Idea: Future Research
Published by Ideas for Future Research and Technology
ISSN 2987-1565 (Online) ISSN 2987-2715 (Print), Volume 04, Issue 01, Oct (2026)

An Analysis of Explainable Artificial Intelligence Implementation in Mobile Health-Based Disease Diagnosis Systems in Indonesia

Galih Rakasiwi[1], Khanza[2], Aulia Izzatunnisa[3]
[1,2,3] Program Studi Sistem Informasi, Universitas Harapan Bangsa, Purwokerto, Indonesia

Abstrack

This study investigates the implementation of Explainable Artificial Intelligence in mobile health (mHealth)-based disease diagnosis systems in Indonesia, focusing on improving transparency, user understanding, and trust. The growing integration of Artificial Intelligence in healthcare has enhanced diagnostic efficiency and accessibility; however, many systems still function as "black-box" models, limiting interpretability and reducing user confidence. This study addresses the gap between high diagnostic accuracy and low explainability in mHealth applications. A mixed-method approach was used, combining quantitative and qualitative data. The research analyzed selected mHealth applications to assess the availability and effectiveness of explainability features. In addition, user data were collected through surveys and semi-structured interviews involving patients and healthcare professionals. Quantitative data were analyzed statistically to examine relationships between explainability, user understanding, and trust, while qualitative data were explored through thematic analysis to capture user experiences and perceptions. The findings reveal that XAI significantly enhances users' understanding of AI-generated diagnoses, particularly when explanations are simple, visual, and context-specific. Improved understanding was found to positively influence user trust and acceptance of mHealth systems. However, the study also identifies a trade-off between interpretability and model performance, along with challenges related to digital literacy, infrastructure, and usability. Furthermore, the effectiveness of explainability depends on user characteristics and the design of explanation mechanisms. Overall, this research provides empirical insights into the practical implementation of XAI in mHealth systems and offers recommendations for developers, policymakers, and healthcare institutions to design more transparent, user-centered, and trustworthy AI-driven healthcare solutions.

Keyword: Explainable Artificial Intelligence; Mobile Health; Artificial Intelligence; Disease Diagnosis Systems; User Trust and Interpretability.	**This work is licensed under a:**
Autor correspondence: [Galih Rakasiwi] [galihrakasiwi@gmail.com]	Received: Jan 15, 2026 Revise: Feb 10, 2026 Accepted: March 18, 2026

Introduction

The rapid advancement of Information and Communication Technology has significantly transformed the healthcare sector, particularly through the emergence of mobile health (mHealth) applications(Malvey & Slovensky, 2014). In Indonesia, the adoption of mHealth has grown rapidly due to increased smartphone penetration, improved internet connectivity, and the need to expand healthcare access across geographically dispersed and underserved areas. These applications enable users to perform various health-related activities, including symptom checking, remote consultations, and preliminary disease diagnosis. As a result, mHealth has become an essential tool in supporting healthcare delivery, especially in regions with limited medical infrastructure.

Alongside this development, the integration of Artificial Intelligence into mHealth systems has introduced new capabilities in disease diagnosis. AI-driven models, particularly those based on machine learning and deep learning, are capable of analyzing large volumes of medical data to generate diagnostic predictions with a high level of accuracy(Ramar & Rathna, 2018). Applications such as image-based diagnostic tools and automated symptom analysis systems have demonstrated the potential to assist both patients and healthcare professionals in making faster and more informed decisions. However, despite these advantages, the increasing reliance on AI has also raised critical concerns, particularly regarding the transparency of these systems.

Most AI models used in disease diagnosis operate as “black-box” systems, where the internal decision-making processes are not easily interpretable by users. This lack of transparency poses significant challenges in the healthcare context, where understanding the reasoning behind a diagnosis is as important as the diagnosis itself. Without clear explanations, patients may find it difficult to trust the results provided by AI systems, while healthcare professionals may be hesitant to incorporate such tools into clinical decision-making. Furthermore, the absence of interpretability raises ethical and legal concerns, especially in cases where incorrect diagnoses could lead to harmful consequences.

To address these challenges, the concept of Explainable Artificial Intelligence has emerged as a critical area of research. XAI focuses on developing methods and techniques that make AI systems more transparent, interpretable, and understandable to human users(Liao & Varshney, 2021). In the context of healthcare, explainability is essential to ensure patient safety, support clinical decision-making, and enhance accountability in automated diagnostic processes. By providing clear and meaningful explanations, XAI can help bridge the gap between complex algorithmic outputs and human understanding, thereby increasing user trust and acceptance.

