



## Effectiveness of Predictive Analytics in Precision Public Health in Strengthening Health System for Future Pandemics

Olatinwo Islamiyyat Adekemi<sup>1\*</sup>, Olajide Damola Sheriff<sup>2</sup>

<sup>1</sup>Tilu Technologies, Abuja, Nigeria

<sup>2</sup>West African Health Organization Headquarters, Bobo-Dioulasso, Burkina-Faso

**Received Date:** August 18, 2022; **Accepted Date:** August 23, 2022; **Published Date:** September 03, 2022;

\* **Corresponding author:** Olatinwo Islamiyyat Adekemi. Tilu Technologies, Abuja, Nigeria.  
Email: [islamiyyat.me@outlook.com](mailto:islamiyyat.me@outlook.com)

### Introduction

The outbreak of the COVID-19 pandemic in 2020 made it evident that to determine effective intervention and achieve positive public health outcomes, there is a need to have a holistic insight and understanding of the patients' or population's health, including their social, biological, and environmental health conditions. Though there is access to a wide range of data, interoperability among these data sources and implementation of these data are essential to strengthen health infrastructures in enabling healthcare systems deliver a greater and effectual quality of health. Difference in mortality, treatment response and number of cases during the pandemic reinforced that there are disparities and inefficiencies in health. The pandemic also prompted the evolution and gave opportunity to several field which include precision public health, health informatics and telemedicine. The emergence of the novel COVID-19 variant due to its virology and mutation, is an awakening call that precision public health is essential without undermining the underpinned value and effectiveness of predictive analytics.

Precision public health has been analogized as an extension of personalized medicine to population that is, it is the method of improving the health of a population by accounting for socioeconomic, biological, and environmental factors in proffering health measures (Kee & Taylor-Robinson, 2020). Its main objective is to correctly identify the best treatment and health measures that suits a specific population of people. According to Lopes, Guimaraes and Santos (2020), predictive analytics involves the use of several digital techniques such as data mining, artificial intelligence, modeling, and statistics to make informed future

forecast. Hence, having diverse, complex, and disassociated data sources is not sufficient to make informed health decision. There is a necessity to make patterns of these data and implement them in an efficient way to improve the delivery of quality healthcare while limiting misuse of expensive resources. Precision public health tools provide actionable intelligence to determine high and low-risk population when enhanced with predictive analytics techniques. More comprehensive information by place, person and time is becoming available in the period of big data and predictive analytics to quantify the public health effect and implementation demands.

Some studies have deployed several predictive analytics models in precision public health and precision medicine, public health is base evidentiary foundation. there is a need to assess the effectiveness of this techniques. According to Centre for Disease Control (2020), millions of deaths can be averted through precise prevention of known risks and targeted intervention. Precision public health is a rapidly developing field which has lots of potential and challenges. There is a need to investigate and evaluate the predictive analytics employed, as well as uncover knowledge gaps and contribute to the body of knowledge for implementation studies.

### Abstract

#### Background

Several mortality causalities are responsible for millions of deaths yearly and decrease in life expectancy. The covid-19 pandemic has continued to increase these numbers since 2020 its

emergence many public health measures have been put in place to flatten the curve. Public health has used data from different source to improve decision and policy making. In this era, precision public health among other developing field of health has shown great potential in strengthening health data systems. However, with predictive analytics been support systems in precision public health there is a need to evaluate the performance of these techniques.

## Method

A systematic review was conducted between November 2011 and January 2022 using studies from nine at database which included PubMed, TRIP, SCOPUS, and Cochrane. Grey literature and google scholar were searched. Eligible studies were selected using inclusion and exclusion criteria and finding from the included studies were summarized.

## Result

17 studies from 11 countries published in English between 2011-2021 were selected demographic, environmental, social, and socio-economic data were gathered by the selected studies. Artificial intelligence with machine learning been the most common, was the major predictive analytics technique used by the research. Communicable and non-communicable diseases, prescription overdose and underdose, neonatal conditions, health disparities, substance abuse and motor vehicle injuries are public health areas in which the techniques were deployed.

## Discussion and conclusion

Studies in this review reported that predictive analytics techniques are effective and produced reasonably accuracies. Although, there are some limitations such as lack specific definition of sub-population and units of inference, use of one-dimensional data by some studies, some bias that can confound randomization predictive analytics in precision public health is a great call that requires more work for evidence-based foundation for its application.

## Literature Review

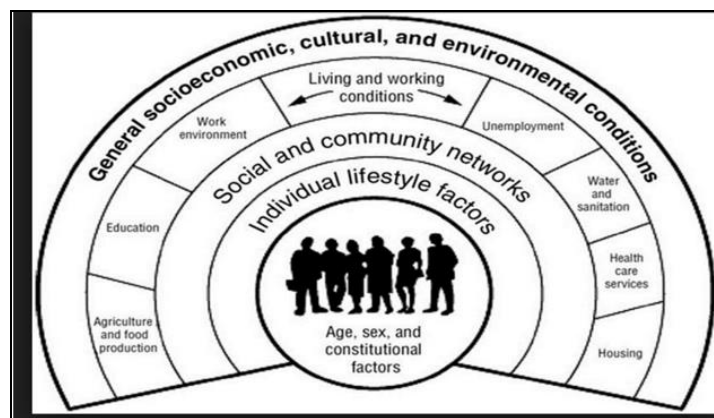
Precision public health is a field that has gained interest over the past years. The term 'precision public health' originated from western Australia in 2013 to match the corresponding developments in precision medicine, the nomenclators where the public health practitioners for the health department (Weeramanthri et al., 2018). According to Center for Disease Control (2021) public health is the "science of protecting and improving the health of people and their communities". Preventive and prompt intervention remains the way in which public health can be strengthened. Precision public health is the study of the relation between individuals' biological, genetical, environmental and social determinants of health to detect disease incidence and proffer interventions within a population (Khoury, 2016). Precision public health involves grouping target population according to certain behaviors, traits, socioeconomic, biological and genetic information to improve health interventions and outcomes. According to

Rychetnik, Frommer, Hawe and Shiell (2002) public health intervention are set of actions having comprehensible objective to generate identifiable outcomes, encourage wellness and prevent illness within a population. Kindig and Stoddart (2012) explained that populations in the context of public health is often described as group of people belonging to the same geographic regions, they also include groups such as ethnic groups, people with similar health conditions, employees, people in similar age brackets, prisoners and many more groups. Precision medicine is about proffering health intervention to the right patient at the right time, it is mostly used interchangeably with genomic medicine (Khoury et al., 2018). Precision public health is the application of precision medicine in population for the benefit of the larger group of people (Chowkwanyum, Bayer and Galea, 2018).

Kee (2020) explained that the ideology/notion of precision public health is that, it is a method of enhancing the health of the population through the implementation of new technologies, which would assist in the provision of individualized measures, treatments and policies. Khoury, Engelagu, Chambers and Mensah (2019) enlightened that precision public health involves taking into accounts the determinants of health in the delivery of specific health measures to a specific set of people at a specific time. According to their study, precision public health can be better simplified into precision by place, time and person. Precision by person involves using data generated from disease subtypes using for example biomarkers or disease susceptibility which could be genetics to make informed health decisions. Precision by place uses environmental health and geographical data while precision by time involves data gathered from wearables (e.g smart watches) and social media platforms.

Dahlgren and Whitehead (1991) describe the layers of determinants of health as, layers that portray a social environmental theory to health, and they outline the connection between individual, environment and diseases. The Dahlgren and Whitehead (1991) model (**figure 1**, below) explains individuals at the middle with a set of genetic and biological factors which they do not have control over, layered above them are: individual lifestyle factors, social and community networks, general socioeconomic, cultural and environmental conditions (housing, working, education, work environment, unemployment, agriculture, health care services, water and sanitation).

Accounting for genetic variation, social and environmental disparity, difference in lifestyles of individuals will aid determination of specific health treatments that would be most effective within a targeted population. Precision public health is an integrated approach to prevent disease and promote health while reducing health disparities with the implementation of new technologies (Kneale et al., 2020). Precision public health involves the use of wide range of data and multilayered information network. Some of these data includes exposome, signs and symptoms, genome, microbiome, epigenome, hydrome (water), nutriome (consumed food & drinks), legome (education status), econome (financial condition), ethnome (cultural, racial and ethnical backgrounds) and other types of patients' data (Gansky & Shafik, 2020).



