Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 246 (2024) 5388–5395

Procedia Computer Science

www.elsevier.com/locate/procedia

28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)

Deployment of Artificial Intelligence Models for Sleep Apnea Recognition in the Sleep Laboratory

Ángel Serrano Alarcón*[a], Maksym Gaiduk[b], Natividad Martínez Madrid[a], Ralf Seepold[b], Juan Antonio Ortega[c]

[a] *Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany*

[b] *HTWG Konstanz, Alfred-Wachtel-Str. 8, 78462 Konstanz, Germany*

[c] *University of Seville, San Fernando 4, 41004 Sevilla, Spain*

Abstract

There are a large number of scientific publications that focus on the development and evaluation of artificial intelligence (AI) models for the detection of various pathologies in the field of sleep medicine. However, most of these publications do not show the process or methodology to be followed for the final deployment of these models in a complete diagnostic system (in terms of software and hardware). This is a major drawback when translating from the development or research environment to the real clinical setting. This work focuses on a methodology for deploying an AI model for sleep apnea detection with the end user in mind: the clinician. For the deployment, the transmission of data between the device, the cloud platform and the machine learning server, as well as the protocols used, were considered. In addition, the storage and visualization of the data has been taken into account so that it can be analyzed accurately by experts.

Peer-review under responsibility of the scientific committee of the 28th International Conference on Knowledge Based and Intelligent information and Engineering Systems

Keywords: *Deep Learning; Sleep/Wake state; Artificial Intelligence; Sleep Disorder*

* Corresponding author. Tel.: +49 7121 271 4094
E-mail address: angel.serrano_alarcon@reutlingen-university.de

Peer-review under responsibility of the scientific committee of the 28th International Conference on Knowledge Based and Intelligent information and Engineering Systems
10.1016/j.procs.2024.09.665

1.Introduction

The development of artificial intelligence models in both academic and industrial contexts is a current priority [1, 2, 3]. The exponential growth of artificial intelligence has the potential to revolutionize numerous aspects of our lives [4]. However, it is in the field of medicine where the most promising results are expected, due to the potential for greater precision, higher speed in the diagnosis of diseases and the development of more effective treatments based on the characteristics of each patient [6,7,8]. However, the development of artificial intelligence models in medicine is one of the fields where it is most challenging for the models to be used in a real clinical environment after their development [5,9,10]. This is due to the fact that the development of the models and validation is done with data from real patients, where privacy should be of paramount importance [11]. Moreover, the accuracy of the outcome of these models is of particular importance, as if the models are developed to assist doctors in decision-making, a poor prediction may result in a misdiagnosis of the disease or an inadequate treatment, which carries significant risk [5,12,13]. Another disadvantage of the utilization of artificial intelligence models in the field of medicine is the heterogeneity of clinical environments [14]. This presents a challenge in obtaining a sufficient quantity and quality of data for the development of models [15].

As stated, the development of AI models presents a number of challenges, yet another technical obstacle hinders the wider deployment of AI models in both a scientific and a clinical environment. A review of the scientific literature reveals an extensive number of models developed for the detection of various diseases [16,17,18,19]. These models demonstrate reasonable accuracy for use in the clinical environment. However, few scientific papers demonstrate how AI models can be deployed in a manner that enables end users, such as doctors, to utilize them [20]. In addition, depending on the medical field in which the model is applicable, there is one crucial aspect when developing artificial intelligence models in medicine: the explainability of the model [5]. The inability to analyze the predictions of AI models renders them unsuitable for use in the medical field. This aspect should be taken into account during the deployment of models. In the field of sleep medicine, where the diseases it harbours can lead to other pathologies, the approach used to generate results is particularly important [21,22].

Conversely, the deployment of AI models for the purpose of disease detection, in particular sleep apnoea, is technically complex. The machine learning model must be deployed and integrated with the device that collects the signals from the patient. Polysomnography is the most commonly employed method for the detection of sleep apnea, although alternative approaches utilizing portable monitors are becoming increasingly prevalent. Consequently, this methodology will be employed to elucidate the methodology employed for the deployment of AI models for apnea recognition. In general, the models should operate with a software platform (which may be in the cloud) for data transmission and management [6].

