Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/rtfqps06

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN PUBLIC HEALTH SURVEILLANCE: APPLICATIONS AND CHALLENGES

Dr Ashwini L H[1*], Dr Vinaykumar L H[2], Dr Hanumanaik L[3]

[1*]Ex-Assistant Professor, JJM Medical College and rc Davanagere, Karnataka
[2]Administrative Medical Officer Phc Aladageri Hirekerur tq, Haveri District
[3]Chief Medical Officer GH Harihara, Harihara tq, Davanagere District

***Corresponding Author:** Dr Ashwini L H
*Ex-Assistant Professor, JJM Medical College and rc Davanagere, Karnataka

Abstract
Artificial Intelligence (AI) and Machine Learning (ML) are already being used in public health surveillance for the analysis of trends, predictions of outcomes and interventions. However, their use in addressing global health inequity is still relatively constrained, especially in LMICs, where other structural factors such as economic and social determinants and governance issues continue to be a challenge. This work uses AI/ML to analyze global public health data, with an emphasis on health inequality, as captured by the Health Access and Quality (HAQ) Index, Mortality-to-Incidence Ratios (MIR), and Risk-Standardized Death Rates (RSD). The data from the Global Burden of Disease Study 2019 was used to predict healthcare trends using Random Forest regression and to categorise countries into meaningful groups for action using K-Means clustering. Clustering was evaluated by silhouette scores, and the predictive accuracy was evaluated by cross-validation. R^2 and Mean Absolute Error (MAE). Results reveal significant disparities: Germany and other Western European countries scored HAQ Index values of 85 and above, while countries in Sub-Saharan Africa, including Chad and Nigeria, scored between 25 and 35. Some countries in the South Asian region such as India have moved up from 45th to 65th place which shows that there is much room for strategic change. Random Forest was more accurate than the baseline models (R^2 =0.94 in Germany and R^2= 0.80 in Afghanistan) and suggested Chad and Afghanistan as the regions where the intervention should be conducted.

Keywords: Artificial Intelligence, Machine Learning, Public Health Surveillance, Global Health Disparities, Predictive Modelling

INTRODUCTION
The use of Artificial Intelligence (AI) and Machine Learning (ML) in public health surveillance has revolutionalized the way that data is collected, analyzed and trends interpreted and interventions planned for. The nature of healthcare inequalities being a worldwide issue requires effective, sustainable, evidence-based approaches to mapping the problem, predicting its dynamics, and guiding the decision-making process. In the last ten years, AI and ML have shown the ability to complement conventional public health systems by providing predictive information and clustering that can optimize the use of resources and improve healthcare delivery (Brownstein et al., 2009; Kass-Hout &

Alhinnawi, 2013). Nevertheless, their use in LMICs, where health disparities are most acute, has been relatively unexamined.

Context and Background

Conventional disease surveillance and health care management information systems have been based on manual or semi-automated data collection procedures. These methods, though useful, are often time-consuming, expensive and have a narrow focus, especially in low resource settings (Ginsberg et al., 2009). New developments in AI and ML have provided possibilities to improve these systems through the use of big data to forecast health care patterns and recognize high risk communities in real time (Salathé et al., 2012; Paul & Dredze, 2011). For instance, while analysing and predicting the flu epidemic, predictive modeling has demonstrated the ability of AI/ML to deal with the challenging and frequently changing health data (Eysenbach, 2006; Chunara et al., 2012).

Still, the inequality in the healthcare system across the globe has not been eradicated completely. The Health Access and Quality (HAQ) Index captures these disparities well, with high-income countries, which generally have HAQ Index values above 85, and low-income countries, where values are often below 40 (Global Burden of Disease Study 2019, 2022). The above inequalities are also compounded by other factors including poor health care facilities, political instability and brittle health care systems and resources in the low performing areas (Nsoesie et al., 2014; Lazer et al., 2014). These disparities cannot be solved using conventional methods that are unable to process large amounts of information and offer recommendations.