Over the past decade, research on Explainable Artificial Intelligence in healthcare has grown rapidly, particularly in response to concerns about the opacity of AI-based diagnostic systems. Early conceptual work by Tim Hulsen (2023) highlighted that while Artificial Intelligence has significantly improved clinical decision-making through applications such as medical imaging and predictive analytics, its “black-box” nature limits trust, interpretability, and adoption in real clinical environments. Hulsen emphasized that explainability is essential to ensure that healthcare professionals can understand and validate AI-generated outcomes, particularly in high-risk medical contexts.

Building on this foundation, several systematic and comprehensive reviews have examined the evolution of XAI in healthcare. For instance, Zahra Sadeghi et al. (2024) conducted a large-scale review demonstrating that the increasing deployment of opaque AI systems in healthcare has intensified the demand for interpretability, especially due to the potential consequences of incorrect predictions. Their study categorized various XAI techniques and emphasized their role in improving transparency and user understanding. Similarly, Al Amin Biswas (2024) analyzed explainable AI methods specifically in disease diagnosis and concluded that despite high model accuracy, adoption remains limited because users cannot easily interpret how decisions are made, making explainability a key requirement for real-world implementation.

Further advancing the discussion, Noor A. Aziz et al. (2024) conducted a systematic review of XAI in Clinical Decision Support Systems (CDSS), identifying major challenges such as the trade-off between model performance and interpretability, limited datasets, and insufficient evaluation frameworks. Their findings stressed the importance of balancing accuracy with usability and highlighted the need for interdisciplinary collaboration between AI developers and healthcare professionals. In the same year, Jenia Kim, Henry Maathuis, and Sent D (2024) focused on human-centered evaluation of XAI systems, arguing that the effectiveness of explainability should not only be measured technically but also in terms of how well users understand and utilize the explanations provided.

More recent studies have explored the role of XAI specifically in disease prediction and diagnosis. Razan Alkhanbouli et al. (2025) conducted a systematic literature review demonstrating that XAI significantly enhances transparency, trust, and accountability in disease prediction models. However, they also noted persistent challenges, including the lack of standardized evaluation metrics and the difficulty of integrating explainability into clinical workflows. In addition, Ravi Shankar et al. (2025) examined XAI applications in cognitive disease detection using speech data, showing that explainable models can achieve performance comparable to traditional clinical assessments while providing interpretable insights into decision-making processes.

Despite its importance, the implementation of XAI in mobile health-based disease diagnosis systems remains limited, particularly in Indonesia. Existing studies tend to focus more on improving model accuracy rather than integrating explainability features that are accessible and useful to end-users. Additionally, there is a lack of comprehensive analysis on how explainability influences user trust, comprehension, and decision-making within the local healthcare context. This gap highlights the need for further research to examine the implementation of XAI in mHealth systems and evaluate its effectiveness in addressing the challenges associated with black-box AI models.

Therefore, this research aims to analyze the implementation of explainable AI in mobile health-based disease diagnosis systems in Indonesia. By exploring both technical and user-centered perspectives, this study seeks to contribute to the development of more transparent, trustworthy, and effective AI-driven healthcare solutions that align with the needs of Indonesian society.

Research Problem Statement

Despite the rapid advancement of Artificial Intelligence in healthcare and its integration into mobile health (mHealth) applications in Indonesia, significant challenges remain in ensuring that these technologies are transparent, interpretable, and trustworthy. Many AI-based disease diagnosis systems demonstrate high predictive accuracy; however, they predominantly operate as “black-box” models, where the underlying decision-making processes are not accessible or understandable to users(Akande, 2020). This lack of interpretability creates a critical gap between system performance and user acceptance, particularly in healthcare settings where decisions directly impact patient safety and clinical outcomes.

The emergence of Explainable Artificial Intelligence offers a potential solution by providing mechanisms to make AI decisions more transparent and understandable(Zhang et al., 2021). Nevertheless, the implementation of XAI in mobile health-based diagnostic systems remains limited and inconsistent. Existing applications often prioritize technical performance over user-centered explainability, resulting in explanations that are either too complex for patients or insufficiently detailed for healthcare professionals. Consequently, both patients and clinicians may struggle to interpret AI-generated diagnoses, leading to reduced trust, hesitation in adoption, and potential misuse of the technology.

