

Harnessing Explainable AI for Enhanced Decision-Making in Healthcare: Bridging the Gap Between Clinical Expertise and Machine Learning Models

ABSTRACT

There is a favorable forecast on the possibilities of using Artificial Intelligence (AI) in medicine. It can assist in medical diagnosis, planning of treatment and ensure better patient care. Nonetheless, the earlier effusive praise of the potential of AI in clinical practice is tempered by concerns of the black box nature of many algorithms, particularly those that are deep learning models. This paper considers the use of Explainable AI (XAI) in addressing the dilemma facing clinicians who want to act on the well-known facts of AI with the help of machine learning algorithms. Based on the qualitative research methodology, with emphasis on literature review, we explore the requirements, advantages, and barriers associated with the use of XAI in medicine attending to improve the delivery of AI in medicine, evaluate the usability and safety of such innovations as well as generalize about the reasons behind the introduction of these technological novelties. XAI can offer significant enhancement towards trust of Artificial Intelligence systems for patients and their guardians, make relaxed accountability obligations towards healthcare practitioners, and improve the fairness in health-related services provided to patients. Nonetheless the other requirement of balancing the complexity of the model and explainability demands great focus alongside adherence to the regulations. Finally, the study emphasizes that XAI can propel improvement in the quality of tertiary medical care of patients, however, for practical implementation of this type of technologies into healthcare systems much more translational research shall be applied into development.

Introduction

Background

Effective integration of Artificial Intelligence (AI), particularly its application to machine learning (ML), as far as the health care domain is concerned, makes for an interesting area, which is actively developing. Such technologies are used in such areas of health care as medical imaging, prediction, and even treatment customization. Most importantly, AI has the most

promising aspect wherein it enhances the quality of medical services delivered through precision and increased access. As an example, Saraswat et al. (2022) note that ML applications assist in development of highly accurate prediction models, automate analysis of medical images and personal treatment strategies. In the same line of thoughts, Castiglioni et al. (2021) claimed that AI helps to increase the accuracy of the diagnosis mainly in the area of medical imaging and helps in faster diagnosis and earlier detection of a certain disease.

However, these achievements have not reached their full potential as there are problems such as the ‘black box’ factor associated with many AI models, and deep learning in particular. The overriding concern is the situation where health professionals find it hard to embrace such systems because they function as “black boxes” or are tools that they do not comprehend or have confidence in. As Jeyaraman et al. (2023) assert, introduce that explicitness must be a component of any AI system that is to be applied in medicine if acceptance is to be achieved. Furthermore, Ahmed et al. (2022) say that irrelevant AI makes opaque decisions and such a level of opacity hinders the assimilation of AI technologies into clinical practice. In response, however, an understandable inclination has been to object to the informal nature of conventional AI and advocate for Explainable AI (XAI) instead. According to Yang et al. (2021), XAI provides a solution to this challenge through transparent and interpretable presentation of AI, enabling medical professionals to make use of AI in clinical practice.

Objectives of the Study

This study is dedicated to the exploration of the application of XAI in a healthcare context focusing on how such systems may help further collaborative approaches by making the analysis conducted by an AI jointly with the domain expertise of healthcare providers. The specific objectives are:

To examine the factors that affect the explainability of AI-driven insights in health

To evaluate the extent of XAI inclusion and application in the medical diagnosis process.

To highlight the implementation gaps in the deployment of XAI in clinical practice and opportunities like managing complexity in explainable artificial intelligence.

To seek for XAI's "education" effect when applied to AI and potential as well as risks that can be encountered in the context of biased models.

To consider how organizations implementing XAI systems can guarantee compliance with norms and standards relevant to organizational transparency and ethical practices, in the context of the application of artificial intelligence in healthcare.

Current AI Applications in Healthcare

Healthcare is being transformed completely with the utilization of Artificial Intelligence and Machine Learning in treatments, diagnosis, and patient management. Such technologies have taken over the conventional ones by promoting accuracy, speed, and customization. In the case of AI, the most predominant usage is in medical diagnostics. As Rudra Tiwari (2023) observes, efficient algorithms based on deep learning, for example, are good at recognizing distortion of an abnormal image within a mammogram with a probable tumor as well as the early features of diabetic retinopathy from retinal pictures. This is because such applications leverage AI's capability to train on large datasets and improve on them over time. While this application is largely for use in diagnostics, AI also has applications in predictive analytics. Here, according to Muhammad Asif Naveed (2023), AI evaluates data obtained from electronic clinical records and comes up with targeted treatment strategies to enhance predictive accuracy. Furthermore, smart wearables and other remote monitoring systems that utilize AI technology help monitor the health status of patients remotely, as reviewed by Vaishnavi Mishra (2023).

