SCIENTIFIC OASIS

Journal of Operations Intelligence

Journal homepage: www.jopi-journal.org
ISSN: 3009-4267

Transforming Preventive Healthcare with Machine Learning Technologies

Mumtaz Karatas[1,*], Zahra Zare[1], Yu-Jun Zheng[2,3]

[1] Biomedical, Industrial and Human Factors Engineering Department, Wright State University, Dayton, USA
[2] Affiliated Hospital of Hangzhou Normal University, Hangzhou Normal University, Hangzhou, China
[3] School of Information Science and Technology, Hangzhou Normal University, Hangzhou, China

ARTICLE INFO

Article history:
Received 11 December 2024
Received in revised form9 February 2025
Accepted 30 March 2025
Available online 7 April 2025

Keywords:
Machine learning; Preventive healthcare; Early disease detection; Risk prediction; Health analytics

ABSTRACT

In this paper, we explore the transformative role of machine learning (ML) in preventive healthcare (PHC), a proactive approach to health management that aims to prevent diseases and promote overall well-being. We begin by providing an overview of PHC, its importance, and its applications across various healthcare settings. The manuscript then presents a brief review of ML techniques in this field, examining their potential to revolutionize early disease detection, personalized risk assessment, and targeted interventions. We review key studies that demonstrate the capabilities of ML in areas such as cancer screening, cardiovascular risk prediction, population health management, and traditional medicine. By synthesizing current research and identifying future directions, this work aims to enhance the understanding of how ML is reshaping the PHC domain, potentially improving health outcomes and reducing healthcare costs.

1. Introduction

Preventive healthcare (PHC), also known as prophylaxis, is a proactive approach to healthcare management that aims to prevent diseases and promote overall well-being before they occur or progress. PHC includes a wide range of strategies and interventions, including regular check-ups, flu shots, screenings, blood tests, antismoking counseling, and lifestyle modifications. PHC is applied across various healthcare settings, from primary care clinics to public health departments. By focusing on early detection and risk reduction, PHC not only improves individual health outcomes but also has the potential to reduce overall healthcare costs and alleviate the burden on healthcare systems.

Although PHC is a crucial approach for minimizing patient risks and improving quality of life, several

*Corresponding author.
E-mail address: mumtaz.karatas@wright.edu

https://doi.org/10.31181/jopi31202538

challenges negatively impact its effective implementation. One major issue is the limited ability to predict and prevent diseases before they actually occur. Traditional risk assessment models often rely on generalized population data, leading to suboptimal individual recommendations. Another challenge is patient engagement, i.e., many individuals do not adhere to preventive measures due to a lack of personalized guidance and motivation. Additionally, resource allocation in PHC is often overlooked, making it even more difficult to target high-risk populations effectively.

To address these challenges, various analytical techniques and data-driven approaches have been widely used by researchers and practitioners. Artificial intelligence (AI) and Machine Learning (ML) applications and big data analytics play a crucial role in a variety of applications, such as environmental monitoring, transportation, smart city, education, banking and security, energy, and healthcare applications [1]. These methods have proven their potential in tackling these issues by leveraging vast amounts of health data to generate useful insights. The advances in AI/ML have enabled the healthcare community and medical practitioners to execute intelligent disease management [2].

ML, lying at the intersection of statistics, optimization, and computer science, enables the development of complex statistical models from big data, potentially identifying new variables and variable combinations that predict patient outcomes. Advanced ML techniques, such as deep learning (DL), have demonstrated the ability to extract intricate, previously unrecognized features from images [3]. Applying these techniques to challenges in PHC could lay the foundation for improved quantitative decision-making and predictive analytics.

For example, ML algorithms can analyze electronic medical and health records, wearable device data, and genetic information to identify patterns and predict the likelihood of diseases such as diabetes, cardiovascular conditions, and cancer. State-of-the-art diagnostic models can now identify serious conditions in their earliest stages, sometimes surpassing standard health screenings in accuracy. For instance, a ML model developed in [4] is used to identify individuals with a high-risk of colorectal cancer and recommend supplement conventional screening. Personalized risk assessments based on ML models can enhance early intervention strategies, improving patient outcomes. Analyzing individual data from various sources, including public health records, genomic results, and environmental factors, ML methods make it possible healthcare providers to create detailed risk profiles and offer personalized advice on lifestyle modifications, screening, and preventive measures [5]. ML-driven recommendation systems can increase patient engagement by recommending preventive measures to individual needs. These techniques are also used in diagnostic imaging to identify patterns and trends that may elude human practitioners. In terms of resource management, they are used for optimizing screening programs and allocating healthcare resources effectively by identifying high-risk populations. These advances in several domains demonstrate that ML methods have revolutionized PHC implementations. This transformative stream of research is still evolving and reshaping how healthcare providers approach patient care, offering new opportunities to improve outcomes and reduce costs [6].

Motivated by the growing impact of ML in PHC, our main ambition is to explore its applications further and assess its potential in addressing current challenges. As ML continues to evolve, a comprehensive understanding of existing research is essential to identify key trends, limitations, and future research directions. Therefore, in this study, we provide a brief review of relevant studies, highlighting how ML techniques have been applied in PHC and discussing their effectiveness in improving disease prevention, early diagnosis, and resource optimization.

The remainder of this paper is structured as follows: In Section 2, we provide a brief description of the PHC concept and discuss the transformative research trends in this domain. Section 3 reviews current studies and applications of machine learning in early disease detection, highlighting key algorithms and their performance across various medical domains. Section 4 presents a discussion as well as the challenges and limitations. Finally, Section 5 concludes the paper.

2. Preventive Healthcare: A Paradigm Shift in Medicine

PHC implementations represent a fundamental shift in traditional medicine and public health applications. Unlike traditional reactive healthcare, which primarily focuses on treating illnesses after they occur, PHC aims to proactively maintain health and prevent diseases from developing or progressing. Therefore, moving from a reactive to a proactive model of care, PHC approaches encompass a wide range of strategies and techniques that are designed to identify and mitigate health risks before they pose serious medical risks to patients.

PHC can be broadly categorized into three level acts as a guideline for health personnel to prevent and intervene in chronic illnesses [7].

- Primary Prevention: Aims to prevent diseases or injuries before they occur. Examples include vaccinations, health education programs, and lifestyle modifications.
- Secondary Prevention: Focuses on early detection and treatment of diseases before they cause significant problems. This includes screening programs for cancers, cardiovascular diseases, and diabetes.
- Tertiary Prevention: Involves managing chronic diseases to prevent further deterioration or complications. This includes rehabilitation programs and management strategies for conditions like diabetes or heart disease.

