Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 241 (2024) 476–481

Procedia Computer Science

www.elsevier.com/locate/procedia

The 1st International Workshop on Intelligent Mobile Systems based on Internet of Things August August 5-7, 2024, Marshall University, Huntington, WV, USA

MetaHospital: implementing robust data security measures for an AI-driven medical diagnosis system

Hari Mohan Rai[a] , Dana Tsoy[b*], Yevgeniya Daineko[b]

[a]*Gachon University, 1342 Seongnam-daero, Gyeonggi, South Korea*
[b]*International Information Technology University, Manas St. 34/1, Almaty 050040, Kazakhstan*

Abstract

The paper is dedicated to the security measures within the concept of MetaHospital. The idea of MetaHospital is an answer to the modern challenges of time where interaction through digital media is no longer a fantasy. In this paper authors describe the idea of MetaHospital and its modules as well as the ways of personal data protection.
The data protection model is given and with a detailed description of each stage. Through the incorporation of rigorous data security protocols, MetaHospital guarantees the safeguarding of patient information, cultivating a trustworthy environment for medical care and research endeavours. Committed to progress and patient well-being, MetaHospital persistently refines its practices, leveraging technology and data-driven perspectives to enhance healthcare delivery and optimize patient outcomes.

Keywords: MetaHospital; cybersecurity, Metaverse; privacy; data security

1. Introduction

MetaHospital stands as a beacon of modern healthcare infrastructure, designed to cater to the evolving needs of patients in today's dynamic medical landscape. As a premier medical facility, MetaHospital prioritizes patient well-being and strives to provide comprehensive healthcare solutions encompassing diagnostics, treatment, and preventive care [1]. With state-of-the-art medical equipment, specialized departments, and a team of skilled healthcare

* Corresponding author.
E-mail address: d.tsoy@iitu.edu.kz

10.1016/j.procs.2024.08.067

professionals, MetaHospital offers a holistic approach to healthcare delivery, ensuring patients receive the highest standard of medical attention and personalized treatment plans tailored to their individual needs.

Diagnostic errors pose a formidable challenge within healthcare systems globally, exerting profound implications on patient welfare and mortality rates. As delineated in a study featured in BMJ Quality & Safety, diagnostic inaccuracies afflict an estimated 5% of adults annually in the United States alone, culminating in approximately 12 million adults encountering diagnostic discrepancies on an annual basis [2], [3]. Furthermore, findings from the Institute of Medicine (IOM) underscore the gravity of this issue, attributing diagnostic errors to an alarming 10% of patient fatalities within the United States. Consequently, the imperative for adopting AI-driven diagnostic methodologies becomes paramount, serving as a pivotal strategy to mitigate the prevalence and impact of manual diagnostic errors, thereby safeguarding patient well-being, and enhancing healthcare outcomes [4].

In parallel, the integration of artificial intelligence (AI) into medical diagnosis systems has revolutionized the field of healthcare, empowering healthcare providers with powerful tools to enhance diagnostic accuracy, streamline workflows, and improve patient outcomes. AI-driven medical diagnosis systems leverage machine learning algorithms to analyze vast amounts of medical data, including patient medical records, diagnostic images, and genomic information, to assist clinicians in making timely and accurate diagnoses [5], [6]. By harnessing the power of AI, MetaHospital's medical diagnosis system augments the capabilities of healthcare professionals, enabling them to detect diseases at earlier stages, predict treatment responses, and optimize treatment plans for better patient outcomes. The proliferation of AI-driven diagnostic methodologies is experiencing a notable surge on a worldwide scale, as hospitals and healthcare establishments harness the capabilities of artificial intelligence (AI) and machine learning (ML) technologies to augment diagnostic precision, efficacy, and patient prognoses [7], [8]. As elucidated in a comprehensive report by Grand View Research, the global market for AI in healthcare surpassed a substantial valuation of $4.9 billion in 2020, marking a significant milestone in the trajectory of technological integration within the healthcare domain. Projections indicate a robust compound annual growth rate (CAGR) exceeding 41% from 2021 to 2028 [9], underscoring the burgeoning momentum and profound potential of AI-driven solutions in revolutionizing diagnostic paradigms and reshaping the landscape of healthcare delivery on a global scale.