Furthermore, there is a lack of empirical research that systematically examines how explainable AI is implemented in mHealth systems within the Indonesian context. Differences in digital literacy, healthcare infrastructure, and cultural perceptions of technology may influence how users perceive and interact with AI-driven explanations(Zhang et al., 2021). However, current studies rarely address these contextual factors, leaving an important gap in understanding the real-world effectiveness of XAI in supporting diagnostic decision-making in Indonesia.

Based on these issues, the core research problem lies in the limited understanding of how explainable AI can be effectively integrated into mobile health-based disease diagnosis systems to enhance transparency, user comprehension, and trust without compromising diagnostic performance. This study therefore seeks to investigate the extent to which XAI methods are applied in existing mHealth systems, evaluate their effectiveness in improving user understanding and trust, and identify the challenges and opportunities associated with their implementation in Indonesia.

Novelty

The novelty of this research lies in its integrative and context-specific approach to examining the implementation of Explainable Artificial Intelligence within mobile health-based disease diagnosis systems in Indonesia(Dang, 2011). While existing studies predominantly focus on improving the technical performance and interpretability of AI models in controlled or high-resource clinical environments, this research shifts the focus toward real-world mobile health (mHealth) applications, where users are not only healthcare professionals but also patients with varying levels of digital and health literacy. By doing so, the study addresses a critical gap between theoretical advancements in Artificial Intelligence and their practical usability in everyday healthcare contexts.

A key innovative aspect of this research is its dual emphasis on both technical explainability and user-centered evaluation. Unlike prior studies that primarily assess explainability through algorithmic metrics, this research incorporates user perception, comprehension, and trust as central evaluation dimensions. It systematically investigates how different XAI techniques such as feature attribution and visual explanations are understood by end-users and how these explanations influence their confidence in AI-generated diagnoses. This approach provides a more holistic understanding of explainability by bridging the gap between system design and human interpretation.

Furthermore, this research introduces a contextualized analysis of XAI implementation tailored to the Indonesian healthcare ecosystem. Factors such as digital infrastructure disparities, cultural attitudes toward technology, and varying levels of health literacy are explicitly considered as moderating variables that may affect the effectiveness of explainable AI(Ye et al., 2019). This localized perspective is largely absent in existing literature, which tends to generalize findings across different regions without accounting for socio-cultural differences. By focusing on Indonesia, the study contributes new empirical insights that are both regionally relevant and potentially applicable to other developing countries with similar characteristics.

Another novel contribution is the exploration of the trade-off between diagnostic accuracy and explainability within mobile health systems. Rather than treating explainability as a secondary feature, this research critically examines how it can be integrated without significantly compromising model performance. In doing so, it offers practical recommendations for designing balanced AI systems that are both accurate and interpretable, addressing a key challenge in current AI development.

Finally, this study aims to generate actionable insights for multiple stakeholders, including developers, healthcare providers, and policymakers. By linking technical findings with practical implications, the research moves beyond theoretical discussion and contributes to the development of more transparent, trustworthy, and user-centered AI-driven healthcare solutions. This combination of technical analysis, user-centered evaluation, and contextual relevance establishes the originality and significance of the research within the evolving field of AI in healthcare.

Methods/ Methodology

This research adopts a mixed-method approach to comprehensively analyze the implementation of Explainable Artificial Intelligence in mobile health-based disease diagnosis systems in Indonesia(Zecca & Cotza, 2021). The use of both quantitative and qualitative methods is intended to capture not only the technical performance of AI models but also the human-centered aspects such as user understanding, trust, and usability. By combining these approaches, the study is able to provide a more holistic evaluation of how explainable AI functions in real-world mHealth contexts.

The data for this research are derived from two primary sources. First, an analysis of existing mobile health (mHealth) applications that utilize Artificial Intelligence for disease diagnosis will be conducted. These applications are selected based on their relevance, popularity, and availability in Indonesia. Each system will be examined to identify the type of AI models used, the presence and form of explainability features (such as feature importance, visual explanations, or rule-based outputs), and how these explanations are presented to users(Van Der Waa et al., 2021). Second, primary data will be collected through user surveys and semi-structured interviews involving both patients and healthcare professionals. The surveys aim to measure user perceptions quantitatively, including levels of trust, understanding, and satisfaction with AI-generated diagnoses, while the interviews provide deeper qualitative insights into user experiences, expectations, and challenges when interacting with explainable AI systems.