As already mentioned, there are not many scientific works in the literature that show a way to deploy artificial intelligence models in the field of medicine [23]. This is not only something that should be done at the end of the development stage of a complete system, but also during the development stages of each component of the system to check if the system's performance is in accordance with the requirements that were established before starting development. Our approach in this work also aims to provide the opportunity to perform an evaluation of the performance of the models in a real simulation, where raw data is received, and the generated prediction is sent back to a cloud platform.

The first objective of this paper is to describe the process by which an artificial intelligence model previously developed and evaluated for sleep apnea detection has been deployed and integrated for use by the end user. This process has been designed to demonstrate the model's potential for use in a real clinical setting in the future. The second aim of this scientific work is not only to show the integration of a deep learning model in a system whose purpose is to be used in a real clinical environment, but also to provide support during the software development stages. In addition, this paper emphasizes the functionality of the model within the system, since it is useless for the artificial intelligence model to be very precise if it cannot be correctly integrated into the system. In addition to the deployment of the deep learning model, the operation of the rest of the system components should be in accordance with the machine learning server.

2. Materials and Methods

The following section outlines the details pertaining to the data set utilized, the data preprocessing techniques employed, the software tools used in the process, and the deployment of the deep learning models

2.1 Dataset

The main objective of this project is to deploy the deep learning models for use with data collected by a handheld device. However, due to the concurrent development of the models and the portable monitor, it was not feasible to make the data available. Consequently, data from existing repositories was utilised. Two data sets were employed to identify the optimal model for sleep/wake state calculation. The data from patients in the Sleep Heart Health Study (SHHS) repository were used for the training and validation of the model [24,25]. The Sleep Heart Health Study (SHHS) is a multicentre cohort study conducted by the National Heart, Lung, and Blood Institute (NHLBI) with the objective of determining the cardiovascular and other consequences of sleep-disordered breathing. Two datasets were generated from the SHHS: a first polysomnogram (SHHS Visit 1) was obtained in 3,295 participants, and a second polysomnogram (SHHS Visit 2) was obtained in 3,295 of the same participants. The server's performance with the deep learning models was evaluated using eight hours of patient data.

2.2 Infrastructure

Although this work is focused on the details regarding the deployment of deep learning models for apnoea detection, it is also important to know the complete system where these deep learning models will be deployed. Three main parts of the system can be distinguished:

- Hardware: this part of the system concerns the device that collects the physiological signals.
- Cloud Platform: this part of the system is in charge of all the logic of the medical system as a whole with the transmission of data between the medical device and the machine learning server.
- Machine learning server: it is the component of the system that includes the deep learning models that have been deployed and that receive the signals from the portable monitor and generate the prediction.

2.3 Software tools

In this section we do not cover the hardware and cloud platform. However, we analyse the machine learning server in detail. For this stage, several Python libraries were used. This was done for two reasons: first, all the development of the models was carried out in Jupyter Notebook and Keras was the library used for the development of the models. This library also facilitates the subsequent saving and reuse of the models. In addition, Jupyter Notebook allows us to easily test the communication between the cloud platform and the machine learning server without spending large amounts of development time and avoiding possible problems by setting up a conventional server. Therefore, the already trained model was loaded into a Jupyter Notebook and the following libraries were used for the development of the machine learning server through an Message Queuing Telemetry Transport (MQTT) client such as Paho (2.1.0), a library for handling JSON files such as jsonlib (1.6.1), a library for manipulating data structures such as NumPy (1.26.4) and libraries for file compression and decompression such as gzip , zipfile and zlib for Python (3.10)

2.4 Preprocessing of the signals

In order to reduce the amount of memory occupied by the signals sent from the cloud platform to the machine learning server, the data was compressed before being sent by using one of the libraries for compression stated in Section 2.3. Therefore, the first task that is done in the machine learning system is the decompression of the data. The data collected by the patient is not standardized and large artifacts were not removed. This stage occurs on the machine learning server. The signals were standardized by applying statistical procedure z-score. Large artefacts were removed by interpolation when SpO2 values were below 80% and HR values were not between 150 or 40 bpm. After preprocessing the signals, the model then is fed with them and generates predictions. In this case, it determines whether the data points of the signals belong to an apnea event or not. The distinction is made by sending 0s in cases of no apnea and 1s in cases of apnea.