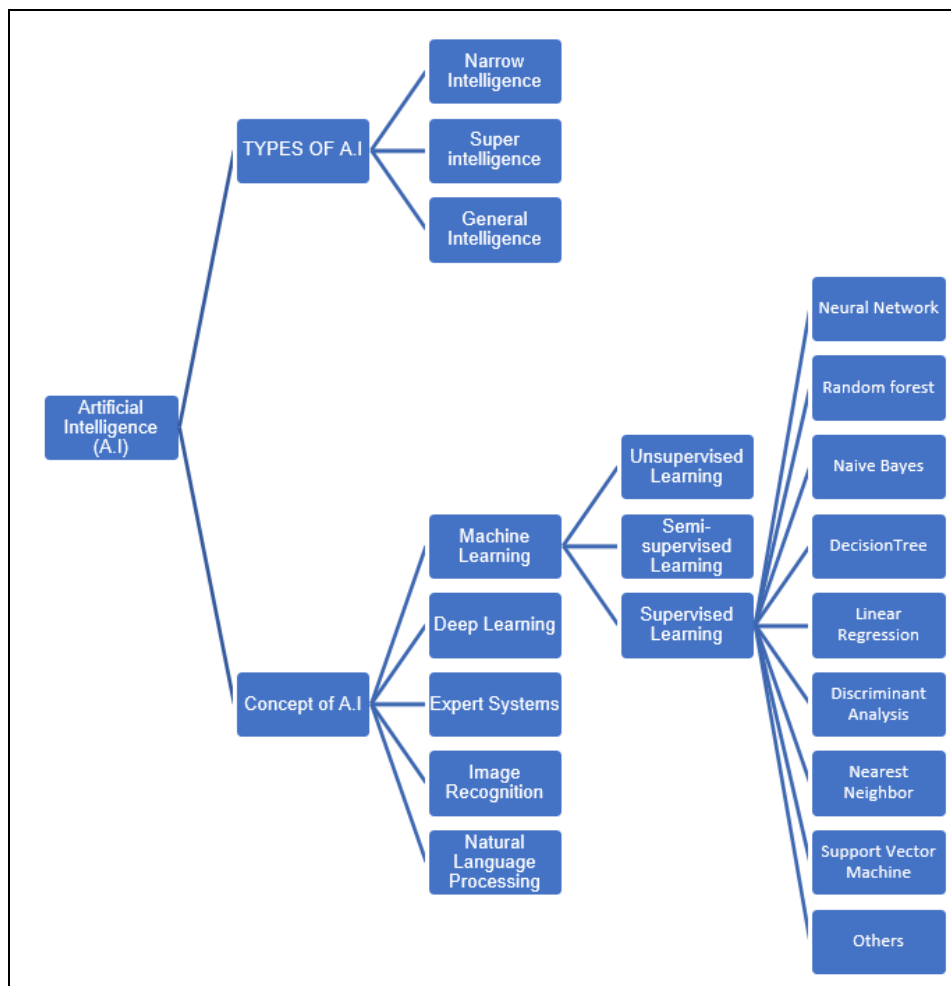
**Figure 1:** Social Determinants of Health Inequalities (Dahlgren & Whitehead, 1991).

Given the availability of data, emergence of computing and smart technologies such as patient generated health data (PGHD) and electronic health records (EHR), the public health domain is evolving. In the past, the only data available in precision public health, were collated from patients' health records, treatment histories, past diagnosed diseases, inhibitions and treatment interventions but the new technologies have broadened the digital data accessible (Kohane, 2015). Predictive analytics is conceptualized as a means of envisaging the future through data extraction and combination process, which are based upon numerous concepts depending on the function required (Lopes, Guimaraes & Santos, 2020). With the evolvement of smart and emerging technologies, data such as patient generated health data which include blood pressure, heart rate, history of physical activities, number of sleeping hours have been combined with various social data to achieve a more holistic insight of a communal lifestyle and conditions (Dolley, 2018; Leff & Yang, 2015). Predictive analytics have shown promising result in its application in influencing health and health delivery, identifying high-risk individuals/population, infectious disease surveillance and environmental surveillance (Escobar et al., 2014; Madan, Bebian, Lazer & Pentland, 2010; Oh et al, 2018; Ganasegeran & Abdulrahman, 2020). The World Health Organization (WHO) (2021), defines public health surveillance as "an ongoing, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice".

For instance, vital information such as respiration rate (generated by radar system), heart rate and facial temperature of patients were used as input to rapidly diagnose and effectively categorize patients into high risk H5N1 influenza, low risk H5N1 influenza and no flu clusters (Sun, Matsui, Hakozaiki & Abe 2015). A 30% to 60% decrease in antibiotics administration to neonates in the treatment of neonatal sepsis was achieved in a study that

implemented predictive analytics to stratify the risk of early incidence of sepsis in newborn greater or equal 34 weeks' gestation (Escobar et al. 2014).

The major predictive analytics used in precision public health is artificial intelligence, it has been acknowledged as the most powerful analytical tool in existence because it has the frame of computing that permits machines to act or respond to input with a cognitive capacity almost similar to humans (Silver et al., 2017). AI is a field of computer science that utilizes computer systems to imitate human intelligence through iterative, complex design coordination usually at a speed and scale that surpass human capability (Stead, 2018). The advent of AI, bringing together gigantic data stream from diverse specialties, has climaxed in the rise of Big Data with frameworks integration over the Internet of Things (IOT) and Cloud Computing systems (Ganasegeran & Abdulrahman, 2019). From a precision public health point of view, AI can utilize modern algorithms to learn patterns from an expansive volume of healthcare information, and subsequently use the acquired knowledge to help clinical practice while using its learning and self-correcting capacities to improve its precision based on feedbacks (Jiang et al., 2017). Mesko, Hetenyi and Gyorffy (2018) explained that almost all modern AI application frameworks use artificial narrow intelligence, which is a system whose functionality is to execute a single assignment extremely well. To provide an understanding on the types and concepts of artificial intelligence, **Figure 2** below; shows a detailed chart. Aspects in which predictive analytics can improve the accomplishment of precision public health include: infectious disease and environmental surveillance, tailoring treatment and interventions, disease epidemiology, predicting risk and outcomes, tackling antimicrobial resistance and many more (Dolley, 2018; Ganasegeran & Abdulrahman, 2020; Sun et al., 2015).



**Figure 2:** Types and Concepts of Artificial Intelligence (Mesko, Hetenyi & Gyorffy, 2018; Wahl, Cossy-Gantner, Germann & Schwalbe, 2018; Lopes, Guimaraes & Santos, 2020).

According to World Health Organization (2020) there were 55.4 million deaths worldwide in 2019 and the life expectancy at birth globally was 73.4 years. The WHO fact sheet highlighted, ischemic heart disease, stroke, chronic obstructive pulmonary disease, lower respiratory infections, neonatal conditions, trachea, bronchus and lung cancers, Alzheimer’s disease and other dementias, diarrheal diseases, diabetes mellitus and kidney diseases as the leading global causes of death respectively. These conditions accounted for 30.5 million out of the overall 55.4 million deaths reported. Other cause of death includes other non-communicable diseases, communicable diseases, suicide, maternal death and many more. These deaths have the potentials of been prevented if known risk factors can be reduced. The COVID-19 pandemic in 2020 was responsible for over 2,5 million deaths globally (CDC, 2022). WHO declared it a “public health emergency of international concern” on the 30<sup>th</sup> of January 2020 (WHO, 2022). Though the pandemic gained lots of attention, there are other public health issues that are continuously responsible for mortality. Centre for Disease Control and Prevention (2022) reported that COVID-19, mental conditions, alcohol and substance abuse, food safety and insecurity, healthcare associated infections, heart disease and stroke, HIV, motor vehicle injuries, physical activity and obesity, prescription drug overdose are all issues of public health concern

that continually impact the world. Precision public health and precision medicine are about providing the best treatment at the right time, According to Khoury, Lademarco and Riley (2016), as the area of precision public health is fast developing and it isn’t just about genes, medications and diseases; it encompasses precision prevention amongst other priority areas.

The aim of this study is to explore the effectiveness of predictive analytics methods in precision public health in relation to the aforementioned causes of death and issues of public health concern. The objectives of this review are; to identify the types of predictive analytics techniques that have been used in previous precision public health studies, report the public health condition of interest, report the techniques’ effectiveness and outcome of the study; and build on the findings of these previous studies to contribute to the body of knowledge.

### Research Methodology

Quantitative research is research that can be used to gather measurable (numerical) data that can be further analysed to answer a specific hypothesis while, qualitative research can be used to seek and understand opinions and attitude, it offers insight to



fundamental issues and can be used to generate a hypothesis (Bryman, 2012). Mixed research method is a combination of both quantitative and qualitative research (De Vaus, 2002).

For the purpose of this study, a systematic review method was implemented to explore the effectiveness of predictive analytics methods in precision public health. According to Robinson and Lowe (2015) systematic review creates evidence to buttress a piece of research through the use of precise questions. It is a way of making a comprehensive study and explanation relating the study to a particular topic.

This research method is less expensive, consumes lesser time, poses no ethical issues, allows inferences to be drawn from existing data which could either be quantitative, qualitative or mixed method. Biases such as moderator bias, response bias, recall bias, cognitive bias that occur in focus group, one-to-one, ethnography and other experimental study designs are mitigated when using the literature review method.