PHC offers several benefits at both individual and societal levels. By addressing health issues early or preventing them altogether, PHC can significantly improve quality of life and longevity. It is often more cost-effective than treating diseases. U.S. Centers for Disease Control and Prevention estimate that chronic diseases, which are often preventable, account for 90% of the nation's $4.5 trillion in annual health care expenditures [**cdc2021**]. By preventing diseases and promoting health, PHC can reduce the strain on healthcare systems, particularly in the context of an aging population and increasing chronic disease prevalence.

PHC field has been significantly improved by the application of various analytical techniques. Epidemiological analysis uses statistical methods to study disease patterns in populations, helping to identify risk factors and inform public health strategies. Risk stratification models help categorize individuals based on their risk levels for various diseases, allowing for more targeted preventive interventions. Predictive modeling techniques, including but not limited to AI/ML, are employed to predict future health outcomes based on current data, enabling proactive interventions. Health economics analyses help in evaluating and prioritizing different preventive strategies at a population level.

These analytical approaches have transformed PHC from a one-size-fits-all model to a more personalized and data-driven paradigm. They enable healthcare providers, planners, decision-makers, and policymakers to make more informed decisions about resource allocation, intervention strategies, and individual patient care. However, PHC faces several challenges, including ensuring equitable access to preventive services, addressing social determinants of health that influence preventive care uptake, balancing the benefits of early detection with the risks of overdiagnosis and overtreatment, and integrating PHC seamlessly into existing healthcare systems. As the field evolves, there is growing interest in incorporating advanced technologies to enhance the effectiveness and reach of PHC.

3. Machine Learning in Preventive Healthcare

In this section, we explore various studies that implement ML techniques in PHC applications such as early disease detection, risk assessment, and proactive health management. With this, our objective is to better understand which real-world implementations benefit from ML techniques and how they are reshaped by them.

Before we present the work on this domain, we refer the reader to Xu and Xu's work [8], where the authors investigated the use of ML techniques for PHC with a focus on comorbidity risk prediction and early diagnosis of chronic disease comorbidity. In this comprehensive systematic literature review, the authors identified four main predictive analytics tasks: disease comorbidity data extraction, clustering, network analysis, and risk prediction. The review highlighted how ML approaches can address inherent data deficiencies in healthcare datasets and provide interpretable models that identify significant risk factors for comorbidity development.

We also refer the interested reader to [9] for a review of how DL in computer vision, natural language processing (NLP), and reinforcement learning impacted healthcare applications. Additionally, in [6], the authors discuss how AI/ML impacted medicine and patient–doctor relationships.

We classify studies on ML applications in PHC into four key areas: (*i*) early disease detection, (*ii*) risk assessment and prediction, (*iii*) personalized preventive care, and (*iv*) population health management. Within each category, we investigate how ML techniques have been utilized to enhance PHC effectiveness. We highlight the most prominent studies in a separate subsection, providing a focused review of significant contributions in the field.

3.1 Early Disease Detection

One of the most promising areas of ML in PHC is early disease detection. The ability to identify diseases in their nascent stages can significantly improve treatment outcomes and reduce healthcare costs. Several studies have demonstrated progress in this field, leveraging ML algorithms to analyze medical imaging, genetic data, and electronic health records for early diagnosis. These advancements have been particularly impactful in detecting conditions such as cancer, cardiovascular diseases, and diabetic retinopathy detection, where early intervention plays a critical role in patient prognosis.

3.1.1 Cancer Detection

Cancer detection remains a critical area where ML has shown significant potential. Akselrod-Ballin et al. [10] proposed a combined ML and DL model developed for breast cancer prediction. Their model was trained on 38,444 mammography images and integrated health data from 9,611 women. The proposed model achieved accuracy comparable to expert radiologists in predicting biopsy malignancy, potentially reducing missed breast cancer diagnoses. The authors also stated that the proposed combined approach has the potential to substantially reduce missed diagnoses of breast cancer.

Building on this work, Lång et al. [11] conducted a study where an AI system analyzed initial screening mammograms of 429 women diagnosed with interval cancer. The system detected 19% of interval cancers in prior mammograms, accurately locating and ranking them as "high risk", which could significantly reduce missed diagnoses in screening programs. The authors concluded that the use of AI/ML in screen reading proved to be effective in reducing the interval cancer rate without supplementary screening modalities.

In another study which considers prostate cancer and breast cancer, Litjens et al. [12] investigated the use of DL to enhance objectivity in histopathological slide analysis. Using deep neural networks, they achieved high accuracy in detecting prostate cancer from digitized H&E-stained histopathology slides, improving diagnostic precision. The authors mentioned that DL methods have the potential to improve the efficacy of prostate cancer diagnosis and breast cancer staging. Similarly, Trister et al. [3] stated that ML holds significant promise for improving the accuracy and effectiveness of breast cancer screening by leveraging advanced algorithms to analyze complex imaging and clinical data. They also mention that its successful integration into clinical practice will require large-scale, high-quality datasets, rigorous validation, and careful consideration of ethical, logistical, and policy challenges to ensure that the benefits outweigh the harms.

Extending the application of ML in cancer detection, McKinney et al. [13] developed a DL-based model for breast cancer screening that outperformed human experts. Their model reduced both false positives and false negatives, demonstrating the potential of DL to enhance the accuracy and efficiency of cancer screening programs. Similarly, Ardila et al. [14] proposed a DL-based model for lung cancer detection that outperformed radiologists when prior computed tomography imaging was unavailable. This study highlights the potential of ML implementations to improve early detection rates for one of the deadliest forms of cancer.

In [15], the authors provided a review of the history, current state, and future of AI/ML research in breast cancer screening and diagnosis. The authors stated that AI/ML has been advancing in breast cancer screening, enhancing human interpretation by improving efficiency and reducing subjectivity. With the growth of computational power and DL, analytical methods has shown promising performance in detection and classification, with potential to perform integrated tasks. They also mentioned that challenges like explainability, repeatability, and generalizability must be addressed before it can function as an autonomous reader in clinical practice.

The consistent outperformance of ML models in cancer detection across various types (breast, prostate, lung) suggests that integrating these systems into clinical practice could significantly improve early detection rates. However, it's crucial to consider how these systems can be implemented alongside human expertise rather than as a replacement, to ensure the highest level of patient care and safety.