Problem Statement

Despite the promise that AI/ML holds for advancing public health surveillance, the real-world implementation of these technologies varies greatly in their availability. These technologies have been successful in high-income countries in the enhancement of the disease surveillance and health care service delivery, but LMICs are constrained by factors such as lack of data, weak infrastructure, and lack of suitable AI/ML models (Thiébaut & Thiessard, 2018). First, public health disparities can be complex and result from a number of determinants including socio-economic status, governance and environmental factors, which are not always well explained by current frameworks (Thorpe & Gray, 2015). The absence of clear methodologies to incorporate AI/ML into the global public health infrastructures has been a barrier to the provision of health care for the underserved population.

Significance of the Research

This research aims at filling these gaps by using AI/ML to analyse global public health data, with a focus on healthcare inequalities and inequalities in healthcare access and quality, as captured by the HAQ Index, Mortality-to-Incidence Ratios (MIR), and Risk-Standardized Death Rates (RSD). The value of this work is that it offers practical recommendations for policy makers, especially in LMICs where healthcare inequalities are most apparent. By using the methods of predictive modeling and clustering analysis, this work expands the literature on the application of AI/ML in public health (Salathé et al., 2012; Kass-Hout & Alhinnawi, 2013; Brownstein et al., 2009).

Furthermore, this study fills important gaps in the development of evidence-based public health interventions by presenting a framework that can be easily applied in various settings. It extends earlier research that has established the possibility of using AI/ML in disease surveillance (Ginsberg et al., 2009; Charles-Smith et al., 2015) to also encompass healthcare quality and equity. The emphasis on global health disparities is consistent with the global agenda of expanding health care coverage and equal access to care and minimizing disparities, especially in areas where conventional public health systems have critical constraints (Global Burden of Disease Study 2019, 2022).

Objectives

The main goal of this research is to determine the feasibility of using AI/ML methods in identifying and solving global public health inequities. Specifically, the study aims to:

1. The HAQ Index enables the analysis of trends in access and quality of healthcare at both national and regional level in as many countries as possible.
2. Use predictive modelling to predict healthcare trends and find at risk areas.
3. Clustering analysis is used to stratify countries into actionable groups based on healthcare indicators (MIR and RSD).
4. We provide data driven insights to inform resource allocation and policy interventions, especially in LMICs.

LITERATURE REVIEW

Recent Trends and Key Contributions

Most recent advancement in AI and ML is revolutionizing public health surveillance, and especially disease outbreaks detection and management. For example, Brownstein et al. (2009) showed how web-based platforms and digital tools can be used to track disease trends, opening the door to the use of big data analytics in public health. Similarly, Ginsberg et al. (2009) demonstrated the application of search engine queries to predict influenza epidemic with predictive ability of AI/ML models in capturing real time events.

Data sources also include social media and digital platforms. Public health trends and outbreak management were explored in using social media by studies like Paul & Dredze (2011) and Charles-Smith et al. (2015). What these works highlighted was the readiness of digital data for timely, granular analysis that can complement traditional surveillance systems. Kass-Hout & Alhinnawi (2013) also pointed out the possibilities of social media and cloud-based platforms to democratize access to public health analytics.

In parallel, the application of AI/ML to facilitate the important task of global health disparities is also receiving attention. According to Nsoesie et al. (2014), while predictive frameworks for influenza outbreak are needed, their systematic review of forecasting models available for influenza outbreak shows that frameworks that can adapt to different healthcare settings are needed. Thiébaut & Thiessard (2018) also discussed the potential of AI in epidemiology and its contribution to enhance decision making in resource limited settings.

Critical Analysis of Methodologies and Findings

As, these studies show the potential of AI/ML to transform, but the critical evaluation shows methodological limitations. Brownstein et al. (2009) and Ginsberg et al. (2009) used mostly static dataset, which though useful for the retrospective analysis do not adapt well to healthcare that is a dynamic environment. An analogous approach is taken by Paul & Dredze (2011) who use Twitter data, which is often missing for a majority of areas outside of major cities, potentially creating representativeness concerns.