Challenges with AI in Healthcare

Even with these improvements, the use of AI is still limited in the field of healthcare. To begin with, the first hurdle is the clear lack of understanding of the algorithms used, predominately in the so-called "black box" deep learning systems. The problem of transparency, as pointed out by the Date (2023), poses a great challenge and a reason why most doctors will back out in embracing and using AI LeCun. Trust issues are even worse to patients regarding ethical issues such as privacy of health records and risks of injustice arising from biased algorithms, what A. Bhardwaj (2022) suggests must be curbed to avoid health inequity bias. Lastly, the combination of AI and the current healthcare systems has shown to be extremely challenging especially in developing countries where the whole process of integrating and utilizing AI is too expensive and complex.

Advancements in Explainable AI

The goal of this task is to rewrite the text, however, without changing its meaning, making the text of a higher quality, adding synonyms, and using complex and uncomplicated language simultaneously and in an interesting way. Increasing the explanatory potential of AI has been one of the goals of facilitating XAI detail. As a result of this monstrous task, scientists came up with technology to explain how AI works. Recently, Javad Iqbal (2023) shows how these strategies are also being implemented in clinical decision-support systems with the aim of understanding the underpinning factors of a recommendation made through an AI system. Such transparency promotes appropriate application of AI-assisted decision making in complement with clinical expertise and judgement and builds up the desire to trust AI systems in health care. On the other hand, as observed by Z. Angehrn (2020), strategies such as attention and feature importance score techniques are making complex models more interpretable thus, boosting trust by clinicians on AI diagnoses and recommendations. In addition, as Farida Sabry (2022) points out, AI systems are being designed with better user interfaces to make such information comprehensible and user-friendly to clinician within short time for timely decisions.

The role of AI in healthcare is critical but so far it is still limited because of trust, interpretability, ethical and such issues. The recent progress made in the field of XAI is of significant importance towards addressing these problems which in turn offers a potential landscape whereby AI systems are not just effective in performance but also effective in a very understandable manner. Further, it can be noted that with the passage of time the technologies which fall within the category of AI have the capability to reorganize and readjust the whole healthcare system in an optimal way which offers healthcare that is tailored kneadle atthe patients' ages.

Methodology

The use of artificial intelligence (AI) and machine learning (ML) technology in healthcare has led to a dramatic improvement in clinical decision-making and in the way patients are handled. These technologies, however, operate on specific principles that enhance their efficiency and enable their usage in an acceptable way. The introduction and advancement of Explainable AI (XAI) are also important to consider because it aims to promote the understanding of the AI models in healthcare systems.

MDI and ML tools available in clinical practice aim at sifting through enormous amounts of data, searching for correlation and predicting outcomes which in turn assists the medical

personnel. Mandala (2023) avers that the incorporation of AI in healthcare provisions entails respecting certain principles like accuracy, interpretability and fairness, so as to win the heart of the providers and the patients. It is important for AI models to be trained on high-quality representative datasets so that reliable predictions can be made. Thalpage (2023) appreciates that unlike many undertakings within the discipline of medicine, bias is a huge risk in the field of diagnostic imaging as it is often the case that AI models must be trained on a variety of datasets.

Transparent AI decision processes, which are barely feasible because of black-boxing nature of many AI applications employed in that sphere, are a must have in medicine (Dwivedi et al., 2022). This is crucial especially in instances where decisions have dire effects, including life and death outcomes. Equity is also a fundamental principle for instance because the design of AI systems must be such that all groups of patients are treated without discrimination. Guleria et al. (2022) point out that accuracy isn't everything. Development of fair ensemble classifiers in addition to high accuracy needs to be addressed to all the patients given fair treatment.

That said, the scope of application of these principles is commendable. The spread of AI incorporating the explanation from Antoniadis et al. (2021), where this model provides data that can assist in making the final decision exclusively based on AI models is critical for enhancing clinical decisions in healthcare. Transparency as such can also help mitigate the fear for the systems and the recommendations made. They argue that in medical imaging, XAI solutions extend the scope of clinicians by explaining how AI-based diagnoses are made van der Velden et al., 2021.

However, the integration of XAI in healthcare also has its challenges. Sadeghi et al. (2023) observe that it is challenging to find a trade-off between interpretability of model results and complexity of the models because operating with simple models can deprive the study of accuracy present in complex ones. Another challenge is ensuring that the XAI systems being used are ethical, as it is their bias as systems that can impact the health of the patient. McDermid et al. (2021) support the need to be able to justify all actions taken and decisions made in the context of use of systems incorporating AI.

