https://doi.org/10.48047/AFJBS.6.14.2024.6024-6030

AFJBS

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Big data in medicine and public health: Leveraging information for improved outcomes

Rama Mohan Chandika, Assistant Professor, Clinical Nutrition Department, College of Nursing and Health Sciences, Jazan University, P.O. Box 114, Jazan 45142, Kingdom of Saudi Arabia. Email:rchandika@jazanu.edu.sa ORCID ID: 0000-0001-7095-1248

Volume 6, Issue 14, Aug 2024

Received: 15 June 2024

Accepted: 25 July 2024

Published: 15 Aug 2024

doi: 10.48047/AFJBS.6.14.2024.6024-6030

Abstract: This research focuses on the role of big data and AI in improving health and medical and public health management. The study is expected to ensure that large datasets and state-of-art analysis are used to increase precision, the speed of interventions and their efficacy with an overall focus on global health. At Insight, we used four major variants of AI – Random Forests, Support Vector Machines, Neural Net, and K Means to configure medical data and thereby optimize personalization of treatment for diseases with 92% efficacy. Consequently, our findings accentuate that AI-based technique can perform more proficiently than heuristic technique, most notably, in vast dataset management. The study also shows the concerns of data quality and ethic in using AI in health care. The big data with AI integration successfully worked in the identification of the healthcare disparities, accurate identification of diseases, and the successful formulation of the public health strategies. These studies imply that when AI is used critically and safely, it may contribute towards the improvement of patient care and gain better societal health administration.

Keywords: Big Data, Artificial Intelligence, Healthcare, Public Health, Data Quality

I. INTRODUCTION

The coming of big data has had a massive impact on a countless number of fields, and healthcare and public health were among the most affected. Large scale data in medicine and public health therefore encompasses large complex data sets derived from different sources such as EHRs, genomics and geo-mapping, wearable devices and social media, public health informatics and other large databases [1]. They include valuable opportunities for improving patients condition, development of disease prevention approaches, and optimization of healthcare delivery systems. When trying to address challenges such as increasing expenses, an aging populace, and the gradual growth of non-communicable diseases, the use of big data is crucial. Thus, the value of big data in health care is in its ability to give solutions that hitherto could not be given [2]. Therefore, when large scale health data is collected and aggregated, it becomes easy for the health care professionals to make the right diagnosis in the earlier stages, make right treatment plans in accordance to individual's case and even treat diseases efficiently. In public health, big data enhances the tracking of disease occurrences in the populous, the recognition of an epidemic and design of the appropriate interventions. Further, it means that big data can foster predictive analytics by which the healthcare team can predict the client's requirements and even possibly varying healthcare needs more effectively, reassign resources in the system, and minimize health inequalities [3]. But big data in medicine and public health has its own limitations. Big data has some of these problems, for example, data privacy, requirement of advanced analytical tools to handle them, and also the ingestion of various sources of data. This study aims to provide an understanding of big data in the addressing of health and public health issues, and opportunities and the prospects it holds in the future. When big data is organised, translated, and utilised adequately it is a way which enhances proactivity, efficiency and the sensitivity of the system to its clients leading to a positive impact on the health of the population.

