Journal of Computing Theories and Applications ISSN:3024-9104

Review Article

A Scoping Literature Review of Artificial Intelligence in Epidemiology: Uses, Applications, Challenges and Future Trends

Kamal Bakari Jillahi [1,*] and Aamo Iorliam [1,2]

[1] Department of Computer Science, American University of Nigeria; e-mail : kamal.bakari@aun.edu.ng
[2] Department of Data Science & Analytics, American University of Nigeria; e-mail : aamo.iorliam@aun.edu.ng
* Corresponding Author : Kamal Bakari Jillahi

Abstract: Artificial Intelligence (AI) has been applied to many human endeavors, and epidemiology is no exception. The AI community has recently seen a renewed interest in applying AI methods and approaches to epidemiological problems. However, a number of challenges are impeding the growth of the field. This work reviews the uses and applications of AI in epidemiology from 1994 to 2023. The following themes were uncovered: epidemic outbreak tracking and surveillance, Geo-location and visualization of epidemics data, Tele-Health, vaccine resistance and hesitancy sentiment analysis, diagnosis, predicting and monitoring recovery and mortality, and decision support systems. Disease detection received the most interest during the time under review. Furthermore, the following AI approaches were found to be used in epidemiology: prediction, geographic information systems (GIS), knowledge representation, analytics, sentiment analysis, contagion analysis, warning systems, and classification. Finally, the work makes the following findings: the absence of benchmark datasets for epidemiological purposes, the need to develop ethical guidelines to regulate the development of AI for epidemiology as this is a major issue impeding it's growth, a concerted and continuous collaboration between AI and Epidemiology experts to grow the field, the need to develop explainable and privacy retaining AI methods for more secured and human understandable AI solutions.

Keywords: AI; Epidemiology; Infodemiology; Infoveillance; Machine Learning.

Received: March, 8th 2024
Revised: April, 4th 2024
Accepted: April, 9th 2024
Published: April, 14th 2024

1. Introduction

Epidemiology is the study and analysis of the prevalence, patterns, and causes of disease conditions within a specific population, while artificial intelligence (AI) is an area of science and technology interested in the design, implementation, and use of intelligent computers to carry out tasks that would normally require human intelligence[1]. As a significant technical development, AI has allowed people to work with computers in a diverse range of industries that previously required higher mental and intellectual capacities [2], [3]. AI uses relevant information sources like websites, media posts, mobile phone network data, global positioning system data, big data, and the like to infer seemingly unapparent trends that the human eye may not notice, achieving superior performance for specific tasks [4]. In order to support policy and decision-making, AI, at its most basic level, performs a significant level of comparative analysis using digital data, allowing public health and epidemiological incidence information to be used to infer prevalence, trends, and nature of occurrences to help mitigate escalation of incidences [5], [6]. Recently, there has been some stagnation in the area of AI application in Epidemiology due to some challenges. Thus, this study presents a review to unravel the challenges and forecast the future in tracking, monitoring, predicting, and ameliorating occurrences of epidemiological incidences from local to country to regional and global levels. To the best of our knowledge, there is no review on the subject matter that covers the time span under our review. The remaining parts of the paper are structured as follows: after presenting the research methodology that outlines the paper selection process for the review, an overview of the current use of AI in epidemiology is given, followed by lessons learned from these applications, then expected future trends based on empirical evidence is presented, subsequently, conclusions drawn from the review is rendered. Finally, a summary of the issues highlighted is presented, and concluding remarks are made.

2. Methodology

The terms "artificial intelligence," "machine learning," "deep learning," "epidemiology," "infodemiology," "infoveillance," and "infectious diseases" were used on Google Scholar search to find research publications published between 1994 and 2023 in order to collect literature for this review. Out of the 16,123 publications returned, a total of 11,212 were removed from the initial result for not applying AI or machine approaches or not related to epidemiology. Of the 4,911 remaining results, a total of 4,313 were further eliminated after abstract review for not using AI approaches or not being concerned with epidemiological problems. The results were filtered by reviewing their titles for relevance in AI applications for epidemiology. Consequently, only 143 met the inclusion criteria for the review. The following themes were found for AI use in Epidemiology: Tracking Outbreak and Epidemic Analysis, Infodemiology and Infoveillance, Geo-location data and visualization: GIS, Tele-Health Network, Vaccine Resistance and Hesitancy Sentiment Analysis, Quick diagnosis, Treatment Plans Analysis and Comparison, Predicting and Monitoring Recovery and Mortality, Decision Support System in Healthcare, and Knowledge Representation. While these purposes were found for AI use in epidemiology: Prediction, GIS, Trends, Modelling, Decision Support System, Knowledge Representation, Diagnosis, Analytics, Review, Tele-Health, Detection, Ethics, Sentiment Analysis, Contagion Analysis, Warning System, and Classification. The following Data Sources were found to be used in the publications: Web Access Logs, Web Search, Social Media Feeds, Online Benchmark Dataset, Research Repository, Online Data Repository, Mobile Phone Network Data, Wearable Devices, and Primary Data. The studies used the following algorithms: correlation analyses, regression analyses, machine learning, artificial neural networks, statistical learning, natural language processing, decision trees, random forests, gradient boosting trees, logistic regression, support vector machines, discriminant analysis, declarative programming, K-Nearest neighbor and descriptive analyses.

3. Surveillance, Outbreak Tracking and Epidemic Analysis

In 1854, John Snow used cartographic maps of London to show how cholera outbreaks, prevalence, and deaths were closely related to contaminated water sources in the Soho area of the city. This helped raise public awareness and called on public policymakers to recognize the significance of urban water systems sanitation, and it inspired action to prevent future outbreaks. This is thought to be the earliest use of computational methods in epidemiology and effectively demonstrated the role that computing might play in epidemiological research. Consequently, computational methods, especially AI approaches, have witnessed an incremental use in epidemiology over time. This study presents some of these uses from 1994 to 2023. This is not intended as an exhaustive mention but an indication of empirical evidence of AI usage in these areas.

3.1. Early Warning System

3.1.1. HIV/AIDS Pandemic from 1994

Using geographic information systems (GIS) and HIV/AIDS prevalence data, [7] predicted high-risk populations, hot-bed locations, and an individual's infection status in relation to their socio-temporal data and living location. Furthermore, using predictive models, [8] examined the high-risk populations and medical needs related to HIV/AIDS in the US and Canada, respectively. Studies [8], [9] discovered that low service levels and an uneven distribution of medical resources were prevalent in the study areas and that less proportionately represented populations in both countries had trouble accessing services related to HIV/AIDS.

3.1.2. Malaria Pandemic from 1994

A malaria computer detection system dating back to 1992 was developed in Israel. In minor outbreaks, it can automatically assess the sickness's etiology and determine the transmission risk [10]. This system used intelligent algorithms to anticipate the occurrences of malaria. The study also employed predictive algorithms to conclude that malaria epidemics were more likely to develop in places with a relative humidity of 60% and above [11], which was quite near to the realized rates.

3.1.3. Cholera Pandemic from 1994

Using a matched case method and GIS data, [12] analyzed the mode of transmission, distribution of cases, and risk factors underlining cholera epidemics in Lusaka. The findings revealed a significant correlation between the high incidence of cholera and the absence of a latrine and drainage system surrounding houses. They suggested that hand washing with soap and chlorinating drinking water could help prevent the epidemic [12].

3.1.4. The Influenza a H1N1 Pandemic of 2009

In order to determine actual disease prevalence and track the rapidly evolving of public concerns about H1N1 or swine flu, [13] used Twitter-embedded data and a number of AI approaches. The researchers investigated public concerns through the collection of tweets pertaining to H1N1 activity using pre-specified search phrases with other keywords linked to spread of disease, illness countermeasures, and food consumption in the United States. Using the supervised learning technique, they used surveillance data on influenza-like illnesses to create an estimating model. The findings demonstrated the significance of using Twitter to measure public interest in or worry over H1N1-related health-related incidents. These consist of periodic upsurges in user activity on Twitter that have been related to travel and food consumption patterns, hand cleanliness and mask use as preventive measures, and drug-related tweets regarding the use of particular antivirals and vaccines. They concluded that Twitter data can be utilized with AI algorithms to predict influenza outbreaks accurately[13].

3.1.5. Dengue Outbreaks of 2009

Across South-East Asia, dengue is extremely endemic, thus, using spatiotemporal perspectives, [14] carried out an infodemiology and infoveillance study to investigate the links between dengue fever in Manila and weekly Google Trends (GT) analytics about the disease from the Internet between 2009 and 2014. They further used demographic search queries connected to dengue to determine how people looked for health-related services. Based on their findings, dengue incidence reports and weekly temporal GT patterns were almost identical. The terms "dengue," "symptoms and signs of dengue," "treatment and prevention of dengue," "mosquito," and "other diseases" are among the themes that were found when searching for information about the condition. The study concluded that GT is a helpful tool in addition to traditional disease surveillance techniques.

3.1.6. Ebola Outbreak of 2014

The majority of West African nations had an Ebola outbreak in 2014, which raised worldwide concern about and impending global pandemic and unified preparedness and prevention measures. Consequently, millions of search records on Google Trends about Ebola were used by [15] to investigate outbreak at the global level and across countries where primary disease cases were reported. The study further examined the correlation between the number of new cases and the weekly overall search web hits associated with the disease. The West African nations most impacted by the Ebola outbreak had the largest search traffic, producing inquiries about the disease. Furthermore, their capital cities had the most concentrations of this traffic[15]. Nonetheless, the dispersion of web searches across national borders stayed constant in Western nations. Among the Ebola-affected African nations, there were varying degrees of correlation between the weekly Google Trends index and the total number of new disease cases. A comparatively strong correlation was found between the Google Trends index and the total number of Ebola cases reported worldwide [15]. The researchers posited that the data from Google Trends and global epidemiology data showed a substantial correlation. Additionally, they concluded that international organizations may use such data to precisely detect epidemics and quickly develop suitable, implementable strategies for disease prevention[15].

3.1.7. The Zika Pandemic of 2015

A vast majority of tropical populations are endemic to common diseases spread by mosquitoes. The greatest documented outbreak of Zika occurred in 2015 and 2016, when it reached a "red-alert" warning level, indicating numerous complications necessitating international public health actions. In such cases, the population's health-seeking behaviors are indispensable pointers for any efficient potential control methods. A study [16] concluded that internet user activity data has proven to be useful in tracking people's behavior and health-seeking behaviors, particularly during disease epidemics. To determine and characterize self-reporting of a person's behavioral change during the spread of a disease, [16] analyzed user tweets between 2015 and 2016. The study examined 1567 Twitter users and discovered differences in their individual traits, social network qualities, and linguistic preferences. As a result of the Zika virus, these users altered or considered altering their trip plans. This research suggests that using AI principles could help us understand how the general public views and responds to the dangers of an infectious disease outbreak.

3.1.8. Chikungunya Outbreaks of 2017

The 2017 Chikungunya outbreak in Italy raised significant public health concerns. In order to better understand peoples' health-seeking behaviors, [17] first looked into potential correlations between epidemiological data and internet traffic. Furthermore, a structural equation model representing public responses to Chikungunya was generated using data from Google Trends, Google News, Twitter traffic, PubMed publications, and Wikipedia visits and revisions. The work concluded that Chikungunya-related web searches acted as a mediator in the interactions between autochthonous cases, tweet productions, and overall cases. However, in the allochthonous case model, web searches posed major mediating tweets rather than epidemiological numbers, considerably mediating tweet productions. They came to the conclusion that new technology for gathering public concerns, raising awareness, and avoiding false information could help health authorities become instantly aware of such outbreaks [17].