However, to reduce biases in this review, inclusion and exclusion criteria were set and strictly adhered to. Restrictions were also put in place to filter irrelevant studies that can lead to misinformation or misinterpretation of data. Selection bias and information bias were reduced by using proper search terms in conjunction with Boolean operators and truncation across appropriate databases. Grey literature database was searched to avoid omitting fundamental evidences that could result in publication bias. Ethical issues did not arise in this study as all data used were secondary data. The process of this review is; identify research question, conduct an evidence-based review by searching relevant databases, critically analyse findings, identify relevant studies based on inclusion and exclusion criteria, chart the studies, collate all relevant data, summarize and report findings.

- Identifying research question

To accomplish the aforementioned aim of this study, the selected papers answered the following question: (i) what

predictive analytics technique used? (ii) what issue of public health concern or medical condition was addressed (iii) what was the effectiveness of the predictive analytics method and outcome of the study reported?

- Identifying relevant studies and search strategy

Nine electronic databases known for offering high-quality medical publications were searched to identify relevant articles. These databases are;

PubMed, Cochrane, TRIP, Scopus, SAGE, APA PsycNet, IEEE, Web of Science and Google scholar.

A PICO question is a question framing process in evidence-based practice, it involves highlighting the information needed to form a question and categorizing the information under P- population or problem, I- intervention or indicator C- comparison and O- outcome of interest (Heneghan & Badenoch, 2002).

The PICO for this review was:

Population (P) - Precision Public Health

Population (P) - Epidemiology

Intervention (I) - Predictive Analytics

Outcome (O) - effectiveness

**Table 1** below contains the search terms and alternative terms that were used to conduct the free-text search. The search terms in **table 1** were used alongside Boolean operators such as “OR” to widen the search and “AND” to narrow it. Truncation was also used to avoid missing out any relevant evidence. The following keywords were the MeSH (medical subject heading) term used: “precision public health”, “predictive analytics” and “effectiveness”. A manual search was conducted on google scholar using the phrase “effectiveness of predictive analytics methods in precision public health”. For the search process the filters used were: ‘human studies’ and ‘English publications.’

Population	Population	Intervention	Outcome
“Precision public health”	Epidemiology	“Predictive analytics”	Effectiveness
<b>Alternative terms</b>			
“Precision health” “Precision medicine” “Precision healthcare” “Personalized medicine” “Population health”	“Environmental Epidemiology” “Infectious disease Epidemiology” “Public health surveillance” “Population surveillance” “Environmental surveillance” “Infectious disease surveillance”	“Machine learning” “Deep learning” “Artificial intelligence” “Expert systems” “Image recognition” “Natural language processing”	

**Table 1:** Search terms for the PICO.

- Study selection and eligibility criteria

Upon completion of the search process duplicates were removed. The following inclusion and exclusion criteria were used to select relevant articles.

**Inclusion criteria:**

1. All studies that included all forms of predictive analytics method in relation to precision public health
2. Articles with the aforementioned keywords in their titles or abstracts were selected
3. Only research and review articles were selected
4. All studies that addressed effectiveness or comparison of predictive analytics in precision medicine and any of its mentioned alternative terms
5. Only human studies in written in English was included

**Exclusion criteria:**

1. Studies that didn’t address or focus on predictive analytics and its association to precision public health or any of its aforementioned alternative terms
2. Studies that addressed effectiveness of predictive analytics in other field other than in human health
3. Studies that solely addressed data without association or implementation to predictive analytics in precision public health
4. Studies in which full text are not available

Eligible studies were: (i) original studies that addressed the population, intervention and outcome of interest (ii) used appropriate research methodology (iii) met all inclusion criteria. The CASP tool (Critical Appraisal Skills Programme, 2013) was used to screen both qualitative and quantitative studies. This tool was chosen because it provides a highly

respected and well-established appraisal process. It is also clear and easy to use. Quality of the studies were judged based on personal discretion using CASP tool as a yardstick.

- Data extraction and analysis

After selecting the eligible studies, information was extracted based on the objectives of this systematic review. Data gathered included (i) characteristics of the study which are: study author and year, country, aim, data source, data attributes and issue of public health concern addressed (ii) predictive analytics technique used, validation set, evaluation and performance measures and outcome of the study reported (iii) best technique of each study and its evaluation and performance scores. A standard tabular template on Microsoft Word was used to summarize the gathered data.

**Results**

**Study selection process**

The literature search produced 5235 potentially useful studies. Duplicates were removed, 2645 studies remained. 1800 studies were screened using title and abstracts, 1750 were excluded. 50 full-text studies remaining were further full screened with focus on the effectiveness of predictive analytics methods in precision public health in relation to the aforementioned causes of death and issues of public health concern and adherence to the inclusion criteria, this yielded 17 eligible studies which were English publication and human studies. The **diagram 1** below is a PRISMA flow chart (PRISMA checklist, 2015), which gives a detailed representation of the study selection process. Findings from the included studies were summarised using tables (**table 2, 3 and 4**). Using the Delphi method, the quality assessment criteria based on the primary objective of this study were adjusted to identify the final quality of the selected studies and the threshold

for inclusion of any study was set at five (Dalkey, Helmer & Helmer,1963). All selected studies scored minimum of 6 points which is equivalent to 50%; and above when ‘yes’

which carried the weight of ‘1’ and ‘no’ which carried the weight of ‘0’ where used for the evaluation of included studies; therefore, they were all included in the review.

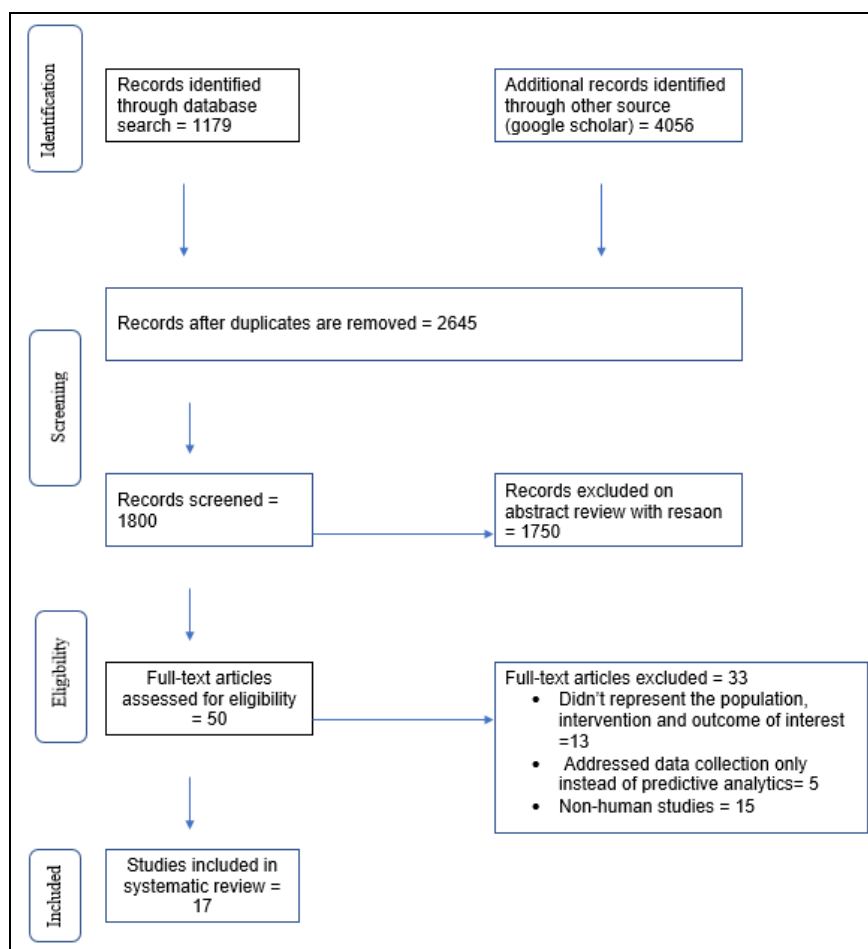


Diagram 1: Prisma Flow Diagram.

Study characteristics of selected studies reporting the effectiveness of predictive analytics methods in precision public health in relation to the issues of public health concern.