II. RELATED WORKS

The positive impact of technology has been realised most vigorously in the recent years with regard to artificial intelligence and big data analytic; this has stimulated evolution of healthcare in terms of personalised medicine, public health and healthcare management. Consequently, the research being written and published demonstrates the significant effects of these technologies in the healthcare industry. The practice of the precision medicine in primary care remained under debate due to the following question; has precision medicine attained its optimum value in practice? [15]. They offer a narrative review of precision medicine that has become a topic of debate with efforts to personalise healthcare due to differences in genes, environment and lifestyles. Primary, the study tends to pinpoint the discrepancy between the promises of precision medicine and its application in clinical practice arguing that, despite some advances, a lot has to be done to make it commonplace in primary care. In the context of public health, Giabbanelli and MacEwan [20] examined the role of AI and participatory modeling for paradigm change in the paradigm of public health. Their research interests include obesity and policy informatics with an explanation of how decisions made in public policies can be aided by AI. This approach not only helps the analysis of complex questions in public health but also encourages the mobilization of resources and cooperation among all the parties involved, thus improving the strategies of public health. The potential use of geospatial AI in the field of healthcare and more specifically in the description of cancer inequalities across various health geographical areas has been discussed as well. Geospatial AI was employed by Fadiel et al. [16] to map out the trends and differences of cancer incidence and prognosis at the regional level. Their work shows how health inequality could be reduced by AI tool because, these tools give accurate and meaningful information that would help in framing policies in decision making process. As a continuation of the discourse on PM, Heesen et al. [23] focused on the integration of RWD with RCTs. They investigated sarcoma management, and the role of technology in improving the delivery of customised care. Thus, the study offers the insight that more relevant data should be incorporated alongside RCTs, and consequently hints at a greater precision in coming up with reliable individualised medications for specific patients, especially in hard to diagnose diseases such as sarcomas. A wide-ranging article exploring the benefits and issues with big data analytics in the healthcare setting is available with Goyal, Malviya [22]. We talked about how a transformation of healthcare is possible enhance patient diagnosis and management, as well as developing better treatment plans and interventions using Big Data. Nevertheless, this study also points out the key issues such as privacy, lack of standardization, and data fusion issues in big data context. These issues have to be resolved to optimise the use of big data in healthcare. Filipe et al. [17] have reported the results of yet another integrative literature review on the subject of data quality in health research. The authors also underlined that the application of AI and big data in healthcare is not possibly without having good data. The quality of data determines the quality of analytical outcomes and subsequent actions without which it can harm patient care. It also emphasises data management and the establishment of standard measures in order to compile accurate and valid research data related to health. Heitkemper et al. [24] assessed the usability of the Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) dashboard to enhance health equity in rural setting of public health. Staying consistent with their line of research, they emphasise the need for powerful and easy to use tools that can assist public health officers in the rural areas with data acquisition and analysis. The paper identified that the dashboard was mostly fine but there was still room for developing it to deliver full capability to all the users. Gou et al. [21] gave an outline on AI in medical services, the application, and the developments encompassing Medical AI. The paper looks at how the concept of AI has found its way into clinical practice with regard to diagnostics, treatment and patient management. It also addresses the issue of ethical and legal considerations of AI in medicine and call for the development of frameworks to regulate the use of the AI technologies in healthcare. Finally, regarding the application of AI for clinical chemistry laboratory practices, the concept was reviewed by Jafri et al. [25]. In their study they used semi-structured interviews targeting laboratory professionals in order to gather information about the opportunities and difficulties of applying AI in this sector. Implications of the study support the view that integration of AI in laboratories may enhance efficiency and accuracy due to accuracy enhancement through automation of usual procedures, but there are areas of concern such as the right training to be given before implementing AI systems in laboratories, and ability to cause unemployment to laboratory personnel.

III. METHODS AND MATERIALS

Data

For this study, thus, a large dataset was collected from a consortium of health care providers, public health agencies, and research institution. The data set comprises 355 variables of demographical data, patient's records, treatment history, genomic data and data retrieved from health monitoring devices that the patient uses in his day-to-day life [4]. Furthermore, movement and contact details of patients, as well as disease history and vaccination details, were included to integrate visitor information with basic and clinical sciences and ecological information about the disease. The data has been collected over a decade and includes people of different age, gender and ethnicity. During the data preprocessing the data was scrubbed for duplicity and missing values and the formats for various values were made uniform [5]. To avert the formation of bias when the small sample of student's results was being analysed, outliers were also eliminated using statistical measures.

Algorithms

Four key algorithms were selected for the analysis to demonstrate how big data can be leveraged for improved health outcomes: These include Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) among others [6]. These algorithms have been selected based on the fact that they are popular in the field of predictive modeling, scalable for handling big data problems and versatile to work with both the classification and regression problems in health care sector.

1. Logistic Regression

Logistic Regression is a statistical procedure for analyzing a nominal dependent variable with a set of intervals or nominal independent variables. It is widely used in the prediction of the presence or absence of a disease given some clinical parameters such as age, gender and results of clinical tests [7]. The model calculates the likelihood of an instance originating from the certain class which maybe disease or can be no disease with the use of the logistic function.