3. Results

In order to have an overview of the system and to be able to analyze the organization of the system components and their mutual relationships and dependencies, the system is shown in Figure 1. There are three completely differentiated components. A component called hardware that represents the medical device that is responsible for collecting physiological signals, applying signal preprocessing to improve quality and subsequently storing the signals in an internal storage system in the device itself. This internal storage can be a removable SD card.

On the other hand, we have the component called Cloud Platform. From the Hardware component, the signals are sent to the Cloud Platform for the processing and management of the data in various ways. Firstly, in the Cloud Platform the data will be sent to the Machine Learning Server. The original data coming from the Hardware component is stored in the Cloud Platform.

Once the response is received with the predictions made in the Machine Learning Server, these predictions are stored along with the original signals previously stored. A data visualization tool is provided so that doctors can analyse the data. The most important part and the one that this manuscript focuses on is the Machine Learning Sever in which the task of predicting apnea events takes place. Data transmission is carried out through the MQTT protocol. The data is compressed in the Cloud Platform and sent to an MQTT broker under a certain topic. The Machine Learning Server that acts as an MQTT client is subscribed to that topic and when there is data to make predictions, it receives such data.

In the Machine Learning Server, the data is decompressed, large artifacts are eliminated, down sampling or up sampling the signals to 1Hz and finally the standardization of the signal. The data then feeds the deep learning algorithms and generates predictions of 0s (no apnea) and 1s (apnea) that are sent back to the Cloud Platform. Both the arrays containing the signals and the predictions have the same dimensions. Since the model predictions are at a resolution of 1s and all signals have a frequency of 1HZ. Once the prediction is complete, it is sent back through the MQTT protocol to the MQTT broker. The Cloud Platform is also subscribed to a certain topic to name the predictions, so it receives the predictions to be able to be displayed alongside the original signals.

Although it is not visible as a component in Figure 1, one of the most important parts of the system is the MQTT communication between the Cloud Platform and the Machine Learning Server. For this communication, an MQTT broker is required, which can either be integrated into the cloud platform or be an external broker to which the Cloud Platform and the Machine Learning Server publish and subscribe to the signals and predictions, respectively.