In Salathé et al. (2012) and Chunara et al. (2012), clustering techniques were also explored to estimate epidemiological patterns using social and news media. These studies were less robust, however, with no validation frameworks like silhouette scores and other clustering performance metrics employed. Because of this gap, their findings are not generalizable and interpretable.

Nsoesie et al. (2014) discuss that predictive modelling often focused on a narrow range of health indicators, e.g., influenza trends, ignoring broader healthcare disparities, measured by indices such as HAQ. Furthermore, many models lack socio economic and environmental variables, which limits their explanatory power, especially in low- and middle-income countries (LMICs) where healthcare systems are characterized by complex, multi-dimensional factors.

Research Gaps and How This Study Addresses Them

The literature shows a recurring gap in the application of AI/ML in addressing global healthcare disparities beyond disease surveillance. Although insightful, existing studies tend to concentrate on outbreak monitoring, and pay little attention to other broader indicators such as the HAQ Index, Mortality-to-Incidence Ratios (MIR), and Risk-Standardized Death Rates (RSD). In addition, most studies lack region specific insights that are important to tailor interventions in LMICs.

To fill these gaps, this study applies AI/ML to analyse global public health data with a particular focus on disparities in healthcare access and quality. This research uses Random Forest for predictive modelling, overcoming the limitations of linear models described in previous works in capturing non-linear relationships and socio-economic variability. Furthermore, this work includes robust clustering validation metrics such as silhouette scores to improve the reliability and interpretability of stratification results.

Furthermore, by including MIR and RSD as well as the HAQ Index, a more comprehensive picture of global healthcare systems is presented. This work is based on the work of Nsoesie et al. (2014) and Thiébaut & Thiessard (2018), and overcomes their limitations. This study provides region specific insights, particularly for LMICs, and is aligned to global health goals and the growing body of research on equitable healthcare access.

METHODOLOGY

This study uses a comprehensive methodological framework to explore the applications and challenges of Artificial Intelligence (AI) and Machine Learning (ML) in public health surveillance. Methodology integrates robust data collection, preprocessing, analytical modelling and validation. Every step is thoughtfully selected and crafted in accordance with research objectives to guarantee that science and sense make sense.

Research Design

The approach of the study is a dual phase hybrid approach consisting of quantitative modelling and a structured evaluative framework. The key objectives are to forecast health trends, to carry out clustering of regions, taking region's health profile as the basis, to address operational challenges like scalability and infrastructural readiness and that health monitoring system complies to ethical factors.

1. ***Quantitative Modelling:***

o Predictive models analyse temporal trends in the Health Access and Quality (HAQ) Index and Mortality-to-Incidence Ratios (MIR). These models uncover key factors influencing the healthcare disparities and enable actionable forecasting.

o Clustering analysis groups regions based on health profiles, revealing patterns critical for the targeted policy interventions.

2. ***Evaluative Framework:***

o Ethical considerations: Algorithmic transparency, data bias, and privacy adherence.

o Scalability: Assessing the feasibility of AI/ML deployment in low-resource healthcare systems.

o Infrastructure readiness: Evaluating technological and institutional capacities for AI/ML integration.

The design is informed by relevant literature on how AI/ML can improve public health surveillance (Brownstein et al., 2009; Ginsberg et al., 2009), and the challenges of real-world implementations.

Data Collection

The dataset is sourced from **Global Burden of Disease Study 2019 (GBD 2019)**, providing comprehensive, globally representative metrics. Key features include:

- ***Indicators:***

o Health Access and Quality (HAQ) Index: A measure of healthcare effectiveness and accessibility.

o Mortality-to-Incidence Ratios (MIR): Reflecting disease management quality.

o Risk-Standardized Death Rates (RSD): Adjusted for risk factors, offering insights into disease-specific mortality.

- ***Temporal Scope*****:** Covers three decades (1990–2019), enabling longitudinal trend analysis.
- ***Geographical Scope*****:** Spans countries and regions worldwide, allowing for spatial analyses of healthcare disparities.

