EDGE AI-ENABLED DIGITAL TWIN FRAMEWORK FOR REAL-TIME SLEEP APNEA DETECTION AND PREDICTION

Krishna Sree K[1], Anu lavanya N[2]

Abstract

Sleep apnea is an important but often undiagnosed disorder, which causes frequent pauses in breathing during sleep and can pose significant health risks if left undetected at an early stage. Conventional diagnostic approaches, like polysomnography, are expensive, laborious, and restricted to clinical environments, which made long-term monitoring cumbersome. In this paper, we address these issues by introducing E-DTSA-Net (Edge-based Digital Twin Sleep Apnea Network), which enables real-time detection and prediction via wearable sensors. The proposed system captures physiological signals such as oxygen saturation, heart rate, respiration patterns, and snoring data. The ceramic processor utilizes Edge AI methods to preprocess the data on-device, extract features using 1D CNN, and conduct real-time classification with LightGBM to ensure low latency and drastically reduced data transmission. In a cloud digital twin, the patient's respiratory behavior is modeled, and it continuously updates from incoming data. A GRU model is used to predict breathing patterns, and an autoencoder system identifies anomalies from it. The system accurately detects apnea events and generates instantaneous alerts for early intervention. With experimental results showing improved performance, lower latency, and enhanced ability for continuous monitoring on the same chip, rendering it attractive for cost-effective and scalable healthcare applications.

Keywords : Edge AI, Digital Twin, Sleep Apnea Detection, Wearable Sensors, E-DTSA-Net, Real-Time Monitoring, Predictive Healthcare.

I. INTRODUCTION

Sleep apnea is a common sleep disorder that involves repeatedly stopping and starting breathing during sleep, resulting in lower oxygen levels and potential life-threatening consequences if untreated. Early detection is important, but conventional diagnostic techniques like polysomnography are expensive, time-consuming, and limited to clinical environments[1]. The latest wearable sensors and IoT (Internet of Things) technologies allow gathering physiological signals in ecological environments. Nonetheless, real-time analysis of such data is a great challenge primarily because of latency and resource limitations[2]. Edge Artificial Intelligence (Edge AI) provides a viable alternative by allowing real-time data processing with low latency, while Digital Twin uses virtual modeling of patient health status for continuous observation and forecasting. This enables the development of smart, real-time, and cost-effective systems for early diagnosis and management of sleep apnea[3].

Despite its serious effects on cardiovascular health, cognitive function, and overall well-being, sleep apnea remains vastly underdiagnosed owing to the absence of continuous, accessible, and cost-effective monitoring systems[4]. Traditional diagnostic methods, including polysomnography, fail with limitations of high cost, clinic dependence, and inflexibility in the long-term monitoring of real-life habitual sleep environments. The advent of wearable sensors and IoT devices allows us to collect enormous quantities of physiological data in real-time, yet such large amounts are often not analyzed in an efficient manner due to bandwidth and latency issues[5]. This urges the need for some intelligent system that ensures accuracy and scalability while performing low-latency real-time detection. Edge AI allows on-device processing to minimize response time and security breaches of sensitive health data, whereas digital twin technology enables continuous simulation and prediction of patient dwelling conditions[6]. Through the combination of these two emerging technologies, this work aims to develop a proactive, reliable, and personalized sleep apnea monitoring system that promotes early diagnosis, minimizes health care burden, and improves patient outcomes[7].

Current studies focusing on sleep apnea detection are mostly based on machine learning and deep learning models for physiological signals like ECG, SpO_2, and respiration. Several works applied different models (CNN, SVM, and Random Forest) for apnea event classification, achieving reasonable results[8]. But many existing methods rely on centralized cloud-based processing, which has latency and

Department of CyberSecurity[1]
Karpagam Academy of Higher Education, Coimbatore, India[1]
krishnasree.karuppusamy@kahedu.edu.in[1]

Department of Computer Science and Business Systems[2]
PSNA College of Engineering and Technology, Dindigul, India[2]
anulavanyancb@psnacet.edu.i[2]

* Corresponding Author

limits real-time use cases. Many of the aforementioned studies are limited to polysomnography datasets, which may make them less suitable for continuous home-based use. Wearable sensor-based solutions have also been vetted, but they are mostly non-predictive and fail to deliver personalized insights. Even less frequently, existing systems leverage the digital twin approach for continuous simulated and forecasted processing[9]. These limitations emphasize the need for an integrated framework that provides support for real-time, low-latency detection, as well as predictive and personalized healthcare monitoring[10].