Serial number Author and Year	Aim	Country of study	Data source	Data Attributes	Issue of public health concern
1 Bakar, Kefli, Adullah and Sahani, 2011	To develop a predictive model using multiple classifiers to predict dengue outbreak	Malaysia	Valid patient cases reported at the National university of Malaysia Medical centre, 2003-2009	Year, epidemic week, age, gender, address, occupation, type of dengue, incubation period, type of outbreak, repetition case and death code.	Dengue fever. Communicable disease
2 Akbulut, Ertugrul and Topcu, 2018	To build a predictive system using machine learning techniques with assistive e-Health application to detect foetal health status	Turkey	Maternal questionnaire and evaluation. 3 clinicians from RadyoEmar radiodiagnostics centre in Istanbul. Population: 96 pregnant women	Maternal age, blood serotype, foetal age, past delivery numbers, diabetes history of mother, alcohol and smoking status, drug usage during pregnancy, presence of disabled people in parents' family.	Foetal and Neonatal conditions/ infections

3 Allen et al., 2020	To determine whether a machine learning algorithm can minimize racial bias in patient risk predictions as compared with commonly used rules-based methods	United States	Medical information mart for intensive care III (MMIC-III) database Population: 28,460 patients 23,263 whites and 5,197 non-whites	Diastolic and systolic blood pressure, heart rate, temperature, respiratory rate, oxygen saturation, white blood cell, platelets, creatinine, Glasgow coma scale, fraction of inspired oxygen, potassium and sodium levels	Health (racial) disparity in different causes of in-hospital mortality
4 Salim et al., 2021	To detect spatiotemporal dengue hotspots area in Selangor, association of the outbreak with climatic variables and evaluate machine learning models for predicting the dengue fever outbreaks	Malaysia	e-Notifikasi system in the Ministry of health in the year 2013-2017. Six districts were selected in Selangor: Gombak, Hulu Selangor, Hulu Langat, Klang and Petaling.	Age, gender, address, date of birth, date of onset, date of notification, humidity, rainfall, temperature and windspeed	Dengue fever. Communicable disease
5 Balzer et al., 2020	To use population HIV-testing data from rural Kenya and Uganda to construct HIV risk scores and assess their ability to identify seroconversion	Kenya Uganda (East Africa)	Data from 16 communities in the intervention arm of the SEARCH study conducted in rural Uganda and Kenya in 2013-2017	Age, sex, marital status, alcohol use, family planning, male circumcision, education, occupation, mobility and head of house	Human immunodeficiency virus
6 Pei et al., 2018	To conduct a risk predicting model for incident of essential hypertension in North urban of Chinese Han population using support vector machine (SVM)	China	i. Beijing Anzhen physical examination centre ii. Beijing Changping district community health service centre iii. Beijing Changping Huilongguan community health service centre	Environmental and genetic data	Hypertension (high blood pressure)
7 Al Mamlook et al., 2020	To build a machine learning-based model to predict the severity of traffic crash injuries in adults over 60 and identify the main causality factors	Michigan, United States	Michigan traffic crash facts dataset. A subset of 106,274 records of drivers of age 60+	Speed limit, age, car age, traffic volume features, alcohol indicator, crash year, injury	Motor vehicle injuries
8 Wang et al., 2018	To use machine learning method to classify obese and non-obese individuals based on obesity	Taiwan	139 individuals from 2014-2015 genome-wide SNP analysis	Clinicopathological features	Obesity



	associated single-nucleotide polymorphisms (SNP) and evaluate the performance of machine methods				
9 Tao et al., 2018	To develop an automatic, reliable and fast ischemic heart disease detection or localization system using machine learning	China	Subjects were from a retrospective study of cardiovascular disease including multiple hospitals	Raw magnetocardiography signal, maximum cardiac current distribution characteristics and magnetic field map pattern	Ischemic heart disease (coronary heart disease)
10 Ehrentraut, Kvist, Sparrelid and Dalians, 2013	To apply machine learning technique to the problem of detection healthcare associated infections	Sweden	Electronic health records from 120 in patients at a major university hospital in 2012	Structured and unstructured hospital visit data	Healthcare associated infections
11 Nagata et al., 2021	To detect extreme overdose and underdose prescription that occur very rarely in clinical practices using unsupervised machine learning algorithms	Japan	Electronic health records from Kyushu university hospital from January 1, 2014 to December 31, 2019	Age, weight and dose	Prescription drug overdose and underdose
12 Zhang-James et al., 2020	To develop prediction model using machine learning algorithm to identify attention-deficit/hyperactivity disorder (ADHD) youths at risk of substance use disorder (SUD)	Sweden	19,787 children born between 1989-1993 with lifetime diagnosis of ADHD from the National patient registry	Health registers	Substance abuse
13 Nishat et al., 2021	To apply different machine learning algorithms in the detection of chronic kidney disease for the purpose of evaluating their performances and accuracies	California, United states of America	Dataset from the machine learning repository of university of California	Age, blood pressure, red and white blood cells count, anaemia, pedal enema, blood glucose random etc	Chronic kidney disease
14 Muhammad, Algehyne and Usman, 2020	To demonstrate the capability of supervised machine learning algorithms to predict diabetes mellitus type 2.	Nigeria	383 diagnostic data of diabetes mellitus type 2 Muritala Mohammed specialist hospital, Kano	Age, family history, glucose, cholesterol, blood pressure, high density lipoprotein, triglyceride, body mass and diagnosis result	Diabetes
15 So, Hooshyar,	To develop a two-layer model using	Seoul, Korea	Clinical data from people who	Demographic data and mini mental state	Dementia

Park and Lim, 2017	machine learning technique for early diagnosis of dementia and classification of dementia		visited the Gangbuk-Gu dementia centre	examination (MMSE-KC)	
16 Li et al., 2020	To develop a fully automatic framework to detect CoVID-19 using chest CT and evaluate its performance	Multicentre	All positive cases of CoVID-19 with RT-PCR were acquired from December 31-February 17, 2020	3D volumetric chest scan	Lower respiratory disease. Lung disease
17 Myers et al., 2019	To accelerate early diagnosis of familial hypercholesterolemia patients at risk of heart attack by applying machine learning to large healthcare encounter dataset	United State of America	Longitudinal data from academic health systems National healthcare encounter database Integrated healthcare delivery system	Electronic health record data	Genetic disorder that increases likelihood of coronary heart disease at a younger age

**Table 2:** Characteristics of the included studies based on aim, country of study, data source, data attributes and issue of public health concern addressed.

The studies were conducted in the following countries: Malaysia, Turkey, Peru, Kenya, Uganda, China, United States, Taiwan, Sweden, Japan, Nigeria and Korea. The studies obtained all necessary approval and consent from study participants. All the included studies used quantitative approach; statistical analysis result was documented. Three studies (Li et al., 2020; Myers et al., 2019; Pei et al., 2018), conducted their research based on data gathered from multiple centres. Only a study gathered its data from multiple countries (Balzer et al., 2020). One study (Al Mamlook et al., 2020) used data from traffic crash facts dataset. Data gathered from previous study was used by two studies (Tao et al., 2018; Wang et al., 2018). Two studies (Pei et al., 2018; Salim et al., 2021) data sources included metrological, environmental and climatic data. Only one study (Akbulut, Ertugrul and Topcu, 2018) included data gathered from survey questionnaires.

All the data attributes used in the selected studies included demographic data based on biological factor (age, gender, ethnicity), individual lifestyle factor, social and community network, general socioeconomic, cultural and environmental

conditions. Li et al., 2020 used 3D- volumetric scan data while Tao et al., 2018, used magnetocardiographic signals data. Two studies (Bakar, Kefli, Adullah and Sahani 2011; Salim et al., 2021) included disease outbreak data.

The issues of public health concern mentioned addressed in the selected studies are; dengue fever, neonatal and foetal anomalies, health disparity and mortality causalities, HIV, hypertension, motor vehicle injuries, obesity, ischemic heart disease, healthcare associated infections, substance abuse, chronic kidney disease, genetic disorder, prescription drug overdose and underdose, dementia, lower respiratory diseases, diabetes mellitus and COVID -19. One study (Allen et al., 2020) reported the mortality caused as a result of an issue of public health concern.

More information as regard the authors, year of study and the aim of all the selected studies can be found in **Table 2**.