3.1.9. The COVID-19 Pandemic of 2019

The SARS-CoV-2 virus strain that emerged in 2019 brought the entire world to a standstill. This witnessed the widespread use of AI techniques for event detection, monitoring, and prediction at local and global scales. These studies cover a wide range of topics, from trend monitoring and prediction to rapid diagnostics. Furthermore, most of these studies focused on DNA analysis, modeling, prediction, diagnosis, and classification. The disease has been diagnosed using machine learning techniques like KNN, ANN, and Naive Bayes algorithms [18]. Research [19]used data from 151 published articles, using NLP and other ML techniques to create a COVID-19 diagnosis model based on patient symptoms and standard test methods. Consequently, the learning algorithm found that COVID-19 patients can be categorized into subtypes according to their symptoms, gender, and serum immune cell levels. The goal in [20] is to improve clinical diagnosis accuracy using an AI approach with data from patients undergoing hemodialysis and forecast the likelihood that the patient has undiagnosed COVID-19. The study built an ML model that can predict HD patients with undiagnosed COVID-19 with an accuracy rate of 95%. In a related work, [21] presented an AI system that diagnoses COVID-19 cases based on CT scans, clinical symptoms, exposure history, and laboratory testing. For CRISPER-based nucleic acid detection, assay designs, and experimental resources are recommended in [22], which can be employed for ongoing monitoring. Using machine learning methods, the work proposed a model for identifying 67 viral species and SARS-CoV-2 subtype. Study [23] uses machine learning and now casting to analyze the underdiagnosis of COVID-19 in Brazil. A machine learning technique is utilized to identify cases that have not yet been diagnosed to produce a new cast. Four machine learning techniques—logistic regression, support vector machines, decision trees, and random forests—are applied in [24]to process patient data and identify COVID-19 cases. The work presents an AI-based cloud-based method that is built as a mobile app that tracks cough patterns to detect COVID-19 instances. In [25], Generative Adversarial Networks (GANs), Extreme Learning Machines (ELM), and Long/Short Term Memory (LSTM) were used to diagnose COVID-19 where these methods assembled data from both structured and unstructured sources to monitor patients' situation and offer augmented curation. Research [26] proposed a framework -the COVID Deep framework, which integrates wearable medical sensors with a DNN to enable widespread virus testing. In [27], blood sample information is utilized with a machine learning system to diagnose the disease. A machine learning-based approach is presented in [28] to identify COVID-19 suspected cases by analyzing blood data as input. A machine learning system identifies the condition using hematochemical values from routine blood examinations, specifically white blood cell counts, platelets, CRP, GGT, ALT, ALP, AST, and LDH plasma levels are used as features. Results show that the proposed method achieves good accuracy compared to other methods. In [29], ANN and machine learning are used in conjunction with a basic statistical test to identify COVID-19 patients based solely on complete blood counts without the need for information about the patients' medical histories or symptoms. A machine learning algorithm to conduct tests based on blood tests is proposed in [30]. According to [31], who examined data from test results of 81 positive and 7,775 negative cases, COVID-19 patients typically had lower platelet counts and higher plasma fibrinogen levels, and about 25% had outright thrombocytopenia. The data were loaded into an extraction system driven by a neural network for analysis.

3.1.10. Monkeypox outbreak of 2022

The virus that causes monkeypox was first identified and documented in 1958 by the eminent Danish virologist von Magnus, who was looking into an outbreak that was impacting monk colonies in Denmark. In May 2022, an outbreak of the disease was reported, with more than 2,100 cases registered globally as of June 16, 2022. Research [32] studied the "possible" transmission dynamics of the virus using an ML approach. It concluded that "the sexual transmission route, which, although not confirmed yet, seems highly likely in the diffusion of the infectious agent". More precisely, the study finalized that "future studies should be able to offer a compelling rationale for some of the unresolved issues that remain, such as the measurement of the monkeypox virus load, its viability and shedding in human male semen, its stability and persistence, and the concentrations at which it can be sexually transmitted"[32].

3.2. Contact tracing and Contagion Network Analysis

A public health strategy in the fight against infectious diseases is contact tracing, predicated on the idea that disease is spread through direct human contact. Healthcare professionals try to break the chain of transmission by identifying other potential patients who may have been exposed to the disease and tracking them down based on their contact history. This allows for monitoring and, if needed, treatment [33]. Contact tracing works best when there are few infected cases or a low reproductive ratio of the disease because it involves a lot of manual labor to interview patients and gather their contact data [33]. Given that mobile phone telecommunication traces offer reliable information about physical human interaction, they are used to model a contagion network using mobile sensed interaction data [34]–[36]. The studies investigate how communication datasets might function as a practical indicator for interactions that take place and how they can be used in modeling epidemiological interactions.

The impact of community structure in illness dynamics was investigated using a large-scale socio-technological network based on Facebook data [37]. Furthermore, a community's fine-grained face-to-face contacts could be mapped utilizing close proximity interactions (CPIs) recorded using wireless sensors [38]. Potentially more effective than random immunization, new preventive techniques were devised by reconstructing the contact network and evaluating the CPIs in a community. Using RFID tags to sense CPIs over two days at a conference, the dynamics of infectious diseases have been mimicked, emphasizing the temporal and heterogeneity aspects[39]. Artificial Intelligence (AI) methods such as NLP and ML have also been applied to enhance surveillance and virus tracking. In [34] , a system that uses artificial intelligence (AI) to evaluate anonymized smartphone data and find high-risk connections of COVID-19 patients for focused tracing was presented. In the same vein, several studies have shown that artificial intelligence (AI) methodologies can be used to track and predict seasonal flu trends in the context of influenza surveillance. These methodologies involve mining online data sources, including social media activity, online user trails, search engine queries, and digital health records [40], [41].

3.3. Diagnosis

Diagnosing an illness is crucial because many policy makers, such as the World Health Organization, believe that testing is essential to successfully managing an epidemic since it yields significant information about local outbreaks that may be contained before they spread. Research [42] offers a machine learning system that records patient data, including coughs, and uses that to train a classifier for disease diagnosis. In [43], a binary ML classifier is proposed to diagnose COVID-19 using only cough and breathing patterns data. The work claimed to have achieved an accuracy comparable to other lab-based scientific procedures.

An online survey is created in [44] to gather information on COVID-19 patients. Next, to forecast possible COVID-19 patients based on their signs and symptoms, the data were fed into several ML prediction algorithms, such as Support Vector Machine, Logistic Regression, and MLP. Results indicate an astonishing accuracy of the proposed method. Textual clinical data are gathered and feature extraction procedures such as Bag of words (BOW), Term frequency/inverse document frequency (TF/IDF), and report length are employed to gather data in [45]. Afterwards, Multinomial Naive Bayes and Logistic Regression were applied to classify the data, a high correlation was found between the expert health practitioner's actual prognosis and the predicted diagnosis. Study [46] provides a diagnosis method based on chest CT images. This approach involves taking radionic features from the region of interest and feeding them into an artificial intelligence segmentation algorithm. The algorithm also receives information for classification in the form of clinical symptoms, epidemiology history, and biological symptoms. Results from this work indicated a promising application of AI in disease diagnosis and management. Other examples of using AI and ML techniques and approaches for diagnosis can be found in [47], [48].

3.4. Infoveillance and Infodemiology for Tracking Human Health Seeking Behaviors

Public health informatics' incorporation of internet data has been a potent tool for studying how people seek medical attention in real time during epidemic outbreaks. As mentioned earlier, Google Trends is a well-liked and often-used tool that uses user-specified keywords and themes to deliver user activities, which are pointers about trends, patterns, and variations of online interests over time [49]. Such adaptations created two nomenclatures: the first been “Infodemiology,” defined as “The distribution and determinants of information in a social group or an electronic medium (here, the Internet) with the ultimate purpose of guiding public health and policy,” while the second “Infoveillance,” defined as “the continuous monitoring of infodemiology metrics for trend analysis and surveillance” [50].

3.5. Predicting and Monitoring Spread, Recovery and Mortality

Numerous research studies have shown how useful artificial intelligence (AI) may be in creating predictive models to predict epidemics. A significant benefit in healthcare is the capacity to forecast and track patient outcomes, such as recovery and fatality rates. This is largely made possible by artificial intelligence (AI) and data-driven techniques, which offer medical professionals critical insights for patient care and treatment modifications. For instance, [51] found that their method could precisely reconstruct the early dynamics and predict future projections of an epidemic. They combined deep neural networks and epidemiological modeling to predict the spread of COVID-19 in China. In a similar vein, [52] created an AI model that combined human mobility data with case data to accurately predict the spread of COVID-19 outside of China 1-2 weeks ahead of time. Other studies have also used neural networks to forecast dengue fever epidemics in China [41] and Zika virus outbreaks in Colombia [53]. AI has demonstrated encouraging potential for utilizing various data sources to produce useful pandemic forecasts. [54] provides an example of how deep learning algorithms might be used to monitor patients continually so that medical personnel can closely follow their progress in real time. Furthermore, by identifying patients with high mortality probability, these technologies enable early interventions and allow healthcare professionals to allocate resources effectively and provide proactive care to those who need it most. What's especially noteworthy is that these predictions are dynamic and can adapt to changing patient conditions.

3.6. Geo-location data and visualization: GIS, Mapping

Knowledge about infectious disease spatial patterns can help identify their causes and preventative measures. Analysis of the geographic patterns of diseases and the inter-play between pathogenic components (zoonotic agents, vectors, infected persons, and system dynamics) in relation to their geographic settings is becoming increasingly common with the use of GIS, AI, and associated technologies. The spatial patterns of these components can be seen and analyzed over time with the aid of basic AI methods and analytical GIS applications in epidemiology. This can reveal spatiotemporal patterns, trends, and connections that would be harder or impossible to discover in other ways. Understanding the spatial spread and dynamics of an outbreak is essential to the design of prevention and control strategies. GIS has proven to be an indispensable tool that assists in meeting these demands. As a result, the use of GIS in public health information management for disease control has grown in importance over the years [55], [56]. The most common use of GIS in epidemic research is the creation of maps that show the locations of cases, probable risk factors, and sources [55]. These maps can be used to explain the disease's spreading trends and pinpoint outbreak outliers. Research [57] used GIS techniques to examine the prevalence of Lyme in Baltimore, Maryland, in 1990. They also conducted logistic regression analysis, compared the epidemic's development to a random setting, created a risk prediction model, and estimated the epidemic's range. Study [58] investigated how the geographic environment affected the prevalence of encephalitis in Nepal, discovered a substantial correlation between the percentage of irrigated area and low precipitation, and profiled the spatial pattern of cases during the 2005 epidemic. Using GIS analysis, [59] investigated the causes of multiple myeloma and acute myeloid leukemia. They discovered that individuals who have lived near six emission points adjacent to significant amounts of waste from petroleum refining for more than ten years may be at a higher risk of developing these diseases. Using GIS, the degree of exposure to dioxin concentrations was determined based on the case and control's addressed in [60]. An investigation of the link between dioxin release and soft tissue tumor revealed the correlation between a patient's address and their vulnerability. Study [61] employed GIS and remote sensing to examine the environmental factors influencing dengue and chikungunya vector breeding habitats. She discovered that environment, climate, and topography all significantly impact the spread of disease throughout a nation, making the use of GIS and remote sensing tools invaluable for managing and controlling epidemics. Using geographic information system (GIS) technology, [19] investigated the spatial distribution pattern of tuberculosis incidence and found a strong correlation between the disease's high incidence and regional geographic data. These findings served as a scientific foundation for developing the tuberculosis control policy.

3.7. Vaccine Resistance and Hesitancy Sentiment Analysis

Doctors and medical experts are starting to worry about vaccination-related reluctance and hesitancy. It is known as vaccine hesitancy when people have second thoughts about vaccinating their children and themselves. Research [62]emphasized that there was enough data to support medical professionals' and academics' efforts to address people's reluctance to use social media posts. This is because of social media's promotion of information sharing among users. Additionally, users are free to express and publish their thoughts thus, information is made freely available. However, it is hard to keep up with the amount of content to monitor, given the speed at which blogs and postings are growing. Automated AI-driven sentiment analysis will be used to inform public health policy about the general public's issues, concerns, and fears regarding vaccines and medications. Given sentiment analysis's first use and the emergence of Web 2.0, it is crucial to emphasize that sentiment analysis could be used with social media data to understand and forestall sentiments against drugs and vaccines in general. Using Twitter data, [63] conducted sentiment analysis on vaccination using the recently developed COVID-19 vaccine and discovered that public opinion is divided between pro- and anti-vaccine. In the same vein, [64] proposed using Facebook to map the anti-vaccination attitude landscape. By observing 197 user accounts, they could visualize the network of Facebook profiles that discussed vaccines and gain insight into the interactions between pro- and anti-vaccine individuals. They were unable to obtain data from user's posts and location information. However, the ethical protocols that must be adhered to when extracting data from social media platforms are a source of worry, as pointed out by [65]. Using a Natural Language Processing program called clean NLP and data from YouTube videos, [66] studied pro- and anti-vaccine sentiment expression online. Its likeability in the immunization advocacy category and word frequency rating detected 275 videos and could discern the growing tendencies among anti-vaccine users. With an emphasis on Twitter users, [67] created two methods for categorizing views on vaccination to ascertain which approach is the most effective for obtaining opinions through data in the future. The study used 95,566 tweets and employed Multinomial Naïve Bayes and Support Vector Machine algorithm to analyze the tweets and concluded that All the used approaches are equally good for use in gathering opinions about vaccine sentiment due to their comparable performances. A study [68] stated that understanding the opinions of both pro- and anti-vaccine Twitter users is crucial to understanding the public's stance toward vaccination. They employed a linear SVM method to create a more structured representation by determining the communicative patterns and mutual influences between the two user groups using 669,136 tweets. They discovered that the number of Twitter users against vaccinations is increasing within a closed network, making it challenging for health institutions to obtain pertinent information. Nevertheless, they claimed that emoticons, which are also crucial for recognizing sarcasm in Tweets, were overlooked during the study. However, emojis are indispensable because individuals use them all the time to convey their beliefs, particularly to indicate whether they support or oppose vaccinations[69]. Research [70] argued that a Twitter user's perspective could be shaped by a particular experience they had recently, especially on the subject of concern. They examined the likelihood that people would have opinions about certain events that occurred over that period using 112,397 tweets from September 1, 2016, to June 30, 2017. Their ten months of study revealed that vaccination-related events can shift the number of Twitter users interested in hearing about new vaccine developments. Every occurrence has influenced Twitter users to share their opinions, whether those opinions are against or favorable to vaccination.