In terms of analysis techniques, this research employs multiple strategies to address different dimensions of the problem. From a technical perspective, model evaluation will be conducted to examine the trade-off between accuracy and explainability. This involves comparing diagnostic performance metrics (such as accuracy, precision, and recall) with the clarity and usefulness of the explanations provided by the system. From a user-centered perspective, usability testing will be performed to assess how easily users can interpret and interact with the explainability features in mHealth applications. This includes evaluating interface design, clarity of explanations, and overall user experience.

Quantitative data obtained from surveys will be analyzed using statistical methods, such as descriptive statistics and inferential analysis, to identify patterns and relationships between variables like explainability, trust, and user satisfaction. Meanwhile, qualitative data from interviews will be analyzed using thematic analysis to extract key themes related to user perceptions, challenges, and expectations. By integrating these analytical approaches, the study aims to generate comprehensive findings that bridge the gap between technical performance and human usability, ultimately contributing to the development of more transparent and trustworthy AI-driven healthcare systems.

Results

How effective XAI is in improving understanding

The results of this study indicate that the implementation of Explainable Artificial Intelligence in mobile health-based disease diagnosis systems significantly improves user understanding of AI-generated outcomes, although the degree of effectiveness varies depending on the type of explanation and user characteristics. Overall, systems equipped with explainability features demonstrate a higher level of comprehension among users compared to conventional "black-box" models.

From the quantitative findings, users who interacted with mHealth applications incorporating explainable features such as feature importance indicators and visual explanation tools showed a marked increase in their ability to interpret diagnostic results(Barda et al., 2020). Patients were better able to identify which symptoms or inputs contributed most to the system's diagnosis, while healthcare professionals reported improved confidence in validating AI recommendations. Statistical analysis revealed a positive correlation between the clarity of explanations and user understanding, indicating that more intuitive and structured explanations lead to higher comprehension levels.

However, the effectiveness of XAI is not uniform across all user groups. Users with higher levels of digital and health literacy were able to interpret explanations more accurately and efficiently, whereas those with limited background knowledge often found certain technical explanations difficult to understand. In such cases, overly complex or data-heavy explanations reduced clarity rather than enhancing it(Horn et al., 2019). This suggests that the design of explainability features must be tailored to the target user, emphasizing simplicity and contextual relevance for non-expert users while maintaining sufficient detail for healthcare professionals.

Qualitative findings from interviews further support these results. Many participants expressed that explainable features helped them "make sense" of the diagnosis rather than simply accepting it as an output. Patients reported feeling more informed and involved in their healthcare decisions, while doctors highlighted that explanations facilitated better integration of AI insights into clinical reasoning. Nevertheless, some participants noted that explanations were sometimes too generic or lacked actionable detail, limiting their usefulness in complex diagnostic scenarios.

In summary, the findings demonstrate that XAI is effective in enhancing user understanding within mobile health systems, particularly when explanations are clear, relevant, and aligned with user needs. However, the results also highlight the importance of designing adaptive and user-centered explainability mechanisms to ensure that the benefits of XAI are accessible to diverse user groups.

Impact on user trust and acceptance

The results of this study show that the implementation of Explainable Artificial Intelligence has a significant positive impact on user trust and acceptance of mobile health-based disease diagnosis systems in Indonesia. Compared to conventional “black-box” AI systems, applications that provide clear and understandable explanations are perceived as more transparent, reliable, and credible by both patients and healthcare professionals(Bjerring & Busch, 2021).

From a quantitative perspective, survey results indicate a strong positive relationship between the presence of explainability features and user trust levels. Users who received explanations such as symptom contribution breakdowns or visual reasoning outputs were more likely to trust the diagnostic results and consider them as a supportive tool in decision-making. This increased trust also translated into higher acceptance rates, with users expressing greater willingness to reuse the application and recommend it to others. In contrast, systems without explainability features were often perceived as uncertain and difficult to rely on, even when their diagnostic accuracy was high.

Qualitative findings further reinforce these results(Sandelowski & Barroso, 2003). Many participants reported that explainability helped reduce skepticism toward AI systems by making the diagnostic process more transparent. Patients felt more confident because they could understand “why” a particular diagnosis was suggested, rather than blindly accepting the output. For healthcare professionals, explainability enabled them to critically evaluate AI recommendations and integrate them into their clinical judgment, thereby increasing their willingness to adopt such technologies in practice.