**Review of predictive analytics techniques in the selected studies**

Serial number Author and Year	Predictive analytics technique	Validation set	Evaluation and Performance measures	Outcome reported
1 Bakar, Kefli, Adullah and Sahani, 2011	Decision tree, Rough set classifier, Associative classification, Naïve bayes classifier	Tenfold cross validation	Accuracy, Average accuracy, ROC	The performance of single and multiple classifiers were able to produce better accuracy with more quality rules compared to the single classifiers. Machine learning techniques gave reasonable accuracy in predicting the

				outbreak of dengue fever.
2 Akbulut, Ertugrul and Topcu, 2018	Average perceptron, Boosted decision tree, Bayes point machine, Decision jungle, Locally-deep support vector machine, Logistic regression (LR), Neural network support vector machine	Tenfold cross validation	Accuracy, Precision, Recall, F-measure	The machine learning technique gave an estimate which was sufficient to give insight of foetal health. It was efficient and assistive in predicting congenital anomalies compared to the traditional methods
3 Allen et al., 2020	XG Boost, modified early warning score (MEWS), simplified acute physiology score (SAPS-II), acute physiologic assessment and chronic health evaluation (APACHE)	Tenfold cross validation	Sensitivity, Specificity, AUROC <sup>c</sup>	The machine learning algorithm showed unbiased predictive performance, which made in an accurate and appropriate tool for identification and stratification of all patients at risk of in-hospital mortality
4 Salim et al., 2021	Decision tree (CART), Artificial neural network (MLP), Bayes network (TAN), Support vector machine (linear, polynomial, RBF)	Wasn't mentioned	Accuracy, Precision, Recall, F-measure	Machine learning models are useful and displayed great potential in classification and prediction of dengue fever outbreak
5 Balzer et al., 2020	Risk based prediction using demographics, Model based, Logistic Regression (LR)	Fivefold cross validation	AUC, Sensitivity	Machine learning resulted in notable gain in sensitivity when controlling the rate of positive prediction hence, it is an effective approach for targeted prevention delivery. ML improved classification compared to model or risk based prediction
6 Pei et al., 2018	Support vector machine with radial kernel function, Support vector machine with Laplacian function	Tenfold cross validation	AUC, Sensitivity, Specificity, Accuracy	The SVM showed high prediction accuracy when dealing with classified problems although, its performance relied on the chosen kernel function
7 Al Mamlook et al., 2020	Naïve Bayesian (NB), Decision tree (DT), Logistic regression (LR), Light-GBM, Random forest (RF)	Tenfold cross validation	Accuracy, Precision, Recall, F-measure	Light-GBM model displayed great accuracy and can be used to effectively identify the key factor causing crashes in injury severity for the elderly
8 Wang et al., 2018	Support vector machine (SVM), K-nearest neighbor (KNN), Decision tree (DT)	Fivefold cross validation	AUC, Sensitivity, Specificity, Accuracy	ML methods demonstrated great performance and they provide feasible means for conducting preliminary analyses of obesity based on genetic factors
9 Tao et al.,	Support vector machine (SVM),	Tenfold cross	Accuracy, Precision,	The SVM-XGBoost was suitable for early and post-surgical diagnosis.

2018	K-nearest neighbor (KNN), Decision tree (DT), XG Boost	validation	Recall, F-score AUC	It was an accurate tool for interpretation of magnetocardiographic data
10 Ehrentraut, Kvist, Sparrelid and Dalians, 2013	Support vector machine (SVM), Random forest (RF)	Tenfold cross validation	Precision, Recall, F-score	RF was effective in making informed decision with good dependence between the class and features. The pre-processing method led to no statistically significant improvement in the performance of the techniques
11 Nagata et al., 2021	One-class support vector machine (OCSVM)	Fivefold cross validation	Precision, Recall, F-measure	OCSVM displayed high performance in synthetic data analysis and detected majority of the clinical overdose and underdose
12 Zhang-James et al., 2020	Cross-sectional mode (Random Forest), Longitudinal model (Recurrent neural network)	Wasn't reported	AUC, Average AUC Low brier score	The techniques yielded significant prediction and notably the longitudinal deep learning model can be used in monitoring and prediction substance use disorder in child development over years
13 Nishat et al., 2021	Naïve Bayes (NB), Decision tree (DT), Logistic regression (LR), Random forest (RF), Support vector machine (SVM), K-nearest neighbor (KNN), Multilayer perceptron (MLP), Quadratic discriminant analysis (QDA)	Random search cross validation	Accuracy, Precision, Sensitivity, F-score ROC-AUC	The ML techniques were effective in the prediction of chronic kidney disease. RF was the most efficient
14 Muhammad, Algehyne and Usman, 2020	Naïve Bayes (NB), Logistic regression (LR), Random forest (RF), Support vector machine (SVM), K-nearest neighbor (KNN), Gradient boost	Wasn't reported	Accuracy, ROC	The ML models were effective in accurately predicting and diagnosing diabetes mellitus type 2. RF had the best performance in accuracy while GB and RF had the best in term of ROC
15 So, Hooshyar, Park and Lim, 2017	Naïve Bayes (NB), Logistic regression (LR), Random forest (RF), Support vector machine (SVM), Bayesian network (BN), Multilayer perceptron (MLP), Bagging	Tenfold cross validation	Accuracy, Precision, Recall, F-measure	MLP showed effectiveness in diagnosis of dementia when SVM was most effective in classification of dementia and mild cognitive impairment.
16	Deep learning model	Wasn't	Sensitivity,	Deep learning model can accurately

Li et al., 2020		reported	Specificity, AUC	detect COVID- 19 and differentiate it from community acquired pneumonia and other lung infections
17 Myers et al., 2019	FIND FH (two consecutive random forest model layers)	Twofold cross validation	Positive predictive value, Recall, Area under the precision recall curve, Area under the receiver operating characteristics curve	The machine learning model was able to swiftly scan large databases and identify individuals with familial hypercholesterolaemia

**Table 3:** outcomes on predictive analytics techniques, validation set, evaluation measure and the effectiveness among the selected articles.

Artificial intelligence (AI) is the most common predictive analytics used in precision public health and machine learning is an essential domain of AI (Silver et al., 2017). Presently, most modern AI application framework are system whose functionality is to effectively execute a single task that is, narrow artificial intelligence (Mesko, Henteyi & Gyorrffy, 2018). Machine learning is a domain of AI, infers learning from existing data, past encounters to predict new outcome (Wahl, Cossy-Gratner, Germann, Schwalbe, 2018). Supervised machine learning uses training data which are forms of identified data while unsupervised machine learning involves data mining that involves finding and learning patterns in large data set (Ganasegeran & Abdulrahman, 2019).

Seven studies (Akbulut, Ertugrul and Topcu, 2018; Al Mamlook et al., 2020; Bakar, Kefli, Adullah and Sahani, 2011; Nishat et al.,2021; Salim et al., 2021; Tao et al., 2018; Wang et al., 2018) used decision tree, it is a supervised machine learning method in which the model is trained to classify a target variable by learning simple decision rules from existing input variables (Ghiasi, Zendejboudi & Mohsenipour, 2020). One study (Nagata et al., 2021) used one class support vector machine (OCSVM) which is an unsupervised machine learning approach, it is a method that learns decision function, fit a hyperplane and detect abnormal data deviation from the decision boundary. Six studies (Akbulut, Ertugrul and Topcu, 2018; Al Mamlook et al., 2020; Balzer et al., 2020; Muhammad, Algehyne and Usman, 2020; Nishat et al.,2021; So, Hooshyar, Park and Lim, 2017) used logistic regression (LR), it is an SML classification model, it estimates the probability of the occurrence of an event within a certain class or event by fitting a curve line between variables (Wright, 1995). Five studies (Al Mamlook et al., 2020; Bakar et al., 2011; Li et al., 2020; Nishat et al.,2021; So et al.,2017) used Naives bayes/ bayesian technique (NB), it is a SML that learns using training data from prior proability of each class and enforces independence of features while classifying data (Jackins, Vimal, Kaliappan & Lee, 2021). Nine studies (Akbulut, Ertugrul and Topcu, 2018; Ehrentraut, Kvist, Sparrelid and Dalians, 2013; Muhammad, Algehyne and Usman, 2020; Nishat et al.,2021; Pei et

al., 2018; Salim et al., 2021; So et al., 2017; Tao et al., 2018; Wang et al., 2018) utilized support vector machine (SVM), it is a SML that identifies classifier of a given data and proffers solution for both regression and classification problems (Cortes & Vapnik, 1995). Random forest (RF) is a SML that is an ensemble learning consisting of many decision trees and provides output class of individual trees (Smith, 2017). RF was used in seven studies (Al Mamlook et al., 2020; Ehrentraut et al., 2013; Muhammad, Algehyne and Usman, 2020; Myers et al., 2019; Nishat et al.,2021; So et al., 2017; Zhang-James et al., 2020). Neural network (NN) was used in three studies (Akbulut, Ertugrul and Topcu, 2018; Salim et al., 2021; Zhang-James et al., 2020), this model produces a network of cells and the connection between the cells are regulated such that the network can learn patterns and structures (LeCun, Bengio & Hinton, 2015). Allen et al., 2020 used deep learning which is a technique that uses neural system to execute numerous capacities. Akbulut, Ertugrul and Topcu, 2018 was the only study that used neural network support vector machine. K-nearest neighbor (KNN), this SML does its classification of unknown events based on its neighbour's classification, it doesn't train dataset (Muncherino, Papajorgji & Pardalos, 2009). KNN was implemented in for studies (Muhammad, Algehyne and Usman, 2020; Nishat et al.,2021; Tao et al., 2018; Wang et al., 2018). Two studies (Salim et al., 2021; So et al., 2017) used Bayesian network or Bayes network, this graphically simulates probabilistic relation between related variables and can address cases where there are missing data (Friedman, Geiger & Goldszmidt, 1997). Multilayer perceptron is a feed forward neural network which must include a proposed response as it aim of learning is to translate data into the proposed response (Gardener & Dorling, 1998). Multilayer perceptron was used in two studies (Nishat et al.,2021; So et al., 2017) while average perceptron was used in a study (Akbulut, Ertugrul and Topcu, 2018). XGBoost and gradient boost combines result from multiple decision trees model to predict result (Chen & Guestrin, 2016). Allen et al., 2020 and Tao et al., 2018 used XGBoost while Muhammad, Algehyne and Usman, 2020 used gradient boost. Quadratic discriminant analysis is a statistical classifier that disjoints multiple classes