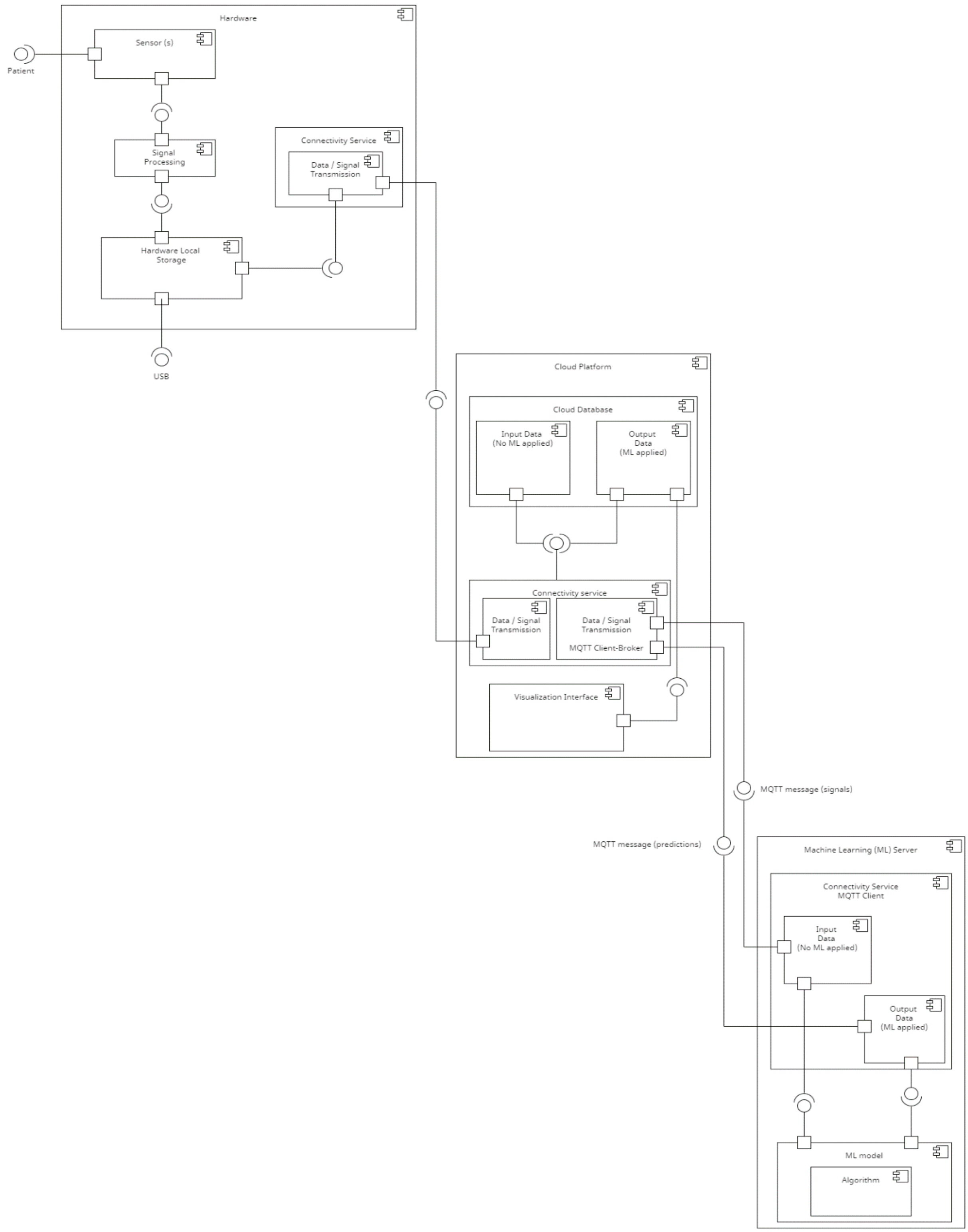

Figure 1. Component diagram of the system including hardware part, cloud platform and machine learning server.

4. Discussion

The majority of artificial intelligence algorithms developed for use in a real clinical environment lack the necessary information to be effectively implemented. Primarily, the majority of scientific work that develops artificial intelligence algorithms are focused on obtaining results with high accuracy. While this is essential, there are still numerous aspects to consider. Even if the level of accuracy is high, without these additional factors, the use of the model would not be possible. One of these factors is the explainability of the model and also its deployment.

In this study, our primary focus was on the deployment stage of the deep learning model within an integrated system. Concurrently, this permits the assessment of the functionality of the model in its integration with a comprehensive apnea detection system. Concurrently, the limited number of deployment stages for the Machine Learning Server renders it suitable for accurate and expeditious testing of its operational functionality. The data transmission protocol selected, namely MQTT, is optimal for the transfer and receipt of data, as it enables the server to remain in a state of "listening" at all times, functioning as an MQTT client and thus capable of transmitting predictions to the Cloud Platform as soon as they are ready.

It is important to consider the workload that the Machine Learning Server can support, as well as the exceptions that can be thrown due to receiving signals with errors or data formats that do not align with the models' input specifications. One of the advantages of utilising this technology for data transmission is that the system can expand by incorporating additional machine learning servers without compromising the efficacy of the solution, as other types of pathologies can be identified using the same signals published from the cloud platform via the MQTT client.