3.8. Treatment Plans Analysis and Comparison

A crucial part of managing healthcare is evaluating and contrasting treatment options, particularly during a pandemic. AI methods—machine learning in particular—have completely changed this procedure. Using these techniques, medical personnel can evaluate and optimize treatment plans based on insights from data about patients' recovery and mortality rates based on comparisons. One good example of a study that goes into treatment plan analysis is [71]. In order to link patient symptoms, clinical information, and test results for COVID-19, machine-learning techniques were employed. This improves diagnosis and helps provide individualized treatment programs. Artificial intelligence (AI) has the potential to differentiate between distinct disease subtypes and enable more individualized treatment plans by spotting trends in patient data. The ability of AI to process enormous volumes of data quickly and accurately is especially remarkable. Patterns and correlations that may not be seen using traditional approaches can be found by analyzing a large number of patient data and medical literature, thanks to the processing capacity of AI algorithms. This ability to uncover hidden insights is a game-changer in healthcare. Maximizing medical resources is possible with this degree of accuracy and personalization in treatment programs as AI enables more personalized treatment plans, avoiding needless surgeries or therapies in scenarios such as pandemics where resources are scarce. This helps the healthcare system as a whole, in addition to helping the patients.

3.9. Decision Support System in Healthcare

Decision support systems (DSS) are very helpful in the healthcare industry, especially during pandemics. These systems give medical practitioners data-driven insights to help them make wise decisions. Underneath, DSS uses artificial intelligence (AI) and data analysis methods to provide therapy, monitoring, and patient care recommendations. Study [72] shows how AI-based DSS functions for COVID-19 patients. Its natural language processing algorithm offers basic healthcare education, information, and guidance. This DSS offers essential advice, healthcare suggestions, interactive counseling sessions, home remedies, and preventive actions. It allows people to control their health and their ailments more effectively. Additionally, [73] presented an AI-based method for evaluating the risk of contracting COVID-19 infection during virtual visits. Using natural language processing, this DSS analyzes data gathered from telehealth encounters to assist medical practitioners in determining the risk of infection and developing appropriate treatment regimens. The ability of AI-driven DSS to adjust to unique patient characteristics is awe-inspiring. Personalized suggestions can be provided based on a patient's medical history, current state of health, and even personal preferences. A higher degree of personalization in healthcare advice can improve treatment and preventative measure compliance while also increasing patient engagement.

3.10. Tele-Health Network

Artificial tele-health systems are incredibly helpful during the pandemic because they enable individuals to obtain the necessary services from the comfort of their own homes, thereby preventing the spread of an epidemy. Research [73]proposed a unique AI-based method for assessing the risk of COVID-19 infection during virtual visits. The algorithm uses a natural language processing method to infer possibilities using data gathered from telehealth visits. A natural language processing system was suggested in [74] to give COVID-19 patients free initial healthcare education, information, and advice. The system offers users interactive counseling sessions, home cures, preventive measures, and health advice.

3.11. Knowledge Representation

Knowledge representation is essential to applying artificial intelligence (AI) in epidemiology to convert enormous volumes of data into insightful understandings. Research on efficient ways to use AI approaches to describe epidemiological knowledge has increased dramatically in the last several years [75]. According to [76], structured knowledge representation organizes data into clearly defined formats that simplify processing and interpretation. Organized representations—like taxonomies and ontologies—have been widely used in epidemiology to classify risk factors, diseases, and population demographics. Thanks to these structured models, AI systems can understand and analyze intricate epidemiological interactions [77].

3.12. Semantic Web Technologies

Epidemiological research has used semantic web technologies, like Resource Description Framework and Web Ontology Language, to generate linked representations of entities, truths, and verified knowledge in the epidemiological domain[78]. RDF makes connecting different data sources easier, making it possible to integrate a range of epidemiological data[79]. To ensure consistency and interoperability in knowledge representation, OWL also permits the construction of ontologies that might be used to verify and validate common understanding agreed by domain experts [80].

3.13. Machine Learning-based Knowledge Representation

According to [81], there is potential for machine learning approaches, especially deep learning models, to identify complex patterns in epidemiological data effectively. Using representational learning algorithms, meaningful clusters and latent associations have been extracted from epidemiological datasets through word embedding and graph-based neural networks [82]. By improving the way epidemiological knowledge is represented, these techniques lead to more precise forecasts and insights.

Table 1: A Summary of Papers Reviewed and their Areas of application in Epidemiology

S/N	Area	Publications
1.	**Detection and Surveillance**	
	Surveillance and outbreak alarm:	[7], [15], [16]
	Early Warning System	[7], [8], [17], [9]–[16]
	Disease and Contact Tracking	[33]–[41]
2.	**Transmission and Diagnosis**	
	Early and alternative Diagnosis	[18]–[28], [30]–[32], [42]–[48]
	Spread, Mortality, and recovery Modelling	[41], [51]–[53], [83]
3.	**Human Behavior**	
	Infodemiology and Infoveillance	[14]–[17], [49], [50]
	Vaccine Resistance and Hesitancy	[62]–[70]
4.	**Visualization and Analytics**	
	GIS and Geo-spatial Analysis	[19], [55]–[61]
	Treatment Plan Analytics	[52]–[54], [71]
	Decision Support System	[72]–[74]
5.	**Knowledge Representation**	
	Semantic Web Tech	[75]–[82]
	ML Representation	[19], [84], [85]

Table 1 above shows that the area of epidemiology with the most reviewed papers is early and alternative diagnosis. This ranges from using cough to diagnose flu and other viral infections, using medical and personal images to diagnose diseases, to other applications of AI algorithms to provide self-diagnosis. Most research in this area achieves very good results due to the availability of AI algorithms that can infer un-seemingly and un-apparent relationships between entities. These algorithms may provide this diagnosis by relying on things that are not intuitive to humans or not in agreement with the common knowledge of epidemiologists [86]. This buttresses the need to develop a more explainable methodology for this epidemiological problem. Furthermore, other areas, such as ML knowledge representation and early outbreak system received minimal research interest, this might be associated with the lack of computational methods to address these problems from the epidemiological viewpoints or the availability of better tools and approaches for solving these problems in the epidemiological community. With promising results in the area of data visualization and analytics, we might witness a surge in the application of these computational methods in the field of epidemiology [87]. Other areas, such as decision support systems, may see further adoption and application in epidemiology due to their centrality in AI and ubiquity in epidemiology.

4. Lessons Learnt from the Review

Below is a summary of all included studies in this review:

Table 2. Summary of Papers included in Review

Ref	Author(s)	Year	Use	Data Source	Algorithm
[88]	Adiga et al.	2021	Modelling	Web Search Logs	Statistical Learning
[15]	Alicino et al.	2015	Analytics	Web Search Logs	Regression
[89]	Aljaaf et al.	2015	Decision Support System	Research Repository	Review
[90]	Antoniou et al.	2018	Classification	Primary Data	Regression
[91]	Ardabili et al.	2020	Prediction	Online Data Repository	Machine Learning
[92]	Avila et al.	2020	Decision Support	Research Repository	Statistical Learning
[93]	Baker & Taylor,	2016	Review	Research Repository	Review
[42]	Belkacem et al.	2021	Classification	Primary Data	Discriminant analysis
[74]	Bharti et al.	2020	Tele-Health	Primary Data	Natural Language Processing
[94]	Brinati et al.	2020	Detection	Primary Data	Decision Tree model
[95]	Brown and Lee	2021	Ethics	Research Repository	Review
[43]	Brown et al.	2020	Diagnosis	Social Media Feeds	machine learning
[11]	Ceccato et al.	2011	Contagion Analysis	Primary Data	Statistical Learning
[5]	Charan et al.	2018	Detection	Primary Data	Artificial Neural Network
[96]	Chen	2018	Decision Support	Online Benchmark Dataset	declarative programming
[97]	Borlase et al.	2017	Review	Research Repository	Natural Language Processing
[46]	Chen et al.	2020	Detection	Primary Data	Machine Learning
[98]	Chen et al.	2020	Knowledge Representation	Research Repository	Ontology and Natural Language Processing

Ref	Author(s)	Year	Use	Data Source	Algorithm
[36]	Cho et al.	2011	Prediction	Social Media Feed	Correlation
[99]	Comito et al.	2020	Review	Research Repository	Review
[35]	Crandall et al.	2010	Prediction	Social Media Feeds	Statistical Learning
[100]	Daltayanni et al.	2012	Trends	Web Search	Data Mining and Natural Language Processing
[70]	D'Andrea, et al.	2019	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[16]	Daughton and Paul	2019	Prediciton	Social Media Feeds	Machine Learning and Natural Language Processing
[101]	Deng et al.	2019	Tele-Health	Primary Data	Artificial Neural Network
[102]	Dirk et al.	2008	GIS	Primary Data	Statistical Learning
[47]	Dutta & Bandyopadhyay,	2020	Detection	Online Data Repository	Machine Learning
[103]	Dutta et al.	2019	Knowledge Representation	Online Data Repository	Natural Language Processing
[104]	Dwivedi et al.	2021	Review	Research Repository	Review
[18]	Elboujnouni	2022	Detection	Primary Data	Statistical Learning and Natural Language Processing
[105]	Eysenbach	2009	Warning System	Social Media Feeds	Natural Language Processing
[50]	Eysenbach	2011	Detection	Social Media Feeds	Natural Language Processing
[44]	Fayyoumi et al.	2020	Prediction	Primary Data	Statistical Learning and Machine Learning
[106]	Feng et al.	2022	Diagnosis	Primary Data	Regression
[48]	Ferrari et al.	2020	Review	Research Repository	Review
[9]	Fulcher and Kaukinen	2005	GIS	Online Data Repository	GIS and Statistical Learning
[107]	Garcia and Martinez	2022	Review	Research Repository	Machine Learning
[23]	Garcia et al.	2020	Diagnosis	Primary Data	Machine learning
[7]	Geanuracos et al.	2007	GIS	Primary Data	GIS
[57]	Glass et al.	1995	GIS	Primary Data	Regression
[55]	Gluskin et al.	2014	Trends	Web Search	Correlation
[108]	Gomes et al.	2020	Diagnosis	Primary Data	Machine Learning
[109]	Gomoi and Stoicu-Tivadar	2010	Detection	Online Data Repository	Artificial Neural Network
[82]	Gupta & Singh,	2019	Classification	Primary Data	Machine Learning
[110]	Gupta and Patel	2023	Review		Deep Learning
[26]	Hassantabar et al.	2021	Classification	Primary Data	Deep Learning

Ref	Author(s)	Year	Use	Data Source	Algorithm
[14]	Ho et al.	2018	Trends	Web Search	Natural Language Learning
[64]	Hoffman, Felter, & Chu	2019	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[111]	Hossen and Karmoker	2020	Prediction	Online Data Repository	Random Forest, Support Vector Machine and K-Nearest Neighbor
[65]	Hunter et al.	2018	Ethics	Research Repository	Review
[112]	Imran et al.	2020	Diagnosis	Primary Data	Machine Learning
[25]	Jamshidi et al.	2020	Diagnosis	Primary Data	Deep Learning
[113]	Johnson et al.	2022	Knowledge Representation	Online Data Repository	Natural Language Processing
[78]	Ferreira et al.	2013	Knowledge Representation	Research Repository	Natural Language Processing
[76]	Jones et al.	2003	Knowledge Representation	Research Repository	Review
[114]	Jones et al.	2013	Prediction	Social Media Feeds	Regression
[115]	Jumper et al.	2021	Prediction	Online Benchmark Dataset	Artificial Neural Network
[4]	Kaplan and Haenlein	2020	Review	Research Repository	Review
[116]	Kapoor et al.	2020	Prediction	Primary Data	Artificial Neural Network
[34]	Dosilovic et al.	2018	Review	Research Repository	Review
[117]	Kaur et al.	2021	Review	Research Repository	Review
[45]	Khanday et al.	2020	Diagnosis	Online Data Repository	ANN, Natural Language Processing
[118]	Davagdorj et al.	2021	Prediction	Research Repository	Machine Learning
[119]	Kim et al.	2023	Warning System	Research Repository	Review
[10]	Kitron et al.	1994	GIS	Primary Data	Regression
[27]	Kukar et al.	2020	Diagnosis	Primary Data	Machine Learning
[67]	Kunneman et al.	2019	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[120]	Kesmodel	2018	Review	Research Repository	Review
[77]	Liu et al.	2018	Analytics	Primary Data	Machine Learning
[19]	Maciel et al.	2010	GIS	Primary Data	Statistical Learning
[121]	Madanan et al.	2021	Decision Support	Primary Data	Artificial Neural Network, Bio-inspired Computing
[122]	Maduri and Krishnan	2020	Diagnosis	Primary Data	IoT
[17]	Mahroum et al.	2018	Review	Social Media Feeds	Correlation and Regression
[61]	Masimalai	2014	GIS	Primary Data	Correlation

