



Integrating Machine Learning, Deep Learning, and Geospatial Data Science for Public Health Surveillance and Health Inequality Assessment: A Scoping Review

Mohammad Ali^{1*}, George Edison², Alexandra Harry³

¹Independent Researcher Iraq

²Independent Researcher, France

³Independent Researcher USA

¹m.ali.m2000m@gmail.com, ²Geogeedison2nice@gmail.com, ³Alaxendraharry37@gmail.com



ABSTRACT

Corresponding Author

Mohammad Ali

m.ali.m2000m@gmail.com

Article History:

Submitted: 01-09-2026

Accepted: 08-10-2026

Published: 13-10-2026

Keywords

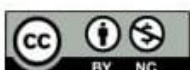
Machine Learning, Deep Learning, Geospatial Data Science, Public Health Surveillance, Disease Prediction, Health Equity.

Global Trends in Science and Technology is licensed under a Creative Commons Attribution-Noncommercial 4.0 International (CC BY-NC 4.0).

Combining machine learning (ML), deep learning (DL) with geospatial data science (GIS) holds potential for improving public health surveillance and the evaluation of health inequalities. The scoping review examines their use in disease prediction, spatial risk mapping, outbreak detection and health facility accessibility analysis. It explores geospatial data sources, analytic methods and the possibilities of using an integrated model to identify vulnerable social groups and geographical health inequalities. Some of the major challenges are data quality, algorithmic bias, privacy, model interpretability, and computational resources. The review identifies research priorities to build on robust, fair, and transparent analytical tools to support evidence-based public health decision-making and mitigate health inequalities.

INTRODUCTION

Public health surveillance is an essential tool for disease surveillance, detection of emerging health threats and evidence-based decision making. Public health surveillance has historically depended on the use of epidemiological records, clinical reports, population surveys and regular health information systems to learn about the patterns of disease and evaluate population health [1]. Today however, with the escalating complexities of healthcare issues such as outbreaks of infectious diseases, chronic illness and environmental risks and socio-economic inequalities, there is a need for more sophisticated analytical techniques [2]. The use of machine learning (ML), deep learning (DL) and geospatial data science in this context provides major opportunities to increase the quality of public





health surveillance and enhance the analysis of health inequalities.

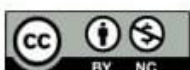
Using machine learning, researchers can find patterns in large, complicated datasets, classify health outcomes and even predict disease risks. This is where deep learning, a specialized type of machine learning, comes into play, as it is able to process high-dimensional, unstructured data such as satellite images, medical images, mobility data and other complex health data [3]. Geospatial data science, meanwhile, offers tools to analyse the spatial patterns of diseases, detect geographic clusters, create maps of healthcare accessibility, and explore environmental conditions and health outcomes. Combining these methods could give a complete picture of disease transmission and the differing levels of health risk in different geographical areas and within population groups [4].

Health inequality is an important public health issue, as disparities in income, education, jobs, housing, environment and health care access can lead to unequal health outcomes. Such differences are frequently geographic (between urban and rural areas or between regions) and/or socioeconomic (between different socioeconomic groups). Predictive modelling and spatial analysis can aid in identifying unserved populations, high-risk geographic areas, and in allocating health care resources equitably [5]. However, key issues of data quality, privacy, algorithmic bias, model interpretability, and the lack of representation of vulnerable populations are still relevant.

The purpose of this scoping review is to explore the current literature related to the integration of machine learning, deep learning, and geospatial data science for public health surveillance and health inequality assessment. It aims to showcase the most important methodological trends, data sources, applications, new trends and research gaps in this interdisciplinary domain. The review aims to explain the potential role of these technologies in relation to disease monitoring, spatial risk prediction and the identification of health inequalities, by summarizing the available evidence. This review should help researchers, public health professionals, and decision-makers become aware of the opportunities and challenges of integrative analytical approaches to improve and inform more effective, equitable, and data-driven public health practice.

DATA SOURCES, SEARCH STRATEGY AND STUDY SELECTION

To conduct a scoping review on how machine learning, deep learning and geospatial data science are integrated in public health surveillance and health inequality assessment a systematic and transparent approach to identifying relevant literature is critical. This section elaborates the source of scientific evidence, literature searching strategy and the procedure of study selections for identifying relevant studies [6]. To make sure that the review is comprehensive of the available knowledge, methodological advances and opportunities for further research in this interdisciplinary area, a structured approach can be used [7].





Major academic databases such as PubMed, Scopus, Web of Science, IEEE Xplore and ScienceDirect are used to find relevant literature. The PubMed biomedical and public health research database, Scopus, and Web of Science have a wide array of interdisciplinary publications [8]. IEEE Xplore is especially helpful in locating technical research involving machine learning algorithms, deep learning architectures, remote sensing and geospatial computing. ScienceDirect can offer more evidence in the fields of spatial epidemiology, environmental health and computational methods. Google Scholar can also be used to find follow-up studies and publications that have been found in the main databases [9].

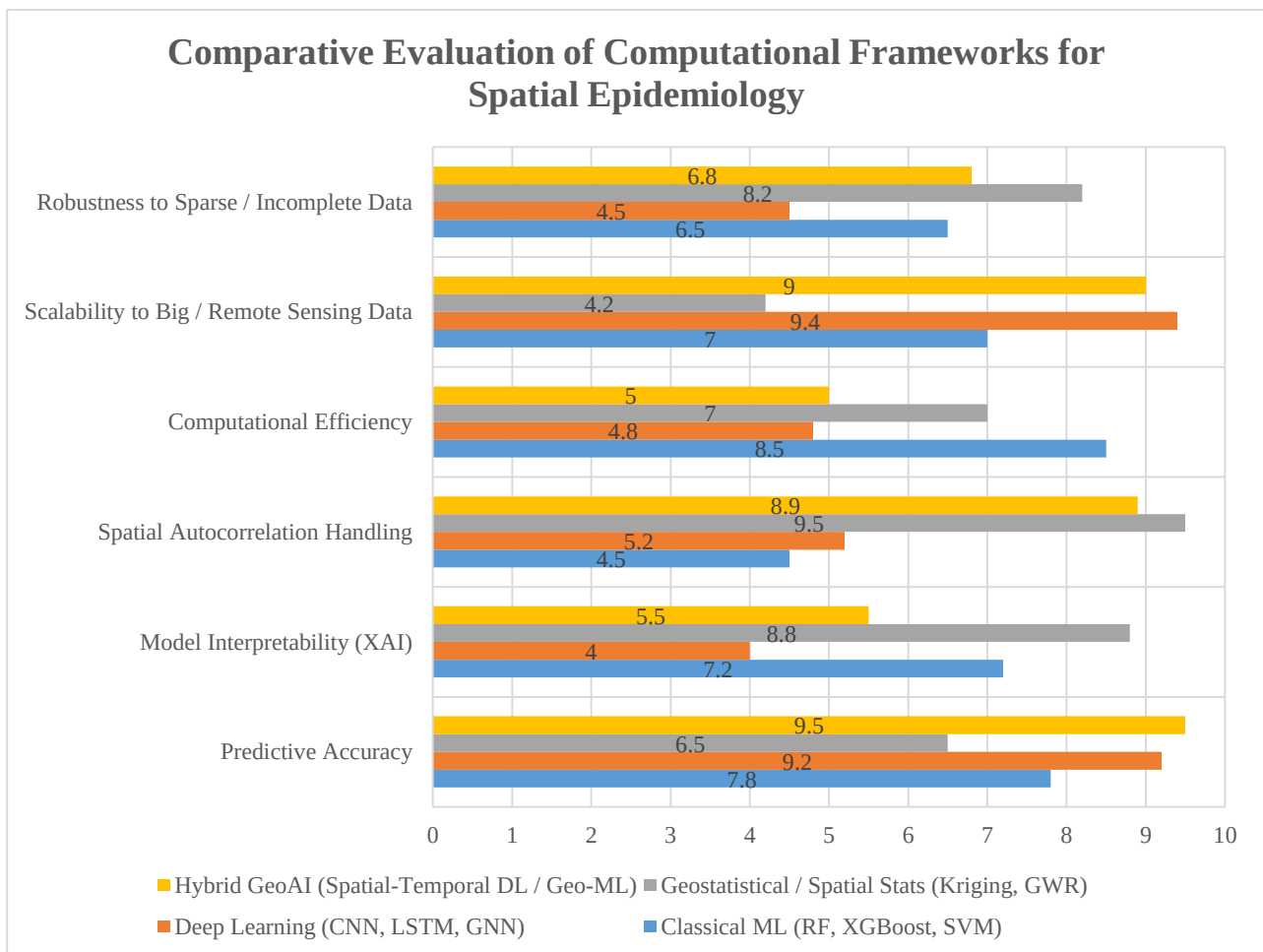
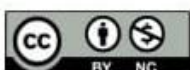