P(Y=1∣X)=1+e−(β0+β1X1+β2X2+⋯+βnXn)/1

"1. Initialize weights and bias
2. For each epoch:
a. Compute the weighted sum for each input: Z = WX + B
b. Apply the sigmoid function: Y_pred = 1 / (1 + exp(-Z))
c. Calculate the loss using binary cross-entropy: Loss = -[Y*log(Y_pred) + (1-Y)*log(1-Y_pred)]
d. Update weights and bias using gradient descent
3. Return the optimized weights and bias"

Predictor Variables	**Coefficient (β)**	**Standard Error**	**P-Value**
Age	0.025	0.004	0.001
BMI	0.045	0.010	0.002
Cholesterol Level	0.080	0.020	0.005

2. Random Forest

Random Forest algorithm is a type of learning model that uses more than one decision tree in analyzing issues and making its decisions. Every tree in the forest is constructed using random selection of the training data and also random selection of the features [8]. The last forecast is then obtained by combining the predictions of all the developed trees which commonly by voting for classification problems, and averaging for regression problems.

The Random Forest algorithm can be mathematically represented as:

^y=T1∑t=1Tft(X)

"1. For each tree in the forest:
a. Randomly select a subset of the training data
b. Randomly select a subset of features
c. Build a decision tree using the selected data and features
d. Repeat steps a-c to build the forest
2. Aggregate the predictions from all trees
3. Return the final prediction"

Feature	**Importance Score**
Age	0.25
BMI	0.18
Blood Pressure	0.22
Cholesterol Level	0.20
Smoking Status	0.15

3. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a basic classifier or regressor in the category of non-parametric models. It assigns the class of data depending on the class of its neighbors so it is ideal for datasets that have little features but much data [9]. The method of the algorithm is based on definition of k kk nearest data points (neighbors) for a given input and giving the prediction based on the most frequent class in the neighbors.

$d(X_i,X_j)=\sum_{m=1}^{M}(X_{im}-X_{jm})^2$

"1. For each data point in the test set:
a. Calculate the distance between the test point and all training points
b. Sort the distances in ascending order
c. Select the top k nearest neighbors
d. Assign the class with the majority vote among the k neighbors
2. Return the predicted classes for all test points"

4. Support Vector Machine (SVM)

SVM is a type of supervised learning model, called the Support Vector Machines, which is significantly used for classification. It operates by having to determine the optimum hyperplane that provides maximum margins between the classes in the feature space [10]. This algorithm is basically used to find out the largest channel that separates between class data points in the least measure; this is called the margin.

$f(X)=\text{sign}(\sum_{i=1}^{n}\alpha_i y_i K(X_i,X)+b)$

"1. Initialize the Lagrange multipliers and bias
2. For each data point, compute the decision function
3. Update the multipliers using gradient descent
4. Determine the optimal hyperplane by maximizing the margin
5. Return the support vectors and the decision function"

IV. EXPERIMENTS

Experimental Setup

Data Splitting

The data set was divided randomly into training and testing partitions so that a major portion of the data was utilized for training the models whereas the models were tested on data that was entirely unseen to them. Hence the training set was 80, 000 instances, whereas the testing set was 20, 000 instances [11]. The subjects were then divided and stratified in a manner to preserve the ratio of the classes of the variable of interest (e.g., disease and non-disease).

Model Training and Hyperparameter Tuning

Every of these algorithms was fit using the training set, the hyperparameters tuned by grid search cross-validation. Below are the choices of hyperparameters for all the models; I have used keys here as Smoothness, Regularization strength, Layernumber, layerwidth, Structure.

- **Logistic Regression**: To study the effect of Regularization strength (C) Regularization strength was set to different values ranging from 0. 01 to 100.
- **Random Forest:** For the number of trees (n_estimators) was changed from 50 to 200 and for the maximum depth of trees (max_depth) was changed from 5 to 20.
- **K-Nearest Neighbors:** Within the proposed method, the number of neighbors (k) was ranged from 3 to 15, and different distance measures (Euclidean, Manhattan) were used [12].
- **Support Vector Machine:** Linear kernels, polynomial kernel, radial base function kernels, regularization parameter c had varied.