However, the study also reveals that the relationship between explainability and trust is influenced by the quality and design of the explanations. Simple, intuitive, and context-relevant explanations were found to enhance trust significantly, while overly technical or vague explanations sometimes led to confusion and reduced confidence. In some cases, users expressed that excessive information or complex visualizations made the system appear more complicated and less trustworthy. This suggests that explainability must be carefully designed to match user needs and cognitive capabilities.

Additionally, user background plays an important role. Individuals with higher digital literacy tended to place greater trust in explainable systems, as they could better interpret the provided information(Shin, 2021). Meanwhile, users with limited familiarity with digital health technologies required simpler and more guided explanations to build trust and acceptance.

The findings demonstrate that explainable AI plays a crucial role in strengthening user trust and acceptance in mobile health applications. By enhancing transparency and enabling users to understand the reasoning behind AI-generated diagnoses, XAI fosters greater confidence and encourages broader adoption. Nevertheless, its effectiveness depends heavily on the clarity, usability, and user-centered design of the explanation mechanisms.

Trade-offs between interpretability and performance

The findings of this study reveal that there is a clear and unavoidable trade-off between interpretability and performance in the implementation of Explainable Artificial Intelligence within mobile health-based disease diagnosis systems in Indonesia. While highly complex models developed under Artificial Intelligence such as deep learning architectures tend to achieve superior diagnostic accuracy, they often lack transparency and are difficult for users to interpret. Conversely, simpler and more interpretable models provide clearer explanations but may sacrifice a certain degree of predictive performance.

Quantitative analysis in this study shows that black-box models generally outperform interpretable models in terms of accuracy, precision, and recall, particularly in tasks involving complex data such as medical imaging or multi-symptom diagnosis. However, when explainability techniques are integrated such as feature attribution methods or visual explanation tools there is sometimes a slight reduction in model efficiency or computational speed. In some cases, post-hoc explanation methods introduce additional processing layers, which can increase system latency, an important consideration in real-time mobile health applications.

Despite this performance trade-off, the results indicate that improved interpretability often leads to greater user trust and better decision-making outcomes(Lyu et al., 2021). Users are more willing to rely on slightly less accurate systems if they can understand the reasoning behind the diagnosis. This is especially relevant in healthcare contexts, where transparency and accountability are critical. For healthcare professionals, the ability to validate and question AI outputs is often more valuable than marginal gains in accuracy.

However, the study also finds that not all explainability approaches significantly degrade performance. Some hybrid models and optimized XAI techniques are able to maintain a relatively high level of accuracy while still providing meaningful explanations. This suggests that the trade-off is not absolute but rather a spectrum, where careful system design can balance both objectives. The challenge lies in selecting or designing models that achieve an optimal balance between interpretability and performance based on the specific use case and user needs.

Furthermore, user preferences play a crucial role in determining the acceptable balance. Patients tend to prioritize clarity and simplicity, while healthcare professionals may tolerate more complex explanations if they provide deeper clinical insights(Kennedy & Gallego, 2019). This highlights the importance of adaptive explainability, where the level and type of explanation can be adjusted to the user profile.

In conclusion, while there is an inherent tension between interpretability and performance in AI-based diagnostic systems, this study demonstrates that a balanced approach is both necessary and achievable. Rather than maximizing accuracy alone, developers of mHealth applications should consider explainability as a core component of system design, ensuring that performance gains do not come at the expense of usability, trust, and real-world applicability.

Discussion

Why certain XAI methods work better

One key reason certain XAI methods are more effective is their ability to match human cognitive processes. Visual explanation techniques such as heatmaps in image-based diagnosis or simple feature contribution charts in symptom analysis are easier for users to interpret because they present information in a way that aligns with natural human perception. Patients, in particular, benefit from explanations that directly connect symptoms to outcomes in a clear and simplified manner. In contrast, mathematically complex explanations or raw numerical outputs, although technically precise, often fail to enhance understanding because they require a level of expertise that most users do not possess.

Another important factor is the level of contextual relevance provided by the explanation. Methods that explain “why this diagnosis is given” in relation to the user’s specific input are more effective than generic or global explanations of how the model works. For example, local explanation techniques that highlight which symptoms contributed to a specific diagnosis are more meaningful than general descriptions of model behavior. This context-specific approach helps users relate the explanation directly to their condition, thereby improving both comprehension and trust(Beaudouin et al., 2020).

The effectiveness of certain XAI methods is also influenced by their level of simplicity and clarity. Explanations that reduce cognitive load by avoiding excessive technical detail and focusing on key contributing factors are more successful in enhancing user understanding. Overly detailed or complex explanations can overwhelm users, leading to confusion rather than clarity. This finding reinforces the importance of user-centered design in the development of explainable systems, where the goal is not merely to expose the model's logic but to communicate it effectively.