Figure 1. Comparative Evaluation of Computational Frameworks for Spatial Epidemiology

These can involve machine learning, deep learning, AI, geospatial data science, spatial epidemiology, public health surveillance, health inequalities, health disparities, disease prediction and spatial analysis [10]. The user can use Boolean operators (AND, OR etc.) to refine their search and receive more relevant publications. Machine learning or deep learning could be part of a search together with geospatial analysis and public health surveillance or health inequality assessment, for example. Search dates, search terms for the databases, and any limitations on language and/or publication year





should be retained to enhance transparency and reproducibility [11].

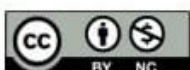
A series of clearly defined stages should guide the choice of studies. First, the duplication should be removed and then titles and abstracts should be screened to identify potentially relevant publications. Full-text assessment should then follow, to check and identify which of the studies selected are eligible for the review [12]. Studies can be included if it is exploring the application of machine learning, deep learning, or geospatial analytical techniques for public health surveillance, disease mapping, spatial risk prediction, healthcare access, or health inequality assessments. Publications may not be included if there is no public health or geospatial application [13].

PUBLIC HEALTH SURVEILLANCE AND HEALTH INEQUALITIES ASSESSMENT

In the twenty-first century, the study of public health relies upon public health surveillance and the assessment of health inequalities as cornerstone activities of public health monitoring population health, shaping the pattern of disease and understanding health inequalities [14]. Public health surveillance is defined as the ongoing and systematic gathering, analysis, interpretation and dissemination of health-related information used for planning, executing and assessing public health responses. It helps public health authorities identify outbreaks at the earliest opportunity, and monitor the epidemiology of communicable and non-communicable diseases, as well as the effectiveness of interventions and populations vulnerable to elevated risk [15].

The traditional surveillance systems are predominantly based on clinical records, laboratory reports, population surveys and administrative health databases. These sources may be limited by reporting delays, incomplete reporting and limited spatial detail, however. To overcome some of the above challenges, machine learning, deep learning, and geospatial data science can be integrated, enabling better disease forecasting, automatic pattern recognition, and spatial risk assessment [16].

The main objective of health inequality assessment is to discover and quantify disparities in health status, access to health services and health outcomes among different populations. The differences may be linked to socio-economic status, gender, age, education, ethnicity, employment, environmental exposure, and geographic location [17]. While there are some health disparities related to biology, there are many others that are associated with preventable, inequitable social, economic and environmental conditions. To understand the inequities in health and to design strategies to achieve equitable health care access and better health outcomes, it is important to assess these inequities [18].



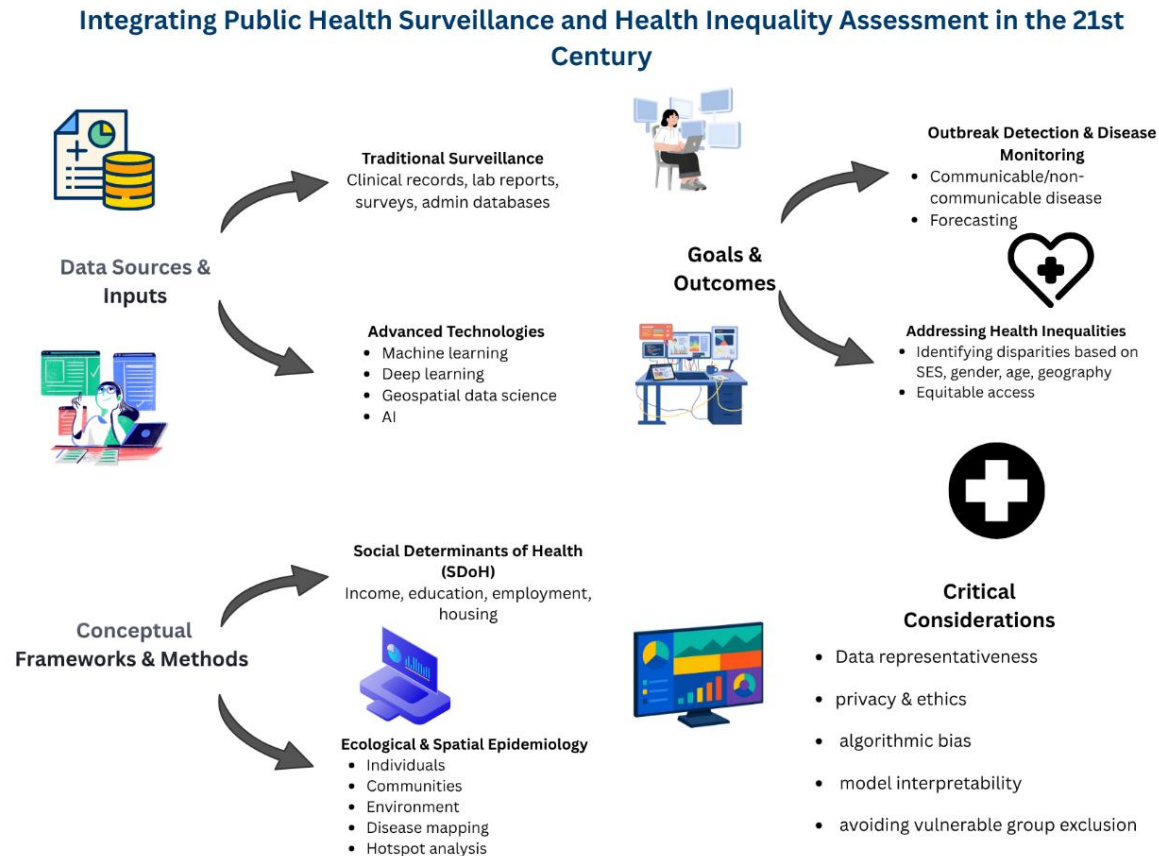


Figure 2. Integrating Public Health Surveillance and Health Inequality Assessment in the 21st Century

Various conceptual frameworks can inform the integration of public health surveillance and health inequality evaluation. The social determinants of health (SDoH) framework focuses on the role that individual social determinants such as living conditions, income, education, employment, housing, and social environment have on health outcomes [19]. The ecological model takes into account the relationship between individuals, communities, institutions and the environment. Spatial epidemiology offers another approach to the study of geographical variation in disease occurrence and the detection of areas at higher risk of disease by using spatial clustering, disease mapping, and hotspot analysis [20].