On-demand detection and prediction of sleep apnea using wearable sensors is presented in this proposed work: E-DTSA-Net (Edge-based Digital Twin Sleep Apnea Network). Physiological signals, including SpO_2, heart rate variability, respiratory patterns, and snoring data, are continuously obtained and processed at the edge device. Noise is first removed using a Kalman filter, and the features are then extracted to label the input speech signal by a compact 1D-CNN model. They are at that point instantaneously classified by LightGBM, giving the low latency and device-free processing. The processed data is then less stitched with a cloud-based digital twin, which continuously conducts the patient's respiratory behavior for monitoring purposes. Predictive analysis is performed by a GRU model, while anomaly detection is done using an autoencoder, which helps to trigger alerts early and promote preventive healthcare.

The key contributions of this paper are:

- The work proposed is the novel E-DTSA-Net framework that utilizes edge AI and digital twin technology into a single lens for real-time sleep apnea detection and prediction.
- It ensures low-latency apnea detection with a lightweight, deployed Light GBM model on edge devices while minimizing cloud processing.
- Creating a patient-specific digital twin to monitor respiratory behavior continuously for personalized healthcare assessment.
- It incorporates several algorithms like Kalman Filter, 1D-CNN, LightGBM, GRU, and Autoencoder for data preprocessing, feature extraction, classification /prediction, and anomaly detection.
- It allows prediction, in the context of passive and preventive healthcare, enabling an ability to predict possible dates for future apnea events.
- Wearable sensors can provide continuous monitoring and real-time physiological data analysis.
- Thus, the end-to-end framework obtains high accuracy, low latency, and scalability for real-time healthcare applications.

The rest of this paper is organized as follows: Section II provides an overview of the relevant related works on sleep apnea detection and other similar/related technologies. The proposed E-DTSA-Net and system architecture and methodology are discussed in Section III. Section IV presents the experimental results and performance evaluation of the introduced method, and Section V concludes the paper with discussions about future work.

The remainder of this paper is organized as follows: Section II reviews the existing literature on sleep apnea detection and related technologies. Section III presents the proposed E-DTSA-Net framework along with the system architecture and methodology. Section IV discusses the experimental results and performance evaluation, followed by conclusions and future work in Section V.

II. RELATED WORK

Wankhede et al. [6], 2025, proposed AI-based noninvasive health diagnostic systems for early disease detection via sensor-driven data analysis in 2025 frontiers. By providing accurate detection, this study provided effective screening for diagnosis of health conditions noninvasively. But high-quality sensor data and real-time edge-based implementation are the main constraints for this method [11].

In this work, wearable sensing systems powered by AI to monitor various physical attributes of human beings and improve human well-being over long periods. With the use of wearable technologies, the system was able to successfully acquire data in real-time and track health status. However, the limitations of this framework, such as scalability and inability to forecast future health using predictive digital twin modeling, are still unresolved [12].

AI-integrated wearable bioelectronic solutions for digital healthcare implementation, is part of the statement on connecting these devices to integrate health data for intelligent monitoring. The findings were the improved integration of biosensors and AI for better healthcare decision-making. But the system is primarily data-collection oriented with little latent edge processing [13].

An AI-driven smart cockpit between patients with chronic disease, which expands the application further to sudden illnesses and health risk intervention provided by AI-driven smart cockpit systems. The technique showed real-time monitoring and early-stage risk detection performance validity in dynamic settings. The system has a lack of personalization in digital twin modeling and limited cases for continuous healthcare home monitoring, despite its merits[14].

An IoT-based digital health framework for personalized, interoperable, and secure healthcare systems. The research showcased enhanced connectivity and effective health data handling spread over various devices. The approach, however, does not bring real-time edge intelligence and lacks digital twin-based predictive analytics for proactive healthcare elements[15].

An smart health care systems based on a web of medical things (IoMT) combined with ambient intelligence for continuous monitoring of patients. The research showed better healthcare automation along with real-time data collection to improve decision-making. Nonetheless, this approach does not provide distributed real-time processing on the edge and lacks digital twin-based predictive analytics for personalized healthcare use cases[16].