To overcome these concerns and improve the practice of healthcare, a number of introduction proper solutions have been put forward. The Consolidated Framework for Implementation Research (CFIR) provides a model that organizes domains that affect or promote the use of new practices or interventions in health care systems (Ilott et al., 2012). The Learning-Adapting-Leveling (LAL) model describes the development of individualized therapies as involving repetitive trial-and-error learning and ongoing modification (Lal et al., 2013). The development of the SEIPS 2.0 framework adds new concepts of configuration, engagement, and adaptation to the study of the activity of healthcare providers (Holden et al., 2013).

The implementation of such theoretical constructs in clinical settings is important with regard to improving the quality of decisions. The same applies to FNCL models in Healthcare described by Harry et al. (2020): as it paints the picture of healthcare practices, it creates an easier path to adoption of devices like CDSs which are embedded into attendants' day to day activities. At the same time, The Normalization Process Theory (NPT) examines the processes of work that are undertaken in order to achieve the incorporation of innovative interventions into the usual clinical flow (May et al., 2018).

Notwithstanding the promised advantages integrating theoretical frameworks to practice has some hurdles such as opposition from the clinicians, absence of clinical variations, and the lack of demonstrable clinical effectiveness (Kyrimi et al., 2021). In order to do away with these challenges, it is important to incorporate clinicians into the design and evaluation processes of such technologies and to prove their impact clinically by way of every possible evidence bunch.

The Theoretical Domains Framework (TDF) helps in identifying barriers to integration at the individual, organizational, and other levels and addresses them in a systematic manner (Waddell et al., 2021). Using TDF also provides, for instance, health care organizations with the ability to plan effective means for addressing barriers and ensuring proper utilization of new innovations.

These adoption of these theoretical frameworks are expected to change the way healthcare professionals make clinical decisions, draft policies and even train. Dimopoulos-Bick et al. (2019) propose that such frameworks may prevent overuse by imposing restricted care variability

management. They can also aid in the development of appropriate training reforms for health workforce using the specific strategies (Lipworth et al., 2013).

Finally, the contribution of theoretical frameworks to the creation of new technologies for the health care sector should also be highlighted. They further strengthen the concepts of how to design and deploy clinical decision support systems (CDSS) to ensure that new technologies fit better into the clinical processes and users' practice (Laka et al., 2020). This alignment is particularly important in seeking the uptake of new technologies and how they would work towards improving patient outcomes.

The interaction of AI and ML with healthcare practices when accuracy, interpretability, and fairness are put at the core would enhance patients' health conditions. The forms associated with the theoretical frameworks of understanding how these structures can be address inefficiencies in practice and help these technologies to be used effectively within the healthcare environment.

Implementing Explainable AI in Healthcare

The evolution of Artificial Intelligence-based systems has highlighted the need for Interpretability of AI models, especially in fields such as healthcare, where patients' well-being depends on complex poorly understood models.

Case Studies:

XAI has been put into good use within the medical field as illustrated in numerous case studies. Such tools include Local Interpretable Model-Agnostic Explanations (LIME) and Anchors, which have been used for breast cancer diagnosis. These provide an understanding of the reason behind the AI decision in a manner that is understandable for both the doctor and the patient. Larasati (2022) points out that even ordinary level individuals without technical knowledge can comprehend how the AI functioned in these scenarios. This sort of explainability is very important in fields such as oncology, where the incorporation of AI recommendations could dramatically change the course of treatment. The application of explainable AI extends to other domains like Mental Health and Cardiology towards ensuring that clinicians make responsible

decisions. Nilsen et al (2022) argues the XAI tools may enable mental health practitioners to utilize more AI tools with less apprehension.

Integration Strategies:

There is a need to devise a strategy for incorporating XAI into healthcare. One of the approaches to accomplishing this is called federated learning, which enables AI models to train on information kept in different locations without needing to reveal any patient's personal details. Saraswat et al. (2022) argue that the use of XAI tools such as SHAP and LIME alongside Federated learning can support the real-time observation and interpretation of AI applications inside clinics. Moreover, Alam et al. (2023) prioritize the basic premise that AI systems should make use of publicly available and visible data in order to gain the confidence of the clinicians and the patients. In this way, all participants, including the patients, appreciate and accept the aura of the decisions made by the AI.