In these contexts, machine learning can be used to detect intricate health outcomes and their potential factors, and deep learning can handle vast and complex data sets. Geospatial data science adds another layer of analysis by including geographic space, spatial dependence, environmental features and accessibility metrics [21]. These methods can be combined to assist in identifying underserved communities, forecast the disease hotspot and evaluate geographical disparities in providing health services.

However, there are considerations of data representativeness, privacy, algorithmic bias, model



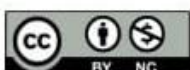
interpretability and the potential of excluding vulnerable groups that must be taken into consideration during effective application [22]. Conceptual frameworks should, therefore, not only influence the technical integration of analytical methods, but also interpret the findings in social and geographical contexts. Integrating these views can enhance the public health's monitoring capacity and inform effective policies and actions to address preventable health disparities [23].

GEOSPATIAL DATA SOURCES & INTEGRATION FOR PUBLIC HEALTH RESEARCH

Geospatial data are vital to public health research because they can be used to capture the geographical distribution of disease, environmental exposures, health care facilities, and population attributes. The linkage of geospatial data with health records and socioeconomic data facilitates the study of spatial variations in health outcomes, mapping disease hotspots and understanding environmental and social determinants of health inequalities [24]. Over time, and thanks to the use of a variety of information sources, and progress in machine learning, deep learning, and geographic information systems (GIS), opportunities for more accurate and comprehensive public health surveillance have increased [25]. There are a number of geospatial data sources that are commonly used in public health research. Satellite imagery and remote sensing data can be used to capture many environmental characteristics that could affect health, including land use, vegetation, surface temperature, air pollution, flooding, and others. Geographic information systems are the tools that allow for the storage, management, visualization and analysis of spatial information [26]. The use of a global positioning system (GPS) and mobility data can provide researchers with information on population movements, epidemiology, and access to health services. Census data, household surveys, electronic health records, hospital disease registries, and disease surveillance systems in addition can offer demographic, socioeconomic and clinical data that are geographically linked to individuals [27].

Environmental datasets are especially useful for the study of association between environmental exposures and health effects. For instance, satellite-derived estimates of air pollution can be linked to health data on respiratory disease to determine the geographical region of overlap between exposure to air pollution and burden of respiratory disease [28]. Likewise, climate data such as temperature, precipitation, and vegetation can aid in research on climate-sensitive diseases, such as malaria and dengue fever. Population density, road networks and the locations of healthcare facilities could also be combined to evaluate geographical barriers to medical services and determine populations that are not being served [29].

Several pre-processing and harmonization steps are necessary to integrate these various datasets. Typical issues are spatial resolution, coordinate reference systems, time coverage, data format, missing data and geographical boundary [30]. The geographic coordinates must be standardized, the





temporal periods must be synchronized, there must be no missing values and the data sets must be compatible before analyses can be made. Spatial aggregation may be used to merge data from multiple sources while retaining meaningful geographic relationships; geographic overlays and spatial joins are other techniques that can be used to do this [31]. But care must be taken when interpreting aggregated data because of the modifiable areal unit problem, ecological fallacy, and possible spatial misclassification.

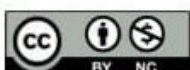
Integrated geospatial data can be enhanced by machine learning and deep learning techniques to increase their analytical value. Machine learning algorithms can be used to discover associations between environmental, demographic, and epidemiological data, and deep learning models can be used to derive complex features from high-dimensional space data provided by satellites [32]. These techniques can be used to assist in the prediction of disease risk, environmental health monitoring and identification of populations who are at risk for multiple health issues.

Yet, there are significant issues in data privacy, availability, quality, representativeness, and computational requirements that should be addressed when considering geospatial data integration. Sensitive location data must be safeguarded and gaps in data that impact rural and disadvantaged communities must be remedied [33]. The integration of accurate geospatial data and sophisticated analytical techniques serves as a powerful tool for enhancing public health monitoring, identifying health disparities, and guiding healthcare planning and decision-making [34].

APPLICATIONS OF MACHINE LEARNING IN PUBLIC HEALTH SURVEILLANCE

Machine learning (ML) is becoming a significant analytical tool in public health surveillance due to its capability of detecting intricate patterns, handling vast amounts of data, and foreseeing health outcomes. Unlike traditional statistics-based approaches which might assume certain patterns and relationships, machine learning algorithms can discover patterns within historical information and utilize that information to make forecasts or categories [35]. These functions help identify early disease outbreaks, predict outbreaks, assess risk, and identify populations that are vulnerable to and could benefit from specific public health interventions. The use of machine learning alongside epidemiological, demographic, and environmental and geospatial data can enhance the quality and speed of public health decision making [36].

A key use case for machine learning is in monitoring and forecasting outbreaks of infectious diseases. Random forests, support vector machines, decision trees, gradient boosting, and logistic regression algorithms are available to identify factors associated with disease occurrence and to estimate a probability of outbreaks [37]. Historical records of cases, rainfall, temperature, humidity, population density and sanitation indicators are some of the factors that can be used to assess the risk of dengue





fever. Likewise, machine learning models can help track influenza transmission and pinpoint locations where it might spread [38]. These predictive capabilities can be used to help prioritize surveillance activities and to allocate public health resources more effectively.

Another application is the evaluation of health inequalities. Socio-economic factors, health care utilization data, environmental exposures and disease burden can all be used in machine learning algorithms to identify populations with disproportionate health burden. The models can be combined with geospatial data to create risk maps which display geographical patterns of disease occurrence and healthcare requirements [39]. Spatially informed machine learning can also be used to identify communities that don't have access to hospitals, vaccination services, diagnostic facilities, and essential medicines. The results of these findings can help in equitable distribution of resources and targeted public health interventions [40].

Machine learning has several public health surveillance challenges. The quality, completeness and representativeness of the data available will affect model performance. Prediction may not be as reliable because of incomplete surveillance data, inadequate training data, or changes in disease patterns. In addition, complex models may be hard to understand and hence not easily accepted by healthcare practitioners and policy makers [41]. It is, therefore, vital that privacy protection, external validation, fairness assessment, and continuous monitoring be present.

