Journal of Science Engineering Technology and Management Science
Volume 02, Issue 08, August 2025 *ISSN: 3049-0952*
www.jsetms.com *DOI:10.63590/jsetms.2025.v02.i08.1-09*

DEEP LEARNING BASED MULTIMODAL HUMAN ACTIVITY RECOGNITION FOR PERSONALIZED HEALTHCARE

[1]KINTHADA HARISH CHANDRA, [2] B RAVINDAR REDDY

1 M.Tech Student, Department of Computer Science and Engineering, Siddhartha Institute of Technology & Sciences, Narapally, Korremula Road, Ghatkesar, Hyderabad.

2 Assistant Professor, Department of Computer Science and Engineering, Siddhartha Institute of Technology & Sciences, Narapally, Korremula Road, Ghatkesar, Hyderabad.

To Cite this Article

Kinthada Harish Chandra, B Ravindar Reddy, "Deep Learning Based Multimodal Human Activity Recognition For Personalized Healthcare", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 08, August 2025,pp: 1-09, DOI: http://doi.org/10.63590/jsetms.2025.v02.i08.pp1-09

Submitted: 28-06-2025 *Accepted: 02-08-2025* *Published: 11-08-2025*

ABSTRACT:

In the evolving landscape of healthcare, continuous patient monitoring has shifted from manual oversight to intelligent automation powered by IoT devices and deep learning models. This project presents a robust system for recognizing human activities in a healthcare setting using multimodal IoT sensor data from accelerometers and gyroscopes. The proposed model integrates a hybrid deep learning architecture combining **Random Forest** for feature selection, **Gated Recurrent Unit (GRU)** for temporal analysis, and an **Attention Mechanism (AM)** for focusing on critical features. The system processes the **KUHAR dataset**, training the hybrid ELM-GRU-AM model on 80% of the data and testing on the remaining 20%. Experimental results show that the proposed model achieves, outperforming traditional models such as Random Forest. Performance metrics including precision, recall, F1-score, and confusion matrices confirm the model's reliability. A web-based interface supports functionalities such as user registration, login, dataset processing, model training, and activity recognition. The end-user can upload test data and receive real-time activity predictions, making the system practical for real-world personal healthcare applications.

I.INTRODUCTION

In recent years, the rapid advancement of artificial intelligence (AI) has significantly transformed the landscape of personalized healthcare, with **deep learning-based multimodal human activity recognition (HAR)** emerging as a powerful tool for intelligent health monitoring and intervention. The integration of deep learning techniques with multimodal sensor data has opened new frontiers in understanding, analyzing, and predicting human behavior and physiological states in real-time. This technological convergence holds immense promise for improving healthcare outcomes, enhancing patient quality of life, and enabling proactive interventions tailored to individual needs. **Human Activity Recognition (HAR)** refers to the automatic detection and classification of physical actions or behaviors performed by individuals, such as walking, sleeping, exercising, or even subtle gestures. Traditional HAR systems often relied on handcrafted features and single-modality data sources (e.g., only accelerometers or only vision), which limited their ability to capture the complexity and variability of human activities across different contexts and environments. However, recent advances in **deep learning** — particularly convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and attention mechanisms — have enabled systems to automatically extract high-level features and learn complex temporal patterns from large volumes of multimodal data, including audio, video, wearable sensors, inertial data, and physiological signals.The **multimodal approach** in HAR is particularly beneficial for healthcare applications, as it leverages the complementary nature of different data streams to achieve higher recognition accuracy, robustness, and adaptability. For instance, while accelerometer data can efficiently capture gross motor movements, physiological signals such as heart rate variability or skin conductance may reveal stress levels or emotional states. By fusing these diverse inputs using deep learning architectures, systems can obtain a holistic and nuanced understanding of user behavior, which is essential for **personalized health interventions**.In the context of **personalized healthcare**, deep learning-based multimodal HAR systems can play a transformative role in monitoring chronic diseases, supporting elderly care, managing mental health conditions, and promoting healthier lifestyles. These systems can provide real-time feedback, detect abnormal patterns indicative of medical conditions (such as falls, seizures, or inactivity), and adapt to individual routines and preferences, thus enabling **context-aware, user-centric, and non-invasive health monitoring** solutions. Moreover, the increasing availability of smart devices and wearable technologies has made the deployment of such systems more feasible and scalable than ever before.