2. IITU concept of MetaHospital

The International Information Technology University (IITU) proposes MetaHospital, a novel concept that integrates three core components to develop essential skills in young medical specialists who lack real-world experience (Fig. 1.).

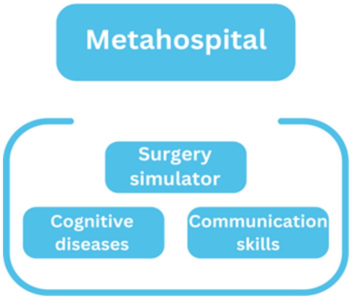

Fig. 1. Diagram of IITU' MetaHospital concept.

These components are:

- **Surgery Simulator**: This module, exemplified by the angioplasty and stenting training simulator, provides a virtual reality environment where users can practice the entire procedure. This allows them to gain practical skills in a safe and controlled setting.
- **Cognitive Disease Diagnostic Tool**: This tool focuses on Alzheimer's disease diagnosis. It utilizes a Leap Motion controller to track a user's specific set of movements, which are then analysed to assess the

patient’s potential cognitive decline. The ability to perform and sequence these movements becomes a diagnostic indicator.

- **Communication Skills Development Application**: Recognizing the importance of effective communication with diverse patients, MetaHospital incorporates a virtual patient interaction module. This application allows users to interact with simulated patients, fostering their communication skills in a realistic yet controlled environment.

3. Data security measures under MetaHospital

As the adoption of digital healthcare solutions increases, concerns regarding data security and patient privacy become more pronounced. The proliferation of data breaches within medical databases poses a grave threat to patient privacy and confidentiality, potentially precipitating a cascade of detrimental repercussions such as identity theft and financial malfeasance [9]. According to the HIPAA Journal, a staggering 642 healthcare data breaches were reported in the United States in 2020 alone, culminating in the exposure of over 30 million healthcare records to unauthorized access [10]. Furthermore, insights gleaned from the Ponemon Institute’s Cost of a Data Breach Report reveal the exorbitant financial toll exacted by such breaches, with the global average cost of a healthcare data breach soaring to $7.13 million. This alarming figure is compounded by an average cost of $429 per breached record, underscoring the dire imperative to fortify data security measures rigorously [11]. Thus, to safeguard patient data integrity and avert the perils of data breaches, the implementation of robust data security protocols is indispensable within healthcare infrastructures. MetaHospital recognizes the importance of safeguarding patient data from unauthorized access, data breaches, and cyber threats. To address these challenges, MetaHospital implements robust data security measures, including encryption protocols, access controls, data anonymization techniques, and regular security audits, to ensure the confidentiality, integrity, and availability of patient information [19]. To address these challenges, MetaHospital implements robust data security measures, including encryption protocols, access controls, and data anonymization techniques, to ensure the confidentiality, integrity, and availability of patient information. These cybersecurity measures include:

- **Security Information and Event Management (SIEM):** This system enables real-time monitoring, detection, and response to security incidents across the network infrastructure.
- **Intrusion Detection Systems (IDS):** These systems actively monitor network traffic and detect any suspicious or unauthorized activity that may indicate a security breach.
- **Incident Response Protocols:** MetaHospital has established protocols to ensure a swift and coordinated response to security incidents.