of data by utilizing quadratic decision surface (Silitonga et al., 2021), it was used in a study by Nishat et al., 2021. Light GBM is a gradient boosting framework, and it was used in a study by Al Mamlook et al., 2020. Bagging was implemented in a study by So et al. (2017), it is an ensemble learning model that improves the accuracy of machine learning used in statistical classification and regression analysis (Quinlan, 1996). Decision jungle was used in a study by Akbulut, Ertugrul and Topcu (2018), it is an ensemble of decision direct acyclic graphs. Rough classifier is a classifier based on the rough set theory and it can manage vagueness in knowledge while associative classification is a classification algorithm that uses the association rule extracting model to determine rules and build classifier (Pal & Jain, 2010; Pawlak, Grzymala-Bussr, Slowinski & Ziarko,1995). Both rough classifier and association classification were used in a study by Bakar et al., 2011.

SVM, decision tree and random forest techniques were the most used among the researcher respectively. Four studies (Allen

et al., 2020; Li et al., 2020; Myers et al., 2019; Nagata et al., 2021) compared different machine learning algorithms while only one study (Pei et al., 2018) compared the different kernel function of the same SVM algorithm.

The evaluation and performance measure used in the selected studies were: sensitivity/recall, specificity, accuracy, precision, F-measure/score, average accuracy, low brier score, positive predictive value, area under the precision curve, area under the receiver characteristics operating curve (ROC-AUC), receiver operating curve (ROC), area under the ROC curve or area under curve (AUC) and average AUC. However, the most commonly used measures were sensitivity, accuracy, precision and F-measure.

More detailed information as regard validation set, evaluation and performance measure can be obtained from **table 3 and 4.**

Serial number Author and Year	Best technique	Specificity (%)	Accuracy (%)	Sensitivity/ Recall (%)	Precision (%)	AUC (%)	Average AUC (%)	ROC-AUC (%)	F-measure (%)	ROC (%)	Low brier score
1 Bakar, Kefli, Adullah and Sahani, 2011	Multiple classifier	-	63.5	-	-	-	-	-	-	68.0	-
2 Akbulut, Ertugrul and Topcu, 2018	Decision forest	-	89.5	75.0	75.0	-	-	-	75.0	-	-
3 Allen et al., 2020	XG boost	65.6	-	75.1	-	-	-	78.0	-	-	-
4 Salim et al., 2021	SVM with linear kernel	95.0	70.0	14.0	56.0	-	-	-	-	-	-
5 Balzer et al., 2020	Logistic regression	-	-	73.0	-	73.0	-	-	-	-	-
6 Pei et al., 2018	SVM with Laplacian function	86.7	80.1	63.3	-	88.6	-	-	-	-	-
7 Al Mamlook et al., 2020	Light GBM	-	87.54	81.40	87.90	-	-	-	83.70	-	-
8 Wang et al., 2018	SVM	63.02	70.77	80.09	-	73.00	-	-	-	-	-
9 Tao et al.,	SVN-XG	-	94.03	94.78	86.56	-	-	-	92.79	-	-

2018	Boost										
10 Ehrentraut, Kvist, Sparrelid and Dalians, 2013	Random forest (RF)	-	-	86.00	80.00	-	-	-	83.00	-	-
11 Nagata et al., 2021	OCS VM	-	-	96.90	98.00	-	-	-	97.30	-	-
12 Zhang-James et al., 2020	Recurrent neural network	-	-	-	-	-	63.00	-	-	-	86.00
13 Nishat et al., 2021	Random forest (RF)	-	99.75	100.00	100.00	-	-	99.00	100.00	-	-
14 Muhammad, Algehyne and Usman, 2020	Random forest (RF)	-	88.76	-	-	-	-	-	-	86.28	-
15 So, Hooshyar, Park and Lim, 2017	Multi layer perceptron	-	97.20	97.00	97.00	-	-	-	97.00	-	-
16 Li et al., 2020	Deep learning	96.00	-	94.00	-	98.00	-	-	-	-	-
17 Myers et al., 2019	Random forest layers	-	-	45.00	-	55.00	-	89.00	-	-	-

**Table 4:** best technique of each study and its evaluation and performance scores.

## Discussion

The aim of this systematic review is to explore the effectiveness of predictive analytic methods in precision public health reviewing eligible studies that have conducted in the past. This systematic review assessed studies in which predictive analytics techniques were used in some issues of public health concern. The studies were extracted based on the set inclusion and exclusion criteria after searching nine databases. The characteristics of included studies such as data attributes, issues of public health concern addressed, predictive analytics technique used, outcome reported, evaluation and performance measures were explored. All studies were conducted in the last 11 years.

From the findings in the result section above, the researchers used several predictive analytics techniques. Artificial intelligence

(AI) was the most common technique used which corroborates with the report of Silver et al., (2017), that AI is the most common predictive analytics in precision public health. All the 17 studies included reported that predictive analytics especially artificial intelligence techniques showed great performance in achieving their studies' aims. The techniques were reported to have improved and assisted in the various areas of public health concerns in which they were utilized.

Omics (demographic data), clinical, social, environmental and economic data were collected in the selected studies. This data showed that public health and precision medicine were related and has some sort of convergence. It also showed the relevance of big data in the merging of the both aforementioned health fields. Velmovitsy, Bevilacqua, Alencor, Cowan and Morita (2020) stated that "big data is a glue that brings precision medicine and public

health together”, this coincides with our findings. Multiple data attributes and sources reported in the selected studies proved the need for stratification and understanding of data. It also gave a reminder that for public health to be precise in its measures, it must rely on a rich evidence-base data from diverse sources. Data collection is just a phase, generating insight through the use of analytics is essential to make meaning of the huge amount of the diverse data. Integrating these ambiguous data requires a swift and efficient method that doesn't just gather but can analyse and forecast at a population level.

Data pre-processing is the process of detecting and removing incomplete, missing or incorrect data from a dataset while data normalization is the process of changing multiple ranges of values to values of the same kind (Muhammad, Algehyne and Usman, 2020). Selected studies used both process to overcome limitation such as missing data. This review revealed that many studies developed datasets based on single data sources. Most of the predictive analytics techniques were developed from a one-dimensional dataset gathered from a single centre or data source. Most of the studies were retrospective. Although diverse data attributes were collated, multidimensional datasets can produce a more accurate and generalizable predictive model. Wolpert and Macready (1995) showed that testing of variety of machine algorithms is the best way of determining the most accurate machine learning based application. Most of the studies compared different types of machine learning algorithms and the algorithm that outperformed was reported. Bagging, boosting and sampling methods were used in the studies that employed supervised machine learning techniques to mitigate class imbalance.

Of the 17 included studies, only four of the studies used neural network and deep learning methods. This infers that machine learning methods were the most common. Although, results from these methods are easier to evaluate and translate, they are still regarded as black box (Wang, 2003), which causes adoption difficulties as understanding of causal pathway and justification of actions are of core value to clinicians in health care related issues. All studies used a part of the selected dataset as a testing set and the other as a training set. Standard evaluation metrics such as accuracy, specificity, sensitivity or recall and precision were used for determining the performance of the developed predictive analytics technique. Study designs were majorly retrospective and prone to bias such as heterogeneity of data, sources, and sampling bias. According to Shickel, Tighe, Bihorac and Rashidi (2017) electronic health records (EHR) data produces good prediction performance models, though, using raw HER data in deep learning models is not very effective because HER has bias such as frequency of health visits, patient population and many more which are capable of confounding randomization.

One of the top priorities of precision public health (PPH) is to reduce health disparities (CDC, 2022). In a study by Allen et al. (2020) machine learning was applied in minimizing bias in in-hospital mortality predictors. The study reported that machine learning algorithm showed an unbiased performance and was accurate in its predictions. It reduced the potential of racial bias between white and non-white groups. Improving ability to prevent diseases by early detection of outbreaks and efficient surveillance

are also objectives of precision public health. Bakar et al. (2011) used valid patient cases to develop a predictive model to forecast dengue fever outbreak while Salim et al. (2021) used demographic and climatic. Data to develop machine learning models to detect spatiotemporal dengue hotspots and predict its outbreak. Though, these studies are 10 years apart, using different models and classifiers, they have a similarity which was their report on the reasonable accuracy and great potential in classification and prediction displayed by the techniques.