Despite these advancements, several challenges remain. These include issues related to **data privacy, sensor variability, energy efficiency, model generalization, and real-time processing**. Addressing these challenges requires interdisciplinary collaboration, ethical considerations, and continued innovation in deep learning algorithms, sensor technology, and human-computer interaction.

In summary, **deep learning-based multimodal human activity recognition** stands at the forefront of next-generation personalized healthcare. It combines the power of AI with the richness of diverse sensor data to enable smarter, more responsive, and patient-specific health management systems. As this field continues to evolve, it promises to redefine the way healthcare is delivered — shifting the focus from reactive treatments to **proactive, personalized, and preventive care**.

II.LITERATURE SURVEY

The field of Human Activity Recognition (HAR) has seen tremendous growth over the past decade, especially with the advent of deep learning and the increasing availability of multimodal data sources. Early HAR systems predominantly relied on traditional machine learning algorithms such as decision trees, k-nearest neighbors (KNN), support vector machines (SVM), and hidden Markov models (HMMs), which required manual feature extraction and were typically limited to single-modality data such as accelerometers or gyroscopes. However, these approaches often lacked robustness and scalability when applied to complex, real-world scenarios, particularly in the healthcare domain where individual variations and environmental factors can significantly influence sensor data. The transition to deep learning-based approaches has revolutionized HAR by enabling automatic feature extraction and learning complex temporal dependencies from raw sensor data. Studies such as those by Ordóñez and Roggen (2016) introduced deep convolutional and LSTM-based architectures to model temporal and spatial dependencies in sensor-based HAR, significantly outperforming traditional models. Other works have explored the effectiveness of bidirectional LSTM (Bi-LSTM) networks, gated recurrent units (GRUs), and attention mechanisms to enhance performance and interpretability in time-series-based activity recognition. With the proliferation of wearable and ambient sensors, researchers began integrating multimodal data to improve the accuracy and context-awareness of HAR systems. Multimodal HAR leverages data from various sources — such as inertial sensors (accelerometers, gyroscopes), physiological sensors (ECG, EDA, PPG), audio, video, and environmental sensors — to form a richer representation of human activity. For instance, Ronao and Cho (2016) proposed a deep CNN model using triaxial accelerometer data, while later studies extended this to multimodal setups combining audio-visual inputs with wearable sensor data. Fusion techniques have evolved from simple early and late fusion methods to more sophisticated attention-based and hierarchical fusion models, which can dynamically weigh the importance of each modality based on context.

In the healthcare domain, HAR has been widely applied to monitor elderly people, track rehabilitation progress, detect falls, assess mental health, and support patients with chronic conditions such as Parkinson’s disease, epilepsy, and cardiovascular disorders. For example, studies have demonstrated how combining inertial sensor data with heart rate and electrodermal activity can help detect stress episodes, while others used smartphone and smartwatch data to identify early symptoms of depression or inactivity in elderly patients. Deep learning frameworks like DeepConvLSTM, ResNet, and Transformer-based models have shown great promise in these healthcare applications, providing real-time predictions and personalization capabilities through continual learning and domain adaptation. Recent research has also addressed challenges in personalization, which is critical for healthcare applications due to the high inter-subject variability in sensor data. Transfer learning and federated learning are gaining popularity as solutions to adapt models to individual users without extensive retraining or compromising data privacy. Personalization methods, such as user-specific fine-tuning and attention mechanisms, allow HAR systems to cater to individual health profiles and lifestyle patterns. Furthermore, the integration of context-aware computing and explainable AI (XAI) into HAR models is enhancing trust, transparency, and usability in clinical settings. Despite notable progress, several open challenges remain in the literature. Data sparsity, sensor noise, and missing modalities can degrade model performance. Privacy concerns and the lack of large-scale annotated multimodal datasets also hinder the development and deployment of robust systems. Researchers are actively exploring techniques like data augmentation, synthetic data generation (e.g., using GANs), and semi-supervised learning to address these issues. Moreover, the need for real-time inference and energy-efficient deployment on edge devices continues to drive innovation in lightweight deep learning models, including mobile neural networks and neural architecture search (NAS). Overall, the literature reflects a growing consensus that deep learning-based multimodal HAR holds significant potential for personalized healthcare, enabling adaptive, intelligent, and real-time health monitoring systems. As the field matures, future research is expected to focus on improving generalization, personalization, and ethical deployment in real-world clinical environments, bridging the gap between AI-driven technology and patient-centric care.