Challenges and Barriers:

Despite its promise, the biggest concern is the barriers to comprehensive integration of Healthcare XAI. One major challenge concerns finding the sweet spot between accuracy of models and their interpretability. Chomutare et al. (2022) describe situations when accuracy of created models is of absolute importance: at its best - such models act as "black boxes" – and nobody knows what happens inside it. On the other hand less complicated models may be too coarse which is not good since this is healthcare. One more problem is advocating for generalizability of the explanatory models beyond the particular healthcare institution since patients, medical practice and data may all be different. Another important reason is the lack of standard approaches for XAI that would enable the integration of these models into the existing healthcare information technology infrastructure as pointed out by Srinivasu et al. (2022).

Ethical and Legal Considerations:

Use of XAI tools in healthcare comes with ethical considerations. Patients should always be protected while the decisions made by AI systems are rendered to be traceable. Jeyaraman et al.

(2023) emphasize that these values should not be overlooked as it would put the future of AI powered healthcare into jeopardy.

From a legal perspective, there are still some significant issues though due to the constantly changing parameters concerning regulation in the field of AI technologies. Owing to the rapid growth of AI, it is quite difficult to come up with existing healthcare regulations, a situation which leads to a gap. Matulionyte et al. 2022 highlight that ensuring compliance of XAI systems with laws and policies has been an enormous challenge, resulting in a lack of clarity for stakeholders in the healthcare ecosystem, which could undermine the determinants needed for the acceptance of XAI. The attributes attributed to XAI in healthcare systems certainly improves efficiency however there are legal issues that pose a dilemma of how these should be handled. Finding solutions for these problems will take real teamwork among developers, physicians, bioethicists and lawyers in order to achieve the right balance of efficacy and ethics when bringing XAI into the healthcare systems.

DISCUSSION

AI has progressed rapidly in recent years. This advance has made it possible to utilize machine learning (ML) models in healthcare settings enabling improved diagnostics, better outcome predictions, and more accurate prescription of therapies. However, the black box problem that is characteristic of these models has emerged as a huge limitation which has created the need for Explainable AI (XAI). XAI's objective is to improve the interpretability of ML models, so that clinical decisions are made taking into consideration the knowledge of a doctor and ML algorithms.

Interpretation of Findings

Such a gap can often lead to poor translation of research findings into clinical applications. The majority of the health studies, including clinical trials, are commonly published without sufficient context to the results such as confidence intervals or clinical importance and therefore create room for confusion in that statistical significance is mistaken for clinical significance (Docter et al 2020). This makes them likely to accept the use of those treatments that do not work and ignore those that do.

Moreover, factors such as bias and generalizability are some of the things which most often affects the conclusion of the study. As it turns out, selection, measurement and publication biases are some major influences on research outcomes (Byrne et al. 2020). These treatment related biases, if not well addressed, may result in either sanguine or disappointingly conservative estimates of treatment effects when such estimates should assist in making clinical choices. Additionally, there are studies in which generalization is not performed because the study subjects are very homogeneous in terms of age, gender, specific ethnic groups within a population, or the presence of comorbidities for example, which will not allow the results to be generalized to other populations.

One such challenge is XAI, which provides insight into the decision-making processes of ML models. Such understanding will assist clinicians in understanding what aspects of AI recommendations are useful and which ones are not useful so that they are not misused in clinical practice. In addition, XAI helps to detect and mitigate similar problems, accelerating the validity and trustworthiness of the use of AI in healthcare.

Comparison with Existing Literature

Interpreting and comparing findings in healthcare research is recognised as particularly difficult, especially when it concerns the adoption of AI and machine learning. Asci et al., (2021) found notable communication gaps, as most of the reports over-relied on p-values, less emphasis was placed on the useful effect size of results (considering clinical usefulness) . This has raised concerns that metrics such as confidence intervals, effect sizes and number needed to treat which may be more clinically relevant were rarely used.

Furthermore, there are differences in report types such as case studies and Randomized Controlled Trials (RCTs) where RCT is the ideal because it shows less bias in causation. Authors such as Rombach et al. (2020) also suggest that reporting practices need to be standardized in order to ensure that outcome statistics can be aggregated and compared. In contrast, explainable AI (XAI) internalizes model interpretation and consistency to clinical data and patient care in order to enhance implementation. Traditional and modern approaches to the examination of

model explainability and outcomes such as feature importance and model agnostic methods provide proposals for reasonable changes concerning previously published guidelines.

****Implications for Clinical Decision-Making****

The role of explainable artificial intelligence (XAI) in healthcare raises numerous possibilities for its respective area of practice. One of them is the improvement of the inter-professional trust, especially with regard to the horror movies referred to as the “black box” (machine learning) (Ghassemi et al. 2021). How it does this is simply by helping in the opening up of these models which then makes them acceptable to users.