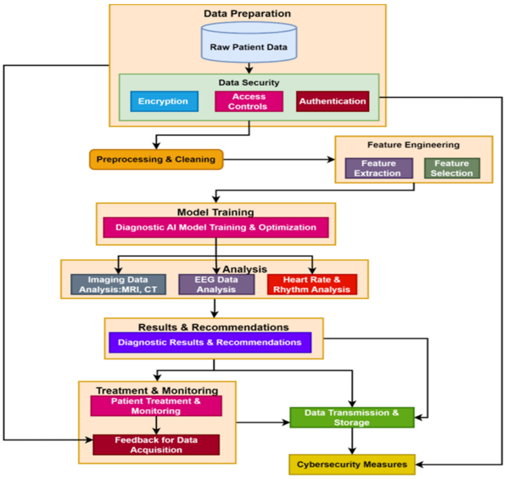
Data Preparation
Raw Patient Data
Data Security
Encryption
Access Controls
Authentication
Preprocessing & Cleaning
Feature Engineering
Feature Extraction
Feature Selection
Model Training
Diagnostic AI Model Training & Optimization
Analysis
Imaging Data Analysis:MRI, CT
EEG Data Analysis
Heart Rate & Rhythm Analysis
Results & Recommendations
Diagnostic Results & Recommendations
Treatment & Monitoring
Patient Treatment & Monitoring
Feedback for Data Acquisition
Data Transmission & Storage
Cybersecurity Measures

Fig. 2. The methodology of the proposed project.

4. Methodology

The methodology of the proposed work is delineated through a comprehensive block diagram, illustrated in Fig. 2. This diagram encapsulates the essence of the MetaHospital project, which harnesses an AI-driven Medical Diagnosis system fortified with robust data security measures.

The initial stage of MetaHospital, data acquisition, is critical for building a strong foundation. Hospitals often use different data collection and storage methods, leading to inconsistencies in format, coding standards, and terminology. Missing or inaccurate data entries can also negatively impact the quality and reliability of the datasets [20]. MetaHospital addresses these challenges through several strategies. Standardized data acquisition protocols will be established across participating hospitals to minimize inconsistencies. Patient data will be anonymized using robust techniques, while secure storage and encryption protocols will safeguard sensitive information. Furthermore, data validation and cleaning procedures will be implemented to identify and address missing values, inconsistencies, and errors within the datasets.

To safeguard the acquired datasets against unauthorized access and potential breaches, MetaHospital will implement a robust framework of data security measures. Only authorized team members with the necessary credentials and permissions will be granted access to the data repository [6]. Access privileges will be carefully managed and monitored to prevent unauthorized individuals from gaining entry to the datasets.

In the subsequent stage, the raw patient data undergoes preprocessing and cleaning to ensure data quality and integrity. By meticulously preparing the data before analysis, the accuracy and reliability of subsequent diagnostic procedures are significantly enhanced [21]. Firstly, noise is removed from the dataset. Secondly, missing values within the dataset are addressed through imputation or deletion, depending on the extent and nature of the missing data. Thirdly, the data is standardized and normalized to ensure consistency and comparability across different variables and datasets [22]. This involves scaling the data to a common range or distribution, thereby facilitating more meaningful comparisons and analyses.

In the fourth stage, advanced feature extraction and feature selection techniques will be employed on the cleaned dataset. This pivotal step aims to extract relevant features from the preprocessed data, which encapsulate important characteristics of the patient's health status. Additionally, feature selection techniques may be applied to mitigate dimensionality and enhance model performance. During feature extraction, various methodologies will be explored to extract pertinent features from the preprocessed data. Deep learning-based approaches, such as convolutional neural networks (CNNs) [21] or recurrent neural networks (RNNs), may be utilized to automatically extract meaningful features from complex medical datasets. Alternatively, traditional feature extraction methods such as Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), statistical features, and morphological features can be employed to capture distinctive patterns and structures within the data. Following feature extraction, feature selection techniques will be applied to refine the dataset further. Dimensionality reduction techniques, such as Principal Component Analysis (PCA), can be utilized to identify the most relevant features while minimizing redundancy and noise in the dataset. Additionally, other feature selection methods, such as Recursive Feature Elimination (RFE), L1 regularization, or tree-based feature selection, may be employed based on the characteristics of the dataset and the specific requirements of the AI-driven medical diagnosis system [23].