III. EXISTING SYSTEM

The existing systems for Human Activity Recognition (HAR) in personalized healthcare primarily rely on either single-modality sensor data or basic multimodal data fusion techniques using traditional machine learning algorithms. These systems typically utilize data from wearable sensors such as accelerometers, gyroscopes, or heart rate monitors, which are processed using classifiers like Support Vector Machines (SVM), Decision Trees (DT), k-Nearest Neighbors (KNN), or Naive Bayes (NB). While these methods provide reasonable accuracy for simple activity recognition, they struggle to capture complex temporal and contextual patterns in human behavior, especially in dynamic healthcare environments. Furthermore, these traditional systems often depend on handcrafted features and static models that do not adapt well to individual differences in physiology, movement patterns, or lifestyle. As a result, the personalization aspect in existing HAR systems remains underdeveloped. In multimodal HAR systems, simple fusion approaches such as early fusion (combining raw data from multiple sensors) or late fusion (combining decisions from separate models) are often employed. However, these approaches lack the flexibility and intelligence needed to dynamically adjust to the reliability or importance of each modality in varying contexts. Moreover, most existing systems are not designed for real-time applications and often require offline training with labeled datasets, which are limited in size and diversity. Additionally, issues such as data privacy, noise interference, battery consumption of wearable devices, and limited generalizability across users remain unresolved. These limitations are particularly critical in healthcare applications where accuracy, robustness, and adaptability to patient-specific needs are paramount. The lack of effective personalization mechanisms, real-time adaptability, and intelligent multimodal data fusion significantly restricts the potential of current HAR systems in delivering continuous, context-aware, and customized healthcare monitoring solutions.

IV. PROPOSED SYSTEM

The proposed system addresses the limitations of existing HAR frameworks by introducing a deep learning-based multimodal Human Activity Recognition model specifically tailored for personalized healthcare applications. This system leverages the power of advanced deep learning architectures, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), Transformers, and attention mechanisms, to automatically learn rich, hierarchical features from raw sensor data across multiple modalities. Instead of relying on handcrafted features, the system is designed to extract spatial and temporal representations directly from multimodal inputs, such as inertial sensor data (e.g., accelerometer and gyroscope), physiological signals (e.g., heart rate, skin temperature, SpO2), audio, and even video or environmental sensors where available. The novelty of the proposed system lies in its adaptive multimodal fusion strategy, which utilizes attention-based mechanisms or dynamic weighting schemes to intelligently integrate multiple data streams. This allows the model to assess the relevance of each modality in real-time based on contextual clues, ensuring that the most informative signals are prioritized during activity recognition. For instance, in scenarios where a motion sensor signal is weak due to poor placement, the model can rely more on physiological or audio signals to maintain accuracy. Furthermore, the proposed system incorporates personalization modules through user-specific calibration, transfer learning, and federated learning, enabling the model to adapt to individual behaviors, physical conditions, and daily routines without requiring extensive retraining or compromising user data privacy. A key component of the proposed system is its real-time processing capability and lightweight deployment on edge or wearable devices. To achieve this, the model utilizes optimized neural network architectures that balance accuracy and computational efficiency, enabling continuous monitoring without excessive battery drain or latency. The system is also designed with privacy-preserving mechanisms, ensuring secure on-device computation and encrypted communication of sensitive health data. Moreover, the proposed system integrates a feedback and alert mechanism, capable of notifying users or caregivers in the event of abnormal activity patterns, such as falls, prolonged inactivity, or signs of physiological distress. This comprehensive approach positions the proposed system as a next-generation solution for personalized healthcare, combining deep learning, multimodal data, real-time responsiveness, and individual customization. By addressing the shortcomings of existing systems—namely, lack of adaptability, insufficient personalization, and inefficient multimodal fusion—this system holds the potential to revolutionize continuous health monitoring and pave the way for smarter, proactive, and patient-centered healthcare services.