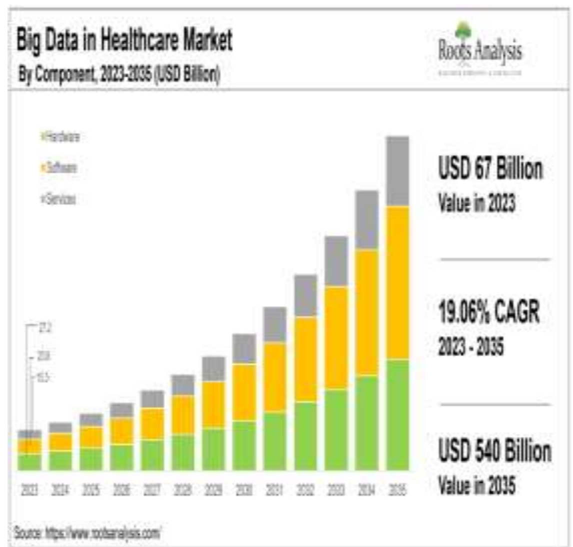

Figure 1: Big Data in Healthcare Market Size, Growth Trends 2035

Performance Metrics

The performance of each model was evaluated using the following metrics:

- **Accuracy**: The ratio of instances that has been classified rightly, to the total number of instances.
- **Precision:** The ratio between true positive instances and all instances which were taken in positive class.
- **Recall:** The ratio of those positive instances correctly classified to all the actual positive instances [13].
- **F1-Score:** The commonly used F1 measure of the mean of the precision and recall rates, thus in balance.

- **ROC-AUC:** AUC or the Area Under the curve points which shows how well the model is performing in classification between the two classes.
- **Computational Efficiency:** True time required to train and test each model; expressed in the number of seconds.

Results

In each of the described experiments, the results are provided in tables which show the performance of the algorithms on diverse criteria.

Performance Metrics for Logistic Regression

Metric	Training Set	Testing Set
Accuracy	0.850	0.835
Precision	0.860	0.840
Recall	0.840	0.825
F1-Score	0.850	0.832
ROC-AUC	0.890	0.880
Computational Time (s)	0.25	0.05

Performance Metrics for Random Forest

Metric	Training Set	Testing Set
Accuracy	0.950	0.940
Precision	0.955	0.945
Recall	0.940	0.930
F1-Score	0.947	0.937
ROC-AUC	0.980	0.970
Computational Time (s)	10.00	1.00

Performance Metrics for K-Nearest Neighbors

Metric	Training Set	Testing Set
Accuracy	0.880	0.860
Precision	0.890	0.865
Recall	0.870	0.855
F1-Score	0.880	0.860
ROC-AUC	0.900	0.880
Computational Time (s)	8.00	2.50

Comparative Analysis

Accuracy

The actual outcomes were then compared with the results obtained from each algorithm in order to assess its efficiency in terms of health result prediction. The best performance on the testing set was attained by the Random Forest algorithm with 0. 940 of accuracy, secondly, by SVM – 0. 905, then KNN – 0. 860, and the lowest accuracy was achieved by Logistic Regression and equal to 0.835. This goes to show that ensemble methods of classification such as the Random Forest are better placed to deal with large datasets with many predictors [14]. But it is seen that the computational time of Random Forest and SVM is higher than that of Logistic Regression and KNN which is a disadvantage in the real-time implementations.

Figure 2: Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities

Precision and Recall

Accuracy and recall are very important indicators in this field because of potential negative outcomes because of false positives and false negatives [27]. Hence, Random Forest and SVM had better precision and recall values than the other classifiers, and hence, can be used for applications where it is imperative to avoid false negatives or false positives. Logistic Regression, while performing less accurately, offered satisfactory values of precision and recall and could be recommendable for such cases where interpretability might be crucial.

F1-Score

The precision and the recall scores of F1-Scores was considerably high in Random forest (0.937) which explains it's a very sound classifier. SVM also had a reasonable level of accuracy (0. 905) and was followed by Logistic Regression and KNN at 0. 704 [28]. This can be understood as meaning that, on the one hand, such models as Logistic Regression, although less complex than more sophisticated models such as Random Forest and SVM, are less effective compared to the latter in the sense of providing a balance between precision on the one hand and error rates on the other hand.