XAI has also enhanced the relative accuracy and reliability of healthcare information and the decisions made from that information. Since it can explain the factors that affect a model’s decision, a clinician can offer a patient specific therapies, accounting for their history and preferences. This is important in complicated situations where protocols are hard to adhere to because of various developments. Also, Objectives XIA fairing model recommendations relative to culturally or linguistically diverse patients, it contributes to an AI system which seeks to ameliorate the variety of biases embedded in ML frameworks.

Moreover, XAI can further improve new research in medicine by finding new patterns and relations in clinical data which could result in new ideas and treatment plans. This cycle of amodeling, refinement and clinical validation enables constant evolvement of AI applications and processes in practice, in a way that results in improved patient outcomes.

CONCLUSION

The application of Explainable AI (XAI) in the domain of healthcare brings a promising solution to the challenge of incorporating advanced machine learning techniques into real-world settings. The current work has investigated the possible role of XAI in aiding the decision-making processes in the context of healthcare, in particular tackling some important barriers for the use and implementation of AI.

Our analysis asserts that XAI aims at improving transparency and interpretability of AI generated insights in healthcare. XAI can not only improve the confidence that healthcare professionals have in these tools by making them less mysterious but also enhance the incorporation of such tools in healthcare provision owing to increased levels of understanding. This increased accountability does not only improve the quality of the medical diagnosis processes, but also enhances the quality of evidence based decisions made by the healthcare providers.

The introduction of XAI to the day-to-day executors in healthcare institutions presents a few drawbacks. The tradeoff between complexity of the model and explainability of the model or the rules of engagement still poses a challenge like all other principles assuring compliance of XAI implementation with internal and external requirements such as cross border ethics and regulations. However, this understudy elaborates that XAI may have an important contribution towards the minimization of biases in AI models, hence achieving a better health status of diverse populations.

Continued efforts of technologists, clinicians, ethicists and legal experts are needed due to growing ethical and legal concerns associated with XAI within the context of health care. Given the rapid advances in AI technologies, it is necessary that policies are sophisticated enough to prevent landmark AI technologies being misused inside of healthcare facilities.

The working of XAI is remarkable in terms of supporting clinical performance and improving the health of the patients. However, achieving such an objective would depend greatly on research activities, development and collaboration of different disciplines in the health care systems. XAI further limits the possible scope of the future progression of ill-structured problems both in applications and implementations. Crisis corridors within the deepening crisis of structure within health care systems are being considered, with XAI defining it, broadening its framework for future situations.

In the future studies, increasing attention should be given to how human-AI interaction can be made more beneficial, effective and free of biases and errors in healthcare. Moreover, the research into the ethical and legal concerns regarding XAI also falls short; example of concerns

may be invasion of privacy, data misuse, or discriminatory algorithms. Also, meeting the sophistication and ease of use in interpretability methods will be important for further adoption of AI in the clinical practitioner practices.

In this view, XAI is extremely important in development of healthcare as efficient and trustworthy provision of care via intelligent support will make it more acceptable to healthcare professionals. But there is an ever-growing space of work needed to be undertaken to address ethical issues, enhance interpretability, and improve collaboration, whilst making sure that AI remains effective in healthcare, rendering all patients care needed, while being ethically responsible.

REFERENCES

- Ahmed, S., Solis-Oba, R., & Ilie, L. (2022). Explainable-AI in Automated Medical Report Generation Using Chest X-ray Images. *Applied Sciences*. <https://doi.org/10.3390/app122211750>.
- Alam, G., McChesney, I., Nicholl, P., & Rafferty, J. (2023). Improving Prospective Healthcare Outcomes by Leveraging Open Data and Explainable AI. *2023 3rd International Conference on Computing and Information Technology (ICCIT)*, 486-492. <https://doi.org/10.1109/ICCIT58132.2023.10273878>.
- Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20. <https://doi.org/10.1186/s12911-020-01332-6>.
- Antoniadi, A., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B., & Mooney, C. (2021). Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences*, 11, 5088. <https://doi.org/10.3390/APP11115088>.
- B, S. (2023). Advancing Healthcare through Artificial Intelligence: Innovations at the Intersection of AI and Medicine. *Tuijin Jishu/Journal of Propulsion Technology*. <https://doi.org/10.52783/tjjpt.v44.i2.131>.
- Bhardwaj, A. (2022). Promise and Provisos of Artificial Intelligence and Machine Learning in Healthcare. *Journal of Healthcare Leadership*, 14, 113 - 118. <https://doi.org/10.2147/JHL.S369498>.
- Byrne, B., Rooshenas, L., Lambert, H., & Blazeby, J. (2020). A mixed methods case study investigating how randomised controlled trials (RCTs) are reported, understood and interpreted in practice. *BMC Medical Research Methodology*, 20. <https://doi.org/10.1186/s12874-020-01009-8>.
- Byrne, B., Rooshenas, L., Lambert, H., & Blazeby, J. (2020). A mixed methods case study investigating how randomised controlled trials (RCTs) are reported, understood and

interpreted in practice. *BMC Medical Research Methodology*, 20. <https://doi.org/10.1186/s12874-020-01009-8>.