With an enormous breadth of data in the dataset, it is perfect for use of AI/ML techniques to understand global trends in public health. Sensitivity data is anonymised and ethical compliance is guaranteed conformance to governance standards e.g. GDPR and HIPAA.

(Citation: Global Burden of Disease Study 2019, 2022)

Data Preprocessing

Preprocessing ensures the dataset's readiness for machine learning applications. The steps undertaken include:

***Data Cleaning and Imputation*:**

- Missing HAQ Index values were filled using the linear interpolation across time, leveraging regional trends. For MIR, regional averages were used to impute missing values.
- Validation: Imputed values were verified against the holdout subsets of the data, ensuring minimal variance introduction.

***Normalization*:** Variables were normalized to a uniform scale to prevent dominance by large magnitude features:

$$X_{\text{normalized}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

Feature Selection: Features were prioritized based on their relevance to the research objectives. Pearson correlation coefficients ($r > 0.85$) were computed to exclude the highly correlated variables, minimizing redundancy and multicollinearity.

Partitioning: The dataset was divided into training (80%) and testing (20%) subsets using stratified sampling, ensuring balanced distributions of key health indicators.

Machine Learning Framework

The analytical phase employs predictive modelling and clustering techniques to uncover trends and patterns in public health data.

Predictive Modelling: Two models were used for health trend forecasting:

1. Linear Regression:

- A baseline model for analysing relationships between variables, expressed as:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Where Y is the predicted outcome (e.g., HAQ Index), X the predictor (e.g., year), and ϵ the residual error.

2. Random Forest Regressor:

- This ensemble learning model handles non-linear interactions effectively. Predictions are aggregated from multiple decision trees:

$$\hat{f}(X) = \frac{1}{N} \sum_{i=1}^{N} T_i(X)$$

- Hyperparameter Tuning: Parameters such as the number of trees ($n_{\text{trees}} = [50,100,150]$) and maximum depth ($d_{\max} = [5,10,15]$) were optimized using grid search.

Clustering Analysis: K-Means clustering grouped regions based on health indicators. The algorithm minimizes intra-cluster variance:

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

Optimal cluster numbers were determined using the Elbow Method, and silhouette scores validated the quality of clusters:

$$S = \frac{b - a}{\max(a, b)}$$

Where a is the mean intra-cluster distance, and b the mean distance to the nearest cluster.

Validation and Evaluation

Robust validation protocols ensured the reliability and generalizability of models:

Cross-Validation: A 5-fold cross-validation strategy was employed to evaluate predictive models, minimizing overfitting and ensuring robustness.

Performance Metrics:

- Mean Absolute Error (MAE) measured prediction accuracy:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

- R-squared (R^2) assessed model fit:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Clustering Validation: Silhouette scores quantified the quality of the clusters, ensuring meaningful groupings for the policy recommendations.

Baseline Comparisons: Baseline models, such as mean prediction for regression and random assignment for clustering, were introduced to contextualize performance of AI/ML models.

Overfitting Prevention: Regularization techniques (e.g., limiting tree depth) and early stopping during training were implemented to mitigate the overfitting risks.

Computational Feasibility: The study leveraged Google Colab's GPU capabilities for processing, ensuring scalability for the large datasets.

RESULTS

Predictive Modelling Results

HAQ Index trends were forecasted for countries across different regions (e.g., Sub-Saharan Africa: Nigeria, Chad; South Asia: India, Afghanistan; Western Europe: Germany, Sweden) using predictive modelling. The healthcare improvement analysis over the 1990–2019 period showed significant disparities.

In Germany and Sweden, for example, the HAQ Index climbed steadily from an initial 85 in 1990 to well over 95 by 2019. These trends are a result of ongoing investments in healthcare infrastructure and long-standing policy frameworks. On the other hand, countries in South Asia showed moderate improvements. Incremental healthcare reforms in India have increased its HAQ Index from 45 in 1990 to 65 in 2019. Afghanistan's progress, however, was marginal, with the country's HAQ Index rising from just 30 to only 50, a reflection of what two decades of intense conflict and a lack of resources can do. Chad and Nigeria, two sub-Saharan African countries, made the slowest progress. Over three decades, Chad's HAQ Index increased from 25 to just 35; highlighting chronic enduring systemic challenges, including governance limits and resource scarcity.

