5.

[25] F. Yavari Nejad and K. D. Varathan, "Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction," BMC Med. Inform. Decis. Mak., vol. 21, no. 1, pp. 1–12, 2021, doi: 10.1186/s12911-021-01493-y.

[26] 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," J. R. Soc. Interface, vol. 18, no. 179, 2021, doi: 10.1098/rsif.2020.1006.

[27] J. Ong et al., "Mapping dengue risk in Singapore using Random Forest," PLoS Negl. Trop. Dis., vol. 12, no. 6, pp. 1–12, 2018, doi: 10.1371/journal.pntd.0006587.

[28] S. Atek et al., "A predictive decision support system for coronavirus disease 2019 response management and medical logistic planning," Digit. Heal., vol. 9, 2023, doi: https://doi.org/10.1177/20552076231185475.

[29] M. Syaripuddin, R. Ayu, D. Sartika, M. Nadjib, and M. Syaripuddin, "Qualitative Study on the Drug Availability for National Health Insurance (JKN) in Hypertensive Patients : the Perspective of the Tender Winner Manufacturer," East African Sch. J. Med. Sci., vol. 4421, no. 11, pp. 674–678, 2019, doi: 10.36349/EASMS.2019.v02i11.018.

[30] D. Koala, Z. Yahouni, G. Alpan, and Y. Frein, "Factors influencing Drug Consumption and Prediction Methods," CIGI-Qualita Conférence Int. Génie Ind. Qual. - Grenoble, no. May 2021, pp. 1–8, Sep. 2021, doi: https://doi.org/10.48550/arXiv.2109.11779.

[31] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," J. Big Data, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00329-2.

[32] J. Y. S. Ng, R. V. Ramadani, D. Hendrawan, D. T. Duc, and P. H. T. Kiet, "National Health Insurance Databases in Indonesia, Vietnam and the Philippines," PharmacoEconomics - Open, vol. 3, no. 4, pp. 517–526, 2019, doi: https://doi.org/10.1007/s41669-019-0127-2.

[33] A. Burinskiene, "Forecasting Model: The Case of the Pharmaceutical Retail," Front. Med., vol. 9, no. August, pp. 1–16, 2022, doi: 10.3389/fmed.2022.582186.

[34] M. A. Al-Gunaid et al., "Analysis of drug sales data based on machine learning methods," Proc. 2018 Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2018, pp. 32–38, 2018, doi: 10.1109/SYSMART.2018.8746968.

[35] A. Manna, K. Kolpe, A. Mhalungekar, S. Pattewar, P. Kaloge, and R. Patil, "Comparative Study of Various Machine Learning Algorithms for Pharmaceutical Drug Sales Prediction," in 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT), May 2023, pp. 1–6, doi: 10.1109/IConSCEPT57958.2023.10169964.

[36] J. Feizabadi, "Machine learning demand forecasting and supply chain performance," Int. J. Logist. Res. Appl., vol. 25, no. 2, pp. 119–142, 2022, doi: 10.1080/13675567.2020.1803246.

[37] B. K. Almentero, J. Li, and C. Besse, "Forecasting pharmacy purchases orders," in 2021 IEEE 24th International Conference on Information Fusion (FUSION), Nov. 2021, pp. 1–8, doi: 10.23919/FUSION49465.2021.9627017.

[38] X. Zhu, A. Ninh, H. Zhao, and Z. Liu, "Cross - Series Demand Forecasting using Machine Learning: Evidence in the Pharmaceutical Industry," Prod. Oper. Manag., no. March 2019, pp. 1-54, 2021, doi: 10.1111/poms.13426.

[39] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2018th ed. Lean Publishing, 2018.

[40] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning, vol. 19. MIT Press book, 2017.

[41] Y. K. Ever, K. Dimililer, and B. Sekeroglu, "Comparison of Machine Learning Techniques for Prediction Problems," in Web, Artificial Intelligence and Network Applications - WAINA 2019. Advances in Intelligent Systems and Computing, 2019, vol. AISC, 927, pp. 713–723, doi: https://doi.org/10.1007/978-3-030-15035-8_69.

[42] B. A. Mousa and B. Al-Khateeb, "Predicting medicine demand using deep learning techniques: A review," J. Intell. Syst., vol. 32, no. 1, Aug. 2023, doi: 10.1515/jisys-2022-0297.