Visualizing how Machine Learning integrates diverse data types to improve health outcomes

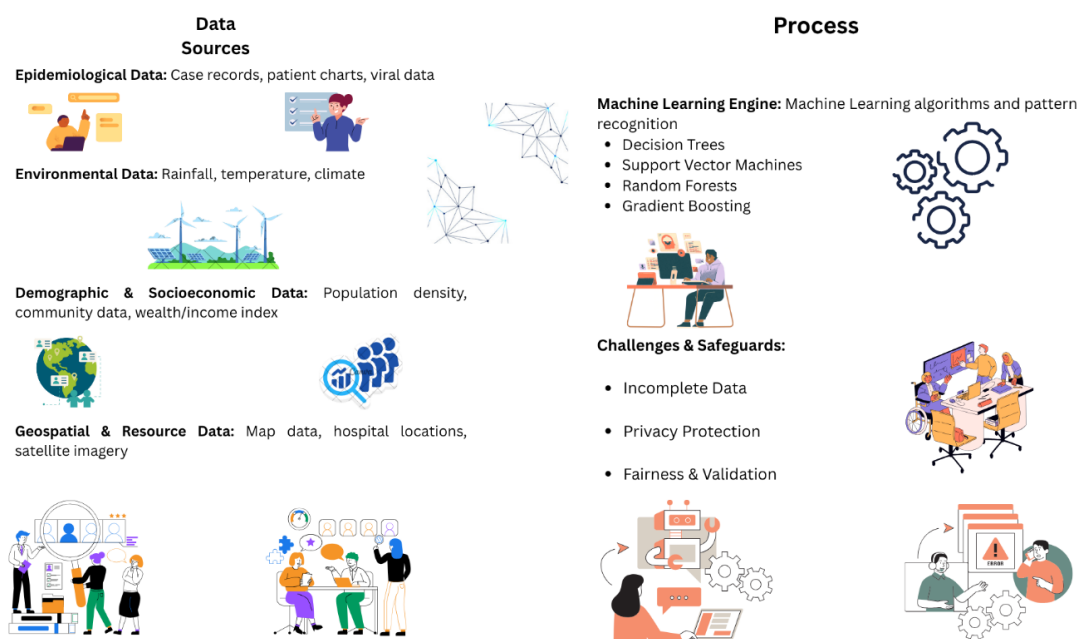


Figure 3. Visualizing how Machine Learning integrates diverse data types to improve health outcomes

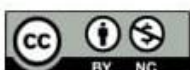


APPLICATIONS OF DEEP LEARNING IN SPATIAL EPIDEMIOLOGY AND PREDICTION OF DISEASE

Deep Learning (DL) is an advanced field of machine learning which involves the use of multilayered Artificial Neural Networks to detect complex patterns in large and high dimensional datasets. It can automatically extract relevant features from raw data sets, which makes it very useful for spatial epidemiology, disease prediction and public health surveillance [42]. Deep learning models can be trained to recognize meaningful information from satellite data, medical images, environmental data, demographic data and more. If combined with geospatial data science, these methods can enhance the ability to locate and identify disease hotspots, to anticipate the spatial distribution of disease, and to evaluate the spatial health risk to a population [43].

A prominent use of deep learning is predicting an infectious disease given the environmental and geographical data. Models like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short term memory (LSTM) networks can be used to identify spatial and temporal patterns related to disease transmission [44]. For instance, CNNs can be used to detect land-use patterns, vegetation types, and environmental conditions from satellite imagery that are related to mosquito breeding habitats. The LSTM networks can predict future disease incidence in time by reading a sequence of measurements of weather parameters (temperature, rainfall, humidity) and disease history [45]. These approaches can be used to aid surveillance of dengue, malaria, influenza and other infectious diseases provided sufficient training and validation data can be made available.

A second area in which deep learning is making a contribution to environmental determinants of health is in the analysis. Another area where deep learning is making a contribution to understanding environmental determinants of health is the analysis [46]. Environmental exposures, such as changes in vegetation, urban heat, air pollution, and flooding, can be estimated based on satellite-based observations and remote sensing data. These environmental indicators can be used in combination with epidemiological data to explore associations between environment and health [47]. Moreover, a deep learning model may be used to identify geographical regions where environmental hazards and disease burdens are co-located, as this will yield valuable information for public health interventions. Another significant use is spatial risk mapping and measuring health inequality. Deep learning models can combine population density, socioeconomic factors, healthcare access, environmental factors and disease surveillance data to pinpoint potentially vulnerable regions [48]. These models can be used to inform an understanding of underserved communities, for example, that have fewer healthcare facilities available or communities that have a lack of environmental and socioeconomic benefits. This data can help policymakers focus on key healthcare infrastructure, prevention measures and





emergency response resources [49].

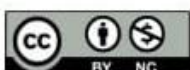
While deep learning has many benefits, it also has some drawbacks. They tend to be data-intensive simulations that demand significant computing power and a good amount of data. The complicated structures can also make the systems hard to understand and interpret, as to where a particular space might be considered to be at high risk [50]. Additional factors such as spatial data leakage, spatial differences between training and testing data, and data availability across geographic regions can also impact the model's performance. Thus, location verification, uncertainty analysis, privacy protection, and explainable artificial intelligence (XAI) methods are crucial [51].

INCORPORATION OF MACHINE LEARNING, DEEP LEARNING, AND GEOSPATIAL DATA SCIENCE SKILLS

Machine learning (ML), deep learning (DL), and geospatial data science (GDS) is an emerging interdisciplinary public health surveillance and health inequality assessment tool. Every discipline offers unique analytical skills which can be complementary to the other disciplines. Machine learning is used to detect patterns and relationships between structured health and social data, and deep learning is used to uncover complex features from large and unstructured data [52]. Geospatial data science brings spatial analysis, geographic visualization and location-based modeling to understand disease and health determinant variation across geographical areas [53]. An approach that melds these can help to provide a more holistic view of population health and enable more precise, timely, and place-specific public health interventions.

One of the critical elements in this integration is the creation of analytical tools that bring together epidemiological data and information on the environment, demographics, socioeconomics and geography. For example, disease surveillance can be linked to information on location of health facilities, population density, land use, etc, as well as data on temperature (which comes from satellites) and rainfall [54]. Machine learning algorithms such as random forests and gradient boost can then be used to identify the factors that are important in the presence of a disease. Satellite imagery can be analysed using deep learning architectures like convolutional neural network (CNN) and recurrent neural network (RNN) along with time-dependent environmental observations [55]. Geographic information systems (GIS) and spatial statistical methods can then be applied to visualize the predictions, determine disease hotspots and explore geographical variations in health risks.

This comprehensive approach is especially useful in the field of infectious disease monitoring. For example, deep learning can be used to identify environmental characteristics that are linked with the presence of mosquito vectors in dengue surveillance, and machine learning can be used to predict the risk of disease based on climatic, demographic, and epidemiological parameters [56]. This can then





be utilized to identify high-risk neighbourhoods and help prioritize vector-control activities. Malaria surveillance, respiratory disease monitoring and environmental health hazard assessment can be done using similar frameworks [57]. Combining spatial and temporal data can also enhance the monitoring of disease spread and detection of new public health risks.

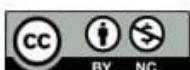
These technologies when applied together can further enhance the assessment of health inequality. An integrated model can be developed by adding other indicators such as household income, education, population density, travel time to health care facilities, and prevalence of disease, to allow for the identification of communities that are facing multiple disadvantages [58]. Spatial risk maps may identify geographic regions with high disease burdens, low health care access, and poor socioeconomic status. This information can guide policymakers in targeting the allocation of health resources and intervention development to the needs of local populations [59].