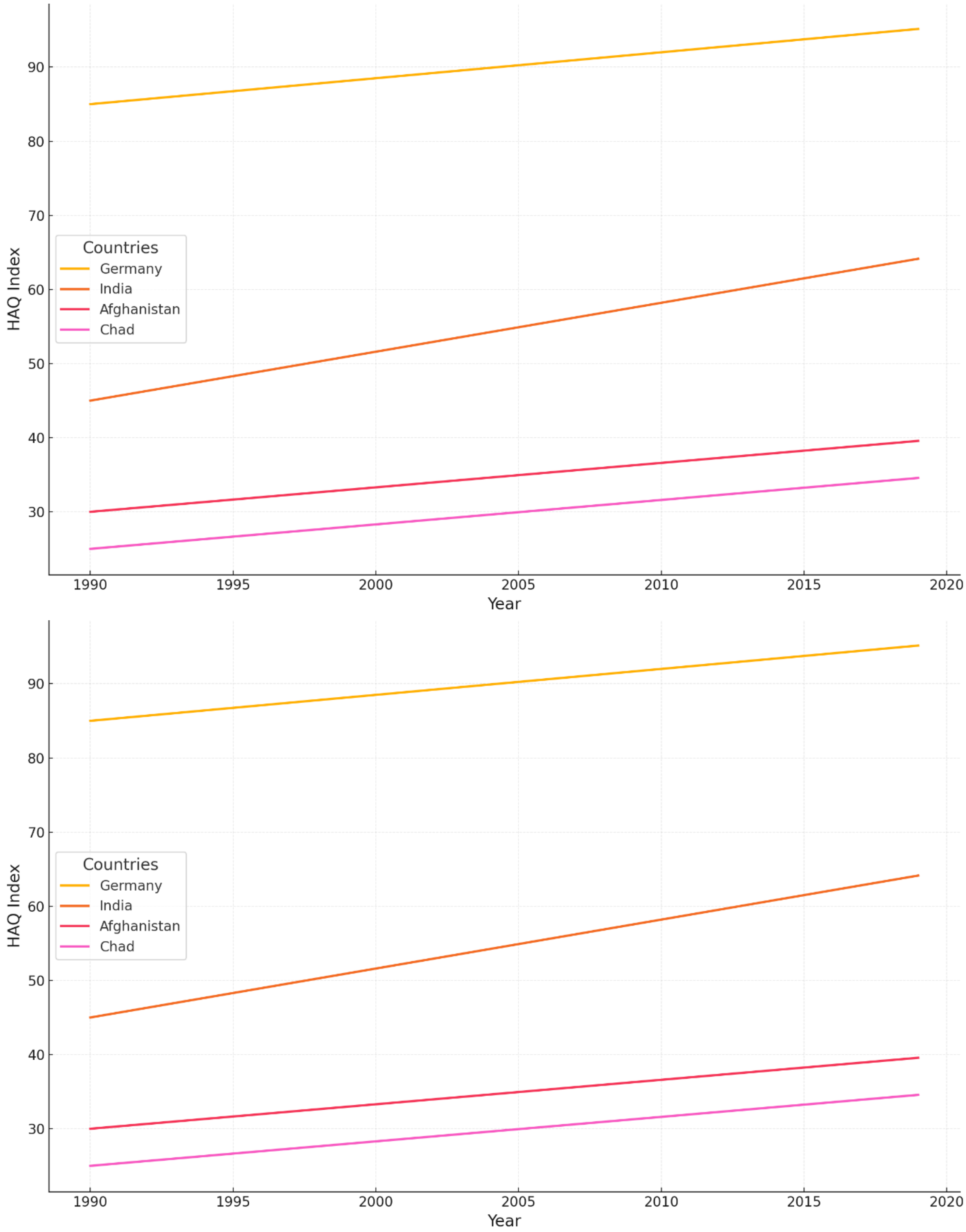

Figure 1: HAQ Index Trends (1990–2019)