Castiglioni, I., Rundo, L., Codari, M., Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N., & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning.. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics*, 83, 9-24 . <https://doi.org/10.1016/j.ejmp.2021.02.006>.

Chomutare, T., Tejedor, M., Svenning, T., Marco-Ruiz, L., Tayefi, M., Lind, K., Godtliessen, F., Moen, A., Ismail, L., Makhlysheva, A., & Ngo, P. (2022). Artificial Intelligence Implementation in Healthcare: A Theory-Based Scoping Review of Barriers and Facilitators. *International Journal of Environmental Research and Public Health*, 19. <https://doi.org/10.3390/ijerph192316359>.

Ciapponi, A., Belizan, J., Piaggio, G., & Yaya, S. (2021). There is life beyond the statistical significance. *Reproductive Health*, 18. <https://doi.org/10.1186/s12978-021-01131-w>.

Ciapponi, A., Belizan, J., Piaggio, G., & Yaya, S. (2021). There is life beyond the statistical significance. *Reproductive Health*, 18. <https://doi.org/10.1186/s12978-021-01131-w>.

Clement, T., Kemmerzell, N., Abdelaal, M., & Amberg, M. (2023). XAIR: A Systematic Metareview of Explainable AI (XAI) Aligned to the Software Development Process. *Mach. Learn. Knowl. Extr.*, 5, 78-108. <https://doi.org/10.3390/make5010006>.

Date, P., Pimprale, V., & Mandke, S. (2023). Explorative study on potential of machine learning and artificial intelligence for improved healthcare diagnosis and treatment. *Journal of Autonomous Intelligence*. <https://doi.org/10.32629/jai.v7i3.1084>.

Dimopoulos-Bick, T., Osten, R., Shipway, C., Trevena, L., & Hoffmann, T. (2019). Shared decision making implementation: a case study analysis to increase uptake in New South Wales.. *Australian health review : a publication of the Australian Hospital Association*. <https://doi.org/10.1071/AH18138>.

- Docter, S., Fathalla, Z., Lukacs, M., Khan, M., Jennings, M., Liu, S., Dong, S., Getgood, A., & Bryant, D. (2020). Interpreting Patient-Reported Outcome Measures in Orthopaedic Surgery: A Systematic Review.. *The Journal of bone and joint surgery. American volume*. <https://doi.org/10.2106/JBJS.20.00474>.
- Docter, S., Fathalla, Z., Lukacs, M., Khan, M., Jennings, M., Liu, S., Dong, S., Getgood, A., & Bryant, D. (2020). Interpreting Patient-Reported Outcome Measures in Orthopaedic Surgery: A Systematic Review.. *The Journal of bone and joint surgery. American volume*. <https://doi.org/10.2106/JBJS.20.00474>.
- Dwivedi, R., Dave, D., Naik, H., Singhal, S., Rana, O., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., & Ranjan, R. (2022). Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Computing Surveys*, 55, 1 - 33. <https://doi.org/10.1145/3561048>.
- Elul, Y., Rosenberg, A., Schuster, A., Bronstein, A., & Yaniv, Y. (2021). Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning–based ECG analysis. *Proceedings of the National Academy of Sciences*, 118. <https://doi.org/10.1073/pnas.2020620118>.
- Erdeniz, S., Veeranki, S., Schrepf, M., Jauk, S., Tran, T., Felfernig, A., Kramer, D., & Leodolter, W. (2022). Explaining Machine Learning Predictions of Decision Support Systems in Healthcare. *Current Directions in Biomedical Engineering*, 8, 117 - 120. <https://doi.org/10.1515/cdbme-2022-1031>.
- Escoffery, C., Lebow-Skelley, E., Udelson, H., Böing, E., Wood, R., Fernandez, M., & Mullen, P. (2019). A scoping study of frameworks for adapting public health evidence-based interventions.. *Translational behavioral medicine*, 9 1, 1-10 . <https://doi.org/10.1093/tbm/ibx067>.
- Gao, Z., Lou, L., Wang, M., Sun, Z., Chen, X., Zhang, X., Pan, Z., Hao, H., Zhang, Y., Quan, S., Yin, S., Lin, C., & Shen, X. (2022). Application of Machine Learning in Intelligent Medical Image Diagnosis and Construction of Intelligent Service Process. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/9152605>.