Furthermore, the replacement of the Machine Learning Server with another server in the event of unavailability of the initial server represents another advantage. Consequently, the overall functionality of the system would remain unimpaired. However, this methodology for deploying deep learning models is not without its disadvantages. In our case, we have employed the use of a Jupyter Notebook to facilitate the testing of the system. However, it is recommended that this methodology be redefined in the future, with the utilisation of a server that exhibits greater functionality and reliability. The Jupyter Notebook is an effective tool for data analysis, but it is prone to kernel crashes when run for extended periods.

A further study is necessary to ascertain whether the Machine Learning Server is capable of processing a large volume of data. In addition, it will be necessary to investigate whether it is more profitable to increase the number of models that can be used to predict different outputs within the same Machine Learning Server, or whether it would be more feasible to make the system scalable by adding a greater number of Machine Learning Servers containing different deep learning models. Although this appears to be the optimal solution, it is also necessary to consider the resources consumed in terms of computation and energy, given the high consumption of graphics processing units (GPUs) that perform the inference task.

5. Conclusion

The development of artificial intelligence (AI) models for use in any area of industry is a widespread phenomenon. One of the areas where advances in AI have the greatest appeal is in medicine. Although the focus is often on the results in terms of accuracy obtained by the models, one of the keys for the models to be widely used in a real clinical environment is their deployment. This work presents a software solution where a machine learning server is configured for the detection of sleep apnoea by creating an MQTT client. This enables artificial intelligence models to receive signals from medical devices to be evaluated and generate a prediction, in this case whether the patient has apnoea or not. Future work should focus on developing the number of models for the detection of other pathologies.

Acknowledgements

This research was partially funded by the German Federal Ministry for Economic Affairs and Energy, ZiM project "Sleep Lab at Home" (SLaH) grant: ZF4825301AW9 and Carl Zeiss Foundation: Project "Non-invasive system for measuring parameters relevant to sleep quality" (project number: P2019-03-003).

References

[1] I. Ahmed, G. Jeon and F. Piccialli, "From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where," in IEEE Transactions on Industrial Informatics, vol. 18, no. 8, pp. 5031-5042, Aug. 2022, doi: 10.1109/TII.2022.3146552.

[2] L. Xiaojun, Z. Xiande, Z. Kexi, D. Zhenli, Z. Kai and W. Xi, "Study on the Application Fields and Development Prospects of Artificial Intelligence," 2021 2nd International Conference on Artificial Intelligence and Education (ICAIE), Dali, China, 2021, pp. 101-105, doi: 10.1109/ICAIE53562.2021.00029.

[3] S. Guo, "On the influence of AI technology on accounting industry and its Countermeasures," 2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), Guangzhou, China, 2021, pp. 214-218, doi: 10.1109/AIEA53260.2021.00053.

[4] S. Yadav et al., "The Role of AI in Healthcare Policy Development and Management," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-6, doi: 10.1109/ICAIIHI57871.2023.10489810.

[5] Alarcón ÁS, Madrid NM, Seepold R, Ortega JA. (2023)."Obstructive sleep apnea event detection using explainable deep learning models for a portable monitor". Front Neurosci.;17:1155900. doi: 10.3389/fnins.2023.1155900. PMID: 37521695; PMCID: PMC10375719.

[6] Maksym Gaiduk et al. (2021) "Design of a sleep apnoea detection system for a home environment", Procedia Computer Science, Volume 192, Pages 3225-3234, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.09.095.

[7] Hamet P, Tremblay J. Artificial intelligence in medicine. Metabolism. 2017 Apr;69S:S36-S40. doi: 10.1016/j.metabol.2017.01.011. Epub 2017 Jan 11. PMID: 28126242.

[8] Bonkhoff AK, Grefkes C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. Brain. 2022 Apr 18;145(2):457-475. doi: 10.1093/brain/awab439. PMID: 34918041; PMCID: PMC9014757.