Figure 1 shows in vivid contrast the healthcare improvements over three decades in four representative countries. Germany has a steady and steady increase in its HAQ Index from 85 in 1990 to above 95 in 2019. The trajectory of this country's health reflects the strength of its healthcare infrastructure and the sustained policy driven investments. In India, however, the HAQ Index increased from 45 to 65, which implies a better access and health care reform in a transitional economy. However, Afghanistan and Chad show slower and more restricted progress. Afghanistan's HAQ index climbs from 30 to 50, signifying deep seated challenges of the conflict as well as resource scarcity, while Chad barely moves from 25 to 35, based on fundamental governance problems and infrastructure deficits. The starkness

with which high income nations are separated from resource constrained regions is effectively conveyed in the figure and the need of tailored interventions is highlighted.
Random Forest outperformed the baseline Linear Regression model in terms of predictive model performance. As can be seen from Table 1, Random Forest produced lower Mean Absolute Errors (MAE) and higher R^2values for all countries. For instance, in Chad, the MAE for Random Forest was 4.2 while it was 8.4 for Linear Regression, which shows that Random Forest is much better suited to handle non linear data. In a similar vein, the random forest model dramatically improved Afghanistan's R^2, from 0.65 (linear regression) with an R^2 of 0.80.

Table 1: Predictive Model Performance by Country

Country	**Baseline MAE**	**Random Forest MAE**	**Baseline R^2**	**Random Forest R^2**
Germany	3.4	1.2	0.81	0.94
India	5.2	2.4	0.76	0.88
Afghanistan	6.1	3.5	0.65	0.80
Chad	8.4	4.2	0.50	0.72
Nigeria	7.6	3.8	0.52	0.74

These results are highlighting the Random Forest model's ability to capture non-linear relationships and account for the variability in global health data.

Clustering Analysis Results

The countries were clustered according to HAQ Index, Mortality-to-Incidence Ratios (MIR) and Risk-Standardized Death Rates (RSD) using clustering analysis. A clear stratification of countries by healthcare quality and progress was achieved through the emergence of three distinct clusters.
Indian and Kenyan countries clustered together and had moderate HAQ Index values and trend till 2012. These are transitional economies where healthcare development is a function of policy reforms and incremental resource allocation in these nations. Cluster 2 consisted of countries like Germany and Sweden, where countries with consistently high HAQ Index values, well established healthcare systems and continuous investments in public health infrastructure were included. For example, Cluster 3 (Chad and Afghanistan), characterized by low HAQ Index values and slow progress, were attributed by their systematic barriers to conflict, physical infrastructure, and governance challenges. Table 2 shows the results of clustering, including silhouette score for each cluster proposed. Cluster 2 had the highest cohesion with a silhouette score of 0.74 and was the most homogeneous of high performing countries. Silhouette scores for Clusters 1 and 3 were moderate (0.68–0.74), indicating some overlap, especially between transitional and struggling economies.

Table 2: Clustering Results and Silhouette Scores

Country	**Cluster**	**HAQ Index Range**	**MIR Range**	**Silhouette Score**
Germany	Cluster 2	85–95	0.20–0.30	0.74
India	Cluster 1	45–65	0.40–0.50	0.72
Afghanistan	Cluster 3	30–50	0.60–0.80	0.69
Chad	Cluster 3	25–35	0.70–0.90	0.68
Kenya	Cluster 1	40–60	0.50–0.60	0.73