- Glasgow, R., Battaglia, C., McCreight, M., Ayele, R., & Rabin, B. (2020). Making Implementation Science More Rapid: Use of the RE-AIM Framework for Mid-Course Adaptations Across Five Health Services Research Projects in the Veterans Health Administration. *Frontiers in Public Health*, 8. <https://doi.org/10.3389/fpubh.2020.00194>.
- Guleria, P., Srinivasu, P., Ahmed, S., Almusallam, N., & Alarfaj, F. (2022). XAI Framework for Cardiovascular Disease Prediction Using Classification Techniques. *Electronics*. <https://doi.org/10.3390/electronics11244086>.
- Harry, M., Saman, D., Truitt, A., Allen, C., Walton, K., O'Connor, P., Ekstrom, H., Sperl-Hillen, J., Bianco, J., & Elliott, T. (2020). Pre-implementation adaptation of primary care cancer prevention clinical decision support in a predominantly rural healthcare system. *BMC Medical Informatics and Decision Making*, 20. <https://doi.org/10.1186/s12911-020-01136-8>.
- Holden, R., Carayon, P., Gurses, A., Hoonakker, P., Hundt, A., Ozok, A., & Rivera-Rodriguez, A. (2013). SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. *Ergonomics*, 56, 1669 - 1686. <https://doi.org/10.1080/00140139.2013.838643>.
- Ilott, I., Gerrish, K., Booth, A., & Field, B. (2012). Testing the Consolidated Framework for Implementation Research on health care innovations from South Yorkshire.. *Journal of evaluation in clinical practice*, 19, 5, 915-24. <https://doi.org/10.1111/j.1365-2753.2012.01876.x>.
- Iqbal, J., Jaimes, D., Makineni, P., Subramani, S., Hemaida, S., Thugu, T., Butt, A., Sikto, J., Kaur, P., Lak, M., Augustine, M., Shahzad, R., & Arain, M. (2023). Reimagining Healthcare: Unleashing the Power of Artificial Intelligence in Medicine. *Cureus*, 15. <https://doi.org/10.7759/cureus.44658>.
- Jesso, S., Kelliher, A., Sanghavi, H., Martin, T., & Parker, S. (2022). Inclusion of Clinicians in the Development and Evaluation of Clinical Artificial Intelligence Tools: A Systematic Literature Review. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.830345>.

- Jeyaraman, M., Balaji, S., Jeyaraman, N., & Yadav, S. (2023). Unraveling the Ethical Enigma: Artificial Intelligence in Healthcare. *Cureus*, 15. <https://doi.org/10.7759/cureus.43262>.
- Khakurel, U., & Rawat, D. (2022). Evaluating explainable artificial intelligence (XAI): algorithmic explanations for transparency and trustworthiness of ML algorithms and AI systems. , 12113, 121130A - 121130A-11. <https://doi.org/10.1117/12.2620598>.
- Kyrimi, E., Dube, K., Fenton, N., Fahmi, A., Neves, M., Marsh, W., & McLachlan, S. (2021). Bayesian networks in healthcare: What is preventing their adoption?. *Artificial intelligence in medicine*, 116, 102079 . <https://doi.org/10.1016/J.ARTMED.2021.102079>.
- Laka, M., Milazzo, A., & Merlin, T. (2020). Factors that impact the adoption of clinical decision support systems (CDSS) in healthcare settings for evidence-based decision making. . <https://doi.org/10.21203/rs.3.rs-34358/v1>.
- Lal, J., Vaidya, A., Gutiérrez-Ibarluzea, I., Dauben, H., & Brand, A. (2013). The Learning-Adapting-Leveling model: from theory to hypothesis of steps for implementation of basic genome-based evidence in personalized medicine.. *Personalized medicine*, 10 7, 683-701 . <https://doi.org/10.2217/pme.13.72>.
- Larasati, R. (2022). Explainable AI for Breast Cancer Diagnosis: Application and User's Understandability Perception. *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, 1-6. <https://doi.org/10.1109/ICECET55527.2022.9872950>.
- Lipworth, W., Taylor, N., & Braithwaite, J. (2013). Can the theoretical domains framework account for the implementation of clinical quality interventions?. *BMC Health Services Research*, 13, 530 - 530. <https://doi.org/10.1186/1472-6963-13-530>.
- Mandala, S. (2023). XAI Renaissance: Redefining Interpretability in Medical Diagnostic Models. *ArXiv*, abs/2306.01668. <https://doi.org/10.48550/arXiv.2306.01668>.
- Matulionyte, R., Nolan, P., Magrabi, F., & Beheshti, A. (2022). Should AI-enabled medical devices be explainable?. *Int. J. Law Inf. Technol.*, 30, 151-180. <https://doi.org/10.2139/ssrn.4140234>.