In the fifth stage, the AI diagnostic algorithm is applied to conduct the analysis and prediction of the utilized dataset. This stage represents the culmination of the preceding steps, where the insights gleaned from data preprocessing, feature extraction, and selection are leveraged to develop robust diagnostic models tailored to specific patient data types. For instance, algorithms specialized in detecting arrhythmias are applied to heart rate data, while those geared towards identifying abnormalities in electroencephalogram (EEG) data are utilized for brain-related disorders [24]. Similarly, algorithms capable of detecting tumors are deployed for medical imaging data, such as X-rays, MRIs, or CT scans. To facilitate accurate diagnosis and prediction, a variety of deep learning models are employed, each uniquely suited to handle different types of data and medical conditions. CNNs are commonly utilized for image-based data analysis, enabling the detection of features and patterns within medical images. Long Short-Term Memory (LSTM) [25] networks are particularly effective for sequential data analysis, making them well-suited for time-series data like EEG signals. Additionally, hybrid models, such as Unet-ResNet and Unet-EfficientNet, combine the strengths of multiple architectures to enhance diagnostic accuracy and robustness.

In the sixth stage, the diagnostic results and recommendations are presented based on the analysis conducted by the diagnostic algorithms. Following the comprehensive analysis of the patient's health data, the diagnostic algorithms generate actionable insights and recommendations aimed at improving patient care and treatment outcomes. The diagnostic results encompass a wide range of information, including the identification of specific medical conditions, the severity of the condition, and the likelihood of disease progression [26]. Leveraging advanced machine learning algorithms, the system can accurately detect abnormalities, anomalies, and patterns indicative of various medical conditions. In addition to providing diagnostic results, the system offers personalized recommendations tailored to the individual patient's needs.

In the subsequent 7th stage, patient treatment and monitoring is required, marking a pivotal phase in the healthcare journey following diagnostic analysis. This stage is dedicated to patient care and ongoing monitoring, leveraging the diagnostic results obtained from earlier stages to inform treatment decisions and track patient progress over time.

After successful processing, diagnosis, and treatment, the next step involves data compression and transmission to facilitate seamless communication between healthcare providers and patients. This phase is crucial for ensuring the secure and efficient transfer of medical information while maintaining patient confidentiality. By employing robust security measures, MetaHospital upholds the highest standards of data privacy and security, safeguarding patient confidentiality always. Additionally, MetaHospital implements Remote Monitoring and Alerting systems to enable continuous monitoring of patients' health status, even outside traditional healthcare settings.

At the end of the process, MetaHospital prioritizes the integrity and confidentiality of patient data by implementing robust cybersecurity measures. These measures are essential for safeguarding sensitive medical information against evolving cyber threats and ensuring compliance with regulatory requirements [30]. MetaHospital integrates advanced security information and event management (SIEM) systems, which enable real-time monitoring, detection, and response to security incidents across the network infrastructure.

5. Conclusion

This paper has presented MetaHospital, a novel AI-driven framework designed to revolutionize healthcare delivery. MetaHospital prioritizes patient privacy and security throughout the entire diagnostic process, from data acquisition to treatment planning. By leveraging advanced AI algorithms and secure data management practices, MetaHospital facilitates accurate diagnoses, personalized treatment plans, and real-time patient monitoring. This holistic approach, centered on the individual patient, fosters improved communication, and ultimately leads to better healthcare outcomes. Meta Hospital's potential extends beyond the concepts presented here. Future work will involve exploring the integration of specific data types and tailoring data protection approaches to address evolving security threats. Nevertheless, the groundwork laid by MetaHospital demonstrates the immense potential of AI technology to transform healthcare delivery and empower patients to take an active role in their health journey.

Acknowledgements

This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP14871641).