The challenges of successful integration involve ensuring data comparability, geographic resolution, temporal compatibility, computational demands, and model understanding. Quality of data and geographical coverage can lead to bias and complicated models can make results unclear. Appropriate spatial and temporal validation, uncertainty assessment, protection of sensitive health information and model fairness assessment across population groups should therefore be applied to researchers [60].

EVALUATING AND ADDRESSING HEALTH INEQUALITIES THROUGH SPATIAL ANALYSIS AND PREDICTIVE MODELLING

Health inequality assessment is a vital part of public health research since it aids in pinpointing the disparities in health outcomes, access to health services and susceptibility to health hazards in geographic units and population groups. These inequalities often go hand-in-hand with socioeconomic status, education, income, employment, environmental factors, and access to health care [61]. Advanced approaches to these differences can be achieved through spatial analytics and predictive modelling, which can help to identify those in need of support and inform where health burdens are likely to be located [62]. The use of machine learning, deep learning, and geospatial data science can enhance the identification and assessment of health inequalities even further.

Health outcome mapping by neighborhood, district, region and country can be done using spatial analytics. Geographic information systems (GIS) techniques and spatial clustering, spatial autocorrelation and hotspot analysis can be used to detect areas with above average prevalence or lack of health care access [63]. For instance, spatial analysis can show that in remote rural areas, communities may suffer from higher rates of preventable diseases due to the fact that they have to travel long distances to the hospital, lack medical staff or have limited diagnostic facilities. Likewise,





the health effects in neighborhoods in urban areas may be disproportionately affected due to the poor housing standards, high air pollution, and poor air quality [64]. The mapping process allows researchers to detect patterns geographically, and target areas for intervention with resources and focus.

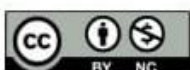
In addition to spatial analysis, predictive modeling can be used to estimate the probability of poor health outcomes from demographic, socioeconomic, environmental and health care factors. These factors can be used to predict health outcomes using machine learning algorithms such as random forests, support vector machines, and gradient boosting [65]. Satellite imagery, large-scale environmental data, and other high-dimensional data can also be used to feed into deep learning models to detect patterns related to population health risks. If these forecasts are combined with spatial information, spatial risk maps can be created showing vulnerable people and places at risk of increased inequalities [66].

One critical use is to evaluate health care access and use inequalities. The combination of population distribution, road network, transportation services, healthcare facilities, and socio-economic factors all enable researchers to gauge the equality of access to key health services between different population groups. These analyses can provide information for the location of hospitals, vaccination centers, maternal care centers and mobile medical services [67]. Predictive models can also be used to determine communities at risk of lack of service in disease outbreaks or during public health crises. While powerful, spatial analytics and predictive modeling have some drawbacks. Findings can be influenced by incomplete data, inaccurate geographical records, biased training data, and differences in populations. Spatial linkages do not imply causal linkages, and a range of data may mask inequalities faced by specific persons or communities [68]. Scientists therefore need to apply the right statistical techniques, test predictive models in different geographic regions, guard against disclosure of sensitive information and analyze results in a social and environmental framework [69].

CHALLENGES, RESEARCH GAPS AND FUTURE DIRECTIONS

While the application of machine learning (ML), deep learning (DL), and geospatial data science for public health surveillance and health inequality assessment continues to be a field with tremendous future potential, some methodological, technical, ethical, and practical questions remain. These constraints need to be addressed to achieve reliable, interpretable and equitable results from integrated analytical approaches [70]. Existing gaps also can inform future research efforts to improve public health surveillance systems and effective strategies to address health disparities.

Availability and quality of health and geospatial data is one of the key challenges. Public health data can be incomplete, have inconsistencies, be stale, and may be only available at a small level of





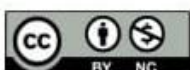
geographic detail. In resource poor settings, these issues are especially important where surveillance systems and digital health systems could be underdeveloped [71]. Epidemiological, environmental and socioeconomic data can be further complicated to integrate because of the differences in spatial resolution, data collection methods and reporting periods. Standardized data collection, interoperable data systems, better data sharing and ways to deal with missing and incomplete data should be future research priorities [72].

Model interpretability, reliability and generalizability are another important challenge. While the complex deep learning architectures might be able to recognize complex patterns, the decisions made by these architectures are often hard to explain. This constraint can diminish confidence in the model and in the public health efforts and/or policy decisions using it, especially if used to guide resource allocation or interventions for vulnerable populations [73]. Additionally, the performance of a trained model may vary across different geographical areas due to variations in the environment, population characteristics, healthcare infrastructure, and prevalence of diseases [74]. Explainable AI, external validation, spatial cross-validation, uncertainty estimation, and continuous monitoring of the model should be emphasized in the future studies.

Ethical issues are also an important research gap. Health surveillance systems often rely on the use of highly personal data such as medical records, demographic details, mobility data, and precise geographical locations. Lack of adequate privacy protection can leave an individual or community vulnerable to confidentiality issues [75]. Biased data sets may lead to biased predictions of marginalized or underrepresented groups that can create or magnify current health disparities. Researchers need to embed privacy preserving methods, clear data governance, evaluating fairness in research, and the community in the research process [76].

Another constraint is the lack of social determinants of health in more sophisticated predictive models. Some studies focus on algorithmic performance and have limited regard for issues of poverty, education, housing, discrimination, healthcare affordability, and other structural issues. Development of interdisciplinary studies that integrate epidemiology, geospatial analysis, computer science, and social science should be pursued in the future [77]. Longitudinal studies, real-time surveillance, multimodal data integration, and prospective performance testing of the models in operational public health environments also should be strengthened.

Barriers to implementation such as limited computing infrastructure, lack of technical expertise, financial constraints, and access to reliable datasets could hinder the use of advanced analytical systems in the LDCs. Some of these barriers can be addressed through capacity-building programs, open source tools of analysis, research networks and the development of models appropriate to the





context [78]. Addressing data limitations, ethical issues, methodological shortcomings, and implementation obstacles are critical in moving forward with integrated public health analytics. Further studies need to go beyond modelling with accuracy and develop well validated, user-friendly and scalable systems that will reduce the health inequity gap in various populations and geographical locations through better disease surveillance [79].

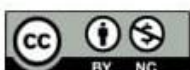
CONCLUSION

The new scope of machine learning (ML), deep learning (DL), and the use of geospatial data science (GDS) in public health surveillance and health inequality assessment is a promising development. This scoping review identifies the promise of these cross-disciplinary methods for better disease monitoring, prediction of health risks, geographical disease patterns, and research on access to and outcomes from healthcare. The integration of spatial analysis and computational intelligence allows researchers and public health professionals to gain a deeper understanding of the intricate interplay between environmental factors, socioeconomic determinants, geographical characteristics, and population health.

Machine learning offers useful tools for pattern learning for epidemiological, demographic, and socioeconomic data, while deep learning allows processing of complex and high-dimensional information such as satellite images, environmental observations, and temporal disease data. These approaches are complemented by the use of geographic information systems, spatial clustering, hotspot identification, and spatial risk mapping, all of which are elements of geospatial data science. Together, they can enhance infectious disease surveillance and monitoring, aid in forecasting outbreaks of infectious disease, evaluate environmental health hazards, and identify communities where health outcomes are disproportionately affected. Moreover, by incorporating a predictive model, geographical inequalities in access to health care services can be better identified and medical resources can be more effectively allocated based on population demand.