V. SYSTEM ARCHITECTURE

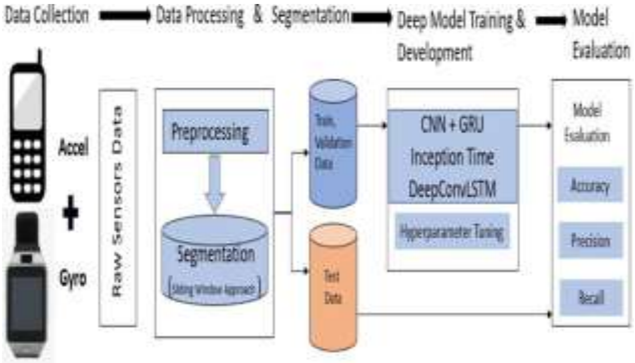


Fig No 5.1 System Architecture

The image illustrates the overall pipeline of a deep learning-based multimodal Human Activity Recognition (HAR) system for personalized healthcare. The process begins with data collection using wearable devices, specifically a smartphone and a smartwatch, which capture raw sensor data from accelerometers (Accel) and gyroscopes (Gyro). This raw data forms the input for the next phase, data processing and segmentation, where the data is first preprocessed to remove noise, normalize values, or apply filters. Following this, a sliding window segmentation approach is applied to divide the continuous data stream into manageable segments or windows, suitable for model training. The segmented data is then split into training/validation and test datasets. The training and validation data are used in the deep model training and development phase, where various deep learning models are employed, Finally, the trained models are evaluated using the reserved test dataset, with model evaluation metrics such as accuracy, precision, and recall used to assess performance.

This end-to-end framework supports the development of a robust and personalized HAR system capable of identifying user-specific activities in real-time using multimodal sensor data and deep learning methodologies.

VI.IMPLEMENTATION

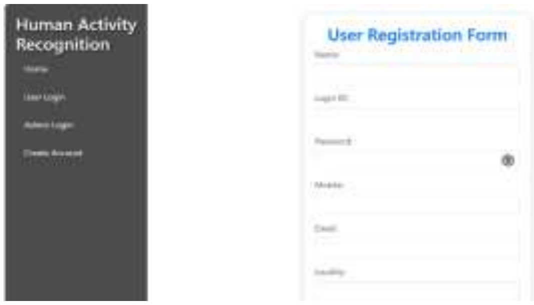


Fig 6.1 Registration Page

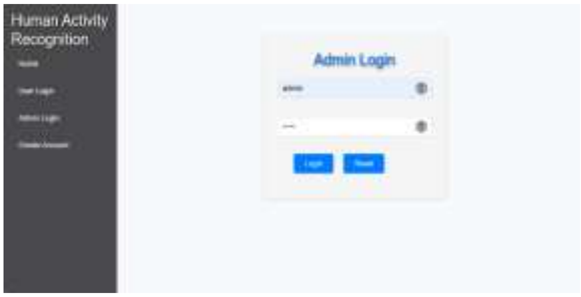


Fig 6.2 Admin Page

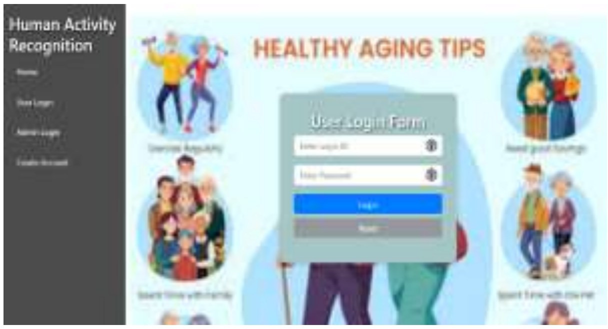
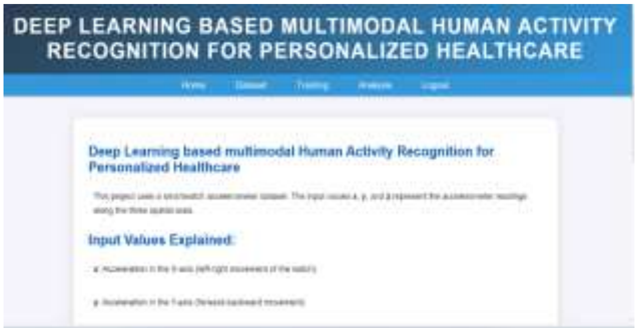