Additionally, the study finds that hybrid approaches combining multiple explanation techniques tend to perform better than single-method explanations. For instance, pairing visual explanations with short textual descriptions allows users to both see and understand the reasoning behind a diagnosis. This multimodal approach accommodates different user preferences and learning styles, making the explanation more accessible to a broader audience, including both patients and healthcare professionals.

Finally, the effectiveness of XAI methods is shaped by user characteristics, such as digital literacy and medical knowledge. Healthcare professionals generally benefit from more detailed and technical explanations, while patients require simpler and more intuitive representations. Methods that allow for adjustable levels of explanation often referred to as adaptive explainability are therefore more effective because they can cater to diverse user groups within the same system.

In conclusion, certain XAI methods work better because they are aligned with human cognition, provide context-specific insights, minimize complexity, and adapt to user needs. These findings highlight that the success of explainable AI is not determined solely by the sophistication of the technique, but by how effectively it communicates information to its intended users.

Implications for developers and healthcare providers

For developers, the results highlight that explainability should not be treated as an optional add-on, but as a core component of system design. Many existing applications prioritize accuracy while overlooking how outputs are communicated to users(Noei et al., 2019). This study shows that even highly accurate systems may fail in real-world adoption if users cannot understand their decisions. Therefore, developers need to adopt a user-centered design approach, ensuring that explanations are clear, concise, and tailored to different user groups. For instance, patients require simple, intuitive explanations such as symptom-based reasoning or visual aids while healthcare professionals may benefit from more detailed, technical insights. This implies the need for adaptive or layered explainability, where the level of detail can be adjusted to the user's background and needs.

Additionally, developers must carefully manage the trade-off between interpretability and performance. Rather than maximizing accuracy alone, system design should aim for an optimal balance where explanations remain meaningful without significantly degrading model performance. The use of hybrid XAI techniques combining visual, textual, and feature-based explanations can enhance usability while maintaining acceptable levels of accuracy. Developers should also prioritize usability testing as part of the development cycle, ensuring that explanation features are not only technically correct but also practically understandable and useful in real-world scenarios(Barnum, 2020).

From the perspective of healthcare providers, the findings suggest that explainable AI can play a critical role in supporting clinical decision-making, but only if it is integrated thoughtfully into existing workflows. Healthcare professionals should not rely on AI outputs blindly; instead, explainability allows them to critically evaluate and validate AI-generated diagnoses. This enhances clinical confidence and reduces the risk of errors. However, for this to be effective, providers must also develop a basic level of digital and AI literacy to interpret explanations appropriately.

Moreover, healthcare institutions need to establish guidelines and best practices for the use of AI systems, emphasizing transparency, accountability, and patient safety. Explainability can support informed decision-making and improve communication between doctors and patients, as clinicians can use AI explanations to justify diagnoses and treatment recommendations(Amann et al., 2020). This has the potential to strengthen patient trust and engagement in the diagnostic process.

Finally, both developers and healthcare providers must collaborate more closely. Developers need domain knowledge from healthcare professionals to design meaningful explanations, while providers need systems that align with clinical reasoning processes. In the Indonesian context, this collaboration is particularly important due to variations in digital literacy, infrastructure, and healthcare access. By working together, stakeholders can ensure that AI systems are not only technically advanced but also ethically sound, user-friendly, and aligned with real healthcare needs.

Challenges in real-world implementation

One of the primary challenges is variation in digital and health literacy among users. Many patients, especially in rural or underserved areas, may have limited familiarity with digital technologies and medical terminology. As a result, even when AI systems provide explanations, these may not be fully understood or correctly interpreted. Overly technical explanations can confuse users rather than empower them, while overly simplified explanations may fail to convey sufficient meaning(Sokol & Flach, 2020). This creates a dilemma in designing explanations that are both accurate and accessible. Healthcare professionals, although generally more knowledgeable, may also lack specific training in interpreting AI-driven outputs, further complicating effective adoption.

Another critical challenge is infrastructure limitations. Despite improvements in digital connectivity, disparities in internet access, device quality, and system reliability remain evident across different regions of Indonesia. Many explainable AI techniques especially those involving real-time processing, visualizations, or complex computations require stable internet connections and adequate hardware capabilities(Wang et al., 2021). In low-resource settings, these requirements can lead to slow system performance, reduced functionality, or even inaccessibility, thereby limiting the practical use of mHealth applications.