Ref	Author(s)	Year	Use	Data Source	Algorithm
[49]	Mavragani and Ochoa	2019	Trends	Web Search	Review
[53]	McGough et al.	2017	Prediction	Social Media Feeds and Online Data Repository	Statistical Learning
[123]	McGregor	2021	Review	Research Repository	Review
[21]	Mei et al.	2020	Diagnosis	Primary Data	Regression
[22]	Metsky et al.	2020	Diagnosis	Online Benchmark Dataset	Machine Learning
[124]	Minz and Mahobiya	2017	Classification	Online Benchmark Dataset	Decision Tree
[20]	Monaghan et al.	2020	Classification	Online Benchmark Dataset	Machine Learning
[24]	Sun et al.	2020	Prediction	Online Data Repository	Machine Learning
[73]	Obeid et al.	2020	Warning System	Primary Data	Statistical Learning
[125]	Oyedepo et al.	2012	GIS	Primary Data	GIS
[84]	Pang and Lillian	2009	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[31]	Pawlowski et al.	2020	Diagnosis	Online Data Repository	Statistical Learning
[126]	Pee et al.	2019	Knowledge Representation	Research Repository	Review
[127]	Pearce et al.	2007	Review	Online Data Repository	Review
[128]	Peng et al.	2020	Review	Online Data Repository	Review
[40]	Pollett et al.	2020	Trends	Web Search Logs	Correlation
[60]	Poulstrup and Hansen	2004	GIS	Online Data Repository	Statistical learning
[129]	Ren et al.	2015	GIS	Primary Data	Correlation
[130]	Ribbens et al.	2014	Classification	Online Benchmark Dataset	Statistical Learning
[62]	Salmon et al.	2015	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[12]	Sasaki et al.	2008	GIS	Online Data Repository and Primary Data	Statistical Learning
[85]	Shaban-Nejad et al.	2021	Review	Research Repository	Review
[13]	Signorini et al.	2011	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[131]	Guo et al.	2021	Review	Research Repository	Review
[56]	Smith et al.	2015	Review	Research Repository	Review
[28]	Soares	2020	Detection	Online Benchmark Dataset	Machine Learning
[59]	Speer et al.	2002	GIS	Primary Data	Correlation
[132]	Sqalli and Al-Thani	2019	Tele-Health	Primary Data	Artificial Neural Network

Ref	Author(s)	Year	Use	Data Source	Algorithm
[133]	Strachna and Asan	2020	Decision Support	Primary Data	Machine Learning
[54]	Wagner et al.	2020	Diagnosis	Online Data Repository	Statistical Learning
[83]	Wahl et al.	2018	Review	Research Repository	Review
[79]	Rao et al.	2019	Knowledge Representation	Research Repository	Natural Language Processing
[134]	Giuste et al.	2023	Review	Research Repository	Review
[135]	Wang et al.	2020	Prediction	Primary Data	Artificial Neural Network
[80]	Wang et al.	2021	Review	Research Repository	Review
[136]	Wesolowski et al.	2012	Modelling	Primary Data	Modelling
[137]	Woo et al.	2021	Classification	Online Benchmark Dataset	Machine Learning
[138]	Wu et al.	2020	Detection	Online Data Repository	Machine Learning
[139]	Xie et al.	2020	Knowledge Representation	Primary Data	Survey
[140]	Xu et al.	2017	Prediction	Web Search Logs	Natural Language Processing
[66]	Yiannakoulias et al.	2019	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[141]	Yu and Zhou	2021	Tele-Health	Primary Data	IoT
[68]	Yuan and Crooks	2018	Sentiment Analysis	Social Media Feeds	Natural Language Processing
[142]	Zhao et al.	2024	Prediction	Primary Data	Machine Learning
[143]	Zhang et al.	2014	Review	Research Repository	Review
[144]	Boehm et al.	2022	Review	Research Repository	Review
[145]	Zhou et al.	2018	Classification	Primary Data	Artificial Neural Network
[87]	Zhou et al.	2021	Decision Support System	Primary Data	Natural Language Processing
[146]	Zoabi and Shomron	2020	Prediction	Primary Data	Machine Learning

4.1 Rate of Publication Growth

Figure 1 shows that AI application in Epidemiology has received little or less prior research interest in the early years considered by this review. However, substantial spikes in interest correspond to major epidemic out breaks. For example, the rise in interest between 2009 and 2012 can be associated with influenza H1N1 and Dengue outbreaks of 2009 and 2010, respectively. Furthermore, the spike between 2013 and 2016 can be associated with the Ebola and Zika epidemics of 2014 and 2015, respectively, while the major spike between 2018 and 2022 can be associated with the chikungunya and COVID-19 epidemics of 2018 and 2019, respectively, and the growth in 2020 can be attributed to the aftermath of COVID-19 pandemic. Thus, it is safe to conclude that AI application in epidemiology research tends to be reactionary to epidemic or pandemic outbreaks. Hence, there is a need for the research community to have concerted and continuous research in this area. This will ensure a steady growth of the discipline and prevent any outbreaks at their onset or early stage. Furthermore, this growth in AI application in epidemiology can also be associated with the growth and development of AI methods. Although AI has been around for a little longer, but its popularity and wide adoption have only risen at the onset of the 21st century. So, one safely concludes that as the AI community grows and matures, it finds more applications and adoption in other disciplines (epidemiology inclusive).

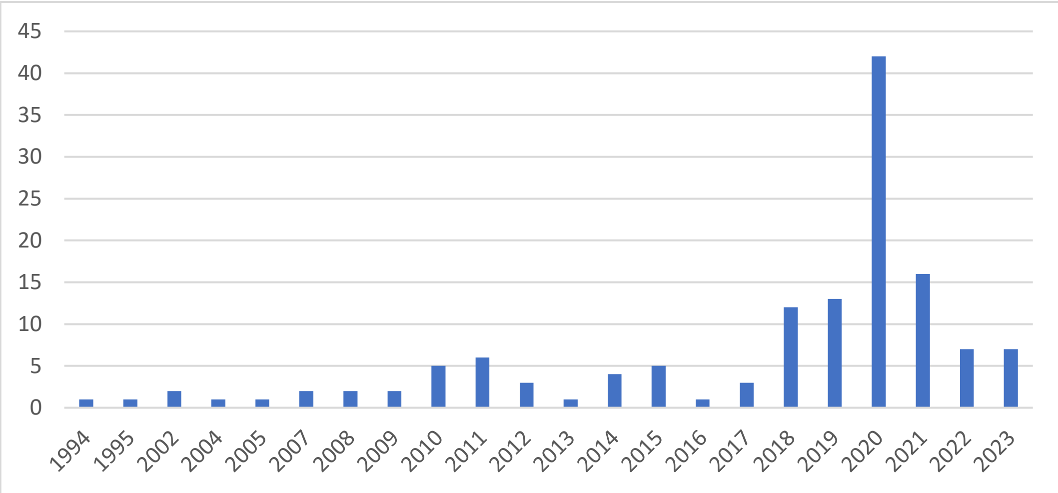

Figure 1: Number of Publications per Year From 1994 to 2023

4.2 Sources of Data

Form Figure 2, it can be seen that most research uses primary data sources followed by Online data repositories and social media feeds. This may be due to the absence or inadequate benchmark datasets exclusively captured and prepared for epidemiological purposes as most research uses open datasets, which are initially not meant for epidemiological purposes.

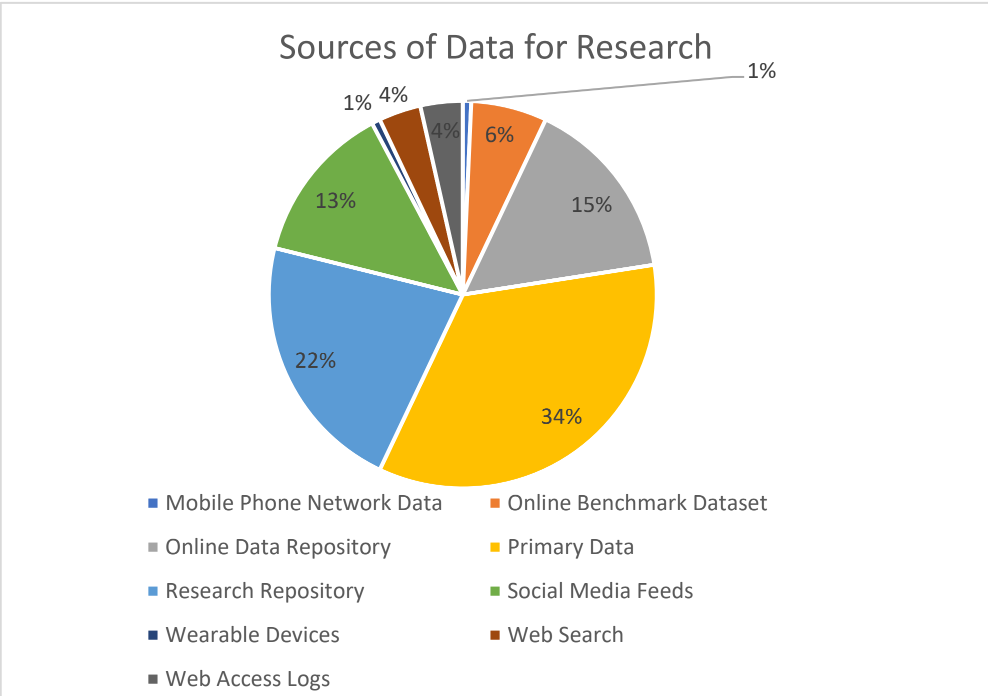

Figure 2. Sources of Data for AI Application in Epidemiology

Furthermore, any meaningful AI method heavily depends on the data quality used to make accurate inferences or predictions. Consequently, the availability and easy access to open data repositories exclusively captured and prepared for epidemiological purposes is crucial for the growth and development of the field. This calls for a major effort to provide digital datasets for different aspects of infodemiology, ranging from surveillance to outbreak to modelling to analytics and other endeavors to help researchers and practitioners build and apply AI concepts and methods toward solving epidemiological problems. It is equally important to mention that the quality of data used in any computational approach varies directly with the quality and volume of data used. Thus, the research community should not only make data available and accessible but should provide a benchmark for the quality of data provided by the community, as this will invariably affect the quality of solutions provided using these AI methods. Another vital area of data provision is epidemiological knowledge provision in the form of ontologies and RDFs that represent the domain knowledge, which will help AI experts mitigate the effects of black box AI models by building explainable models.

4.3. AI Methodology used for Epidemiology

From Figure 3, it can be seen that prediction (15%) – without considering review papers since they are not an area of AI - is the area with major research interest in AI application in epidemiology. This can be attributed to the availability of algorithms, tools, and resources for this task in the AI ecosystem. Moreover, other areas such as Contagion Analysis, Knowledge Representation, Ethics, and Warning systems, which are equally important, have attracted less research attention, which may be due to a lack of interest by the AI community or a lack of matured tools and methodologies for developing solutions in these areas. Thus, for the epidemiological community to have an overall healthy application of AI methods in solving epidemiological problems, there is a need for these equally important areas to receive adequate research attention.

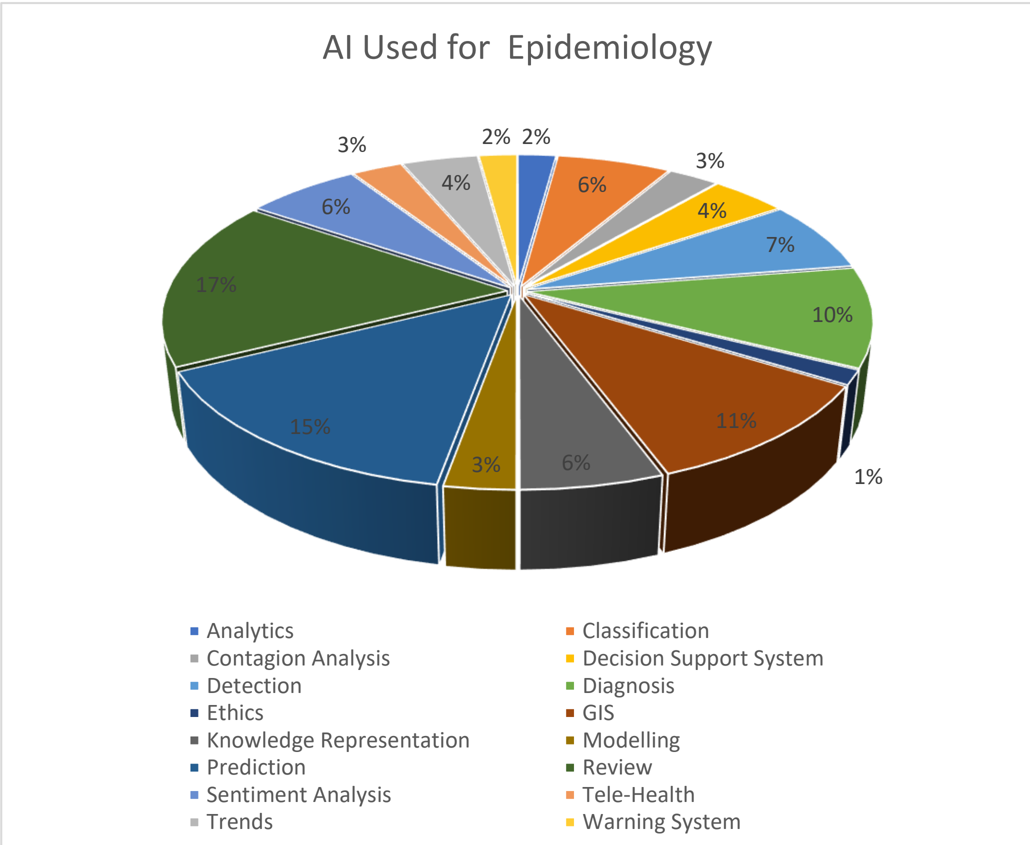

Figure 3. AI Approaches Applied in Epidemiology

In this regard, governments, financial donors, non-governmental organizations, and other stakeholders are advised to support research in these seemingly less important fields to have an overall growth of this field of endeavor. It is particularly worth mentioning that areas such as zoonotic models, vector dynamics, reservoir and cross-species pathogenic evolution, and social dynamics, which are the underpinnings of any epidemiological incidence, either have eluded AI experts in building and analyzing epidemiological models due to less or little involvement of epidemiological experts in conceptualizing these solutions or underrating the importance in including these crucial aspects. This calls for a holistic approach to building AI models targeted at solving epidemiological problems.