3.1.2 Cardiovascular Disease Risk Prediction

Cardiovascular diseases remain a leading cause of mortality worldwide, making early detection and risk prediction crucial. Weng et al. [16] developed a ML-based method for cardiovascular disease risk prediction using data from over 350,000 individuals. Their system showed a high-level of accuracy compared to traditional methods, identifying more candidates for preventive treatment while reducing unnecessary interventions. Their study revealed that ML models can significantly improve the accuracy of cardiovascular risk prediction and increase the number of patients identified. Those patients can then benefit from PHC services and avoid unnecessary treatment of others.

In another study, Alaa et al. [**alaa2019**] developed an ML framework named "AutoPrognosis" for cardiovascular risk prediction that outperformed established clinical risk scores. Their model incorporated a wider range of risk factors and provided personalized risk assessments, demonstrating the potential of ML to offer more nuanced and individualized risk predictions.

In a novel approach, Attia et al. [17] used AI/ML to analyze ECG data to identify patients with asymptomatic left ventricular dysfunction, a precursor to heart failure. This study revealed how ML can extract additional value from existing medical tests, potentially enabling earlier interventions.

Mathur et al. [18] provide a review of AI/ML applications in cardiovascular medicine. The authors state that those methods have revolutionized cardiovascular medicine by enabling advancements in imaging, risk prediction, and drug discovery. Despite the potential, the authors also highlight challenges such as data privacy, selection bias, and interoperability that must be addressed for its effective clinical integration.

Our review reveals that the performance of ML models in cardiovascular risk prediction suggests that these tools should be integrated into routine clinical practice. However, it's important to ensure that these models are regularly updated with new data and validated across diverse populations to maintain their accuracy and relevance.

3.1.3 Diabetic Retinopathy Detection

Diabetic retinopathy is a leading cause of vision loss in diabetic patients, making early detection crucial. Gulshan et al. [19] conducted a pivotal study evaluating the effectiveness of an autonomous DL-based system in detecting diabetic retinopathy from retinal images. Their DL model demonstrated high sensitivity and specificity, outperforming traditional diagnostic methods. The authors concluded that further research is necessary to evaluate the applicability of this algorithm in the clinical setting. In a follow-up study, Ting et al. [20] developed a DL system for detecting multiple eye diseases, including diabetic retinopathy, glaucoma, and age-related macular degeneration. Similarly, their proposed model achieved high sensitivity and specificity across all conditions, demonstrating the potential of ML to provide comprehensive eye disease screening.

In a large-scale implementation study, Abràmoff et al. [21] evaluated an AI-based system for diabetic retinopathy screening in a real-world clinical setting. The proposed model consists of two algorithms, named image quality algorithm and diagnostic algorithm. The numerical results revealed that the two step model demonstrated high performance both in terms of accuracy and efficiency. Results also suggested that AI has potential to increase access to diabetic retinopathy screening, particularly in underserved areas.

The high performance of ML models in diabetic retinopathy detection across multiple studies suggests that these systems could be valuable tools for expanding screening programs, particularly in areas with limited access to ophthalmologists. However, it's crucial to ensure that the implementation of these systems doesn't lead to a de-skilling of healthcare professionals and that there are clear protocols for when human expert review is necessary. For a detailed review of diabetic retinopathy detection through DL techniques, the interested can refer to [22]. This review states that DL-based screening systems have the potential to reduce the time required to determine diagnoses and minimize associated costs. The authors also highlighted that convolutional neural networks (CNNs) are widely used for the classification and detection of diabetic retinopathy (DR) images.

3.2 Risk Assessment and Prediction

ML techniques have also demonstrated significant potential in assessing and predicting health risks, allowing for more proactive and personalized preventive care. By analyzing vast amounts of health data, ML models can identify patterns and risk factors, facilitating early interventions and tailored health strategies. The advancements in the field have been particularly significant in comorbidity prediction and mental health risk prediction.

3.2.1 Comorbidity Prediction

Understanding and predicting comorbidities is crucial for effective preventive care, particularly for patients with chronic conditions. A systematic review by Xu and Xu [8] highlighted the potential of ML techniques in facilitating the prevention and early diagnosis of chronic disease comorbidity. The study emphasized the importance of uncovering latent comorbidity patterns for improved patient care.

In [23], the authors review existing ML methods for comorbidity prediction, highlighting their potential to improve precision medicine and patient outcomes. The authors identify 22 studies incorporating 61 ML models, with many achieving high predictive accuracy (AUC 0.80–0.89). However, the review underscores the challenges in interpretability, noting that the lack of standardized evaluation frameworks for explainable artificial intelligence (XAI) limits transparency and clinical adoption. The authors conclude that further development of XAI techniques and richer data sources will be crucial for enhancing comorbidity prediction and facilitating more personalized healthcare interventions.

[24], on the other hand, investigated the prediction of chronic disease comorbidity and multimorbidity using a ML and network analytics-based approach. The authors extracted patient networks from healthcare administrative data, where nodes represent patients and edges indicate shared diseases. Additionally, they incorporate features from patients' health trajectories to enhance prediction accuracy. The study evaluated five different ML models and two DL models. Among these, Extreme Gradient Boosting achieves the highest accuracy (95.05%), followed by Convolutional Neural Networks (91.67%). The authors concluded that their approach can aid healthcare stakeholders and policymakers in mitigating the adverse impacts of chronic disease comorbidity and multimorbidity.

In [25], the authors proposed a comorbidity portfolio design to improve the prediction of asthma treatment costs by categorizing frequently occurring comorbidities into distinct cost groups. Using ML models trained on real-world patient data, they demonstrated that this approach enhances predictive accuracy and provides valuable insights into the cost impact of different comorbidities. In a recent study, Choi et al. [26] used graph neural networks to model the progression of multiple chronic conditions over time. Their model provided insights into the temporal relationships between different diseases, potentially enabling more targeted preventive interventions.

The ability of ML models to predict and understand complex comorbidities suggests that these tools could be valuable for developing more holistic and personalized preventive care strategies. However, our review also reveals that it's important to ensure that these models are interpretable and that their predictions can be translated into actionable insights for healthcare providers.

3.2.2 Mental Health Risk Prediction

Mental health is an increasingly important aspect of PHC. Kessler et al. [27] applied ML algorithms to predict the onset of post-traumatic stress disorder (PTSD) among trauma survivors, demonstrating the potential of ML in mental health risk assessment and early intervention. In particular, the authors developed an actuarial risk algorithm using administrative data to predict suicides in soldiers within 12 months after inpatient treatment for psychiatric disorders. The study identifies key risk factors and highlights the potential for targeted posthospitalization interventions for high-risk individuals to reduce suicide and other adverse outcomes. Walsh et al. [28] developed a ML model to predict suicide attempts using electronic health records. Their model identified high-risk individuals with greater accuracy than traditional clinical assessments, potentially enabling more timely interventions.