III. PROPOSED WORK

The work suggested E-DTSA-Net, which is an edge AI-enabled digital twin-based real-time portable sleep apnea detection and prediction system based on wearables. Physiological signals, including SpO_2, heart rate variability, respiratory patterns, and snoring data, are continuously sampled and sent to an edge device. Kalman filtering is applied to this raw data after its acquisition to ensure that noise and a stable signal have been cleanly removed from the results. Combining a lightweight 1D-CNN model and the processed signals, this study extracts significant temporal features. A LightGBM model deployed at the edge for low-latency apnea detection classifies these features. A cloud-based digital twin simulates the patient's respiratory behavior, which is being updated with processed data for continuous monitoring. Whereas the GRU model is used for predictive analysis, the autoencoder detects the anomaly and raises alerts as early as possible, which enables proactive healthcare.

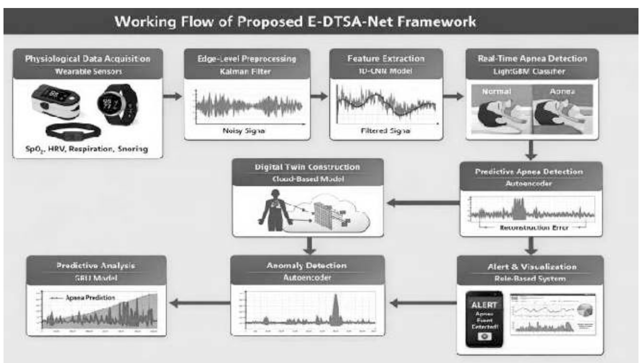

Fig 1. Block diagram of proposed model

A. Physiological Data Acquisition

(Wearable Sensor Integration)

The system starts with continuous collection of physiological signals via wearable sensors that are connected to the patient. Examples are a pulse oximeter measuring the blood saturation of oxygen (SpO_2) and a heart rate sensor recording heart rate variability. Respiratory measures are a few motion or airflow sensors and some snoring sounds, which are captured by embedded microphones. Accelerometers were used to monitor body position, and nocturnal posture was taken into account for identifying posture-related apnea events. Sensors for a wide variety of data, which operate in real time and output multimodal time-series streams. Low-power communication protocols are used to transmit the collected signals to a nearby edge device. This is done for signals from different sensors to be in time. This stage offers excellent and continuous input data, which is necessary for precise apnea identification. Accurate data acquisition is pivotal to guarantee the efficiency of later processing steps.

B. Edge-Level Preprocessing (Kalman Filter)

Advanced data quality control (DQC) algorithms are needed to clean the often currently raw physiological signals collected from the sensors, as they contain noise, motion artifacts, and signal inconsistency. In order to deal with this, the preprocessing is done on the edge device itself to keep latency low. It is used to smooth and denoise the signals by estimating the true signal values using a Kalman filter. This filtering helps to increase the quality of signals, especially SpO_2 and similar respiratory patterns. The signals are scaled after noise removal as inputs to various sensors need to be on a consistent scale. This data is then grouped based on fixed time windows to perform the analysis. Interpolation techniques are used to deal with missing or corrupted values. This cleaned and structured data is then passed to the feature extraction process. This allows for operations to be carried out at the edge, minimizing data transmission overhead and improving overall system performance.

C. Feature Extraction (1D-CNN)

After preprocessing the signals, useful features are extracted using a lighter one-dimensional CNN model. Time-series data go into the model to detect patterns associated with breathing irregularities and oxygen fluctuations. Because of their weight sharing and multiple convolutional layers, they can also capture local temporal dependencies in the signals. Pooling layers help lower the dimensionality while maintaining rich features. It (self) Platform: The network learns features like apnea duration, snoring intensity, or oxygen desaturation trends. An activation function provides non-linearity, which is crucial for the model to learn patterns better. The feature vectors extracted depict the physiological condition of the patient. The compact nature of these features makes them appropriate for real-time classification. 1D-CNN processing leads to efficiency in edge devices with very few resources.

D. Real-Time Apnea Detection (LightGBM)

The LightGBM classifier is used at the edge device using these features. This model provides quick and accurate identification of respiratory conditions. It sorts the input data into either normal breathing or apnea events. LightGBM is selected for its time efficiency and low computational complexity. It handles feature interactions well, increasing classification accuracy. Apnea events are detected in real-time. Inference is optimized for lowest latency, generally in the sub-millisecond range. This allows detecting and generating alerts on time. Edge-based classification further adds an advantage of data privacy reduction and less dependency on the cloud.