4.4. AI Algorithms applied in Epidemiology

Figure 4 shows that reviews are the most prevalent research from the period under review (27). Next are machine learning algorithms (19), followed by natural language processing (16); other algorithms, such as Deep learning, support vector machine, Ensemble methods, and declarative programming, received less substantial application in epidemiology. This may be associated with the fact that these fields are less developed or are currently developed even for mundane tasks; therefore, their application in other areas, especially epidemiology, where there is a need for high accuracy, will take some time. Thus, the future will see more applications of these methods, especially as they are ripe and available for adoption. In this vein, most literature has recommended developing a more explainable AI approach where methods applied in solving problems can be explained based on human reasoning [147]. Hence, it is expected that as these areas of AI grow, more applications will be witnessed in epidemiology.

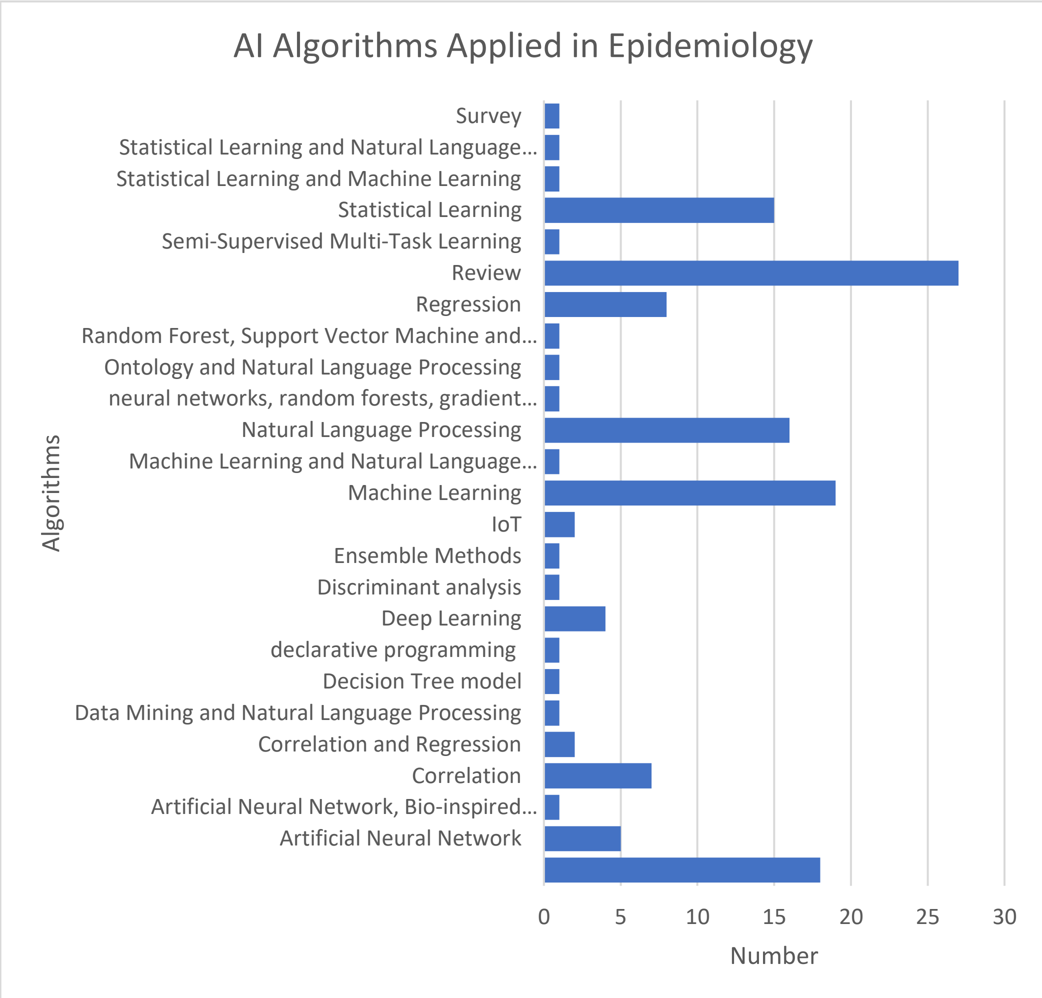

Figure 4. AI Methods Applied in Epidemiology

In the same vein, the AI ecosystem has recently seen some drastic evolution with the introduction of the right to explainability by the European Parliament, and this will have a bearing effect on the AI research community and, consequently, on the algorithms used in solving epidemiological problems. Thus, going to the future, more white box algorithms and explainable AI approaches have to be developed, adapted or extended for epidemiological uses as this will be a requirement for the feasibility of any AI solution that uses or stores people's data or has direct influence or how they or other people perceive them. In this regard, it is worth mentioning that the major algorithm –ANN, used for epidemiological purposes based on this review, is more or less a black-box algorithm. Therefore, these algorithms must either be adapted or replaced by more white-box algorithms.

5. Challenges and Future Directions in Artificial Intelligence in Epidemiology

5.1. Challenges in Implementing AI in Epidemiology

There are numerous obstacles to successfully applying artificial intelligence (AI) in epidemiology. A major obstacle is the availability and quality of data. It is challenging for AI systems to uncover significant patterns from epidemiological datasets since they are frequently heterogeneous, unstructured, and incomplete [131]. Furthermore, personal data privacy problems and ethical concerns arise, particularly when handling private medical records [95]. In order to fully utilize AI in epidemiology research, researchers must strike a careful balance between protecting data security and privacy on one hand and data availability on the other hand.

While artificial intelligence (AI) can offer suggestions on diagnosis, intervention, mitigation of health decline, and proactive measures to prevent conditions from getting worse, it can also improve outcomes at different stages of diagnosis and illness and manage public health incidents. However, certain AI techniques and methods are not understandable to humans. So, a major challenge here is to apply these approaches and methods in a way that it will make sense to humans, thereby solving the explainability challenge. Furthermore, by observing epidemic data on a daily basis, people's behavioral patterns may be able to establish trustworthy forecasts. However, gathering, organizing, preserving, and disseminating this data presents significant challenges to the epidemiological and computing communities. Developing methodologies for collecting, fussing, storing, retrieving, and sharing such data is a major challenge that needs to be overcome to enhance both fields.

Furthermore, there are persistent challenges with epidemiological knowledge representation. One of these challenges is the dynamic nature of epidemiological data, which makes it necessary to update knowledge representations frequently. This makes it difficult to keep AI models accurate and current [97]. Therefore, there is a need for a fluid framework for updating domain knowledge to ensure the accuracy of solutions.

It can be difficult to understand the decisions and predictions made by AI models, particularly deep learning algorithms, often called "black boxes" [134]. For epidemiologists, it is essential to understand the rationale behind AI-generated findings. Therefore, research on interpretable AI models and techniques for elucidating intricate AI-driven epidemiological forecasts is needed to mitigate this shortcoming [144].

While wearable digital sensing devices (CPIs) are advantageous because they continuously record large-scale interactions, providing a more accurate estimate of human interactions in reality, they also have some drawbacks, including user willingness, data aggregation, and structuring, the perception of a link between usage and certain diseases, and a lack of regulation surrounding their development and utilization [134]. A concerted effort to solve these problems will help in their utility for capturing and making human interaction data available for epidemiological purposes. Moreover, online social networks depict virtual social behaviors that diverge from in-person interactions that facilitate the spread of disease and might not adequately convey the intricate dynamics of in-person interactions pertinent to the transmission of disease [114]. Wearable devices for CPI monitoring are currently expensive and scarce and require users to carry additional sensing devices, which may discourage an average person from using them. As a result, CPIs may not be easily extended to larger scales in the near future [38], [39].

Concerns about bias in algorithms and decision-making processes are among the ethical and social issues that the use of AI in epidemiology brings up [120]. Vulnerable populations may be disproportionately impacted by biased data and results [107]. Maintaining the integrity and validity of research findings in AI-driven epidemiological investigations requires addressing these biases and maintaining fairness and equity.

Other issues that have emerged include patient harm due to errors in the system and who is to be held responsible and liable [39], [89], [104], [121] patient privacy issues restricting access to data; and the moral, legal, and medical difficulties associated with wrong usage AI by individuals and care givers adds to the current discourse of AI application in epidemiology [77], [85].

5.2. Future Trends and Research Directions

Overcoming the aforementioned challenges and expanding the discipline to new heights will determine the fate of AI applications in epidemiology. One exciting area is creating AI-powered early warning systems for disease epidemics [119]. These systems have the ability to generate fast alarms through the use of predictive modeling and real-time data analysis. This facilitates prompt containment and quick responses. AI integration with digital epidemiology—incorporating data from wearables, social media, and online platforms—is another potential future route [80]. Artificial intelligence (AI) methods for analyzing these various data sources can improve disease surveillance, monitor public opinion, and spot possible epidemics before they become serious.

Furthermore, collaborative research between epidemiologists and AI specialists is becoming more and more important [110]. Multidisciplinary cooperation can close the knowledge gap between domain-specific expertise and cutting-edge AI techniques, resulting in more significant and successful AI applications in epidemiology.

Governments should also establish legislative guidelines to regulate what and where AI is applied in epidemiology, which will ensure a healthy marriage of the two disciplines [109], [148].

Regardless of socioeconomic status, people carry smart phones everywhere and at all times, making them potentially global physical proximity sensors. Furthermore, their ubiquity in developing nations underscores their viability as a means of recording and studying human movement and migration for epidemiological purposes [136]. Though little work has been done thus far to fully utilize these phone communication and interaction traces for contact tracing and epidemic modeling, doing so would greatly advance the field [33].

Furthermore, the integration of AI with cutting-edge technologies like blockchain and the Internet of Things (IoT) may pave the way for novel approaches to safe, decentralized, and instantaneous epidemiological data processing. Additionally, funding for educational and training initiatives is essential to ensuring that the future generation of epidemiologists has the know-how to fully utilize artificial intelligence. Through promoting cooperation, creativity, and moral behavior, AI in epidemiology can develop further and significantly improve public health worldwide[120].

The development of hybrid models and the integrating of multimodal data sources hold the key to the future of knowledge representation in AI-driven epidemiology. A more thorough knowledge of epidemiological phenomena will be made possible by multimodal representations that include textual, numerical, and picture data [144].

Furthermore, hybrid models—which fuse machine learning algorithms with symbolic AI techniques—are expected to improve the interpretability and explainability of AI systems in epidemiology[118].

6. Conclusions

Despite current difficulties, the use of AI in epidemiology has enormous promise to revolutionize public health practices. With ethical considerations taken into account, interpretability guaranteed, and interdisciplinary cooperation encouraged AI-driven epidemiological research seems to have a bright future. To overcome these obstacles and maximize the potential of AI to enhance global disease surveillance, preventive, and response initiatives, researchers and practitioners must collaborate.

Notwithstanding the difficulties, one of AI's most significant advantages is its assistance with preventative treatment, which helps the healthcare system encourage everyone to become and stay healthy. The application of AI to epidemiology signifies a paradigm shift in our comprehension, evaluation, and management of public health issues. Although many uses, difficulties, and potential developments of AI in epidemiology have all been examined in this overview it is not intended as an exhaustive mention but a pointer to the multi-faceted nature of this inter-marriage of two divergent fields. Thus, it is clear that AI has the power to completely transform epidemiological research and public health procedures.

Author Contributions: Conceptualization: Kamal Bakari. And Aamo Iorliam.; methodology: Kamal Bakari.; validation: Aamo Iorliam; investigation: Aamo Iorliam.; resources: Kamal Bakari.; writing—original draft preparation: Kamal Bakari.; writing—review and editing: Aamo Iorliam.; visualization: Kamal Bakari.; supervision: Aamo Iorliam.