In another study, Reece and Danforth [29] used ML to analyze social media posts for signs of depression. The authors used Instagram data from 166 individuals and applied ML techniques to identify depression markers through analysis of 43,950 photos. Their models outperformed general practitioners' diagnostic success rates, suggesting potential for early screening and detection of mental illness using computational features rather than human ratings of photo attributes. The model demonstrated the potential of using digital footprints for early detection of mental health issues, although it also raised important privacy concerns.

In [30] the authors conducted a systematic review of ML approaches applied to predicting mental health problems, analyzing 30 relevant research articles. They categorized the studies by mental health issues such as schizophrenia, bipolar disorder, anxiety, depression, PTSD, and mental health problems in children. The paper highlighted the challenges, limitations, and future directions in applyingML to mental health, offering valuable insights and recommendations for further development in the field.

In a more recent study [31], the authors used a two-wave cross-sectional survey design to predict mental health outcomes in Chinese youth aged 14–25. Focusing on factors such as stressful life events, social support, and emotional intelligence, the authors identified six key features significantly associated with mental health, including punishment, self-regulation of emotions, and use of social support. The ML models, particularly the SHAP-optimized backpropagation neural network, demonstrated strong performance in both internal and external validation, indicating their potential for clinical application in predicting mental health outcomes.

The promising results of ML in mental health risk prediction suggest that these tools could play a crucial role in early intervention and prevention. However, the sensitive nature of mental health data and the potential for stigmatization necessitate careful consideration of ethical implications and privacy protections in the implementation of these systems.

3.3 Personalized Preventive Care

ML is also transforming PHC by enabling more personalized approaches, allowing healthcare providers to tailor interventions based on individual patient characteristics and needs. By analyzing large datasets, ML models can identify patterns and predict outcomes, helping to proactively address health risks, optimize treatment plans, and enhance patient outcomes. This shift towards personalized preventive care improves efficiency, reduces healthcare costs, and ensures more effective interventions. Here we discuss the implementations under three categories as lifestyle interventions, medication adherence, and traditional medicine implementations.

3.3.1 Lifestyle Interventions

Lifestyle factors play a crucial role in many chronic diseases, making personalized lifestyle interventions a key aspect of preventive care. Zeevi et al. [32] used ML to develop personalized dietary interventions based on individual gut microbiome composition and other personal factors. Their approach demonstrated superior glycemic control compared to standard dietary advice, highlighting the potential of ML for highly personalized preventive care.

Althoff et al. [33] used ML to analyze smartphone data to understand physical activity patterns and their relationship to health outcomes. Their work demonstrates the potential of using ubiquitous computing devices for continuous health monitoring and personalized lifestyle recommendations. Irandoust et al. [34] evaluated the long-term effectiveness of five lifestyle interventions for individuals with eating disorders using ML techniques. Their approach aimed to predict the impact of various interventions on mental health outcomes, incorporating demographic data, lifestyle factors, and psychological variables. By applying ML models such as Random Forest and Gradient Boosting, we seek to identify the most effective strategies for improving health and well-being, with the potential to tailor interventions based on individual needs and characteristics.

In their study, Islam and Shamsuddin [35] explored the application of ML models for diagnosing hypertension and promoting lifestyle changes to manage the condition. They introduced a CNN model trained on a lifestyle dataset to predict hypertension, focusing on factors like age, gender, smoking, diabetes, and education. Their approach utilized data perturbation simulation to interpret the model's learned features and revealed important lifestyle interventions such as stress relief, diet modification, and smoking cessation. This methodology not only highlighted key predictors but also demonstrated the potential of machine learning to provide actionable insights for managing hypertension, particularly in low-income populations.

Our review once again revealed that ML is becoming a valuable tool for personalizing lifestyle interventions aimed at managing chronic diseases. By analyzing individual data, ML can optimize dietary recommendations, monitor physical activity, and predict the effectiveness of lifestyle changes. However, the implementation of these systems must be balanced with considerations of patient autonomy and the potential for technology dependence. It's crucial to ensure that these tools empower patients rather than creating a sense of constant surveillance or pressure.

3.3.2 Medication Adherence

Medication adherence is a critical factor in the effectiveness of many preventive interventions. Nemati et al. [36] aimed to develop and validate an AI Sepsis Expert algorithm for early prediction of sepsis in ICU patients. Using high-resolution vital signs and electronic medical record data, the algorithm accurately predicted sepsis onset 4–12 hours in advance, achieving high performance across both development and validation cohorts.

In an innovative study, Labovitz et al. [37] used AI and smartphone technology to monitor medication adherence in real-time. Their system used computer vision to confirm that patients had taken their medication, demonstrating a novel approach to adherence monitoring. According to the results, patients using the platform demonstrated a 50% improvement in adherence compared to the control group, with a 67% improvement observed in those receiving direct oral anticoagulants, highlighting the potential of real-time monitoring to enhance adherence and patient outcomes.

Koesmahargyo et al. [38] investigated the accuracy of ML models in predicting medication adherence using real-time data from remote dosing measurements. They utilized data from a clinical trial (n = 4,182) where participants' medication intake was recorded through a smartphone app. The models predicted adherence rates of greater than 80% during the trial, for the following week, and the subsequent day. The study found that prior adherence behaviors were strong predictors, and ML models achieved an accuracy of up to 87% (AUC). The authors concluded that real-time dosing data can effectively predict adherence and help proactively manage patient care.

In [39], the authors provided a review to explore the application of ML in improving medication adherence. They analyzed 43 studies focusing on two main approaches: predicting adherence using data such as self-reports and pharmacy claims and monitoring adherence with sensor-based systems. ML models, including neural networks and random forests, showed high predictive accuracy, with some systems achieving up to 93.75% accuracy. The review concluded that ML has a strong potential to enhance medication adherence, though improvements in data accuracy and system design are needed.

The application of ML in improving medication adherence shows great promise for enhancing the effectiveness of preventive interventions. However, the implementation of these systems must be balanced with patient privacy concerns and the potential for creating anxiety or mistrust. It's important to ensure that these tools are used to support and empower patients rather than to police their behavior.