Findings from this review indicates that the aspect of precision public health that involves applying evolving predictive analytics to measuring, predicting, and detecting diseases, exposures, targeted treatment and prevention have been greatly explored while aspects that involves application of this emerging techniques to evaluate policies and implementation programs seem overlooked. For instance, a study by Muhammad, Algehyne and Usman (2020) demonstrated the capability of supervised machine in predicting diabetes mellitus, but this review could not find a study in which such technique was applied to evaluate if the sugar tax has been beneficial in terms of reducing population obesity. Balzer et al. (2020) used population HIV- testing data from two East African countries to construct HIV risk scores and assess their ability to identify seroconversion. This is a good application of predictive analytics to precision public health. This study provided evidence of the effectiveness of the technique at a population-level. This HIV-risk prediction models enhanced with data can assist in the identification and stratification of people for targeted prevention delivery in an epidemic setting.

An issue in precision public health is the lack of agreement on the units used for inference (Gottlieb, Francis & Beck, 2018). It is unclear if population is measured by geographic area or sub-population based on diseases or medical conditions. Most studies in this review applied predictive analytics technique to sub-population based on health conditions such kidney disease, obesity, substance abuse, dementia, lung disease, neonatal conditions, genetic disorder, heart disease and many more.

Precision public health is an expansive form of precision medicine which introduces the latter to population health. This review revealed that several research in which predictive analytics majorly artificial intelligence have been applied to public health issues, have been conducted. However, it still seems as though there is less focus on precision public health and there are several obstacles in the way of generalizability of the predictive models developed. The performances of the predictive analytics technique in this review were based on the specific application and quality of data, therefore there is a need for a standardized protocol, real - world application for justification, comparative models and multidimensional dataset to further validate the effectiveness of the developed models.

The strengths of this systematic review include the in-depth and broad search strategy used, the variety of methodologies (qualitative and quantitative) used in the included studies. Inclusion of grey journal search to minimize publication bias and the study selection process was presented using PRISMA guideline. Most of the studies trained and cross validated their datasets. Limitations of

this review included that most of the studies used one-dimensional dataset. The study selection and extraction process were conducted by a reviewer, and this can result in bias. MESH and Boolean logic were used to search databases and studies were selected based on reviewer's discretion and study's aim applying inclusion and exclusion criteria. Though, the reviewer adhered strictly to the criteria this could still have led to erroneous exclusion of some relevant articles. Hence, we cannot conclude that this systematic review includes all the studies on the effectiveness of predictive analytics in precision public health.

The COVID-19 pandemic accentuates the role of public health in safeguarding population health. The outbreak of the virus resulted in varying pandemic mortality across countries. According to WHO's COVID-19 dashboard, the virus is responsible for over 6 million deaths globally, over 500,000 deaths in the United States, India, and Brazil, below 10,000 deaths in Nigeria, Hong Kong and China and below 100 deaths in Iceland, Burundi and Liechtenstein. The pandemic uncovered some disparities and inequities in public health. There is need to understand and discern why some people who have same demographic and comorbidities can still exhibit different responses to a certain intervention. With the resurgence of COVID-19 and its different variants, precision in tailoring treatments and interventions to population-level traits is essential. Three studies conducted in 2020 by Ellinghaus et al.; Kuo et al. and Van der Made et al., respectively revealed that gene variants on chromosome 3 (3p21.31), and chromosome 9 (9q34.2), ApoEε4 and loss-of-function variants of X-chromosomal *TLR7* are related to severe COVID-19 infections. Another study by Emeruwa et al. (2020) reported that there is an association between built environment, socioeconomic factors, and COVID-19. Strengthening public health in this age requires an encompassing rather than a bland approach. Understanding this genetic variations, biological susceptibility and the impacts of other determinants of health can assist in proffering targeted interventions. Precision public health goes beyond precision prevention, and it involves transition from response to recovery (Dumit et al., 2021). Precision public health shows potential in tackling the present and shaping the future in the advent of pandemic. Precision public health enhanced with data and predictive analytics can help identify infection hotspots, develop robust tracing systems, lessen economic burdens amongst several opportunities (Zhou et al., 2021). The pandemic has given opportunity to precision public health to evolve, it also established that, an effectual public health measure requires a holistic and intersectoral tactic.

## Conclusion

This systemic review is a comprehensive review of 17 studies that deployed predictive analytic in precision public health. With the studies were published from year 2011 to 2021. Nine databases were systematically searched, and the result of this review was on three aforementioned questions result section includes three aspects: Information on data attribute, issues of public health concern, type of predictive analytics technique used, and effectiveness of the technique were extracted.

All studies included in this review reported that predictive analytics was effective in precision public health instances in which

it was deployed. This review targeted assessing and exploring the effectiveness of predictive analytics methods in precision public health in relation to issues of public health concern. Precision public health which is focused on improving the health of the population through precise prevention and tailored intervention, is still at its infancy and more efforts lie ahead in other develop a robust evidentiary foundation for use. There are numerous gaps and methodological limitations such as missing data, systematic bias, data inaccuracy, sample bias that need to be overcome. Although, predictive analytics techniques have been reported to be efficient that is a need for more advances. There is a need for a consensus on how population and subpopulation are defined. The theory for precision public health enhanced with predictive analytics has shown significant progression, considerations should be given to how these techniques would be implemented and evaluated in practice.

## References

1. Akbulut K, Ertugrul E, Topcu V (2018). Fetal health status prediction based on maternal clinical history using machine learning techniques. *Computer Methods and Programs in Biomedicine*, 163:87-100.
2. Al Mamlook R, Abdulhameed T, Hasan R, Al-Shaikhli H, Mohammed I, et al. (2020) "Utilizing Machine Learning Models to Predict the Car Crash Injury Severity among Elderly Drivers," 2020 IEEE International Conference on Electro Information Technology (EIT),105-111.
3. Allen A, Mataraso S, Siefkas A, Burdick H, Braden G, et al. (2020). A Racially Unbiased, Machine Learning Approach to Prediction of Mortality: Algorithm Development Study. *Jornal of Medical Internet Research public health and surveillance*, 6(4):e22400.
4. Bakar A, Kefli Z, Abdullah S, Sahani M (2011). Predictive models for dengue outbreak using multiple rule base classifiers. *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*, 1-6.
5. Balzer B, Havlir D, Kanya M, Chamie G, Charlebois E, et al (2020). Machine Learning to Identify Persons at High-Risk of Human Immunodeficiency Virus Acquisition in Rural Kenya and Uganda. *Clinical Infectious Diseases*, 71:2326-2333.
6. Bryman A (2102). *Social Research Methods*, 4th edition. Oxford: Oxford University Press.
7. CDC (U.S. Centers for Disease Control Prevention). 2001 Outbreak of Ebola Hemorrhagic Fever, Uganda, August 2000-January 2001 Morbidity and Mortality Weekly Report 50:73-77.
8. Centre for Disease Control. (2020). Deaths and Mortality.
9. Centre for Disease Control. (2021). What is public health? Retrieved December 29, 2021.
10. Centre for Disease Control. (2022). About the prevention status report. Retrieved January 13, 2022.
11. Chen T, Guestrin C (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco.
12. Chowkwanyun M, Bayer R, Galea S (2018). "Precision" public health - between novelty and hype. *New England Journal of Medicine*, 379:1398- 400.