Funding: This research received no external funding**.**

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] R. M. Aiken and R. G. Epstein, "Ethical Guidelines for AI in Education: Starting a Conversation," *Int. J. Artif. Intell. Educ.*, vol. 11, pp. 163–176, 2000.
[2] Z. Chen, K. Marple, E. Salazar, G. Gupta, and L. Tamil, "A Physician Advisory System for Chronic Heart Failure management based on knowledge patterns," *Theory Pract. Log. Program.*, vol. 16, no. 5–6, pp. 604–618, Sep. 2016, doi: 10.1017/S1471068416000429.
[3] P. Kumar, S. K. Sharma, and V. Dutot, "Artificial intelligence (AI)-enabled CRM capability in healthcare: The impact on service innovation," *Int. J. Inf. Manage.*, vol. 69, p. 102598, Apr. 2023, doi: 10.1016/j.ijinfomgt.2022.102598.
[4] A. Kaplan and M. Haenlein, "Rulers of the world, unite! The challenges and opportunities of artificial intelligence," *Bus. Horiz.*, vol. 63, no. 1, pp. 37–50, Jan. 2020, doi: 10.1016/j.bushor.2019.09.003.
[5] S. Charan, M. J. Khan, and K. Khurshid, "Breast cancer detection in mammograms using convolutional neural network," in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Mar. 2018, pp. 1–5. doi: 10.1109/ICOMET.2018.8346384.
[6] M. Somasundaram, K. A. M. Junaid, and S. Mangadu, "Artificial Intelligence (AI) Enabled Intelligent Quality Management System (IQMS) For Personalized Learning Path," *Procedia Comput. Sci.*, vol. 172, pp. 438–442, 2020, doi: 10.1016/j.procs.2020.05.096.
[7] C. G. Geanuracos, S. D. Cunningham, G. Weiss, D. Forte, L. M. Henry Reid, and J. M. Ellen, "Use of Geographic Information Systems for Planning HIV Prevention Interventions for High-Risk Youths," *Am. J. Public Health*, vol. 97, no. 11, pp. 1974–1981, Nov. 2007, doi: 10.2105/AJPH.2005.076851.
[8] S. J. Pierce, R. L. Miller, M. M. Morales, and J. Forney, "Identifying HIV Prevention Service Needs of African American Men Who Have Sex With Men," *J. Public Heal. Manag. Pract.*, vol. 13, no. Supplement, pp. S72–S79, Jan. 2007, doi: 10.1097/00124784-200701001-00012.
[9] C. Fulcher and C. Kaukinen, "Mapping and visualizing the location HIV service providers: An exploratory spatial analysis of Toronto neighborhoods," *AIDS Care*, vol. 17, no. 3, pp. 386–396, Apr. 2005, doi: 10.1080/09540120512331314312.
[10] U. Kitron, U. Shalom, C. Costin, H. Pener, Z. Greenberg, and L. Orshan, "Geographic Information System in Malaria Surveillance: Mosquito Breeding and Imported Cases in Israel, 1992," *Am. J. Trop. Med. Hyg.*, vol. 50, no. 5, pp. 550–556, May 1994, doi: 10.4269/ajtmh.1994.50.550.
[11] P. Ceccato, C. Vancutsem, R. Klaver, J. Rowland, and S. J. Connor, "A Vectorial Capacity Product to Monitor Changing Malaria Transmission Potential in Epidemic Regions of Africa," *J. Trop. Med.*, vol. 2012, pp. 1–6, 2012, doi: 10.1155/2012/595948.
[12] S. Sasaki, H. Suzuki, K. Igarashi, B. Tambatamba, and P. Mulenga, "Spatial Analysis of Risk Factor of Cholera Outbreak for 2003–2004 in a Peri-urban Area of Lusaka, Zambia," *Am. J. Trop. Med. Hyg.*, vol. 79, no. 3, pp. 414–421, Sep. 2008, doi: 10.4269/ajtmh.2008.79.414.
[13] A. Signorini, A. M. Segre, and P. M. Polgreen, "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic," *PLoS One*, vol. 6, no. 5, p. e19467, May 2011, doi: 10.1371/journal.pone.0019467.
[14] H. T. Ho *et al.*, "Using Google Trends to Examine the Spatio-Temporal Incidence and Behavioral Patterns of Dengue Disease: A Case Study in Metropolitan Manila, Philippines," *Trop. Med. Infect. Dis.*, vol. 3, no. 4, p. 118, Nov. 2018, doi: 10.3390/tropicalmed3040118.
[15] C. Alicino *et al.*, "Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes," *Infect. Dis. Poverty*, vol. 4, no. 1, p. 54, Dec. 2015, doi: 10.1186/s40249-015-0090-9.
[16] A. R. Daughton and M. J. Paul, "Identifying Protective Health Behaviors on Twitter: Observational Study of Travel Advisories and Zika Virus," *J. Med. Internet Res.*, vol. 21, no. 5, p. e13090, May 2019, doi: 10.2196/13090.
[17] N. Mahroum *et al.*, "Public reaction to Chikungunya outbreaks in Italy—Insights from an extensive novel data streams-based structural equation modeling analysis," *PLoS One*, vol. 13, no. 5, p. e0197337, May 2018, doi: 10.1371/journal.pone.0197337.
[18] M. El Boujnouni, "A study and identification of COVID-19 viruses using N-grams with Naïve Bayes, K-Nearest Neighbors, Artificial Neural Networks, Decision tree and Support Vector Machine," in *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, May 2022, pp. 1–7. doi: 10.1109/ISCV54655.2022.9806081.
[19] E. L. N. Maciel *et al.*, "Spatial patterns of pulmonary tuberculosis incidence and their relationship to socio-economic status in Vitoria, Brazil.," *Int. J. Tuberc. Lung Dis.*, vol. 14, no. 11, pp. 1395–402, Nov. 2010, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/20937178
[20] C. K. Monaghan *et al.*, "Machine Learning for Prediction of Patients on Hemodialysis with an Undetected SARS-CoV-2 Infection," *Kidney360*, vol. 2, no. 3, pp. 456–468, Mar. 2021, doi: 10.34067/KID.0003802020.
[21] X. Mei *et al.*, "Artificial intelligence–enabled rapid diagnosis of patients with COVID-19," *Nat. Med.*, vol. 26, no. 8, pp. 1224–1228, Aug. 2020, doi: 10.1038/s41591-020-0931-3.
[22] H. C. Metsky, C. A. Freije, T.-S. F. Kosoko-Thoroddsen, P. C. Sabeti, and C. Myhrvold, "CRISPR-based COVID-19 surveillance using a genomically-comprehensive machine learning approach," *bioRxiv*. 2020. doi: 10.1101/2020.02.26.967026.

[23] L. P. Garcia *et al.*, "Estimating underdiagnosis of COVID-19 with nowcasting and machine learning," *Rev. Bras. Epidemiol.*, vol. 24, 2021, doi: 10.1590/1980-549720210047.
[24] N. N. Sun *et al.*, "A prediction model based on machine learning for diagnosing the early COVID-19 patients," *medRxiv*. pp. 1–12, 2020. doi: 10.1101/2020.06.03.20120881.
[25] M. Jamshidi *et al.*, "Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment," *IEEE Access*, vol. 8, pp. 109581–109595, 2020, doi: 10.1109/ACCESS.2020.3001973.
[26] S. Hassantabar *et al.*, "CovidDeep: SARS-CoV-2/COVID-19 Test Based on Wearable Medical Sensors and Efficient Neural Networks," *IEEE Trans. Consum. Electron.*, vol. 67, no. 4, pp. 244–256, Nov. 2021, doi: 10.1109/TCE.2021.3130228.
[27] M. Kukar *et al.*, "COVID-19 diagnosis by routine blood tests using machine learning," *Sci. Rep.*, vol. 11, no. 1, p. 10738, May 2021, doi: 10.1038/s41598-021-90265-9.
[28] F. Soares, "A novel specific artificial intelligence-based method to identify COVID-19 cases using simple blood exams," *medRxiv*. p. 2020.04.10.20061036, 2020. doi: 10.1101/2020.04.10.20061036.
[29] A. Banerjee *et al.*, "Use of Machine Learning and Artificial Intelligence to predict SARS-CoV-2 infection from Full Blood Counts in a population," *Int. Immunopharmacol.*, vol. 86, p. 106705, Sep. 2020, doi: 10.1016/j.intimp.2020.106705.
[30] V. A. de Freitas Barbosa *et al.*, "Heg.IA: an intelligent system to support diagnosis of Covid-19 based on blood tests," *Res. Biomed. Eng.*, vol. 38, no. 1, pp. 99–116, Mar. 2022, doi: 10.1007/s42600-020-00112-5.
[31] C. Pawlowski *et al.*, "Longitudinal laboratory testing tied to PCR diagnostics in COVID-19 patients reveals temporal evolution of distinctive coagulopathy signatures." May 21, 2020. [Online]. Available: http://arxiv.org/abs/2005.10938
[32] N. L. Bragazzi, J. D. Kong, and J. Wu, "Is monkeypox a new, emerging sexually transmitted disease? A rapid review of the literature," *J. Med. Virol.*, vol. 95, no. 1, Jan. 2023, doi: 10.1002/jmv.28145.
[33] K. T. D. Eames and M. J. Keeling, "Contact tracing and disease control," *Proc. R. Soc. London. Ser. B Biol. Sci.*, vol. 270, no. 1533, pp. 2565–2571, Dec. 2003, doi: 10.1098/rspb.2003.2554.
[34] F. K. Dosilovic, M. Brcic, and N. Hlupic, "Explainable artificial intelligence: A survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, pp. 0210–0215. doi: 10.23919/MIPRO.2018.8400040.
[35] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg, "Inferring social ties from geographic coincidences," *Proc. Natl. Acad. Sci.*, vol. 107, no. 52, pp. 22436–22441, Dec. 2010, doi: 10.1073/pnas.1006155107.
[36] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2011, pp. 1082–1090. doi: 10.1145/2020408.2020579.
[37] M. Salathé and J. H. Jones, "Dynamics and Control of Diseases in Networks with Community Structure," *PLoS Comput. Biol.*, vol. 6, no. 4, p. e1000736, Apr. 2010, doi: 10.1371/journal.pcbi.1000736.
[38] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proc. Natl. Acad. Sci.*, vol. 107, no. 51, pp. 22020–22025, Dec. 2010, doi: 10.1073/pnas.1009094108.
[39] B. Srivastava and F. Rossi, "Rating AI systems for bias to promote trustable applications," *IBM J. Res. Dev.*, vol. 63, no. 4/5, pp. 5:1-5:9, Jul. 2019, doi: 10.1147/JRD.2019.2935966.
[40] S. Pollett *et al.*, "Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection," *Clin. Infect. Dis.*, vol. 64, no. 1, pp. 34–41, Jan. 2017, doi: 10.1093/cid/ciw657.
[41] Q. Xu, Y. R. Gel, L. L. Ramirez Ramirez, K. Nezafati, Q. Zhang, and K.-L. Tsui, "Forecasting influenza in Hong Kong with Google search queries and statistical model fusion," *PLoS One*, vol. 12, no. 5, p. e0176690, May 2017, doi: 10.1371/journal.pone.0176690.
[42] A. N. Belkacem, S. Ouhbi, A. Lakas, E. Benkhelifa, and C. Chen, "End-to-End AI-Based Point-of-Care Diagnosis System for Classifying Respiratory Illnesses and Early Detection of COVID-19: A Theoretical Framework," *Front. Med.*, vol. 8, Mar. 2021, doi: 10.3389/fmed.2021.585578.
[43] C. Brown *et al.*, "Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Aug. 2020, pp. 3474–3484. doi: 10.1145/3394486.3412865.
[44] E. Fayyoumi, S. Idwan, and H. AboShindi, "Machine Learning and Statistical Modelling for Prediction of Novel COVID-19 Patients Case Study: Jordan," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 5, 2020, doi: 10.14569/IJACSA.2020.0110518.
[45] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *Int. J. Inf. Technol.*, vol. 12, no. 3, pp. 731–739, Sep. 2020, doi: 10.1007/s41870-020-00495-9.
[46] H. J. Chen *et al.*, "Machine learning-based CT radiomics model distinguishes COVID-19 from non-COVID-19 pneumonia," *BMC Infect. Dis.*, vol. 21, no. 1, p. 931, Dec. 2021, doi: 10.1186/s12879-021-06614-6.
[47] S. Dutta and S. K. Bandyopadhyay, "Machine learning approach for confirmation of COVID-19 cases: positive, negative, death and release," *Iberoam. J. Med.*, vol. 2, no. 3, pp. 172–177, May 2020, doi: 10.53986/ibjm.2020.0031.
[48] D. Ferrari *et al.*, "Machine learning in predicting respiratory failure in patients with COVID-19 pneumonia—Challenges, strengths, and opportunities in a global health emergency," *PLoS One*, vol. 15, no. 11, p. e0239172, Nov. 2020, doi: 10.1371/journal.pone.0239172.
[49] A. Mavragani and G. Ochoa, "Google Trends in Infodemiology and Infoveillance: Methodology Framework," *JMIR Public Heal. Surveill.*, vol. 5, no. 2, p. e13439, May 2019, doi: 10.2196/13439.
[50] G. Eysenbach, "Infodemiology and Infoveillance," *Am. J. Prev. Med.*, vol. 40, no. 5, pp. S154–S158, May 2011, doi: 10.1016/j.amepre.2011.02.006.
[51] A. Vespignani *et al.*, "Modelling COVID-19," *Nat. Rev. Phys.*, vol. 2, no. 6, pp. 279–281, May 2020, doi: 10.1038/s42254-020-0178-4.
[52] F. Bloise and M. Tancioni, "Predicting the spread of COVID-19 in Italy using machine learning: Do socio-economic factors matter?," *Struct. Chang. Econ. Dyn.*, vol. 56, pp. 310–329, Mar. 2021, doi: 10.1016/j.strueco.2021.01.001.