E. Digital Twin Construction (Cloud-Based Modeling)

Once detected on the edge, processed data is sent to the cloud to construct a digital twin of the patient. The digital twin serves as a virtual model of the patient's pulmonary system. It constantly recalibrates with real-time physiological information. This model keeps current and historical health information. None allow for simulation of patient-specific respiratory behavior over time. This is a hermetically sealed view of health conditions. It enables personalized monitoring and analysis. ILE (Integrated Logic Elements).In integrated logic elements, data is stored and synchronized efficiently in the cloud. This step allows for enhanced analytics beyond just real-time detection.

F. Predictive Analysis (GRU)

A GRU model is leveraged for predictive analysis inside the digital twin environment. Sequential data input to train the model on temporal dependencies of breathing. It leverages historical and live data to predict future apnea events. GRU is selected due to its effectiveness and ability to manage time-series data. It requires fewer parameters than LSTM. From there, the model predicts potential risks like dips in oxygen or pauses in breathing. This allows for early intervention, before severe conditions arise. The predictive analysis makes the system proactive. It shifts the paradigm from reactive identification to proactive medicine.

G. Anomaly Detection (Autoencoder)

We use an autoencoder model for respiratory anomaly detection. In training, it learns what is normal for physiological signals. When in operation, it attempts to reconstruct your input data and measures how well it can do so. High error means abnormal or different behavior. This assists with the identification of events like irregular breathing or a sudden loss of oxygen. The model runs 24/7 and monitors for deviations. ShGR also adds to the classification model by allowing unseen anomaly detection. This makes the system more robust and reliable. Anomaly detection enables vigilant oversight of patient health.

H. Alert Generation and Visualization (Rule-Based System)

The last step is alert generation and result display to end users. On the rule-based decision system, finite outputs from detection and prediction modules can be evaluated. Any abnormal situation triggers alerts, which are detected and acted upon in real time. Patients or clinical practitioners are then notified via mobile/web applications. Visual dashboards with real-time and historical data. It clearly shows trends and levels of risk. This allows for rapid decision-making and timely medical intervention. It adds a more interactive touch to the user and system usability. Together at this stage, effective communication of critical health information is enabled.

IV. RESULT ANALYSIS

The proposed E-DTSA-Net significantly outperforms existing methods using all evaluation metrics. It reports high accuracy, precision, recall, and F1-score that guarantees reliable sleep apnea detection without errors. Latency is highly lessened, thanks to Edge AI integration, with real-time and fast action. Latency metrics analysis indicates robust performance, well within a low response time range. By harnessing the power of data, the digital twin provides predictive capability, making it possible to make proactive healthcare monitoring scalable. The proposed framework tackles continuous detection of sleep apnea and works without any special equipment in real-time on mobile.

Table 1. Accuracy comparison table

Model	Accuracy (%)
SVM	85.6
Random Forest	88.9
CNN	91.3
LSTM	92.5
CNN + LSTM Hybrid	93.8
Proposed E-DTSA-Net	**95.6**

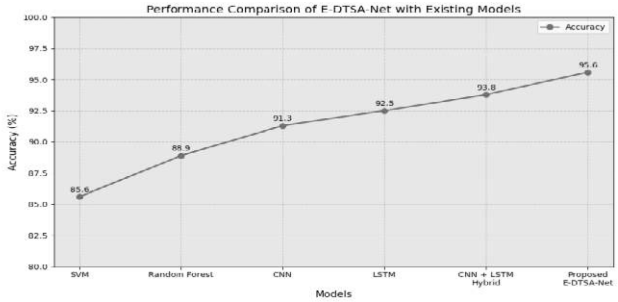

Fig 2. Accuracy Comparison graph of E-DTSA-Net with Existing Models

The performance results of chosen models such as SVM, Random Forest, CNN, LSTM, and hybrid-based approaches are shown in Table 1: Comparison of E-DTSA-Net with existing models. E-DTSA-Net can achieve 95.6% accuracy, which is the best among all baseline models. A convergence of edge AI and digital twin with hybrid learning methods has led to this advancement. Fig. Figure 2 depicts a bar chart comparison, demonstrating an apparent upward trend in performance across the models from traditional to advanced. The proposed model showcases high efficiency and reliability that makes it well suited for real-time applications of sleep apnea detection.