13. Cortes C, Vapnik V (1995). Support-vector networks. *Machine Learning*, 20:273-297.
14. Critical Appraisal Skills Programme. (2013.) Making sense of evidence [Internet]. Retrieved December 8, 2021.
15. Dahlgren G, Whitehead M (1991). Policies and strategies to promote social equity in health. Stockholm: Institute for future studies.
16. Dalkey N, Helmer O, Helmer O (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9:458-467.
17. De Vaus D (2002). *Research Design in Social Research*. London: Sage.
18. Dolley S (2018). Big data's role in precision public health. *Frontiers in Public Health*, 6, 68.
19. Ellinghaus D, et al. Genomewide association study of severe Covid-19 with respiratory failure. *New England Journal of Medicine*, 383:1522-1534.
20. Emeruwa U, Ona S, Shaman J, Turitz A, Wright J, et al. (2020). Associations Between Built Environment, Neighborhood Socioeconomic Status, and SARS-CoV-2 Infection Among Pregnant Women in New York City. *Journal of American Medical Association*, 324:390.
21. Escobar G, Puopolo K, Wi S, et al. (2014). Stratification of risk of early-onset sepsis in newborns 34 weeks' gestation. *Pediatrics*, 133:30-36.
22. Friedman N, Geiger D, Goldszmidt M (1997). Bayesian network classifiers. *Machine Learning*, 29:131-163.
23. Ganasegeran K, Abdulrahman S A (2019). Artificial Intelligence Applications in Tracking Health Behaviors During Disease Epidemics. *Human Behaviour Analysis Using Intelligent Systems*, 6:141-155.
24. Gansky S A, Shafik S (2020). At the crossroads of oral health inequities and precision public health. *Journal of public health dentistry*, 80:14-22.
25. Gardner M, Dorling S (1998). Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmospheric Environment Journals*, 32:2627-2636.
26. Ghiasi M, Zendejboudi S, Mohsenipour A (2020). Decision tree-based diagnosis of coronary artery disease: CART model. *Computer Methods and Programs in Biomedicine*, 192.
27. Ginsburg G, Phillips K (2018). Precision medicine: from science to value. *Health Affairs*, 37:694-701.
28. Heneghan C, Badenoch D (2002). *Evidence-based medicine toolkit*. London: BMJ Books.
29. Horton R (2018). Offline: in defence of precision public health. *Lancet*, 392, 1504.
30. Jackins V, Vimal S, Kaliappan M, Lee M (2021). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77:5198-5219.
31. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, 2:230-243.
32. Kee F, Taylor-Robinson D (2020). Scientific challenges for precision public health. *Journal Epidemiology Community Health*, 74:311-314.
33. Kee F, Taylor-Robinson D (2020). Scientific challenges for precision public health. *Journal of Epidemiology and Community Health*.
34. Khoury M (2016). The Shift From Personalized Medicine to Precision Medicine and Precision Public Health: Words Matter!. *Centers for Disease Control and Prevention*.
35. Khoury M, Engelagu M, Chambers D, Mensah G (2018). Beyond public health genomics: can big data and predictive analytics deliver precision public health? *Public Health Genomics*, 21:244-250.
36. Khoury M, Lademarco M, Riley W (2016) Precision public health for the era of precision medicine. *American Journal of Preventive Medicine*, 50:398-401.
37. Kindig D, Stoddart G (2003). What is population health? *American journal of public health*, 93:380-383.
38. Kohane I (2015). Health Care Policy. Ten things we have to do to achieve precision medicine. *Science*, 349:37-38.
39. Kuo C L, et al. APOE e4 genotype predicts severe COVID-19 in the UK Biobank community cohort. *Journal of Gerontology, Biological Sciences and Medical Sciences*, 75:2231-2232.
40. LeCun Y, Bengio Y, G Hinton (2015). Deep learning. *Nature*, 521:436-444.
41. Leff D, Yang G (2015). Big data for precision medicine. *Engineering*, 1:277-279.
42. Li L, Qin L, Xu Z, Yin Y, Wang X, et al. (2020). Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. *Radiology*, 296:E65-E71.
43. Lopes J, Guimarães T, Santos M F (2020). Predictive and Prescriptive Analytics in Healthcare: A Survey. *Procedia Computer Science*, 170:1029-1034.
44. Madan A, Bebrin M, Lazer D, Pentland A (2010). Social sensing to model epidemiological behavior change. *Proceedings of UBIComp 2010 12th ACM Conference on Ubiquitous Computing*. (pp. 291-300). New York, NY.
45. Mesko B, Hetenyi G, Gyorffy Z (2018). Will artificial intelligence solve the human crisis in healthcare? *BioMed Central Health Services Research*, 18, 545.
46. Mucherino A, Papajorgji P, Pardalos P (2009). " K -nearest neighbor classification " in *Data Mining in Agriculture*, pp. 83-106. New York: Springer.
47. Muhammad L, Algehyne E, Usman S (2020). Predictive Supervised Machine Learning Models for Diabetes Mellitus. *SN Computer Science*, 1, 240.
48. Nagata K, Tsuji T, Suetsugu K, Muraoka K, Watanabe H, et al. (2021). Detection of overdose and underdose prescriptions—An unsupervised machine learning approach. *PLoS one*, 16:e0260315.
49. Oh D, Jerman P, Silvério M S, Koita K, Boparia S, et al. (2018). Systematic review of pediatric health outcomes associated with childhood adversity. *BMC Pediatrics* 18, 83.
50. Pal P, Jain R (2010). Combinatorial approach of associative classification. *International Journal of Advanced Networking and Applications*, 2:470-474.
51. Pawlak Z, Grzymala-Bussr J, Slowinski R, Ziarko W (1995). Rough Set. *Communication of the Association of Computing Machinery*, 38(11).



52. Pei Z, Liu J, Liu M, Zhou W, Yan P, et al. (2018). Risk-Predicting Model for Incident of Essential Hypertension Based on Environmental and Genetic Factors with Support Vector Machine. *Interdisciplinary Sciences: Computational Life Sciences*, 10:126-130.
53. PRISMA checklist [Internet]. (2015). Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Retrieved December 8, 2019.
54. Prospero M, Min J, Bian J, Modave F (2018). Big data hurdles in precision medicine and precision public health. *BioMed Central Medical Informatics Decision Making*, 18:1-15.
55. Public health surveillance, World Health Organization (accessed January 14, 2016).
56. Quinlan J (1996). Bagging, boosting, and C4. 5. *Association of the Advancement of Artificial Intelligence*, 1:725-730.
57. Robinsin P, Lowe J (2015). Literature reviews vs systematic reviews. *Australian and New Zealand Journal of Public Health*, 39:103-103.
58. Rychetnik L, Frommer M, Hawe P, Shiell A (2002). Criteria for evaluating evidence on public health interventions. *Journal of Epidemiology and Community Health* 56:119-127.
59. Salim N, Wah Y, Reeves C, et al. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific Reports* 11, 939.
60. Silitonga P, Bustamam A, Muradi H, Mangunwardoyo W, Dewi B (2021). Comparison of Dengue Predictive Models Developed Using Artificial Neural Network and Discriminant Analysis with Small Dataset. *Applied Sciences*, 11(3).
61. Silver D, Schrittwieser J, Simonyan, Antonoglou I, Huang A, et al. (2017). Mastering the game of Go without human knowledge. *Nature*, 550:354-359.
62. Smith C (2017). *Decision Trees and Random Forests: A Visual Introduction for Beginners*. Vancouver, BC, Canada: Blue Windmill Media.
63. So A, Hooshyar D, Park K, Lim H (2017). *Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques*. *Applied Sciences*, 7(7):651.
64. Stead WW (2018). Clinical implications and challenges of artificial intelligence and deep learning. *Journal of American Medical Association*, 320:1107-1108.
65. Sun G, Matsui T, Hakozaki Y, Abe S (2015). An infectious disease/fever screening radar system which stratifies higher-risk patients within ten seconds using a neural network and the fuzzy grouping method. *The Journal of Infection*, 70:230-236.
66. Tao R, Zhang S, Huang X, Tao M, Ma J, et al. (2018). *Magnetocardiography based Ischemic Heart Disease Detection and Localization using Machine Learning Methods*. *IEEE Transactions on Biomedical Engineering*, 1-1.
67. Van der Made C, et al. (2020). Presence of genetic variants among young men with severe COVID-19. *Journal of American Medical Association*, 324:663-673.
68. Velmovitsky PE, Miranda PADSES, Vaillancourt H, Donovska T, Teague J, Morita PP. Blockchain and IoT: a conceptual framework for a blockchain consent platform in active assisted living. *J Med Internet Res*. (2020) 22:e20832.
69. Wahl B, Cossy-Gantner A, Germann S, Schwalbe R (2018). Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *British Medical Journal Global Health*, 3(000798).
70. Wang H, Chang S, Lin W, Chen C, Chiang S, et al. (2018). Machine Learning-Based Method for Obesity Risk Evaluation Using Single-Nucleotide Polymorphisms Derived from Next-Generation Sequencing. *Journal of Computational Biology*.
71. Wang J (2003). *Data Mining: Opportunities and Challenges*. Monticano, Italy: Idea Group.
72. Weeramanthri T, Dawkins H, Baynam G, Bellgard M, Gudes O, et al. (2018). Precision public health. *Frontiers in Public Health* 6, 121.
73. Wolpert, C. & Macready, W. (1995). No free lunch theorems for search.
74. World Health Organisation (2021). Public health surveillance. Retrieved December 30, 2021.
75. World Health Organisation (2022). Global Health Estimates: Life expectancy and leading causes of death and disability. Geneva: WHO.
76. Wright R (1995). "Logistic regression" in Reading and Understanding Multivariate Statistics. *American Psychological Association*, 217-244.
77. Zhang-James Y, Chen Q, Kuja-Halkola R, Lichtenstein P, Larsson H, et al. (2020). Machine-Learning prediction of comorbid substance use disorders in ADHD youth using Swedish registry data. *Journal of Child Psychology and Psychiatry*, 61:1370-1379.
78. Zhou A, Sabatello M, Eyal G, Lee S, Rowe J, et al. (2021). Is precision medicine relevant in the age of COVID-19? *Genetics in Medicine*, 23:999-1000.

**Citation:** Islamiyyat Adekemi O, Damola Sheriff O (2022) Effectiveness of Predictive Analytics in Precision Public Health in Strengthening Health System for Future Pandemics. *Adv Pub Health Com Trop Med: APCTM-160*.