One of the major benefits of these technologies is that they can help to facilitate more equitable public health decision making. Integrated analytical frameworks that can encompass indicators of income, education, housing, environmental exposure, population density and distance to health care services can help identify vulnerable populations and limited access to necessary services. Such findings can help inform policy decisions for targeted interventions, to improve preventive health and mitigate preventable health disparities. Technological progress, though, is not enough to overcome structural inequalities; effective interventions need to look at social, economic and environmental factors or drivers of health.

While they have significant potential, there are a number of issues that need to be addressed. Issues



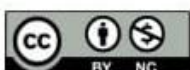


such as data incompleteness, inconsistent reporting, privacy, algorithmic bias, lack of interpretability of model, inadequate computational infrastructure, and the reliability and practical application of integrated approaches could be raised. Geographical differences in areas of operation may also affect the applicability of predictive models. For future research, high-quality data integration, explainable A.I., spatial and temporal validation, privacy preserving analytical approaches, and fairness evaluations in various population groups should be emphasized. Further collaboration between epidemiologists, data scientists, geographers, health care providers, and policy makers is needed to put research results into practice for the benefit of public health.

There is significant potential to harness machine learning, deep learning, and geospatial data science to shift public health surveillance from descriptive monitoring to predictive, spatially informed, and equity-driven decision-making. With the ongoing interdisciplinary research and responsible application, disease prediction can be enhanced, surveillance systems improved and healthcare planning better informed. In sum, these synergistic strategies can help to create more accessible, responsive, and equitable public health systems, especially when designed to serve the needs of underserved and vulnerable populations.

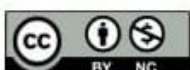
REFERENCES

- [1]. Sri AV, Morampudi MK, Dorsala MR, Bojjagani S, Khan MK. Privacy-Preserving Federated Learning for Retinal Disease Diagnosis using Paillier Homomorphic Encryption with Multiple Encryption Keys. IEEE Access. 2026 Feb 13.
- [2]. Oluwafemi B. Privacy-preserving computation (homomorphic encryption, MPC). Journal of Contemporary Educational Research. 2025 Sep 19;7(1):111-22.
- [3]. Naidu UG, Lakkshmanan A, Krishna JG, Elamathi E, Reddy TS. Federated AI framework for privacy-preserving differential diagnosis across distributed medical networks. In 2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA) 2025 Jun 25 (pp. 932-940). IEEE.
- [4]. Khan MZ, Abbass W, Abbas N, Javed MA, Alahmadi A, Majeed U. Safe-med for privacy-preserving federated learning in iomt via adversarial neural cryptography. Mathematics. 2025 Sep 12;13(18):2954.
- [5]. Shah ST, Ali Z, Waqar M, Kim A. Federated learning in public health: A systematic review of decentralized, equitable, and secure disease prevention approaches. In Healthcare 2025 Oct 30 (Vol. 13, No. 21, p. 2760). MDPI.



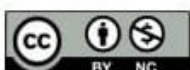


- [6]. Shafik W. Data privacy challenges in artificial intelligence-driven healthcare. In *The Convergence of Federated Learning and Healthcare 5.0 and Beyond: A New Era of Intelligent Health Systems* 2026 Feb 7 (pp. 67-98). Cham: Springer Nature Switzerland.
- [7]. Sri AV, Morampudi MK, Dorsala MR, Bojjagani S, Khan MK. Privacy-Preserving Federated Learning for Retinal Disease Diagnosis using Paillier Homomorphic Encryption with Multiple Encryption Keys. *IEEE Access*. 2026 Feb 13.
- [8]. Veda Sri A, Morampudi MK, Alahari S, Boggavarapu VV, Chennu J, Yakkala S. Privacy-preserving federated learning with homomorphic encryption: Alzheimer's detection use-case. In *Enabling Person-Centric Healthcare Using Ambient Assistive Technology, Volume 2: Personalized and Patient-Centric Healthcare Services in AAT* 2025 Feb 27 (pp. 111-125). Cham: Springer Nature Switzerland.
- [9]. Shah U, Dave I, Malde J, Mehta J, Kodeboyina S. Maintaining privacy in medical imaging with federated learning, deep learning, differential privacy, and encrypted computation. In *2021 6th international conference for convergence in technology (I2CT)* 2021 Apr 2 (pp. 1-6). IEEE.
- [10]. Butt HA, Rashid Z, Ahad A, Yousaf A, Imran I. Privacy-Preserving Machine Learning Models for Medical Data Ensuring Security in Smart Healthcare Systems. In *AI and Blockchain Applications for Privacy and Security in Smart Medical Systems* 2025 (pp. 339-370). IGI Global Scientific Publishing.
- [11]. Oad SK. Role of Artificial Intelligence in Modern Data Science: A Critical Review. *Global Journal of Universal Studies*. 2024 Jun 15;1(1).
- [12]. Singh JP, Aqsa A, Ghani I, Sonani R, Govindarajan V. Privacy-aware hierarchical federated learning in healthcare: integrating differential privacy and secure multi-party computation. *Future Internet*. 2025 Jul 31;17(8):345.
- [13]. Romeo J, Abbass M, Sherif A, Mamun MM, Elersy M, Khalil K. Privacy-preserving machine learning for e-health applications: A survey. In *2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)* 2024 Apr 13 (pp. 1-6). IEEE.
- [14]. Subbarayudu CB, Edwin PB, Hossain RR, Kiran PR, Singh KR, Yogapriya J, Elboyeva MP. AI-Based Federated Learning for Privacy-Preserving Healthcare AI Models. In *2025 IEEE 2nd International Conference on Green Industrial Electronics and Sustainable Technologies (GIEST)* 2025 Oct 11 (pp. 1-5). IEEE.
- [15]. Fouda MM, Fadlullah ZM, Ibrahem MI, Kato N. Privacy-preserving data-driven learning models for emerging communicatio