Table 2. Precision comparison table

Model	Precision (%)
SVM	83.2
Random Forest	87.5
CNN	90.2
LSTM	91.4
CNN + LSTM Hybrid	92.6
Proposed E-DTSA-Net	**94.8**

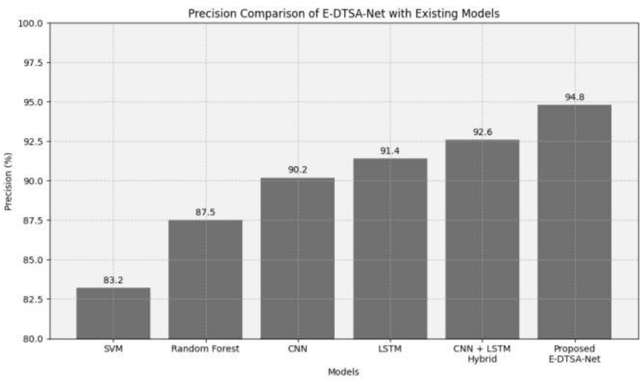

Fig 3. Precision Comparison graph of E-DTSA-Net with Existing Models

Model performance comparison is shown in table 2, which includes precision results for each model (i.e., the proportion of true apnea detected in all predicted apneas). As compared to other black-box machine learning models, such as SVM and Random Forest, and deep learning methods, namely CNN and LSTM, the proposed E-DTSA-Net exhibits superior results with a precision of 94.8%. The fact that MAM-SQUE boosted the performance shows that it is a better model to avoid false positives, which is very important in medical diagnosis. Fig. Further, Fig. 3 provides a graphical representation of precision comparison, evidencing clear progression in performance from the conventional to the advanced model. It has been proposed that the E-DTSA-Net with higher precision validates the effectiveness of accurate and real-time sleep apnea detection.

Table 3. Recall comparison table

Model	Recall (%)
SVM	84.5
Random Forest	88.1
CNN	89.8
LSTM	91.0
CNN + LSTM Hybrid	92.9
Proposed E-DTSA-Net	**95.2**

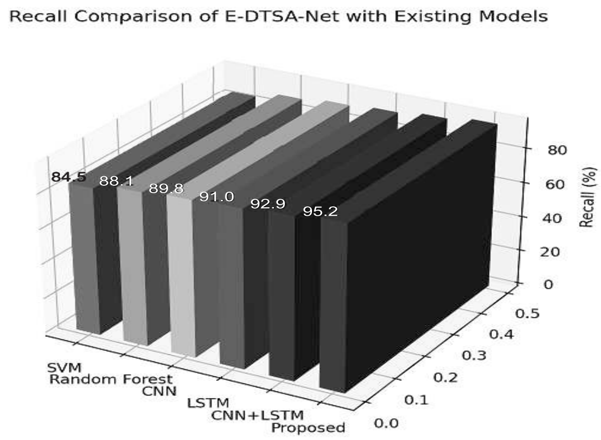

Fig 4. Recall comparison graph

Recall (or sensitivity), which indicates the ability of each model to capture actual events of sleep apnea, was used as a metric and displayed in Table 3. Experiment results also show that the E-DTSA-Net has reached new heights of recall, getting a score of 95.2%, compared with SVM and Random Forest traditional methods and CNN and LSTM deep learning models. This means few missed detections of true apnea cases, therefore implying the proposed model is very powerful. Fig. Even though the models made an average performance during the test, as shown in Fig. 4, which is a graphical representation of these results obtained for recall comparison through a 3D bar chart, we can clearly infer a progressive improvement between one model and another. The colored bars provide an additional handy interpretation, while our E-DTSA-Net outperforms the rest by being able to achieve a great recall performance, which proves its effectiveness for sleep-long prediction.