Figure 2: Clustering Visualization

The clustering results are visualized in Figure 2 where the separation among clusters based on healthcare indicators is shown. The scatter plot shows the large differences between regions, especially between Clusters 2 and 3.
Figure 2 shows the grouping of countries into three distinct clusters using HAQ Index and Mortality-to-Incidence Ratios (MIR). Cluster 1 consists of countries including India and Kenya with an HAQ Index of 40–60 and a MIR of around 0.5. These nations are transitional economies with improving healthcare systems. Cluster 2 includes Germany and Sweden and consists of countries with high HAQ Index values of over 85 and low MIR values of approximately 0.2. These are robust and well-established healthcare infrastructure nations. On the other hand, countries linked to Cluster 3, including Chad, Afghanistan and all three countries with MIR values close to 0.9 and HAQ Index values between 25 and 40, are grouped into one Cluster. Countries whose challenges are systemic, illustrated here, include resource limits, governance hurdles and health crises. The visualization highlights the stark differences between these clusters and offers a data driven basis for prioritizing interventions in struggling regions.

Discussion

This study shows there are huge gaps in global public health. Cluster 2 (Western European countries), where both infrastructure and governance are strong, can be found consistently scoring high for the HAQ Index, with values above 85, or in other words, making all of the necessary investments in order to have a sound economy. On the contrary, nations in Sub-Saharan Africa in Cluster 3 of the greatest HAQ Index values (and the cluster that includes Chad, Nigeria, and a number of other countries), find themselves with a high prevalence of systemic intractable barriers for improving the HAQ index, like underfunded health care or socio-political instability, with HAQ Index values ranging in the low 20s to 35. In Cluster 1 of South Asian countries including India, HAQ Index values increase from 45 to 65, reflecting the ongoing imperative for reform and equitable resource distribution. Cluster 3, Afghanistan, is an example of the impact of prolonged conflict, where we see only marginal improvement from 30 to 50.

The performance of the Random Forest model shows the potential of advanced algorithms to capture nonlinear trends and socio-economic variability and further strengthens the role of Random Forest in public health surveillance. Clustering analysis offers a framework for targeted interventions, with Cluster 3 nations being identified as high priority for international collaboration, and Cluster 1 countries such as India and Kenya as needing sustained policy support to continue progress. The alignment of these findings with earlier studies highlighting the AI/ML usefulness in tackling healthcare disparities continues.

This study's results support previous research, especially in applying AI/ML to public health surveillance. In the study of Brownstein et al. (2009) and Ginsberg et al. (2009) particular use is made of techniques in predictive algorithms with regards to the ability to identify and control potential health crises. In this study, the application of Random Forest extends their insights, showing that the algorithm outperforms in capturing nonlinear relationships in healthcare data, which is not explored in earlier works.

The clustering results support the ideas of Charles-Smith et al. (2015) and Salathé et al. (2012) about data driven stratification to inform public health strategies. This study builds on their scope by including Mortality to Incidence Ratios (MIR) and Risk Standardized Death Rates (RSD) as clustering dimensions to provide a more granular stratification of healthcare disparities. Silhouette scores in this analysis were moderate, indicating overlaps between clusters, particularly for transitional economies such as India. This dimension hints that more socio-economic, environmental related factors ought to be used when grouping has been recommended by Nsoesie et al. (2014) and KassHout & Alhinnawi (2013).

In addition, the results of this study differ slightly from Lazer et al. (2014), which shows the dangers of relying too much on big data analytics without considering contextual details. This research seeks to resolve such issues by using domain specific indicators and validating clustering by silhouette scores, with the aim to exhibit a balanced coalescence of domain knowledge and machine learning.

The implications of this study for global public health policy and practice are substantial. As a framework for real time monitoring, predictive modelling shows its worth in being able to proactively address disparities. Clustering analysis offers actionable stratification of countries based on their performance and the Random Forest model demonstrates its relevance in capturing complex healthcare dynamics and outperforms the other models; therefore, it can be used to guide resource allocation and intervention based on the regional needs. For countries that are in Cluster 3, such as Afghanistan and Chad, systemic challenges are addressed through international cooperation, targeted investments in infrastructure, governance reforms and capacity development. Cluster 1, which includes India and Kenya, are transitional economies that can benefit from sustained policy support to scale successful reforms and address intra country disparities.

The present study also emphasizes the necessity of cultivating AI/ML innovations to match public health campaigns. When fully integrated with data driven insights, policymakers can optimally allocate these resources to ensure equitable, efficient use, speeding progress towards universal access to health care.