[53] S. F. McGough, J. S. Brownstein, J. B. Hawkins, and M. Santillana, "Forecasting Zika Incidence in the 2016 Latin America Outbreak Combining Traditional Disease Surveillance with Search, Social Media, and News Report Data," *PLoS Negl. Trop. Dis.*, vol. 11, no. 1, p. e0005295, Jan. 2017, doi: 10.1371/journal.pntd.0005295.
[54] T. Wagner *et al.*, "Augmented curation of clinical notes from a massive EHR system reveals symptoms of impending COVID-19 diagnosis," *Elife*, vol. 9, Jul. 2020, doi: 10.7554/eLife.58227.
[55] R. T. Gluskin, M. A. Johansson, M. Santillana, and J. S. Brownstein, "Evaluation of Internet-Based Dengue Query Data: Google Dengue Trends," *PLoS Negl. Trop. Dis.*, vol. 8, no. 2, p. e2713, Feb. 2014, doi: 10.1371/journal.pntd.0002713.
[56] C. M. Smith, S. C. Le Comber, H. Fry, M. Bull, S. Leach, and A. C. Hayward, "Spatial methods for infectious disease outbreak investigations: systematic literature review," *Eurosurveillance*, vol. 20, no. 39, Oct. 2015, doi: 10.2807/1560-7917.ES.2015.20.39.30026.
[57] G. E. Glass, B. S. Schwartz, J. M. Morgan, D. T. Johnson, P. M. Noy, and E. Israel, "Environmental risk factors for Lyme disease identified with geographic information systems.," *Am. J. Public Health*, vol. 85, no. 7, pp. 944–948, Jul. 1995, doi: 10.2105/AJPH.85.7.944.
[58] D. E. Impoinvil *et al.*, "The Spatial Heterogeneity between Japanese Encephalitis Incidence Distribution and Environmental Variables in Nepal," *PLoS One*, vol. 6, no. 7, p. e22192, Jul. 2011, doi: 10.1371/journal.pone.0022192.
[59] S. A. Speer, J. C. Semenza, T. Kurosaki, and H. Anton-Culver, "Risk factors for acute myeloid leukemia and multiple myeloma: a combination of GIS and case-control studies.," *J. Environ. Health*, vol. 64, no. 7, pp. 9–16; quiz 35–6, Mar. 2002, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/11901667
[60] A. Poulstrup and H. L. Hansen, "Use of GIS and Exposure Modeling as Tools in a Study of Cancer Incidence in a Population Exposed to Airborne Dioxin," *Environ. Health Perspect.*, vol. 112, no. 9, pp. 1032–1036, Jun. 2004, doi: 10.1289/ehp.6739.
[61] M. Palaniyandi, "The environmental aspects of dengue and chikungunyatransmissionin India : Remote sensing and GIS for epidemic control," *Int. J. Mosq. Res.* ·, vol. 1, no. 2, pp. 35–40, 2014.
[62] D. A. Salmon, M. Z. Dudley, J. M. Glanz, and S. B. Omer, "Vaccine Hesitancy," *Am. J. Prev. Med.*, vol. 49, no. 6, pp. S391–S398, Dec. 2015, doi: 10.1016/j.amepre.2015.06.009.
[63] K. H. Saglani and N. J. Janwe, "Machine Learning Based Sentiment Analysis on Twitter Data," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4413–4419, Aug. 2020, doi: 10.30534/ijeter/2020/60882020.
[64] B. L. Hoffman *et al.*, "It's not all about autism: The emerging landscape of anti-vaccination sentiment on Facebook," *Vaccine*, vol. 37, no. 16, pp. 2216–2223, Apr. 2019, doi: 10.1016/j.vaccine.2019.03.003.
[65] R. F. Hunter *et al.*, "Ethical Issues in Social Media Research for Public Health," *Am. J. Public Health*, vol. 108, no. 3, pp. 343–348, Mar. 2018, doi: 10.2105/AJPH.2017.304249.
[66] N. Yiannakoulias, C. E. Slavik, and M. Chase, "Expressions of pro- and anti-vaccine sentiment on YouTube," *Vaccine*, vol. 37, no. 15, pp. 2057–2064, Apr. 2019, doi: 10.1016/j.vaccine.2019.03.001.
[67] F. Kunneman, M. Lambooij, A. Wong, A. van den Bosch, and L. Mollema, "Monitoring stance towards vaccination in twitter messages," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 33, Dec. 2020, doi: 10.1186/s12911-020-1046-y.
[68] X. Yuan and A. T. Crooks, "Examining Online Vaccination Discussion and Communities in Twitter," in *Proceedings of the 9th International Conference on Social Media and Society*, Jul. 2018, pp. 197–206. doi: 10.1145/3217804.3217912.
[69] G. Guibon, M. Ochs, and P. Bellot, "From Emoji Usage to Categorical Emoji Prediction," in *Computational Linguistics and Intelligent Text Processing*, 2023, pp. 329–338. doi: 10.1007/978-3-031-23804-8_26.
[70] E. D'Andrea, P. Ducange, A. Bechini, A. Renda, and F. Marcelloni, "Monitoring the public opinion about the vaccination topic from tweets analysis," *Expert Syst. Appl.*, vol. 116, pp. 209–226, Feb. 2019, doi: 10.1016/j.eswa.2018.09.009.
[71] L. Li *et al.*, "Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, Aug. 2020, doi: 10.1148/radiol.2020200905.
[72] A. Bharti, S. Krishnan, and S. K. Bharti, "Envisioning the Veracity of Digital Ecosystem in Improvising Effective Pandemic Response," *Front. Blockchain*, vol. 3, Feb. 2021, doi: 10.3389/fbloc.2020.599428.
[73] J. S. Obeid *et al.*, "An artificial intelligence approach to COVID-19 infection risk assessment in virtual visits: A case report," *J. Am. Med. Informatics Assoc.*, vol. 27, no. 8, pp. 1321–1325, Aug. 2020, doi: 10.1093/jamia/ocaa105.
[74] U. Bharti, D. Bajaj, H. Batra, S. Lalit, S. Lalit, and A. Gangwani, "Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, Jun. 2020, pp. 870–875. doi: 10.1109/ICCES48766.2020.9137944.
[75] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in the diagnosis of COVID-19: challenges and perspectives," *Int. J. Biol. Sci.*, vol. 17, no. 6, pp. 1581–1587, 2021, doi: 10.7150/ijbs.58855.
[76] T. F. Jones, R. F. Benson, E. W. Brown, J. R. Rowland, S. C. Crosier, and W. Schaffner, "Epidemiologic Investigation of a Restaurant-Associated Outbreak of Pontiac Fever," *Clin. Infect. Dis.*, vol. 37, no. 10, pp. 1292–1297, Nov. 2003, doi: 10.1086/379017.
[77] J. Liu, J. Ma, J. Li, M. Huang, N. Sadiq, and Y. Ai, "Robust Watermarking Algorithm for Medical Volume Data in Internet of Medical Things," *IEEE Access*, vol. 8, pp. 93939–93961, 2020, doi: 10.1109/ACCESS.2020.2995015.
[78] J. D. Ferreira, D. Paolotti, F. M. Couto, and M. J. Silva, "On the usefulness of ontologies in epidemiology research and practice," *J. Epidemiol. Community Health*, vol. 67, no. 5, pp. 385–388, May 2013, doi: 10.1136/jech-2012-201142.
[79] R. R. Rao, K. Makkithaya, and N. Gupta, "Ontology based semantic representation for Public Health data integration," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Nov. 2014, pp. 357–362. doi: 10.1109/IC3I.2014.7019701.
[80] Q. Wang, M. Su, M. Zhang, and R. Li, "Integrating Digital Technologies and Public Health to Fight Covid-19 Pandemic: Key Technologies, Applications, Challenges and Outlook of Digital Healthcare," *Int. J. Environ. Res. Public Health*, vol. 18, no. 11, p. 6053, Jun. 2021, doi: 10.3390/ijerph18116053.
[81] C. Bellinger, M. S. Mohomed Jabbar, O. Zaïane, and A. Osornio-Vargas, "A systematic review of data mining and machine learning for air pollution epidemiology," *BMC Public Health*, vol. 17, no. 1, p. 907, Dec. 2017, doi: 10.1186/s12889-017-4914-3.

[82] A. Gupta and R. Katarya, "Social media based surveillance systems for healthcare using machine learning: A systematic review," *J. Biomed. Inform.*, vol. 108, p. 103500, Aug. 2020, doi: 10.1016/j.jbi.2020.103500.
[83] B. Wahl, A. Cossy-Gantner, S. Germann, and N. R. Schwalbe, "Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings?," *BMJ Glob. Heal.*, vol. 3, no. 4, p. e000798, Aug. 2018, doi: 10.1136/bmjgh-2018-000798.
[84] B. Pang and Lillian Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*. 2008.
[85] A. Shaban-Nejad, M. Michalowski, J. Brownstein, and D. Buckeridge, "Guest Editorial Explainable AI: Towards Fairness, Accountability, Transparency and Trust in Healthcare," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 7, pp. 2374–2375, Jul. 2021, doi: 10.1109/JBHI.2021.3088832.
[86] T.-Y. Wang, S.-L. Chen, H.-C. Huang, S.-H. Kuo, and Y.-J. Shiu, "The development of an intelligent monitoring and caution system for pressure ulcer prevention," in *2011 International Conference on Machine Learning and Cybernetics*, Jul. 2011, pp. 566–571. doi: 10.1109/ICMLC.2011.6016779.
[87] R. Zhou, X. Zhang, X. Wang, G. Yang, N. Guizani, and X. Du, "Efficient and Traceable Patient Health Data Search System for Hospital Management in Smart Cities," *IEEE Internet Things J.*, vol. 8, no. 8, pp. 6425–6436, Apr. 2021, doi: 10.1109/JIOT.2020.3028598.
[88] A. Adiga, D. Dubhashi, B. Lewis, M. Marathe, S. Venkatramanan, and A. Vullikanti, "Mathematical Models for COVID-19 Pandemic: A Comparative Analysis," *J. Indian Inst. Sci.*, vol. 100, no. 4, pp. 793–807, Oct. 2020, doi: 10.1007/s41745-020-00200-6.
[89] A. J. Aljaaf, D. Al-Jumeily, A. J. Hussain, P. Fergus, M. Al-Jumaily, and K. Abdel-Aziz, "Toward an optimal use of artificial intelligence techniques within a clinical decision support system," in *2015 Science and Information Conference (SAI)*, Jul. 2015, pp. 548–554. doi: 10.1109/SAI.2015.7237196.
[90] Z. C. Antoniou, A. S. Panayides, M. Pantzaris, A. G. Constantinides, C. S. Pattichis, and M. S. Pattichis, "Real-Time Adaptation to Time-Varying Constraints for Medical Video Communications," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 4, pp. 1177–1188, Jul. 2018, doi: 10.1109/JBHI.2017.2726180.
[91] S. Ardabili *et al.*, "COVID-19 Outbreak Prediction with Machine Learning," *Algorithms*, vol. 13, no. 10, p. 249, Oct. 2020, doi: 10.3390/a13100249.
[92] E. Avila, A. Kahmann, C. Alho, and M. Dorn, "Hemogram data as a tool for decision-making in COVID-19 management: applications to resource scarcity scenarios," *PeerJ*, vol. 8, p. e9482, Jun. 2020, doi: 10.7717/peerj.9482.
[93] D. Baker and H. Taylor, "Inequality in health and health service use for mothers of young children in south west England. Survey Team of the Avon Longitudinal Study of Pregnancy and Childhood Team.," *J. Epidemiol. Community Heal.*, vol. 51, no. 1, pp. 74–79, Feb. 1997, doi: 10.1136/jech.51.1.74.
[94] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, and F. Cabitza, "Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study," *J. Med. Syst.*, vol. 44, no. 8, p. 135, Aug. 2020, doi: 10.1007/s10916-020-01597-4.
[95] A. Borda, A. Molnar, C. Neesham, and P. Kostkova, "Ethical Issues in AI-Enabled Disease Surveillance: Perspectives from Global Health," *Appl. Sci.*, vol. 12, no. 8, p. 3890, Apr. 2022, doi: 10.3390/app12083890.
[96] Z. Chen *et al.*, "An AI-Based Heart Failure Treatment Adviser System," *IEEE J. Transl. Eng. Heal. Med.*, vol. 6, pp. 1–10, 2018, doi: 10.1109/JTEHM.2018.2883069.
[97] A. Borlase, J. P. Webster, and J. W. Rudge, "Opportunities and challenges for modelling epidemiological and evolutionary dynamics in a multihost, multiparasite system: Zoonotic hybrid schistosomiasis in West Africa," *Evol. Appl.*, vol. 11, no. 4, pp. 501–515, Apr. 2018, doi: 10.1111/eva.12529.
[98] Y. Chen *et al.*, "An Interpretable Machine Learning Framework for Accurate Severe vs Non-Severe COVID-19 Clinical Type Classification," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3638427.
[99] C. Comito, D. Falcone, and A. Forestiero, "Current Trends And Practices In Smart Health Monitoring And Clinical Decision Support," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Dec. 2020, pp. 2577–2584. doi: 10.1109/BIBM49941.2020.9313449.
[100] M. Daltayanni, C. Wang, and R. Akella, "A Fast Interactive Search System for Healthcare Services," in *2012 Annual SRII Global Conference*, Jul. 2012, pp. 525–534. doi: 10.1109/SRII.2012.65.
[101] Y. Deng *et al.*, "A New Framework to Reduce Doctor's Workload for Medical Image Annotation," *IEEE Access*, vol. 7, pp. 107097–107104, 2019, doi: 10.1109/ACCESS.2019.2917932.
[102] D. U. Pfeiffer, T. P. Robinson, M. Stevenson, K. B. Stevens, D. J. Rogers, and A. C. A. Clements, *Spatial Analysis in Epidemiology*. London, England: Oxford University Press, 2008.
[103] B. Dutta and P. Das, "Semantic Annotator for Knowledge Graph Exploration : Pattern-Based NLP Technique," *J. Inf. Manag.*, vol. 60, no. 1, pp. 49–62, Mar. 2023, doi: 10.17821/srels/2023/v60i1/170889.
[104] Y. K. Dwivedi *et al.*, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manage.*, vol. 57, p. 101994, Apr. 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
[105] G. Eysenbach, "Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet," *J. Med. Internet Res.*, vol. 11, no. 1, p. e11, Mar. 2009, doi: 10.2196/jmir.1157.
[106] C. Feng *et al.*, "A novel artificial intelligence-assisted triage tool to aid in the diagnosis of suspected COVID-19 pneumonia cases in fever clinics," *Ann. Transl. Med.*, vol. 9, no. 3, pp. 201–201, Feb. 2021, doi: 10.21037/atm-20-3073.
[107] K. F. Sellers, E. K. T. Benn, M. Garcia, and M. Kellam, "Addressing Implicit Bias Among Women Statisticians and Data Scientists," *CHANCE*, vol. 30, no. 2, pp. 38–41, Apr. 2017, doi: 10.1080/09332480.2017.1320477.
[108] J. C. Gomes *et al.*, "Covid-19 diagnosis by combining RT-PCR and pseudo-convolutional machines to characterize virus sequences," *Sci. Rep.*, vol. 11, no. 1, p. 11545, Jun. 2021, doi: 10.1038/s41598-021-90766-7.