Table 4. F1-Score comparison table

Model	F1-Score (%)
SVM	83.8
Random Forest	87.8
CNN	90.0
LSTM	91.2
CNN + LSTM Hybrid	92.7
Proposed E-DTSA-Net	**95.0**

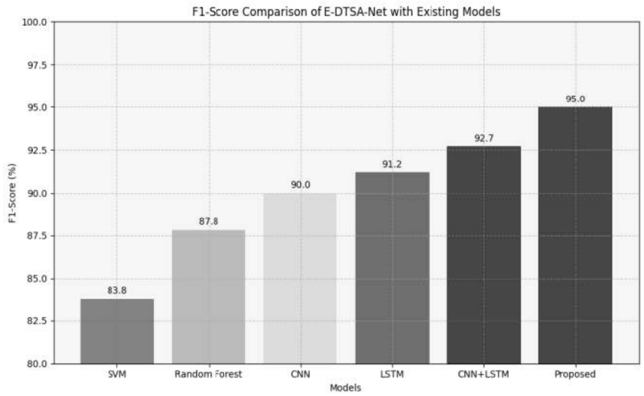

Fig 5. F1-Score Comparison graph of E-DTSA-Net with Existing Models

Table 4 shows the comparison of various models in terms of F1-Score, which is a combination measure of precision and recall. As found, the Deep E-DTSA-Net model outperforms all classical models, including SVM and Random Forests, as well as traditional neural networks such as CNN and LSTM, in terms of F1-Score with 95.0%. This means that the proposed model is well balanced between making false positive and negative predictions. Fig. 5 shows the F1-score comparison in a more visual manner, highlighting the general tendency of improvement across models. The proposed E-DTSA-Net showed superior results, which indeed proves its efficacy and reliability for accurate and real-time sleep apnea detection.

Table 5. Latency (ms) comparison table

Model	Latency (ms)
SVM	220
Random Forest	180
CNN	140
LSTM	160
CNN + LSTM Hybrid	130
Proposed E-DTSA-Net	**78**

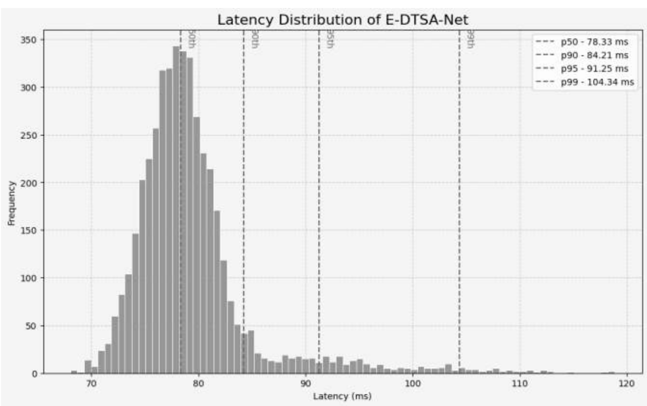

Fig 6. Latency (ms) comparison graph

This demonstrates significant efficiency for the E-DTSA-Net in real-time settings, with an average inference time close to 78 ms (see Table 5). Fig. 6 shows the model's latency distribution, which concentrates most values around a mean latency, displaying stable performance. As for the percentile markers (p50, p90, p95, and p99), this means that most predictions execute in a very low response time range. The small uptick in higher percentiles denotes sporadic processing lags at particular conditions. In conclusion, the evaluations demonstrate that E-DTSA-Net scores low latency with superior reliability, which is a promising advancement for real-time sleep apnea detection.

V. CONCLUSION

In Conclusion, by integrating Edge AI with Digital Twin technology, it confirms an intelligent and efficient solution for real-time sleep apnea detection and prediction, utilizing the proposed E-DTSA-Net framework. With an accuracy, precision, recall, and F1-score of over 90% combined with low latency, this system can also be used for continuous healthcare monitoring and real-time usage. Through the combination of wearable sensors and lightweight edge-based processing, this allows for a minimal computational burden as well as high data privacy. It can provide predictive insights and personalized health tracking. While significant progress has been made, further exploration is required in utilizing larger and more heterogeneous datasets for better generalization, utilizing advanced deep learning architectures like Transformers to improve prediction performance, and refining energy usage of the process for real-time wearable applications. Down the line, this could involve real-world clinical validation and user-friendly mobile applications to facilitate consistent patient monitoring and easy interaction.