3.3.3 Traditional Medicine in Preventive Healthcare

Traditional medicine (TM), encompassing practices like herbal medicine, acupuncture, and traditional massage, plays a significant role in PHC in many cultures. Its holistic approach, often focusing on lifestyle and balance, aligns well with the principles of prevention. According to [**WHO2024**], nearly half of the population in many industrialized countries regularly uses some form of traditional and complementary medicine (T&CM), including the United States (42%), Australia (48%), France (49%), and Canada (70%). Payyappallimana [40] discussed the role and challenges of TM in public health, highlighting its significance in achieving health sector development objectives. The study emphasized that traditional and cultural medical knowledge plays a catalytic role in advancing these goals.

There exist several studies that implemented AI/ML approaches in TM implementations. In [41], the authors reviewed the ML approaches in TCM and discussed the function and feature differences among different ML approaches. In [42], the authors explored advances in patient classification within Traditional Chinese Medicine (TCM) through a ML perspective, highlighting key challenges and computational methodologies. They examined how TCM diagnosis relies on four primary diagnostic methods—inspection, auscultation and olfaction, interrogation, and palpation—each providing critical information for syndrome differentiation and disease classification. By reviewing the existing studies, the study categorized patient classification approaches into sign classification, syndrome differentiation, and disease classification. The authors presented a structured analysis of various ML techniques applied to different TCM diagnostic data types, offering insights into their effectiveness and limitations.

In another study, Zheng et al. [43], proposed an intelligent optimization approach to enhance community-based COVID-19 prevention using TCM. Recognizing the limitations of a unified prevention program for all residents, the authors introduced a diversified strategy that tailors TCM-based PHC methods to different population groups. The study demonstrated the computational efficiency of this approach and reported its successful application in a large region in China, during the peak of the pandemic. Tian et al. [44] investigated the application of ML techniques, specifically naive Bayesian classification and recursive partitioning, to assess the drug-likeness of compounds from the TCM. They discussed that structural fingerprints play a more significant role in distinguishing drug-like from non-drug-like molecules. [45] employed a ML approach to investigate the classification of TCM herbs according to their Meridians, a fundamental concept in TCM theory. By analyzing the molecular features of 646 herbs and their active compounds—including structure-based fingerprints and ADME (absorption, distribution, metabolism, and excretion) properties—they demonstrate that machine learning models can predict Meridian classifications with a top accuracy of 0.83.

Despite its potential, TM faces challenges such as lack of standardization, quality control, and potential interactions with conventional medications. Further research is needed to validate TM practices, establish safety profiles, and develop guidelines for appropriate use. Additionally, cultural sensitivity and ethical considerations are important when incorporating TM into PHC. Addressing these challenges will enable understanding the potential of TM to improve PHC outcomes.

New technologies such as federated learning and edge computing promise to enhance PHC without violating patient confidentiality. Federated learning enables the training of ML models on decentralized devices or servers holding local data samples, preserving data confidentiality when aggregating community knowledge. Edge computing, conversely, allows processing data at or near the source of the data, reducing latency and supporting real-time monitoring of health and decision-making. These technologies have the potential to be a key force in propelling preventive healthcare, especially in resource-limited or off-grid locations.

3.4 Population Health Management

ML is increasingly being applied at the population level for PHC, allowing for more efficient and targeted public health interventions. By analyzing large datasets, these models can identify patterns and predict trends in health outcomes, enabling healthcare providers to allocate resources more effectively. For example, ML can predict disease outbreaks, identify high-risk populations, and optimize vaccination strategies. These applications improve the precision of interventions, ensuring that public health efforts are directed where they are most needed, ultimately enhancing the overall efficiency and impact of health initiatives.

3.4.1 Outbreak Prediction

Early prediction of disease outbreaks is crucial for effective public health responses. To predict influenza outbreaks, Santillana et al. [46] developed ML models using a combination of electronic health records and internet search data. Their study demonstrated the potential for early warning systems in public health. In particular, the core innovation in this study is that the authors combine these independent estimates into a single prediction using ensemble techniques.

Ginsberg et al. [47] proposed a method for early detection of influenza epidemics by analyzing Google search queries related to influenza-like illness (ILI). They showed that the frequency of certain search queries strongly correlates with the percentage of physician visits for ILI symptoms. By processing search data from 2003 to 2008, the authors created a model to estimate weekly ILI activity at regional and state levels in the US, achieving a reporting lag of about one day. While this system faced challenges, it pioneered the use of big data and ML for real-time disease surveillance.

In a more recent study, Liu et al. [**liu2018early**] introduced a novel survival analysis approach, RankSvx, designed to predict the timing of complications in Type 2 diabetes mellitus (T2DM) patients after their initial diagnosis. This method simultaneously optimizes two key objectives: accurately predicting event times and ranking the relative risks of observed and censored events. By incorporating a multi-task framework, the approach captures correlations between multiple T2DM complications, outperforming traditional survival models and offering significant improvements in predictive accuracy for clinical applications.

In [48] the authors reviewed the application of sentiment analysis and ML techniques in predicting disease outbreaks, particularly through the analysis of social media data and official healthcare datasets. They highlighted the growing importance of using platforms like Twitter, where users share sentiments that can be processed to generate valuable health-related information. The paper surveyed various ML models, including regression and classification techniques, employed to predict epidemics such as influenza and COVID-19. It discussed how sentiment analysis, coupled with real-time data, can enhance prediction systems, offering insights into disease trends and aiding healthcare organizations like the CDC and WHO. Additionally, the authors provided an overview of studies from 2010 to 2020, emphasizing the improvements in prediction accuracy brought about by these methodologies, as well as the challenges and opportunities in integrating social media data for more effective disease outbreak prediction.

3.4.2 Health Resource Allocation

Efficient allocation of health resources is crucial for effective preventive care at the population level.

For example, Rajkomar et al. [49] proposed a predictive methodology which uses DL techniques to analyze raw electronic health record (EHR) data in the Fast Healthcare Interoperability Resources (FHIR) format. They demonstrated that DL models, using this unprocessed EHR data, can accurately predict various medical outcomes, such as in-hospital mortality, unplanned readmissions, and prolonged length of stay, without the need for extensive site-specific data harmonization. By incorporating the full spectrum of EHR data, including clinicians' free-text notes, the models achieve superior accuracy compared to traditional predictive models. The authors highlighted the scalability of their approach, as it eliminates the labor-intensive preprocessing required by traditional methods, thus paving the way for more efficient, accurate, and widely applicable predictive systems in healthcare.