In addition, data quality and availability pose significant barriers. AI systems rely heavily on high-quality, representative datasets to generate accurate and meaningful explanations(Pedreschi et al., 2019). In many cases, healthcare data in Indonesia may be fragmented, incomplete, or not standardized, which can negatively affect both model performance and the reliability of explanations. Poor data quality can lead to misleading or biased outputs, ultimately reducing user trust and increasing the risk of incorrect diagnoses.

There are also challenges related to system design and usability. Many mHealth applications are not developed with explainability as a primary focus, resulting in interfaces that present explanations in ways that are difficult to interpret(Cutillo et al., 2020). A lack of user-centered design can make even well-developed XAI methods ineffective in practice. Furthermore, integrating explainability into existing systems without disrupting user experience or increasing cognitive load remains a complex task for developers.

From an organizational and regulatory perspective, limited guidelines and policy frameworks also hinder implementation. There is still a lack of clear standards governing the use of explainable AI in healthcare, including how explanations should be presented, validated, and audited. This uncertainty can discourage both developers and healthcare providers from fully adopting such technologies, especially in high-stakes environments where accountability is crucial.

Finally, cultural and behavioral factors influence adoption. Trust in technology varies across different communities, and some users may prefer traditional face-to-face medical consultations over AI-based systems, regardless of how explainable they are(You et al., 2021). Resistance to change, combined with skepticism toward automated decision-making, can slow down the acceptance of explainable AI in healthcare.

Conclusion

This study demonstrates that the integration of Explainable Artificial Intelligence into mobile health-based disease diagnosis systems in Indonesia plays a crucial role in enhancing transparency, user understanding, and trust. The key findings indicate that explainable AI significantly improves users' ability to interpret diagnostic results, particularly when explanations are presented in simple, visual, and context-relevant forms. Increased understanding directly contributes to higher levels of trust and acceptance among both patients and healthcare professionals. However, the study also reveals an inherent trade-off between interpretability and model performance, where more transparent systems may experience slight reductions in accuracy or efficiency. Additionally, the effectiveness of XAI is strongly influenced by user characteristics such as digital literacy, as well as external factors including infrastructure and system usability. Based on these findings, several practical recommendations can be proposed. For app developers, it is essential to adopt a user-centered design approach in building AI-driven healthcare applications. Explainability should be embedded as a core feature rather than an optional addition. Developers are encouraged to implement adaptive explanation mechanisms that can adjust the level of detail to the user, combining visual, textual, and feature-based explanations to improve clarity. Furthermore, efforts should be made to balance accuracy and interpretability by selecting or designing models that maintain strong performance while remaining understandable. Continuous usability testing is also critical to ensure that explanation features are effective and accessible in real-world conditions. For policymakers, there is a need to establish clear regulatory frameworks and guidelines governing the use of AI and explainability in healthcare. Policies should emphasize transparency, accountability, and patient safety, ensuring that AI systems provide explanations that are both accurate and understandable. Policymakers should also support the development of national standards for evaluating explainable AI systems, as well as promote investments in digital infrastructure to reduce disparities in access. In addition, public education initiatives are necessary to improve digital and health literacy, enabling citizens to engage more effectively with AI-driven healthcare technologies. For healthcare institutions, the adoption of explainable AI should be accompanied by organizational readiness and capacity building. Healthcare providers need training to understand and interpret AI-generated explanations so they can effectively integrate these tools into clinical decision-making. Institutions should also develop internal guidelines for the responsible use of AI systems, ensuring that human oversight remains central in the diagnostic process. Moreover, explainability can be leveraged as a communication tool to enhance patient engagement, allowing clinicians to better explain diagnoses and treatment decisions supported by AI. Overall, this study underscores that the successful implementation of explainable AI in mobile health systems requires not only technological advancement but also collaboration among developers, policymakers, and healthcare providers. By prioritizing transparency, usability, and contextual relevance, stakeholders can ensure that AI-driven healthcare solutions are not only accurate but also trustworthy, ethical, and widely accepted.

Reference

Akande, O. A. (2020). Leveraging explainable AI models to improve predictive accuracy and ethical accountability in healthcare diagnostic decision support systems. *World Journal of Advanced Research and Reviews*, *8*(2), 415–434.

Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., & Consortium, P. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 310.

Barda, A. J., Horvat, C. M., & Hochheiser, H. (2020). A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. *BMC Medical Informatics and Decision Making*, *20*(1), 257.