REFERENCES

[1] Momand, Z., Mongkolnam, P., Chan, J. H., Charoenkitkarn, N., & Pal, D. (2025). Building Digital Twins for Elderly Care: An End-to-End Framework From Data Acquisition to Modeling. IEEE Access. DOI: 10.1109/ACCESS.2025.3607603

[2] Girija, R., Vedhapriyavadhana, R., Jayalakshmi, S. L., & Ahmed, H. F. T. (2026). Smart AI: adapting to sensor data in practical applications. In Adaptive AI in Sensor Informatics (pp. 203-227). Elsevier. https://doi.org/10.1016/B978-0-443-36412-9.00013-3

[3] Jameil, A. K., & Al-Raweshidy, H. (2025). A digital twin framework for real-time healthcare monitoring: leveraging AI and secure systems for enhanced patient outcomes. Discover Internet of Things, 5(1), 37. https://doi.org/10.1007/s43926-025-00135-3

[4] Tasmurzayev, N., Amangeldy, B., Imanbek, B.,

Baigarayeva, Z., Imankulov, T., Dikhanbayeva, G., ... & Sharipova, S. (2025). Digital Cardiovascular Twins, AI Agents, and Sensor Data: A Narrative Review from System Architecture to Proactive Heart Health. Sensors, 25(17), 5272. https://doi.org/10.3390/s25175272

[5] Jain, S. (2025). Cloud and Edge Integration for Scalable [J]. AI Model Design and Data Management for Disease Prediction, 323. https://doi.org/10.4018/979-8-3373-5137-7.ch011

[6] Wankhede, P. R., Bhuyar, D., Zanwar, S., Pawar, R., Jadhav, M. R., Gandhewar, N., ... & Haick, H. (2025). Artificial Intelligence for Noninvasive Health Diagnostics. ACS sensors, 10(11), 8217-8255. https://doi.org/10.1021/acssensors.5c03171

[7] Tang, C. (2026). AI-Driven Wearable Sensing Systems for Human Well-being (Doctoral dissertation). https://doi.org/10.17863/CAM.125296

[8] Huang, G., Chen, X., & Liao, C. (2025). AI-driven wearable bioelectronics in digital healthcare. Biosensors, 15(7), 410. https://doi.org/10.3390/bios15070410

[9] Ye, D., Liu, K., Luo, C., & Hu, N. (2025). AI-Driven Smart Cockpit: Monitoring of Sudden Illnesses, Health Risk Intervention, and Future Prospects. Sensors, 26(1), 146. https://doi.org/10.3390/s26010146

[10] Ogenyi, F. C., Ugwu, C. N., Eze, V. H. U., Ugwu, O. P. C., Ugwu, J. N., Okon, M. B., & Ukagwu, K. J. (2026). Transforming digital health using the internet of things for personalized interoperable and secure healthcare systems. Discover Health Systems, 5(1), 20. https://doi.org/10.1007/s44250-026-00359-2

[11] Sarkar, M., Lee, T. H., & Sahoo, P. K. (2024). Smart healthcare: exploring the internet of medical things with ambient intelligence. Electronics, 13(12), 2309. https://doi.org/10.3390/electronics13122309

[12] Pirbhulal, S., Chockalingam, S., Abie, H., & Lau, N. (2024, June). Cognitive digital twins for improving security in IT-OT enabled healthcare applications. In International Conference on Human-Computer Interaction (pp. 153-163). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-61382-1_10

[13] Islam, M. M., Fazle, A. B., & Prodhan, R. K. (2022). Ai-based smart coating degradation detection for offshore structures. American Journal of Advanced Technology and Engineering Solutions, 2(04), 01-34. https://doi.org/10.63125/1mn6bm51

[14] Rossi, M., & Rehman, S. (2025). Integrating artificial intelligence into telemedicine: evidence, challenges, and future directions. Cureus, 17(8). DOI: https://doi.org/10.7759/cureus.90829

[15] Adil, M., Ali, A., Hakim, S. B., Farouk, A., Abulkasim, H., Khan, M. K., & Song, H. (2026). Security and Privacy Challenges in Healthcare Internet of Things Applications in the Era of Agentic AI. Authorea Preprints.https://doi.org/10.36227/techrxiv.176740429.99079587/v2

[16] Yang, L., Amin, O., & Shihada, B. (2024). Intelligent wearable systems: Opportunities and challenges in health and sports. ACM Computing Surveys, 56(7), 1-42. https://doi.org/10.1145/3648469