[43] X. Zhu, G. Zhang, and B. Sun, "A comprehensive literature review of the demand forecasting methods of emergency resources from the perspective of artificial intelligence," Nat. Hazards, vol. 97, no. 1, pp. 65–82, 2019, doi: 10.1007/s11069-019-03626-z.

[44] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," Artif. Intell. Med., vol. 104, no. February, p. 101822, 2020, doi: 10.1016/j.artmed.2020.101822.

[45] A. M. Aamer, L. P. E. Yani, and I. M. A. Priyatna, “Data analytics in the supply chain management: Review of machine learning applications in demand forecasting,” Oper. Supply Chain Manag., vol. 14, no. 1, pp. 1–13, 2021, doi: http://doi.org/10.31387/oscm0440281.

[46] S. Pizard, F. Acerenza, X. Otegui, S. Moreno, D. Vallespir, and B. Kitchenham, “Training students in evidence-based software engineering and systematic reviews: a systematic review and empirical study,” Empir. Softw. Eng., vol. 26, no. 3, p. 50, May 2021, doi: 10.1007/s10664-021-09953-9.

[47] L. Koval et al., “Unlocking the Potential of Information Modeling for Root Cause Analysis in a Production Environment: A Comprehensive State-of-the-Art Review Using the Kitchenham Methodology,” IEEE Access, vol. 12, no. May, pp. 80266–80282, 2024, doi: 10.1109/ACCESS.2024.3406020.

[48] S. Subudhi et al., “Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19,” NPJ Digit. Med., vol. 4, no. 1, p. 87, May 2021, doi: 10.1038/s41746-021-00456-x.

[49] A. H. Nabizadeh, M. M. Ghaemi, and D. Goncalves, “Predicting the pharmaceutical needs of hospitals using machine learning algorithms,” Int. J. Data Sci. Anal., pp. 0–20, 2024, doi: https://doi.org/10.1007/s41060-024-00530-z.

[50] R. Pall, Y. Gauthier, S. Auer, and W. Mowaswes, “Predicting drug shortages using pharmacy data and machine learning,” Health Care Manag. Sci., vol. 26, no. 3, pp. 395–411, 2023, doi: 10.1007/s10729-022-09627-y.

[51] N. Alballa and I. Al-turaiki, “Machine learning approaches in COVID-19 diagnosis , mortality , and severity risk prediction : A review,” Informatics Med. Unlocked, vol. 24, p. 100564, 2021, doi: 10.1016/j.imu.2021.100564.

[52] G. R. Shinde, A. B. Kalamkar, P. N. Mahalle, N. Dey, and J. Chaki, “Forecasting Models for Coronavirus Disease (COVID - 19): A Survey of the State-of-the-Art,” SN Comput. Sci., vol. 1, no. 4, pp. 1–15, 2020, doi: 10.1007/s42979-020-00209-9.

[53] E. Campillo-Funollet et al., “Predicting and forecasting the impact of local outbreaks of COVID-19: use of SEIR-D quantitative epidemiological modelling for healthcare demand and capacity,” Int. J. Epidemiol., vol. 50, no. 4, pp. 1103–1113, Aug. 2021, doi: 10.1093/ije/dyab106.

[54] M. A. A. Al-qaness, A. A. Ewees, H. Fan, L. Abualigah, and M. Abd Elaziz, “Marine Predators Algorithm for Forecasting Confirmed Cases of COVID-19 in Italy , USA , Iran and Korea,” Int. J. Environ. Res. Public Health, vol. 17(10), no. May, p. 3520, 2020, doi: 10.3390/ijerph17103520.

[55] Y. Haoudi, M. A. Yazdani, D. Roy, and S. Hennequin, “Demand Prediction Based on Machine Learning Algorithms for Optimal Distribution of Insulin,” IFAC-PapersOnLine, vol. 56, no. 2, pp. 10174–10179, 2023, doi: 10.1016/j.ifacol.2023.10.893.

[56] M. Engelke et al., “Predicting Individual Patient Platelet Demand in a Large Tertiary Care Hospital Using Machine Learning.,” Transfus. Med. Hemother., vol. 50, no. 4, pp. 277–285, Aug. 2023, doi: 10.1159/000528428.

[57] W. Zhang and X. Li, “A data-driven combined prediction method for the demand for intensive care unit healthcare resources in public health emergencies,” BMC Health Serv. Res., vol. 24, no. 1, 2024, doi: 10.1186/s12913-024-10955-8.