Barnum, C. M. (2020). *Usability testing essentials: Ready, set... test!* Morgan Kaufmann.

Beaudouin, V., Bloch, I., Bounie, D., Clémençon, S., d'Alché-Buc, F., Eagan, J., Maxwell, W., Mozharovskyi, P., & Parekh, J. (2020). Flexible and context-specific AI explainability: a multidisciplinary approach. *ArXiv Preprint ArXiv:2003.07703*.

Bjerring, J. C., & Busch, J. (2021). Artificial intelligence and patient-centered decision-making. *Philosophy & Technology*, *34*(2), 349–371.

Cutillo, C. M., Sharma, K. R., Foschini, L., Kundu, S., Mackintosh, M., Mandl, K. D., & 1, M. I. in H. W. W. G. B. T. 1 C. E. 1 C. C. 1 G. K. 1 G. V. 1 J. R. 8 S. B. 9 S. N. (2020). Machine intelligence in healthcare—perspectives on

trustworthiness, explainability, usability, and transparency. *NPJ Digital Medicine*, *3*(1), 47.
Dang, H. D. (2011). *A latent transition analysis of self-efficacy among men treated for cocaine dependance*. University of California, Los Angeles.
Horn, C., Pitman, D., & Potter, R. (2019). An evaluation of the visualisation and interpretive potential of applying GIS data processing techniques to 3D rock art data. *Journal of Archaeological Science: Reports*, *27*, 101971.
Kennedy, G., & Gallego, B. (2019). Clinical prediction rules: a systematic review of healthcare provider opinions and preferences. *International Journal of Medical Informatics*, *123*, 1–10.
Liao, Q. V., & Varshney, K. R. (2021). Human-centered explainable ai (xai): From algorithms to user experiences. *ArXiv Preprint ArXiv:2110.10790*.
Lyu, D., Yang, F., Kwon, H., Dong, W., Yilmaz, L., & Liu, B. (2021). Tdm: trustworthy decision-making via interpretability enhancement. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *6*(3), 450–461.
Malvey, D., & Slovensky, D. J. (2014). *mHealth: transforming healthcare*. springer.
Noei, E., Zhang, F., Wang, S., & Zou, Y. (2019). Towards prioritizing user-related issue reports of mobile applications. *Empirical Software Engineering*, *24*(4), 1964–1996.
Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful explanations of black box AI decision systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 9780–9784.
Ramar, V. A., & Rathna, S. (2018). AI-Driven Cloud-Based Deep Learning for Predictive Healthcare Analytics: Enhancing Disease Diagnosis with CNNs in Medical Imaging. *Chinese Traditional Medicine Journal*, *1*(5), 12–18.
Sandelowski, M., & Barroso, J. (2003). Creating metasummaries of qualitative findings. *Nursing Research*, *52*(4), 226–233.
Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551.
Sokol, K., & Flach, P. (2020). One explanation does not fit all: The promise of interactive explanations for machine learning transparency. *KI-Künstliche Intelligenz*, *34*(2), 235–250.
Van Der Waa, J., Nieuwburg, E., Cremers, A., & Neerincx, M. (2021). Evaluating XAI: A comparison of rule-based and example-based explanations. *Artificial Intelligence*, *291*, 103404.
Wang, S., Qureshi, M. A., Miralles-Pechuan, L., Huynh-The, T., Gadekallu, T. R., & Liyanage, M. (2021). Applications of explainable AI for 6G: Technical aspects, use cases, and research challenges. *ArXiv Preprint ArXiv:2112.04698*.
Ye, T., Xue, J., He, M., Gu, J., Lin, H., Xu, B., & Cheng, Y. (2019). Psychosocial factors affecting artificial intelligence adoption in health care in China: cross-sectional study. *Journal of Medical Internet Research*, *21*(10), e14316.
You, Y., Kou, Y., Ding, X., & Gui, X. (2021). The medical authority of AI: A study of AI-enabled consumer-facing health technology. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–16.
Zecca, L., & Cotza, V. (2021). Distance Education and Beyond: A Student Voice Research toward an Ecological Perspective. *Book of Abstracts of 7th International Conference on Education (ICEDU 2021)*, 128.
Zhang, Z., Genc, Y., Wang, D., Ahsen, M. E., & Fan, X. (2021). Effect of AI explanations on human perceptions of patient-facing AI-powered healthcare systems. *Journal of Medical Systems*, *45*(6), 64.