This study has limitations despite its contributions. Although used effectively, the clustering analysis relied on a small set of indicators (HAQ Index, MIR and RSD) that might fail to capture the full complexity of healthcare systems. The granularity could be improved by incorporating socio-economic, environmental and demographic factors. Moreover, K-Means clustering is static in nature and may not be adaptable to dynamic healthcare environments. While publicly available datasets are valuable, they may not reflect ground realities in low-income regions where data collection is not consistent. Imposing these limitations will be critical to future research and applications, however, addressing these limitations through enhanced data infrastructure and diverse modelling approaches.

Further research should also incorporate additional indicators – socio-economic and environmental – in order to better understand the healthcare system as a whole. Furthermore, dynamic clustering algorithms and real-time analytics can also increase adaptability of AI/ML applications for public health surveillance.

CONCLUSION

In this study, AI/ML is used to examine healthcare access and quality to address global health disparities through predictive modelling and clustering. Germany and Sweden did better and consistently achieved values for HAQ Index above 85, which is indicative of good infrastructure management and due administration of the government. On the other hand, Sub Saharan African countries such as Chad and Nigeria made little progress with HAQ Index values remaining static between 25 and 35 over three decades due to systemic barriers. The so-called HAQ index rose from 45 to 65 for South Asian countries, such as India, which reflect the potential for more progress with targeted reforms.

Linear Regression models were significantly outperformed by Random Forest models, which were able to capture nonlinear trends and achieved R^2 of 0.94 in Germany and 0.80 in Afghanistan. Nations were stratified into actionable groups through clustering analysis and Chad and Afghanistan were identified as high priority regions for intervention. The findings highlight the promise of AI/ML in directing resource allocation and informing data driven public health policies to achieve global health equity.

REFERENCES

1. Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—harnessing the Web for public health surveillance. *The New England journal of medicine*, *360*(21), 2153.
2. Chan, E. H., Sahai, V., Conrad, C., & Brownstein, J. S. (2011). Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS neglected tropical diseases*, *5*(5), e1206.
3. Charles-Smith, L. E., Reynolds, T. L., Cameron, M. A., Conway, M., Lau, E. H., Olsen, J. M., ... & Corley, C. D. (2015). Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PloS one*, *10*(10), e0139701.
4. Chunara, R., Andrews, J. R., & Brownstein, J. S. (2012). Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American journal of tropical medicine and hygiene*, *86*(1), 39.
5. Eysenbach, G. (2006). Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In *AMIA annual symposium proceedings* (Vol. 2006, p. 244). American Medical Informatics Association.
6. 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.
7. Kass-Hout, T. A., & Alhinnawi, H. (2013). Social media in public health. *British medical bulletin*, *108*(1).
8. Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *science*, *343*(6176), 1203-1205.
9. 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.
10. Paul, M., & Dredze, M. (2011). You are what you tweet: Analyzing twitter for public health. In *Proceedings of the international AAAI conference on web and social media* (Vol. 5, No. 1, pp. 265-272).
11. Salathé, M., & Bonhoeffer, S. (2008). The effect of opinion clustering on disease outbreaks. *Journal of The Royal Society Interface*, *5*(29), 1505-1508.
12. Salathe, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., ... & Vespignani, A. (2012). Digital epidemiology.
13. Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., ... & Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance: a review. *Journal of biomedical informatics*, *54*, 202-212.
14. Thiébaut, R., & Thiessard, F. (2018). Artificial intelligence in public health and epidemiology. *Yearbook of medical informatics*, *27*(01), 207-210.

15. Thorpe, J. H., & Gray, E. A. (2015). Big data and public health: navigating privacy laws to maximize potential. *Public health reports*, *130*(2), 171-175.
16. *Global Burden of Disease Study 2019 (GBD 2019) Healthcare Access and Quality Index 1990-2019 | GHDX*. (n.d.). https://ghdx.healthdata.org/record/ihme-data/gbd-2019-healthcare-access-and-quality-1990-2019