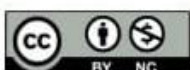


- [16]. Dusi P. Federated learning for next-gen computing applications and privacy-preserving medical diagnosis. Electronics, Communications, and Computing Summit. 2024 Jun 15;2(2):10-8.
- [17]. Kumari P. Privacy-Preserving Federated Learning for Early Prediction of Diabetes and Cardiovascular Disease using Multimodal Data from Rural Primary Health Centers. Natural Resources for Human Health. 2026 Sep 18;6(6):19-38.
- [18]. Jayasurya M. Federated Learning for Privacy-Preserving Healthcare Data Analytics. Peer-Reviewed Journal of Computer Science (PRJCS). 2026 Mar 9;1(3):20-6.
- [19]. Dolezel D, Lalani K, Watzlaf V, Butler-Henderson K, Lambert EV, Morton M, Sand J, Gibbs D, Fenton S. AI Privacy and Security in Healthcare: A Systematic Literature Review. International Journal of Telerehabilitation. 2026 Jun 1;18(1):6747.
- [20]. Adnan A, Kausar F, Shoaib M, Iqbal F, Altaf A, Asif HM. A secure and privacy-preserving approach to healthcare data collaboration. Symmetry. 2025 Jul 16;17(7):1139.
- [21]. Goyal P, Bhattacharyya S. for Privacy-Preserving Healthcare Diagnostics Using Biomedical Signal Analysis. InComputational Intelligence and Soft Computing: First International Conference, CISCom 2025, Melaka, Malaysia, November 26–27, 2025, Proceedings, Part II 2026 Mar 23 (p. 45). Springer Nature.
- [22]. Jamal M, Singh A. Federated and Edge Learning for Real-Time Data Analytics in Clinical Environments. Global Journal of Universal Studies. 2024 Jun 15;1(1).
- [23]. Oad SK. Integration of AI in Data Science: A Systematic Review. Global Trends in Science and Technology. 2026 Mar 18;2(2):64-82.
- [24]. Khan SR, Khan MW. A comprehensive review of privacy-preserving techniques in health data analysis. In2025 International Conference on Cognitive Computing in Engineering, Communications, Sciences and Biomedical Health Informatics (IC3ECSBHI) 2025 Jan 16 (pp. 104-109). IEEE.
- [25]. Kabeer MM. AI in Lean Six Sigma: A Review of Industrial Implementations, Benefits, and Barriers. Global Journal of Universal Studies. 2024 Jan 10;1(2):134-58.
- [26]. Ara J, Roy M, Swarnali SH. Differential privacy and federated learning for secure predictive modeling in healthcare finance. International Journal of Research and Innovation in Applied Science. 2025;10(9):745-59.
- [27]. Ezeogu AO. Homomorphic encryption in healthcare analytics: Enabling secure cloud-based population health computations. Journal of Advanced Research. 2025 Aug 17;1(02):42-60.



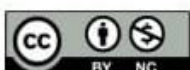


- [28]. Ammal SG, Sumathi R. Privacy-Preserving Healthcare Data Science: A Review of Federated Learning Approaches. In 2026 8th International Conference on Intelligent Sustainable Systems (ICISS) 2026 Mar 4 (pp. 219-225). IEEE.
- [29]. Jayalakshmi R, Tamilvizhi T. Privacy preservation in diabetic disease prediction using federated learning based on efficient cross stage recurrent model. Scientific Reports. 2025 Oct 24;15(1):37258.
- [30]. Müftüoğlu Z, Kızrak MA, Yıldırım T. Privacy-preserving mechanisms with explainability in assistive AI technologies. In Advances in Assistive Technologies: Selected Papers in Honour of Professor Nikolaos G. Bourbakis—Vol. 3 2021 Dec 1 (pp. 287-309). Cham: Springer International Publishing.
- [31]. Sarker D, Sadi MT, Gomes D, Hasan MM. Securing Patient Privacy and Data Protection by Federated Learning in AI-Driven Healthcare. In Enabling Collaborative Health Intelligence With Federated Learning 2026 (pp. 317-352). IGI Global Scientific Publishing.
- [32]. Garg M, Tank S, Tank BK. Federated Learning in Healthcare Finance: A Systematic Review of Privacy-Preserving Models. The Convergence of Federated Learning and Healthcare 5.0 and Beyond: A New Era of Intelligent Health Systems. 2026:743-57.
- [33]. Ay S, Topaloglu U, Zhang W. Safeguarding biomedical AI: a critical scoping review of privacy-enhancing technologies, hybrid approaches, and deployment models. Frontiers in Digital Health. 2026 Aug 17;8:1726771.
- [34]. Koutsoubis N, Waqas A, Yilmaz Y, Ramachandran RP, Schabath MB, Rasool G. Privacy-preserving federated learning and uncertainty quantification in medical imaging. Radiology: Artificial Intelligence. 2025 May 14;7(4):e240637.
- [35]. Butt M, Tariq N, Ashraf M, Alsagri HS, Moqurrah SA, Alhakbani HA, Alduraywish YA. A fog-based privacy-preserving federated learning system for smart healthcare applications. Electronics. 2023 Sep 28;12(19):4074.
- [36]. Chowdhury TK, Kudapa SP. Federated learning models for privacy-preserving data sharing and secure analytics in healthcare industry. International Journal of Business and Economics Insights. 2024 Dec 28;4(4):91-133.
- [37]. Vanangamoudiar ML, Senthilkumar C. Big Data-Driven Explainable AI And Privacy-Preserving Federated Deep Learning For Predictive Healthcare And Collaborative Clinical Intelligence. INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES. 2026 Feb 10:2586-611.



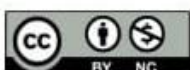


- [38]. Oad VD. Artificial Intelligence in Data Analytics: A Comprehensive Review. Global Journal of Universal Studies. 2024 Jun 15;1(1).
- [39]. Bathrinath S, Wankhede DS. Privacy-Preserving Machine Learning in Clinical Research: Using Federated Learning to Protect Patient Data. Machine Learning in Healthcare: Data-Driven Decisions, Predictive Modelling, Personalized Medicine. 2025 Dec 15:129.
- [40]. Raza H, Erdenetsogt T, Singh A, Farooq M, Kabeer MM, Aslam MS. A Comprehensive Review on Data Science Frameworks for Big Data Analytics. PERFECT: Journal of Smart Algorithms. 2026 Jan 6;3(1):1-0.
- [41]. Kannan N, Balasubramanian K. Privacy-Preserving and Quantum-Resistant Federated Learning Framework for Fog Healthcare Systems. International Journal of Computational Intelligence Systems. 2026 Jul 31.
- [42]. Dey P, Hossain E, Hossain MI, Chowdhury MA, Alam MS, Hossain MS, Andersson K. Comparative analysis of recurrent neural networks in stock price prediction for different frequency domains. Algorithms. 2021 Aug 22;14(8):251.
- [43]. Imtiaz N, Kundu TR, Roy A, Bhuiyan MI, Rahman K, Islam MK. Governance readiness beyond predictive performance: An empirical benchmark for higher-education early warning systems. Frontiers in Computer Science and Artificial Intelligence. 2025;4(5):49-65.
- [44]. Oad SK. AI-Powered Data Science: A Review of Current Developments and Emerging Opportunities. Global Trends in Science and Technology. 2026 May 13;2(2):263-82.
- [45]. Kabeer MM. Utilizing Machine Learning for Continuous Process Improvement in Lean Six Sigma. Global Trends in Science and Technology. 2025 May 7;1(2):49-63.
- [46]. Oad SK. Artificial Intelligence in Data Science: A Comprehensive Review. American Journal of Artificial Intelligence and Computing. 2026 Mar 17;2(1):1-25.
- [47]. Jamshaid M, Muhammad AH, Akbar Z, Niaz S, Siddique MN, Akbar S. Artificial intelligence generated deepfakes as instruments of disinformation: Examining their influence on public opinion, digital trust, and governance. Journal of Information Systems Engineering and Management. 2025;10.
- [48]. Oad VD. AI in Business Data Analytics: A Review of Current Practices and Future Direction. Global Trends in Science and Technology. 2026 Apr 15;2(2):156-74.
- [49]. Kabeer MM. Leveraging AI for process optimization: the future of quality assurance in lean six sigma. American Journal of Artificial Intelligence and Computing. 2025 May 7;1(1):87-103.