[50], on the other hand, developed a predictive model for integrated healthcare and long-term care resource consumption, specifically designed for patients with circulatory diseases. As medical costs and the burden of cardiovascular disease continue to rise, the proposed model, named the Adherence Score for Healthcare Resource Outcome (ASHRO), aimed to improve the efficiency and quality of the healthcare system by incorporating health behaviors into the prediction of medical and long-term care costs. The study utilized a large-scale database containing health insurance claims, long-term care insurance, and health check-up data. The predictive model was built using random forest learning and multiple regression analysis to construct ASHRO scores, which factor in various health behaviors, such as adherence to secondary prevention measures, rehabilitation, and medication use. The model's ability to discriminate and calibrate was evaluated using the area under the curve and

the Hosmer-Lemeshow test. The results showed that patients with higher ASHRO scores, indicating better adherence to health behaviors, had lower mortality rates and lower healthcare costs over a 48-month follow-up period. This approach offered insights into how integrating patient behavior with clinical outcomes can lead to more efficient healthcare resource management and improved patient outcomes.

In their systematic review study, Nsoesie et al. [51] investigated methods used for forecasting influenza outbreaks. They emphasized their potential in decision-making for the allocation of public health resources and the implementation of interventions to mitigate morbidity and mortality. The review concluded that influenza forecasting models can capture key outbreak measures with reasonable accuracy when reliable data and appropriate disease assumptions are in place. However, the authors emphasize that real-time evaluation and performance quantification of these models remain necessary for improving their predictive capabilities.

The application of ML in health resource allocation has the potential to significantly improve the efficiency and effectiveness of PHC at the population level. However, it's crucial to ensure that these systems don't exacerbate existing health disparities and that they are transparent and accountable in their decision-making processes.

4. Discussion, Challenges and Limitations

In this study, we explored the application of ML techniques in PHC, focusing on early disease detection, risk assessment and prediction, personalized preventive care, and population health management. Our review of existing literature reveals the significant potential of ML to revolutionize healthcare by enabling proactive and data-driven approaches to disease prevention and management.

In early disease detection, ML models have demonstrated the ability to analyze complex medical imaging, genetic data, and electronic health records to identify diseases at their earliest stages. This capability holds particular promise for conditions like cancer, cardiovascular diseases, and diabetic retinopathy, where early intervention can substantially improve patient outcomes. The studies reviewed highlight the potential of ML to enhance the accuracy and efficiency of screening programs, reduce false positives and negatives, and ultimately save lives.

ML-driven risk assessment and prediction tools offer the opportunity to move beyond traditional risk assessment models by incorporating diverse data sources, including individual medical history, lifestyle factors, and even environmental exposures. These models can generate personalized risk profiles that enable healthcare providers to tailor preventive interventions to individual needs. By identifying high-risk individuals, ML can facilitate targeted prevention strategies, optimizing resource allocation and improving the effectiveness of preventive care.

Personalized preventive care is another area where ML is making significant strides. By analyzing individual patient data, ML models can provide personalized recommendations for lifestyle modifications, screening schedules, and preventive treatments. Smart health records and behavioral change support systems, driven by ML, can empower patients to take a more active role in their health management. These technologies have the potential to increase patient engagement, improve adherence to preventive measures, and ultimately promote healthier behaviors.

At the population level, ML can play a crucial role in optimizing public health initiatives and resource allocation. By analyzing population-wide data, ML models can identify health disparities, predict disease outbreaks, and inform the development of targeted interventions. These capabilities can help healthcare organizations and policymakers make more informed decisions about resource allocation, program design, and public health strategies.

While ML offers considerable promise for revolutionizing PHC, several challenges and limitations must be acknowledged. One primary concern lies in data quality and availability. ML models are only as good as the data they are trained on, and biases present in the data can lead to skewed or inaccurate predictions. Ensuring access to large, diverse, and representative datasets remains a significant hurdle, particularly for underrepresented populations. Another challenge involves the interpretability of ML models, especially DL approaches. Many complex models operate as "black boxes", making it difficult to understand the reasoning behind their predictions. This lack of transparency can hinder trust and acceptance among healthcare professionals and patients. XAI techniques are emerging, but further research is needed to make these models more transparent and understandable.

Ethical considerations are also important. The use of ML in healthcare raises concerns about data privacy, security, and algorithmic fairness. Safeguarding sensitive patient information and preventing discriminatory outcomes require careful attention to data governance, model development, and deployment practices. Over-reliance on ML models without human oversight could lead to errors or missed diagnoses, emphasizing the need for clinicians to retain their critical judgment. Furthermore, the integration of ML into existing healthcare workflows poses logistical and practical challenges. Many healthcare systems lack the infrastructure and expertise needed to effectively implement and maintain ML-driven solutions. Interoperability with existing electronic health record systems and other data sources can be complex and costly. Addressing these technical and organizational barriers is crucial for realizing the full potential of ML in preventive healthcare.

Finally, the clinical validation and regulatory approval of ML-based diagnostic and predictive tools remain a complex and evolving landscape. Rigorous clinical trials are needed to demonstrate the safety and efficacy of these technologies before they can be widely adopted. Establishing clear regulatory guidelines and standards is essential to ensure that ML-driven healthcare solutions are both effective and safe for patients.

Looking forward, the successful implementation of ML in PHC will require a multidisciplinary approach. Collaboration between data scientists, healthcare professionals, ethicists, and policymakers will be crucial to address the technical, clinical, and ethical challenges identified. Continued research is needed to improve the interpretability of ML models, develop robust validation methods for real-world applications, and establish clear guidelines for the ethical use of AI in healthcare. As the field advances, it will be essential to maintain a focus on patient-centered care, ensuring that ML technologies enhance rather than replace the human elements of healthcare delivery. With careful development and thoughtful implementation, ML has the potential to usher in a new era of proactive, personalized, and highly effective PHC. Subsequent research can explore the application of XAI frameworks to enhance transparency and trust in ML applications in PHC. While several studies have demonstrated high predictive performance, there is still a gap in translating such models into practice since they are not interpretable enough. Developing easy-to-use visualization tools and explanation techniques can bridge such a gap, enabling healthcare professionals to better grasp and trust ML-based recommendations.

5. Conclusion

This brief review of ML applications in PHC highlights the transformative potential of these technologies across various domains of health management and disease prevention. From early disease detection and personalized risk assessment to population health management and resource allocation, ML has demonstrated significant promise in enhancing the accuracy, efficiency, and scope of PHC activities. Most of the studies investigated show that ML models can match or outperform traditional methods and human experts in critical tasks such as cancer detection, cardiovascular risk prediction, and outbreak forecasting. The ability of ML to analyze complex, multidimensional data and uncover subtle patterns offers unprecedented opportunities for early intervention and personalized preventive strategies, potentially revolutionizing our approach to healthcare delivery.