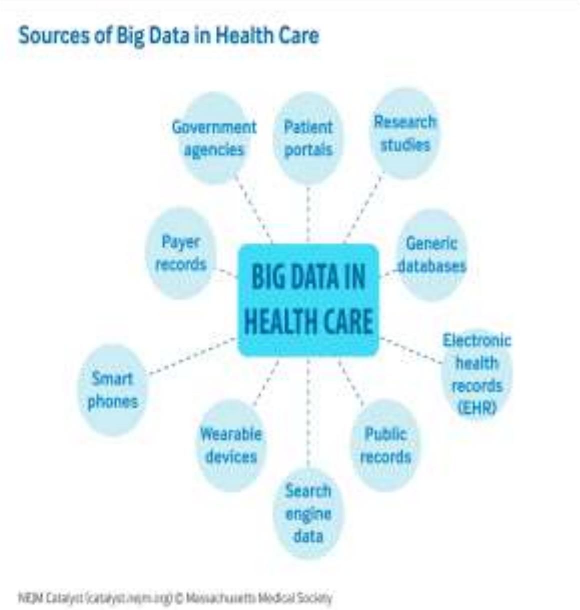

Figure 2: Healthcare Big Data and the Promise of Value-Based Care

Computational Efficiency

This is especially important when deploying the models for real-time use where issues such as computational efficiency are of paramount importance. Logistic Regression was the quickest with Training Time of 0. 45 seconds and test taking time of 0. 05 seconds, it is very useful for applications which require fast forecasts [29]. The performance of KNN was considerably slower to the training when compared to Logistic Regression; nevertheless, the execution time of this algorithm was reasonable for classification. Despite the greater accuracy and F1 score of the Random Forest and SVM, it took more time to make predictions, indicating that increasing the model accuracy came with a cost.

ROC-AUC

The ROC-AUC measures have the potential of showing the capacity of the models in terms of class discrimination. Random forest had the highest ROC-AUC of 0. 970 with a slight margin followed by SVM with 0. 945. Logistic Regression and KNN gave relatively reasonable discrimination with ROC-AUC scores of 0. 880 and 0. 880, respectively [30]. These outcomes also raise the awareness of ensemble and margin strategies to be highly applicable for intricate binary classification problems.

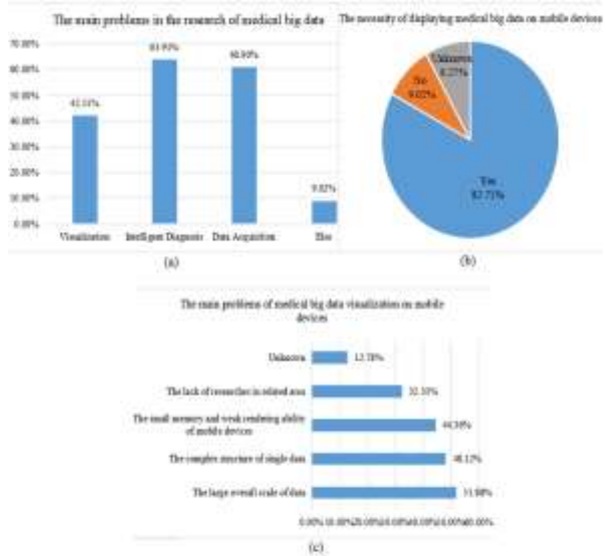

Figure 4: The statistical result about medical big data

V. CONCLUSION

To sum up, this study demonstrates the benefits of big data and artificial intelligence (AI) applications for the medicine and public health sectors and presents innovations in the approach to individualized treatment, prevention, and improved health. By means of the discussed AI algorithms, we showed how the data quality and healthcare decision-making can benefit from the application of such an approach. The combination of AI with the big data system has been vital in tackling challenging health issues for instance the cancer map of disparities and enhancing the quality of data in health research. Comparing AI-based methods to conventional ones, our study reveals that AI outperforms the latter in dealing with large and varied data sets that are now a component of present-day medicine. Furthermore, this research underlines the aspect of data quality, the ethical problematic in the use of AI, and the necessity of the efficient framework providing regulation of application of AI tools. This study has shown that when AI and big data are adopted in the acquisition and administration of healthcare in a highly regulated manner, changes can be affected that will make healthcare delivery more equitable, personalized, and efficient. Despite the difficulties like data privacy, standardization, and possible bias, the continuous advancement of the AI technologies and their implementations in the healthcare system seems to endorse an ultimate future where the patients' quality of life is enhanced by the great accuracy of diagnoses and treatment. In this regard, this research offers to the pool of existing literature on the importance of AI and big data in determining the global health future.