[109]V. Gomoi and V. Stoicu-Tivadar, "A new method in automatic generation of medical protocols using artificial intelligence tools and a data manager," in *2010 International Joint Conference on Computational Cybernetics and Technical Informatics*, May 2010, pp. 243–246. doi: 10.1109/ICCCYB.2010.5491290.
[110]R. T. Gupta, B. Spilseth, N. Patel, A. F. Brown, and J. Yu, "Multiparametric prostate MRI: focus on T2-weighted imaging and role in staging of prostate cancer," *Abdom. Radiol.*, vol. 41, no. 5, pp. 831–843, May 2016, doi: 10.1007/s00261-015-0579-5.
[111]M. S. Hossen and D. Karmoker, "Predicting the Probability of Covid-19 Recovered in South Asian Countries Based on Healthy Diet Pattern Using a Machine Learning Approach," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dec. 2020, pp. 1–6. doi: 10.1109/STI50764.2020.9350439.
[112]A. Imran *et al.*, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app," *Informatics Med. Unlocked*, vol. 20, p. 100378, 2020, doi: 10.1016/j.imu.2020.100378.
[113]M. Johnson, A. Albizri, A. Harfouche, and S. Fosso-Wamba, "Integrating human knowledge into artificial intelligence for complex and ill-structured problems: Informed artificial intelligence," *Int. J. Inf. Manage.*, vol. 64, p. 102479, Jun. 2022, doi: 10.1016/j.ijinfomgt.2022.102479.
[114]J. J. Jones, J. E. Settle, R. M. Bond, C. J. Fariss, C. Marlow, and J. H. Fowler, "Inferring Tie Strength from Online Directed Behavior," *PLoS One*, vol. 8, no. 1, p. e52168, Jan. 2013, doi: 10.1371/journal.pone.0052168.
[115]J. Jumper *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.
[116]A. Kapoor *et al.*, "Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks," Jul. 2020, [Online]. Available: http://arxiv.org/abs/2007.03113
[117]A. Kaur, R. Garg, and P. Gupta, "Challenges facing AI and Big data for Resource-poor Healthcare System," in *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, Aug. 2021, pp. 1426–1433. doi: 10.1109/ICESC51422.2021.9532955.
[118]K. Davagdorj, J.-W. Bae, V.-H. Pham, N. Theera-Umpon, and K. H. Ryu, "Explainable Artificial Intelligence Based Framework for Non-Communicable Diseases Prediction," *IEEE Access*, vol. 9, pp. 123672–123688, 2021, doi: 10.1109/ACCESS.2021.3110336.
[119]B. (Raymond) Kim, K. Srinivasan, S. H. Kong, J. H. Kim, C. S. Shin, and S. Ram, "ROLEX: A Novel Method for Interpretable Machine Learning Using Robust Local Explanations," *MIS Q.*, vol. 47, no. 3, pp. 1303–1332, Jun. 2022, doi: 10.25300/MISQ/2022/17141.
[120]U. S. Kesmodel, "Information bias in epidemiological studies with a special focus on obstetrics and gynecology," *Acta Obstet. Gynecol. Scand.*, vol. 97, no. 4, pp. 417–423, Apr. 2018, doi: 10.1111/aogs.13330.
[121]M. Madanan, N. A. M. Zulkefli, and N. C. Velayudhan, "Designing a Hybrid Artificial Intelligent Clinical Decision Support System Using Artificial Neural Network and Artificial Bee Colony for Predicting Heart Failure Rate," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2021, pp. 1–7. doi: 10.1109/ICCCI50826.2021.9457007.
[122]P. K. Maduri, Y. Dewangan, D. Yadav, S. Chauhan, and K. Singh, "IOT Based Patient Health Monitoring Portable Kit," in *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Dec. 2020, pp. 513–516. doi: 10.1109/ICACCCN51052.2020.9362985.
[123]C. McGregor *et al.*, "Health Analytics as a Service with Artemis Cloud: Service Availability," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Jul. 2020, pp. 5644–5648. doi: 10.1109/EMBC44109.2020.9176507.
[124]A. Minz and C. Mahobiya, "MR Image Classification Using Adaboost for Brain Tumor Type," in *2017 IEEE 7th International Advance Computing Conference (IACC)*, Jan. 2017, pp. 701–705. doi: 10.1109/IACC.2017.0146.
[125]J. A. Oyedepo, O. B. Shittu, T. O. S. Popoola, and E. O. Ogunshola, "Rapid Epidemiological Mapping of Cholera in Some Parts of Abeokuta Metropolis: A GIS-Supported Post-Epidemic Assessment," *Int. J. Public Heal. Epidemiol.*, vol. 4, no. 6, pp. 152–157, 2015.
[126]L. G. Pee, S. L. Pan, and L. Cui, "Artificial intelligence in healthcare robots: A social informatics study of knowledge embodiment," *J. Assoc. Inf. Sci. Technol.*, vol. 70, no. 4, pp. 351–369, Apr. 2019, doi: 10.1002/asi.24145.
[127]N. Pearce *et al.*, "Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC)," *Thorax*, vol. 62, no. 9, pp. 758–766, Sep. 2007, doi: 10.1136/thx.2006.070169.
[128]M. Peng *et al.*, "Artificial Intelligence Application in COVID-19 Diagnosis and Prediction," *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3541119.
[129]Z. Ren *et al.*, "Spatial-Temporal Variation and Primary Ecological Drivers of Anopheles sinensis Human Biting Rates in Malaria Epidemic-Prone Regions of China," *PLoS One*, vol. 10, no. 1, p. e0116932, Jan. 2015, doi: 10.1371/journal.pone.0116932.
[130]A. Ribbens, J. Hermans, F. Maes, D. Vandermeulen, and P. Suetens, "Unsupervised Segmentation, Clustering, and Groupwise Registration of Heterogeneous Populations of Brain MR Images," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 201–224, Feb. 2014, doi: 10.1109/TMI.2013.2270114.
[131]Y. Guo *et al.*, "The application of artificial intelligence and data integration in COVID-19 studies: a scoping review," *J. Am. Med. Informatics Assoc.*, vol. 28, no. 9, pp. 2050–2067, Aug. 2021, doi: 10.1093/jamia/ocab098.
[132]M. T. Sqalli and D. Al-Thani, "AI-supported Health Coaching Model for Patients with Chronic Diseases," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2019, pp. 452–456. doi: 10.1109/ISWCS.2019.8877113.
[133]O. Strachna and O. Asan, "Reengineering Clinical Decision Support Systems for Artificial Intelligence," in *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, Nov. 2020, pp. 1–3. doi: 10.1109/ICHI48887.2020.9374370.
[134]F. Giuste *et al.*, "Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review," *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 5–21, 2023, doi: 10.1109/RBME.2022.3185953.

[135]Y. Wang, M. Hu, Q. Li, X.-P. Zhang, G. Zhai, and N. Yao, “Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner.” Feb. 12, 2020. [Online]. Available: http://arxiv.org/abs/2002.05534

[136]A. Wesolowski *et al.*, “Quantifying the Impact of Human Mobility on Malaria,” *Science (80-.).*, vol. 338, no. 6104, pp. 267–270, Oct. 2012, doi: 10.1126/science.1223467.

[137]Y. Woo, P. T. C. Andres, H. Jeong, and C. Shin, “Classification of diabetic walking through machine learning: Survey targeting senior citizens,” in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Apr. 2021, pp. 435–437. doi: 10.1109/ICAIIC51459.2021.9415250.

[138]J. Wu *et al.*, “Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results,” *medRxiv*. p. 2020.04.02.20051136, 2020. doi: 10.1101/2020.04.02.20051136.

[139]X. Xie, Z. Zang, and J. M. Ponzoa, “The information impact of network media, the psychological reaction to the COVID-19 pandemic, and online knowledge acquisition: Evidence from Chinese college students,” *J. Innov. Knowl.*, vol. 5, no. 4, pp. 297–305, Oct. 2020, doi: 10.1016/j.jik.2020.10.005.

[140]B. Xu *et al.*, “Epidemiological data from the COVID-19 outbreak, real-time case information,” *Sci. Data*, vol. 7, no. 1, p. 106, Mar. 2020, doi: 10.1038/s41597-020-0448-0.

[141]H. Yu and Z. Zhou, “Optimization of IoT-Based Artificial Intelligence Assisted Telemedicine Health Analysis System,” *IEEE Access*, vol. 9, pp. 85034–85048, 2021, doi: 10.1109/ACCESS.2021.3088262.

[142]A. P. Zhao *et al.*, “AI for science: Predicting infectious diseases,” *J. Saf. Sci. Resil.*, vol. 5, no. 2, pp. 130–146, Jun. 2024, doi: 10.1016/j.jnlssr.2024.02.002.

[143]Q. Zhang *et al.*, “The epidemiology of Plasmodium vivax and Plasmodium falciparum malaria in China, 2004–2012: from intensified control to elimination,” *Malar. J.*, vol. 13, no. 1, p. 419, Dec. 2014, doi: 10.1186/1475-2875-13-419.

[144]K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao, and S. P. Shah, “Harnessing multimodal data integration to advance precision oncology,” *Nat. Rev. Cancer*, vol. 22, no. 2, pp. 114–126, Feb. 2022, doi: 10.1038/s41568-021-00408-3.

[145]Y. Zhou *et al.*, “A Radiomics Approach With CNN for Shear-Wave Elastography Breast Tumor Classification,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1935–1942, Sep. 2018, doi: 10.1109/TBME.2018.2844188.

[146]Y. Zoabi, S. Deri-Rozov, and N. Shomron, “Machine learning-based prediction of COVID-19 diagnosis based on symptoms,” *npj Digit. Med.*, vol. 4, no. 1, p. 3, Jan. 2021, doi: 10.1038/s41746-020-00372-6.

[147]A. Iorliam and J. A. Ingio, “A Comparative Analysis of Generative Artificial Intelligence Tools for Natural Language Processing,” *J. Comput. Theor. Appl.*, vol. 2, no. 1, pp. 91–105, Feb. 2024, doi: 10.62411/jcta.9447.

[148]K. Bhaduri, M. D. Stefanski, and A. N. Srivastava, “Privacy-Preserving Outlier Detection Through Random Nonlinear Data Distortion,” *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 41, no. 1, pp. 260–272, Feb. 2011, doi: 10.1109/TSMCB.2010.2051540.