However, the path to fully realizing the benefits of ML in PHC is not without challenges. Issues of data quality, model interpretability, ethical considerations, and integration into existing healthcare systems remain significant hurdles. Addressing these challenges will require a concerted effort from a diverse range of stakeholders, including data scientists, healthcare professionals, ethicists, policymakers, and patients. As we move forward, it is crucial to maintain a balanced perspective that recognizes both the immense potential of ML and the importance of responsible development and implementation. By fostering interdisciplinary collaboration, investing in robust validation methods, and establishing clear ethical guidelines, we can work towards a future where ML enhances PHC, improving health outcomes and quality of life for individuals and populations alike.

Author Contributions

Conceptualization, methodology, software, validation, formal analysis, investigation, M.K., Z.Z., and Y.J.Z.; resources, M.K.; data curation, Z.Z.; writing—original draft preparation, writing—review and editing, M.K., Z.Z., and Y.J.Z.; supervision, M.K., and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

Data will be made available on request.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Karatas, M., Eriskin, L., Deveci, M., Pamucar, D., & Garg, H. (2022). Big data for healthcare industry 4.0: Applications, challenges and future perspectives. *Expert Systems with Applications*, *200*, 116912. https://doi.org/10.1016/j.eswa.2022.116912

[2] Subramanian, M., Wojtusciszyn, A., Favre, L., Boughorbel, S., Shan, J., Letaief, K. B., Pitteloud, N., & Chouchane, L. (2020). Precision medicine in the era of artificial intelligence: Implications in chronic disease management. *Journal of translational medicine*, *18*, 1–12. https://doi.org/10.1186/s12967-020-02658-5

[3] Trister, A. D., Buist, D. S., & Lee, C. I. (2017). Will machine learning tip the balance in breast cancer screening? *JAMA oncology*, *3*(11), 1463–1464. https://doi.org/10.1001/jamaoncol.2017.2763

[4] Kinar, Y., Kalkstein, N., Akiva, P., Levin, M. A., Fields, E. J., Twito, A., Mintz, M., Shashar, M., Shalev, V., Chodick, G., et al. (2021). Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: A binational retrospective study. *Journal of the American Medical Informatics Association*, *28*(1), 79–88. https://doi.org/10.1093/jamia/ocaa203

[5] Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, *19*(6), 1236–1246. https://doi.org/10.1093/bib/bbx044

[6] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[7] Brenner, H., & Chen, C. (2018). The colorectal cancer epidemic: Challenges and opportunities for primary, secondary and tertiary prevention. *British journal of cancer*, *119*(7), 785–792. https://doi.org/10.1038/s41416-018-0264-x

[8] Duo, X., & Zeshui, X. (2024). Machine learning applications in preventive healthcare: A systematic literature review on predictive analytics of disease comorbidity from multiple perspectives. *Artificial Intelligence in Medicine*, 102950. https://doi.org/10.1016/j.artmed.2024.102950

[9] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

[10] Akselrod-Ballin, A., Chorev, M., Shoshan, Y., Spiro, A., Hazan, A., Melamed, R., Barkan, E., Herzel, E., Naor, S., Karavani, E., et al. (2019). Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, *292*(2), 331–342. https://doi.org/10.1148/radiol.2019190383

[11] Lång, K., Dustler, M., Dahlblom, V., Andersson, I., & Zackrisson, S. (2020). Can artificial intelligence reduce the interval cancer rate in mammography screening? *European Radiology*, *30*, 1419–1424. https://doi.org/10.1007/s00330-019-06559-0

[12] Litjens, G., Sánchez, C. I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen-Van De Kaa, C., Bult, P., Van Ginneken, B., & Van Der Laak, J. (2016). Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Scientific reports*, *6*(1), 1–11. https://doi.org/10.1038/srep26286

[13] McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., et al. (2020). International evaluation of an ai system for breast cancer screening. *Nature*, *577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

[14] Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., et al. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, *25*(6), 954–961. https://doi.org/10.1038/s41591-019-0447-x

[15] Baughan, N., Douglas, L., & Giger, M. L. (2022). Past, present, and future of machine learning and artificial intelligence for breast cancer screening. *Journal of Breast Imaging*, *4*(5), 451–459. https://doi.org/10.1093/jbi/wbac049

[16] Weng, S. F., Reps, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PloS one*, *12*(4), e0174944. https://doi.org/10.1371/journal.pone.0174944

[17] Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., Pellikka, P. A., Enriquez-Sarano, M., Noseworthy, P. A., Munger, T. M., et al. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nature medicine*, *25*(1), 70–74. https://doi.org/10.1038/s41591-018-0240-2

[18] Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. *Clinical Medicine Insights: Cardiology*, *14*, 1179546820927404. https://doi.org/10.1177/1179546820927404

[19] Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, *316*(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

[20] Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I. Y., Lee, S. Y., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *Jama*, *318*(22), 2211–2223. https://doi.org/10.1001/jama.2017.18152

[21] Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, *1*(1), 1–8. https://doi.org/10.1038/s41746-018-0040-6

[22] Alyoubi, W. L., Shalash, W. M., & Abulkhair, M. F. (2020). Diabetic retinopathy detection through deep learning techniques: A review. *Informatics in Medicine Unlocked*, *20*, 100377. https://doi.org/10.1016/j.imu.2020.100377

[23] Alsaleh, M. M., Allery, F., Choi, J. W., Hama, T., McQuillin, A., Wu, H., & Thygesen, J. H. (2023). Prediction of disease comorbidity using explainable artificial intelligence and machine learning techniques: A systematic review. *International journal of medical informatics*, *175*, 105088. https://doi.org/10.1016/j.ijmedinf.2023.105088

[24] Uddin, S., Wang, S., Lu, H., Khan, A., Hajati, F., & Khushi, M. (2022). Comorbidity and multimorbidity prediction of major chronic diseases using machine learning and network analytics. *Expert Systems with Applications*, *205*, 117761. https://doi.org/10.1016/j.eswa.2022.117761

[25] Luo, L., Yu, X., Yong, Z., Li, C., & Gu, Y. (2020). Design comorbidity portfolios to improve treatment cost prediction of asthma using machine learning. *IEEE Journal of Biomedical and Health Informatics*, *25*(6), 2237–2247. https://doi.org/10.1109/JBHI.2020.3044156