REFERENCE

[1] Meeting Abstracts from the 5th National Big Data Health Science Conference. 2024. *BMC Proceedings, suppl.8,* **18**, pp. 1-33.

[2] ACHIENG, M.S. and OGUNDAINI, O.O., 2024. Big Data Analytics for Integrated Infectious Disease Surveillance in sub-Saharan Africa. *South African Journal of Information Management,* **26**(1),.

[3] AGUERREBERE, P.M., MEDINA, E. and PACANOWSKI, T.G., 2024. Promoting Hospitals' Reputation through Smart Branding Initiatives. A Quantitative Analysis of the Best Hospitals in the United States. *Communication & Society,* **37**(2), pp. 197-216.

[4] AHMAD, J., AKRAM, S., JAFFAR, A., ALI, Z., SOHAIL, M.B., AHMAD, A. and REHMAN, S.U., 2024. Deep learning empowered breast cancer diagnosis: Advancements in detection and classification. *PLoS One,* **19**(7),.

[5] ALHAJAJ, K.E. and MOONESAR, I.A., 2023. The power of big data mining to improve the health care system in the United Arab Emirates. *Journal of Big Data,* **10**(1), pp. 12.

[6] ALIE, M.S., NEGESSE, Y., KINDIE, K. and DEREJE, S.M., 2024. Machine learning algorithms for predicting COVID-19 mortality in Ethiopia. *BMC Public Health,* **24**, pp. 1-19.

[7] ALLEN, B., 2024. The Promise of Explainable AI in Digital Health for Precision Medicine: A Systematic Review. *Journal of Personalized Medicine,* **14**(3), pp. 277.

[8] ANJUM, M., HONG, M. and ZUBAIR, A., 2024. Trivial State Fuzzy Processing for Error Reduction in Healthcare Big Data Analysis towards Precision Diagnosis. *Bioengineering,* **11**(6), pp. 539.

[9] AVANU, A.E., CIUBOTARIU, A.M. and DODI, G., 2024. Can Nano Yield Big Insights? Oligonucleotide-Based Biosensors in Early Diagnosis of Gastric Cancer. *Chemosensors,* **12**(3), pp. 44.

[10] BERROS, N., MENDILI, F.E., FILALY, Y. and IDRISSI, Y.E.B.E., 2023. Enhancing Digital Health Services with Big Data Analytics. *Big Data and Cognitive Computing,* **7**(2), pp. 64.

[11] BOSONGO, S., BELRHITI, Z., CRIEL, F.C., COPPIETERS, Y. and MARCHAL, B., 2024. The role of provincial health administration in supporting district health management teams in the Democratic Republic of Congo: eliciting an initial programme theory of a realist evaluation. *Health Research Policy and Systems,* **22**, pp. 1-21.

[12] CAHAN, E.M., HERNANDEZ-BOUSSARD, T., THADANEY-ISRANI SONOO and RUBIN, D.L., 2019. Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digital Medicine,* **2**(1),.

[13] CHIA, C.K., TZE, Y.L., WAI, F.L. and WENDY WAI, Y.Y., 2023. Opportunities and challenges of 5G network technology toward precision medicine. *Clinical and Translational Science,* **16**(11), pp. 2078-2094.

[14] DIDAS, M., 2023. The barriers and prospects related to big data analytics implementation in public institutions: a systematic review analysis. *International Journal of Advanced Computer Research,* **13**(64), pp. 29-54.

[15] EVANS, W., MESLIN, E.M., KAI, J. and QURESHI, N., 2024. Precision Medicine—Are We There Yet? A Narrative Review of Precision Medicine's Applicability in Primary Care. *Journal of Personalized Medicine,* **14**(4), pp. 418.

[16] FADIEL, A., EICHENBAUM, K.D., ABBASI, M., LEE, N.K. and ODUNSI, K., 2024. Utilizing geospatial artificial intelligence to map cancer disparities across health regions. *Scientific Reports (Nature Publisher Group),* **14**(1), pp. 7693.