[58] R. Rathipriya, A. A. Abdul Rahman, S. Dhamodharavadhani, A. Meero, and G. Yoganandan, “Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model,” Neural Comput. Appl., vol. 35, no. 2, pp. 1945–1957, 2023, doi: 10.1007/s00521-022-07889-9.

[59] N. Qassrawi, M. Azzeh, and M. Hijjawi, “Drug sales forecasting in the pharmaceutical market using deep neural network algorithms,” Int. J. Syst. Innov., vol. 8, no. 3, pp. 63–83, 2024, doi: https://doi.org/10.6977/IJoSI.202409_8(3).0006.

[60] S. Mirshekari, M. Moradi, M. Ensaf, H. Jafari, and M. Jafari, “Enhancing Predictive Accuracy in Pharmaceutical Sales Through an Ensemble Kernel Gaussian Process Regression Approach,” SSRN Electron. J., vol. 18, no. 5, pp. 255–260, 2024, doi: 10.2139/ssrn.4860667.

[61] F. Mbonyinshuti, J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, “Health supply chain forecasting: a comparison of ARIMA and LSTM time series models for demand prediction of medicines,” Acta Logist., vol. 11, no. 02, pp. 269–280, Jun. 2024, doi: 10.22306/al.v11i2.510.

[62] M. Ankita, A. Srinivas, A. Soni, G. Prajapati, and P. S. Manjunath, “Pharmaceutical Sales Data Prediction Using Time Series Forecasting,” Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 13s, pp. 681–696, 2024, [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/4649.

[63] F. Mbonyinshuti, J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, “Application of random forest model to predict the demand of essential medicines for noncommunicable diseases management in public health facilities,” Pan Afr. Med. J., vol. 42, p. 89, 2022, doi: 10.11604/pamj.2022.42.89.33833.

[64] X. Wang, L. Chen, J. Wang, and D. Zheng, “Machine Learning Approaches for Region-level Prescription Demand Forecasting,” in 2023 IEEE Smart World Congress (SWC), 2023, pp. 1–6, doi: 10.1109/SWC57546.2023.10449058.

[65] 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.

[66] S. Das, S. K. Mahata, A. Das, and K. Deb, “Disease Prediction from Drug Information using Machine Learning,” Am. J. Electron. Commun., vol. 1, no. June, pp. 16–21, 2021, doi: https://doi.org/10.15864/ajec.1403.

[67] M. S. Amalnick, N. Habibifar, M. Hamid, and M. Bastan, “An intelligent algorithm for final product demand forecasting in pharmaceutical units,” Int. J. Syst. Assur. Eng. Manag., vol. 11, no. 2, pp. 481–493, 2020, doi: 10.1007/s13198-019-00879-6.

[68] A. K. Bapatla, S. P. Mohanty, and E. Kougianos, “FortiRx 2 . 0 : Smart Privacy-Preserved Demand Forecasting of Prescription Drugs in,” 2023, doi: 10.1109/OCIT59427.2023.10430944.

[69] K. P. Fourkiotis and A. Tsadiras, “Applying Machine Learning and Statistical Forecasting Methods for Enhancing Pharmaceutical Sales Predictions,” forecasting, vol. 6, no. 1, pp. 170–186, 2024, doi: https://doi.org/10.3390/forecast6010010.

[70] M. B. de Oliveira, G. Zucchi, M. Lippi, D. F. Cordeiro, N. R. da Silva, and M. Iori, “Lead Time Forecasting with Machine Learning Techniques for a Pharmaceutical Supply Chain,” in The 23rd International Conference on Enterprise Information Systems (ICEIS 2021), 2021, vol. 1, pp. 634–641, doi: https://dx.doi.org/10.5220/0010434406340641.

[71] A. Musamih et al., “A blockchain-based approach for drug traceability in healthcare supply chain,” IEEE Access, vol. 9, pp. 9728–9743, 2021, doi: 10.1109/ACCESS.2021.3049920.

[72] M. M. H. Chowdhury and M. Quaddus, “Supply chain readiness, response and recovery for resilience,” Supply Chain Manag., vol. 21, no. 6, pp. 709–731, 2016, doi: 10.1108/SCM-12-2015-0463.

[73] F. Orhan and M. N. Kurutkan, “Predicting total healthcare demand using machine learning: separate and combined analysis of predisposing, enabling, and need factors,” BMC Health Serv. Res., vol. 25, no. 1, p. 366, Mar. 2025, doi: 10.1186/s12913-025-12502-5.