- [50]. Oad VD. Artificial Intelligence for Big Data Analytics: A Review of Trends and Challenges. *Global Trends in Science and Technology*. 2026 Mar 20;2(2):83-102.
- [51]. Jamal A, Raza H, Erdenetsogt T, Singh A, Farooq M, Kabeer MM, Aslam MS. AI and Data Analytics for Precision Agriculture: Current Progress and Future Directions. *JATAED: Journal of Appropriate Technology for Agriculture, Environment, and Development*. 2025 Aug 15;2(2):36-46.
- [52]. Aslam MS. Artificial Intelligence in Product Management: Driving Innovation and Market Success. *Global Journal of Universal Studies*. 2024 Dec 12;1(2):88-112.
- [53]. Oad VD. Exploring AI Techniques in Data Analytics: A Critical Review. *American Journal of Artificial Intelligence and Computing*. 2026 Apr 16;2(1):46-67.
- [54]. Singh A. Harnessing Artificial Intelligence in Product Management and Data Analytics: Trends, Applications, and Future Directions. *Global Trends in Science and Technology*. 2026 Jan 25;2(1):19-35.
- [55]. Abbasi N, Nizamullah FN, Zeb S. Ai in healthcare: Using cutting-edge technologies to revolutionize vaccine development and distribution. *JURIHUM: Jurnal Inovasi dan Humaniora*. 2023 Jun 14;1(1):17-29.
- [56]. Nizamullah F, Fahad M, Abbasi N, Qayyum MU, Zeb S. Ethical and legal challenges in AI-driven healthcare: patient privacy, data security, legal framework, and compliance. *Int. J. Innov. Res. Sci. Eng. Technol*. 2024;13:15216-23.
- [57]. Raza H, Erdenetsogt T, Farooq M, Kabeer MM, Aslam MS. Real-Time Data Analytics: Applications and Challenges in Smart Cities. *Global Journal of Universal Studies*. 2024 Jun 15;1(1).
- [58]. Shaikh MM. Artificial Intelligence and Workforce Performance Optimization in Government Supply Chains. *Global Trends in Science and Technology*. 2026 Jun 12;2(3):53-74.
- [59]. Dubey P, Dubey P, Bokoro PN. Federated learning for privacy-enhanced mental health prediction with multimodal data integration. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. 2025 Dec 31;13(1):2509672.
- [60]. Yuksel BB, Metin AY. Federated learning with homomorphic encryption for secure real time ECG anomaly detection: A multi institutional privacy preserving framework. *Biomedical Signal Processing and Control*. 2026 May 1;116:109557.
- [61]. Raza H, Singh A, Erdenetsogt T, Kabeer MM, Aslam MS, Farooq M. Machine Learning Driven Decision Making in the Modern Data Era. *PERFECT: Journal of Smart Algorithms*. 2026 Jan 6;3(1):11-22.





- [62]. Erdenetsogt T, Jamal M. Machine Learning for Big Data Analytics: A Comprehensive Review. *Global Trends in Science and Technology*. 2026 Apr 25;2(2):199-218.
- [63]. Mitra U, Rehman SU, Podder B, Dutta A, Bhowmick B. Federated Learning in Medical Diagnostics: Advancing Collaborative and Privacy-Preserving AI. In *Adversarial AI and Data Poisoning in Federated Learning 2026* (pp. 339-366). IGI Global Scientific Publishing.
- [64]. Sarraf G, Pal V. Privacy-Preserving Data Processing in Cloud: From Homomorphic Encryption to Federated Analytics. *arXiv preprint arXiv:2601.06710*. 2026 Jan 10.
- [65]. Aslam MS, Jamal M. Machine Learning in Healthcare: Revolutionizing Clinical Decision-Making with Data Analytics. *Global Trends in Science and Technology*. 2026 Mar 18;2(2):23-42.
- [66]. Hartmann SL. Federated Learning and Privacy-Preserving Machine Learning for Secure Distributed Data Analytics. *Main Issues of Pedagogy and Psychology*. 2026 Aug 22;4(2):956-67.
- [67]. Kabeer MM. AI in manufacturing quality management: a review of techniques, tools, and industrial adoption. *Global Journal of Universal Studies*. 2024;1(2):46-64.
- [68]. Abbasi N, Nizamullah FN, Zeb S, Fahad M, Qayyum MU. Machine learning models for predicting susceptibility to infectious diseases based on microbiome profiles. *Journal of Knowledge Learning and Science Technology* ISSN: 2959-6386 (online). 2024 Aug 25;3(4):35-47.
- [69]. Oad SK. From Algorithms to Insights: A Review of AI Applications in Data Science. *American Journal of Artificial Intelligence and Computing*. 2026 May 14;2(1):236-51.
- [70]. Singh A. Human-computer interaction: a review of usability, design, and accessibility trends. *Global Research Repo*. 2025 Sep 9;1(2):362-87.
- [71]. Odera D. Federated learning and differential privacy in clinical health: Extensive survey. *World Journal of Advanced Engineering Technology and Sciences*. 2023 Apr;8(2):305-29.
- [72]. Zeb S, Nizamullah FN, Abbasi N, Fahad M. AI in healthcare: revolutionizing diagnosis and therapy. *International Journal of Multidisciplinary Sciences and Arts*. 2024 Aug 17;3(3):118-28.
- [73]. Sen A, Heng SH, Tan SC. A comprehensive Review of Cryptographic Techniques in Federated Learning for Secure Data Sharing and Applications. *IEEE Access*. 2025 Jul 30.
- [74]. Jahin GT, Maliha F, Hassan FB, Muztaba AM, Patwary MJ, Arefin MS. Privacy-Preserving and Explainable Multi-Hospital CKD Prediction via Federated Learning with Blockchain



- Logging and Differential Privacy. In 2025 IEEE 2nd International Conference on Computing, Applications and Systems (COMPAS) 2025 Oct 23 (pp. 1-6). IEEE.
- [75]. Jamal M. Deep Learning Applications in Medicine: Enhancing Diagnosis and Patient Outcomes. Global Journal of Universal Studies. 2024 Jun 15;1(1).
- [76]. Zeb S, Nizamullah FN, Abbasi N, Qayyum MU. Transforming Healthcare: Artificial Intelligence's Place in Contemporary Medicine. BULLET: Jurnal Multidisiplin Ilmu. 2024;3(4):592385.
- [77]. Raza H, Erdenetsogt T, Kabeer MM, Aslam MS. IoT Generated Big Data Analytics: Trends and Scalability Issues. American Journal of Artificial Intelligence and Computing. 2025;1(2):306-27.
- [78]. Shaikh MM. Enhancing Workforce Efficiency in Public Sector Supply Chains through Predictive Analytics. American Journal of Artificial Intelligence and Computing. 2026 Jun 10;2(1):288-315.
- [79]. Vijendar A, Jain A, Janagam S, Deepthi P, Mannava G, Shankar S. Privacy-Preserving Data Sharing Models Using Homomorphic Encryption and Differential Privacy. In 2026 International Conference on Multidisciplinary Innovations For Smart & Sustainable Future (MISSF) 2026 Apr 17 (pp. 1-6). IEEE.