References

[1] Filippini, Massimo, and Lester C. Hunt. (2011) "Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach." *Energy Journal* **32** (**2**): 59–80.

[2] Filippini, Massimo, and Lester C. Hunt. (2012) "US residential energy demand and energy efficiency: A stochastic demand frontier approach." *Energy Economics* **34** (**5**): 1484–1491.

[3] Weyman-Jones, Thomas, Jùlia Mendonça Boucinha, and Catarina Feteira Inàcio. (2015) "Measuring electric energy efficiency in Portuguese households: a tool for energy policy." *Management of Environmental Quality: An International Journal* **26** (**3**): 407–422.

[4] Saunders, Harry (2009) "Theoretical Foundations of the Rebound Effect"', in Joanne Evans and Lester Hunt (eds) *International Handbook on the Economics of Energy*, Cheltenham, Edward Elgar

[5] Sorrell, Steve (2009) "The Rebound Effect: definition and estimation", in Joanne Evans and Lester Hunt (eds) *International Handbook on the Economics of Energy*, Cheltenham, Edward Elgar

[1] G. Ramkumar, J. Seetha, R. Priyadarshini, M. Gopila, and G. Saranya, “IoT-based patient monitoring system for predicting heart disease using deep learning,” Measurement, vol. 218, p. 113235, Aug. 2023, doi: 10.1016/j.measurement.2023.113235.

[2] M. L. Graber, “The incidence of diagnostic error in medicine,” BMJ Qual Saf, vol. 22, no. Suppl 2, pp. ii21–ii27, Oct. 2013, doi: 10.1136/bmjqs-2012-001615.

[3] M. L. Graber et al., “Cognitive interventions to reduce diagnostic error: a narrative review,” BMJ Qual Saf, vol. 21, no. 7, pp. 535–557, Jul. 2012, doi: 10.1136/bmjqs-2011-000149.

[4] D. E. Newman-Toker et al., “Rate of diagnostic errors and serious misdiagnosis-related harms for major vascular events, infections, and cancers: Toward a national incidence estimate using the ‘big Three,’” Diagnosis, vol. 8, no. 1, pp. 67–84, 2021, doi: 10.1515/dx-2019-0104.

[5] N. Taimoor and S. Rehman, “Reliable and Resilient AI and IoT-Based Personalised Healthcare Services: A Survey,” IEEE Access, vol. 10, pp. 535–563, 2022, doi: 10.1109/ACCESS.2021.3137364.

[6] A. Krishnan, S. Swarna, and B. H. S, “Robotics, IoT, and AI in the Automation of Agricultural Industry: A Review,” in 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/B-HTC50970.2020.9297856.

[7] A. Almalawi, A. I. Khan, F. Alsolami, Y. B. Abushark, and A. S. Alfakeeh, “Managing Security of Healthcare Data for a Modern Healthcare System,” Sensors, vol. 23, no. 7, p. 3612, Mar. 2023, doi: 10.3390/s23073612.

[8] B. Murdoch, “Privacy and artificial intelligence: challenges for protecting health information in a new era,” BMC Med Ethics, vol. 22, no. 1, p. 122, Dec. 2021, doi: 10.1186/s12910-021-00687-3.

[9] “AI In Healthcare Market Size, Share & Trends Analysis Report By Component (Software Solutions, Hardware, Services), By Application (Virtual Assistants, Connected Machines), By Region, And Segment Forecasts, 2024 - 2030.” Accessed: Mar. 20, 2024. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-healthcare-market#

[10] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, “Setting Privacy ‘by Default’ in Social IoT: Theorizing the Challenges and Directions in Big Data Research,” Big Data Research, vol. 25, p. 100245, Jul. 2021, doi: 10.1016/j.bdr.2021.100245.

[11] Steve Alder, “December 2020 Healthcare Data Breach Report,” HIPAA Journal.