- Mcdermid, J., Jia, Y., Porter, Z., & Habli, I. (2021). Artificial intelligence explainability: the technical and ethical dimensions. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 379. <https://doi.org/10.1098/rsta.2020.0363>.
- Mcdermid, J., Jia, Y., Porter, Z., & Habli, I. (2021). Artificial intelligence explainability: the technical and ethical dimensions. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 379. <https://doi.org/10.1098/rsta.2020.0363>.
- Mishra, V., Ugemuge, S., & Tiwade, Y. (2023). Artificial intelligence changing the future of healthcare diagnostics. *Journal of Cellular Biotechnology*. <https://doi.org/10.3233/jcb-230118>.
- Mueller, S., Veinott, E., Hoffman, R., Klein, G., Alam, L., Mamun, T., & Clancey, W. (2021). Principles of Explanation in Human-AI Systems. *ArXiv*, abs/2102.04972.
- Naveed, M. (2023). Transforming Healthcare through Artificial Intelligence and Machine Learning. *Pakistan Journal of Health Sciences*. <https://doi.org/10.54393/pjhs.v4i05.844>.
- Nilsen, P., Svedberg, P., Nygren, J., Frideros, M., Johansson, J., & Schueller, S. (2022). Accelerating the impact of artificial intelligence in mental healthcare through implementation science. *Implementation Research and Practice*, 3. <https://doi.org/10.1177/26334895221112033>.
- Pawar, U., O'Shea, D., Rea, S., & O'Reilly, R. (2020). Explainable AI in Healthcare. *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*, 1-2. <https://doi.org/10.1109/CyberSA49311.2020.9139655>.
- Sabry, F., Eltaras, T., Labda, W., Alzoubi, K., & Malluhi, Q. (2022). Machine Learning for Healthcare Wearable Devices: The Big Picture. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/4653923>.
- Sadeghi, Z., Alizadehsani, R., Cifci, M., Kausar, S., Rehman, R., Mahanta, P., Bora, P., Almasri, A., Alkhalwaldeh, R., Hussain, S., Alatas, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalos, P. (2023). A Brief Review of Explainable Artificial

Intelligence in Healthcare. *ArXiv*, abs/2304.01543.
<https://doi.org/10.48550/arXiv.2304.01543>.

Saraswat, D., Bhattacharya, P., Verma, A., Prasad, V., Tanwar, S., Sharma, G., Bokoro, P., & Sharma, R. (2022). Explainable AI for Healthcare 5.0: Opportunities and Challenges. *IEEE Access*, 10, 84486-84517.
<https://doi.org/10.1109/access.2022.3197671>.

Srinivasu, P., Sandhya, N., Jhaveri, R., & Raut, R. (2022). From Blackbox to Explainable AI in Healthcare: Existing Tools and Case Studies. *Mobile Information Systems*.
<https://doi.org/10.1155/2022/8167821>.

Thalpage, N. (2023). Unlocking the Black Box: Explainable Artificial Intelligence (XAI) for Trust and Transparency in AI Systems. *Journal of Digital Art & Humanities*.
https://doi.org/10.33847/2712-8148.4.1_4.

Tiwari, R. (2023). The Use of AI and Machine Learning in Healthcare and its Potential to Improve Patient Outcomes. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*. <https://doi.org/10.55041/ijrem17582>.

Velden, B., Kuijf, H., Gilhuijs, K., & Viergever, M. (2021). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical image analysis*, 79, 102470 . <https://doi.org/10.1016/j.media.2022.102470>.

Waddell, A., Lennox, A., Spassova, G., & Bragge, P. (2021). Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. *Implementation Science : IS*, 16. <https://doi.org/10.1186/s13012-021-01142-y>.

Yadav, K., & Gaurav, A. (2023). Application and Challenges of Machine Learning in Healthcare. *International Journal for Research in Applied Science and Engineering Technology*. <https://doi.org/10.22214/ijraset.2023.55678>.

Yang, G., Ye, Q., & Xia, J. (2021). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *An*

International Journal on Information Fusion, 77, 29 - 52.
<https://doi.org/10.1016/j.inffus.2021.07.016>.