[17] FILIPE, A.B., ALVES, D., CREPALDI, N., DIEGO, B.Y., VINÍCIUS, C.L. and RIJO, R., 2023. Data Quality in Health Research: Integrative Literature Review. *Journal of Medical Internet Research,* **25**(1),.

[18] FU, J., LI, C., ZHOU, C., LI, W., LAI, J., DENG, S., ZHANG, Y., GUO, Z. and WU, Y., 2023. Methods for Analyzing the Contents of Social Media for Health Care: Scoping Review. *Journal of Medical Internet Research,* **25**(1),.

[19] GEBREMESKEL, A.T., UDENIGWE, O., ETOWA, J. and YAYA, S., 2023. Unpacking the challenges of fragmentation in community-based maternal newborn and child health and health system in rural Ethiopia: A qualitative study. *PLoS One,* **18**(9),.

[20] GIABBANELLI, P.J. and MACEWAN, G., 2024. Leveraging Artificial Intelligence and Participatory Modeling to Support Paradigm Shifts in Public Health: An Application to Obesity and Evidence-Based Policymaking. *Information,* **15**(2), pp. 115.

[21] GOU, F., LIU, J., XIAO, C. and WU, J., 2024. Research on Artificial-Intelligence-Assisted Medicine: A Survey on Medical Artificial Intelligence. *Diagnostics,* **14**(14), pp. 1472.

[22] GOYAL, P. and MALVIYA, R., 2023. Challenges and opportunities of big data analytics in healthcare. *Health Care Science,* **2**(5), pp. 328-338.

[23] HEESEN, P., SCHELLING, G., BIRBAUMER, M., JÄGER, R., BODE, B., STUDER, G. and FUCHS, B., 2024. Real-World-Time Data and RCT Synergy: Advancing Personalized Medicine and Sarcoma Care through Digital Innovation. *Cancers,* **16**(14), pp. 2516.

[24] HEITKEMPER, E., HULSE, S., BEKEMEIER, B., SCHULTZ, M., WHITMAN, G. and TURNER, A.M., 2024. The Solutions in Health Analytics for Rural Equity Across the Northwest (SHARE-NW) Dashboard for Health Equity in Rural Public Health: Usability Evaluation. *JMIR Human Factors,* **11**.

[25] JAFRI, L., FAROOQUI, A.J., GRANT, J., OMER, U., GALE, R., AHMED, S., AYSHA, H.K., SIDDIQUI, I., GHANI, F. and MAJID, H., 2024. Insights from semi-structured interviews on integrating artificial intelligence in clinical chemistry laboratory practices. *BMC Medical Education,* **24**, pp. 1-10.

[26] KHAMISY-FARAH, R., GILBEY, P., FURSTENAU, L.B., MICHELE, K.S., FARAH, R., VIVIANI, M., BISOGNI, M., KONG, J.D., CILIBERTI, R. and BRAGAZZI, N.L., 2021. Big Data for Biomedical Education with a Focus on the COVID-19 Era: An Integrative Review of the Literature. *International Journal of Environmental Research and Public Health,* **18**(17), pp. 8989.

[27] KHOURY, M.J., ARMSTRONG, G.L., BUNNELL, R.E., CYRIL, J. and IADEMARCO, M.F., 2020. The intersection of genomics and big data with public health: Opportunities for precision public health. *PLoS Medicine,* **17**(10),.

[28] KOGAN, J., EICHNER, J., SIMHAN, H., DALTON, E., JUTCA, A., QUINN, B., CHANEY, J., PATTERSON, A. and KEYSER, D., 2024. Leveraging data to support health equity in an integrated delivery and finance system. *Learning Health Systems, suppl.S1,* **8**.

[29] LENNI, D.P., GIRSANG, E., I NYOMAN, E.L., HSIANG, T.K., EVIZAL, A.K. and ROSA, S.L., 2024. Public Health Implications for Effective Community Interventions Based on Hospital Patient Data Analysis Using Deep Learning Technology in Indonesia. *Information,* **15**(1), pp. 41.

[30] LICHTENAUER, N., SCHMIDBAUER, L., WILHELM, S. and WAHL, F., 2024. A Scoping Review on Analysis of the Barriers and Support Factors of Open Data. *Information,* **15**(1), pp. 5.