[9] Gaiduk M, Serrano Alarcón Á, Seepold R, Martínez Madrid N. (2023) "Current status and prospects of automatic sleep stages scoring: Review". Biomed Eng Lett. Jul 10;13(3):247-272. doi: 10.1007/s13534-023-00299-3. PMID: 37519865; PMCID: PMC10382458.

[10] Bhat M, Rabindranath M, Chara BS, Simonetto DA. Artificial intelligence, machine learning, and deep learning in liver transplantation. J Hepatol. 2023 Jun;78(6):1216-1233. doi: 10.1016/j.jhep.2023.01.006. PMID: 37208107.

[11] Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial Intelligence Transforms the Future of Health Care. Am J Med. 2019 Jul;132(7):795-801. doi: 10.1016/j.amjmed.2019.01.017. Epub 2019 Jan 31. PMID: 30710543; PMCID: PMC6669105.

[12] Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull. 2021 Sep 10;139(1):4-15. doi: 10.1093/bmb/ldab016. PMID: 34405854.

[13] Tiwari PC, Pal R, Chaudhary MJ, Nath R. Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. Drug Dev Res. 2023 Dec;84(8):1652-1663. doi: 10.1002/ddr.22115. Epub 2023 Sep 15. PMID: 37712494.

[14] Fennelly O, Grogan L, Reed A, Hardiker NR. Use of standardized terminologies in clinical practice: A scoping review. Int J Med Inform. 2021 May;149:104431. doi: 10.1016/j.ijmedinf.2021.104431. Epub 2021 Feb 25. PMID: 33713915.

[15] Stanfill MH, Marc DT. Health Information Management: Implications of Artificial Intelligence on Healthcare Data and Information Management. Yearb Med Inform. 2019 Aug;28(1):56-64. doi: 10.1055/s-0039-1677913. Epub 2019 Aug 16. PMID: 31419816; PMCID: PMC6697524.

[16] Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E, Dudley JT. Artificial Intelligence in Cardiology. J Am Coll Cardiol. 2018 Jun 12;71(23):2668-2679. doi: 10.1016/j.jacc.2018.03.521. PMID: 29880128.

[17] Contreras I, Vehi J. Artificial Intelligence for Diabetes Management and Decision Support: Literature Review. J Med Internet Res. 2018 May 30;20(5):e10775. doi: 10.2196/10775. PMID: 29848472; PMCID: PMC6000484.

[18] Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. Clin Transl Sci. 2023 Nov;16(11):2106-2111. doi: 10.1111/cts.13619. Epub 2023 Aug 30. PMID: 37646577; PMCID: PMC10651639.

[19] Nayarisseri A, Khandelwal R, Tanwar P, Madhavi M, Sharma D, Thakur G, Speck-Planche A, Singh SK. Artificial Intelligence, Big Data and Machine Learning Approaches in Precision Medicine & Drug Discovery. Curr Drug Targets. 2021;22(6):631-655. doi: 10.2174/1389450122999210104205732. PMID: 33397265.

[20] REDDY, Sandeep, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ health & care informatics, 2021, vol. 28, no 1.

[21] Watson, N. F., & Fernandez, C. R. (2021). Artificial intelligence and sleep: Advancing sleep medicine. Sleep medicine reviews, 59, 101512.

[22] Goldstein, C. A., Berry, R. B., Kent, D. T., Kristo, D. A., Seixas, A. A., Redline, S., & Westover, M. B. (2020). Artificial intelligence in sleep medicine: background and implications for clinicians. Journal of Clinical Sleep Medicine, 16(4), 609-618.

[23] Pouyan Esmaeilzadeh,Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations, Artificial Intelligence in Medicine, Volume 151, 2024, 102861, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102861.

[24] G. Q. Zhang et al. (2018) "The National Sleep Research Resource: towards a sleep data commons.".J. Am. Med. Inform. Assoc. 25, 1351–1358. doi: 10.1093/jamia/ocy064

[25] Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, Rapoport DM, Redline S, Robbins J, Samet JM, Wahl PW.(1997) ”The Sleep Heart Health Study: design, rationale, and methods". Sleep. 20(12):1077-85. PMID: 9493915.