[12] A. Kosvyra, E. Ntzioni, and I. Chouvarda, “Network analysis with biological data of cancer patients: A scoping review,” J Biomed Inform, vol. 120, no. December 2020, p. 103873, 2021, doi: 10.1016/j.jbi.2021.103873.

[13] F. M. Al-Naima, A. H. Ali, and S. S. Mahdi, “Data acquisition for myocardial infarction classification based on wavelets and neural networks,” 2008 5th International Multi-Conference on Systems, Signals and Devices, SSD’08, pp. 3–8, 2008, doi: 10.1109/SSD.2008.4632817.

[14] S. Udipi, “The event data management problem: getting the most from network detection and response,” Network Security, vol. 2021, no. 1, pp. 12–14, Jan. 2021, doi: 10.1016/S1353-4858(21)00008-8.

[15] H. M. Rai, “Cancer detection and segmentation using machine learning and deep learning techniques: a review,” Multimed Tools Appl, 2023, doi: 10.1007/s11042-023-16520-5.

[16] H. M. Rai and J. Yoo, “Analysis of Colorectal and Gastric Cancer Classification: A Mathematical Insight Utilizing Traditional Machine Learning Classifiers,” Mathematics, vol. 11, no. 24, p. 4937, Dec. 2023, doi: 10.3390/math11244937.

[17] H. M. Rai and J. Yoo, “A comprehensive analysis of recent advancements in cancer detection using machine learning and deep learning models for improved diagnostics,” Journal of Cancer Research and Clinical Oncology, vol. 149, no. 15. Springer Science and Business Media Deutschland GmbH, pp. 14365–14408, Nov. 01, 2023. doi: 10.1007/s00432-023-05216-w.

[18] H. M. Rai and K. Chatterjee, “Hybrid adaptive algorithm based on wavelet transform and independent component analysis for denoising of MRI images,” Measurement (Lond), vol. 144, pp. 72–82, 2019, doi: 10.1016/j.measurement.2019.05.028.

[19] H. M. Rai and K. Chatterjee, “Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data,” Applied Intelligence, Aug. 2021, doi: 10.1007/s10489-021-02696-6.

[20] R. Zhang, J. Jia, and R. Zhang, “EEG analysis of Parkinson’s disease using time–frequency analysis and deep learning,” Biomed Signal Process Control, vol. 78, Sep. 2022, doi: 10.1016/j.bspc.2022.103883.

[21] H. M. Rai and K. Chatterjee, 2D MRI image analysis and brain tumor detection using deep learning CNN model LeU-Net, vol. 80, no. 28–29. Springer US, 2021. doi: 10.1007/s11042-021-11504-9.

[22] I. Keshta, “AI-driven IoT for smart health care: Security and privacy issues,” Inform Med Unlocked, vol. 30, p. 100903, 2022, doi: 10.1016/j.imu.2022.100903.

[23] K. Swanson, E. Wu, A. Zhang, A. A. Alizadeh, and J. Zou, “From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment,” Cell, vol. 186, no. 8, pp. 1772–1791, 2023, doi: 10.1016/j.cell.2023.01.035.

[24] J. Swift, “Assessment and treatment of patients with acute unstable bradycardia,” Nurs Stand, vol. 27, no. 22, pp. 48–56, 2013.

[25] A. Hilbert et al., “Data-efficient deep learning of radiological image data for outcome prediction after endovascular treatment of patients with acute ischemic stroke,” Comput Biol Med, vol. 115, Dec. 2019, doi: 10.1016/j.compbiomed.2019.103516.

[26] H. Xu and G. Zhai, “ECG data compression based on wave atom transform,” MMSP 2011 - IEEE International Workshop on Multimedia Signal Processing, pp. 1–5, 2011, doi: 10.1109/MMSP.2011.6093793.

[30] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” Cybersecurity, vol. 2, no. 1, p. 20, Dec. 2019, doi: 10.1186/s42400-019-0038-7.