[26] Choi, E., Bahadori, M. T., & Sun, J. (2020). Learning low-dimensional representations of medical concepts. *AMIA Summits on Translational Science Proceedings*, *2020*, 41. https://doi.org/10.1101/19004903

[27] Kessler, R. C., Warner, C. H., Ivany, C., Petukhova, M. V., Rose, S., Bromet, E. J., Brown, M., Cai, T., Colpe, L. J., Cox, K. L., et al. (2015). Predicting suicides after psychiatric hospitalization in us army soldiers: The army study to assess risk and resilience in servicemembers (army starrs). *JAMA psychiatry*, *72*(1), 49–57. https://doi.org/10.1001/jamapsychiatry.2014.1754

[28] Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science*, *5*(3), 457–469. https://doi.org/10.1177/2167702617691560

[29] Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), 1–12. https://doi.org/10.1140/epjds/s13688-017-0110-z

[30] Chung, J., & Teo, J. (2022). Mental health prediction using machine learning: Taxonomy, applications, and challenges. *Applied Computational Intelligence and Soft Computing*, *2022*(1), 9970363. https://doi.org/10.1155/2022/9970363

[31] Ding, H., Li, N., Li, L., Xu, Z., & Xia, W. (2025). Machine learning-enabled mental health risk prediction for youths with stressful life events: A modelling study. *Journal of Affective Disorders*, *368*, 537–546. https://doi.org/10.1016/j.jad.2025.01.123

[32] Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., et al. (2015). Personalized nutrition by prediction of glycemic responses. *Cell*, *163*(5), 1079–1094. https://doi.org/10.1016/j.cell.2015.11.001

[33] Althoff, T., Sosic, R., Hicks, J. L., King, A. C., Delp, S. L., & Leskovec, J. (2017). Large-scale physical activity data reveal worldwide activity inequality. *Nature*, *547*(7663), 336–339. https://doi.org/10.1038/nature23018

[34] Irandoust, K., Parsakia, K., Estifa, A., Zoormand, G., Knechtle, B., Rosemann, T., Weiss, K., & Taheri, M. (2024). Predicting and comparing the long-term impact of lifestyle interventions on individuals with eating disorders in active population: A machine learning evaluation. *Frontiers in Nutrition*, *11*, 1390751. https://doi.org/10.3389/fnut.2024.1390751

[35] Islam, M. M., & Shamsuddin, R. (2021). Machine learning to promote health management through lifestyle changes for hypertension patients. *Array*, *12*, 100090. https://doi.org/10.1016/j.array.2021.100090

[36] Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the icu. *Critical care medicine*, *46*(4), 547–553. https://doi.org/10.1097/CCM.0000000000002936

[37] Labovitz, D. L., Shafner, L., Reyes Gil, M., Virmani, D., & Hanina, A. (2017). Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. *Stroke*, *48*(5), 1416–1419. https://doi.org/10.1161/STROKEAHA.116.016368

[38] Koesmahargyo, V., Abbas, A., Zhang, L., Guan, L., Feng, S., Yadav, V., & Galatzer-Levy, I. R. (2020). Accuracy of machine learning-based prediction of medication adherence in clinical research. *Psychiatry research*, *294*, 113558. https://doi.org/10.1016/j.psychres.2020.113558

[39] Bohlmann, A., Mostafa, J., Kumar, M., et al. (2021). Machine learning and medication adherence: Scoping review. *JMIRx med*, *2*(4), e26993. https://doi.org/10.2196/26993

[40] Payyappallimana, U. (2010). Role of traditional medicine in primary health care: An overview of perspectives and challenging. = *Yokohama journal of social sciences*, *14*(6), 57–77. https://doi.org/10.1177/2158244010387431

[41] Chen, H., & He, Y. (2022). Machine learning approaches in traditional chinese medicine: A systematic review. *The American journal of Chinese medicine*, *50*(01), 91–131. https://doi.org/10.1142/S0192415X22500042

[42] Zhao, C., Li, G.-Z., Wang, C., & Niu, J. (2015). Advances in patient classification for traditional chinese medicine: A machine learning perspective. *Evidence-Based Complementary and Alternative Medicine*, *2015*(1), 376716. https://doi.org/10.1155/2015/376716

[43] Zheng, Y.-J., Yu, S.-L., Yang, J.-C., Gan, T.-E., Song, Q., Yang, J., & Karatas, M. (2020). Intelligent optimization of diversified community prevention of covid-19 using traditional chinese medicine. *IEEE Computational Intelligence Magazine*, *15*(4), 62–73. https://doi.org/10.1109/MCI.2020.3019875

[44] Tian, S., Wang, J., Li, Y., Xu, X., & Hou, T. (2012). Drug-likeness analysis of traditional chinese medicines: Prediction of drug-likeness using machine learning approaches. *Molecular Pharmaceutics*, *9*(10), 2875–2886. https://doi.org/10.1021/mp300222d

[45] Wang, Y., Jafari, M., Tang, Y., & Tang, J. (2019). Predicting meridian in chinese traditional medicine using machine learning approaches. *PLoS computational biology*, *15*(11), e1007249. https://doi.org/10.1371/journal.pcbi.1007249

[46] Santillana, M., Nguyen, A. T., Dredze, M., Paul, M. J., Nsoesie, E. O., & Brownstein, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS computational biology*, *11*(10), e1004513. https://doi.org/10.1371/journal.pcbi.1004513

[47] Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012–1014. https://doi.org/10.1038/nature07634

[48] Singh, R., & Singh, R. (2023). Applications of sentiment analysis and machine learning techniques in disease outbreak prediction–a review. *Materials Today: Proceedings*, *81*, 1006–1011. https://doi.org/10.1016/j.matpr.2023.04.567

[49] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., et al. (2018). Scalable and accurate deep learning with electronic health records. *NPJ digital medicine*, *1*(1), 18. https://doi.org/10.1038/s41746-018-0029-1

[50] Takura, T., Hirano Goto, K., & Honda, A. (2021). Development of a predictive model for integrated medical and long-term care resource consumption based on health behaviour: Appli-

cation of healthcare big data of patients with circulatory diseases. *BMC medicine*, *19*, 1–16. https://doi.org/10.1186/s12916-021-02041-1

[51] Nsoesie, E. O., Brownstein, J. S., Ramakrishnan, N., & Marathe, M. V. (2014). A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and other respiratory viruses*, *8*(3), 309–316. https://doi.org/10.1111/irv.12226