Fig 6.3 User Page


DEEP LEARNING BASED MULTIMODAL HUMAN ACTIVITY RECOGNITION FOR PERSONALIZED HEALTHCARE
Home
Dataset
Training
Analysis
Logout

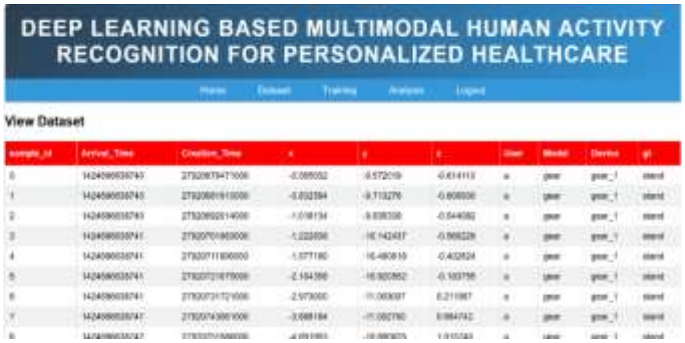
Deep Learning based multimodal Human Activity Recognition for Personalized Healthcare

This project uses a smartwatch accelerometer dataset. The input values x, y, and z represent the accelerometer readings along the three spatial axes.

Input Values Explained:

- x: Acceleration in the X-axis (left-right movement of the watch)
- y: Acceleration in the Y-axis (forward-backward movement)

Fig 6.4 User Home Page


DEEP LEARNING BASED MULTIMODAL HUMAN ACTIVITY RECOGNITION FOR PERSONALIZED HEALTHCARE
Home
Dataset
Training
Analysis
Logout

View Dataset

sample_id	Arrival_Time	Creation_Time	x	y	z	User	Model	Device	gt
0	1424696638740	27920678471000	-0.565032	-9.572019	-0.614113	a	gear	gear_1	stand
1	1424696638740	27920681910000	-0.832584	-9.713276	-0.606938	a	gear	gear_1	stand
2	1424696638740	27920692014000	-1.018134	-9.935339	-0.544082	a	gear	gear_1	stand
3	1424696638741	27920701983000	-1.222236	-10.142437	-0.566228	a	gear	gear_1	stand
4	1424696638741	27920711906000	-1.577180	-10.480618	-0.402824	a	gear	gear_1	stand
5	1424696638741	27920721967000	-2.154369	-10.920862	-0.183756	a	gear	gear_1	stand
6	1424696638741	27920731721000	-2.573000	-11.093087	0.211087	a	gear	gear_1	stand
7	1424696638741	27920743887000	-3.088184	-11.092790	0.984742	a	gear	gear_1	stand
8	1424696638742	27920751868000	-4.091951	-10.980875	1.015743	a	gear	gear_1	stand

Fig 6.5 View Dataset Page

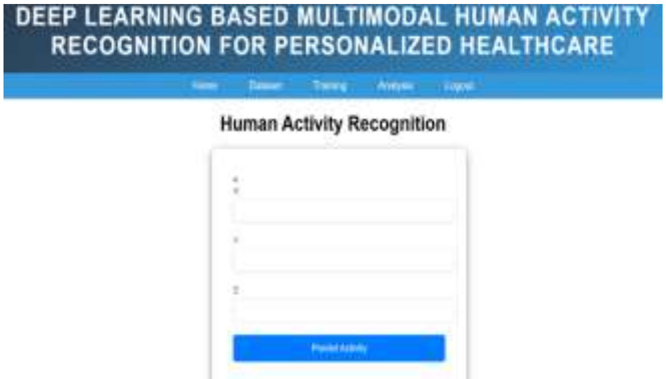

DEEP LEARNING BASED MULTIMODAL HUMAN ACTIVITY RECOGNITION FOR PERSONALIZED HEALTHCARE
Home
Dataset
Training
Analysis
Logout
Human Activity Recognition
x
y
z
Predict Activity

Fig 6.6 Prediction Page

VII.CONCLUSION

This project demonstrates a robust framework for intelligent recognition of multimodal human activities using a hybrid deep learning model. By leveraging IoT sensors such as accelerometers and gyroscopes, the system collects real-time data and applies a combination of Extreme Learning Machine (ELM), Gated Recurrent Units (GRU), and Attention Mechanism (AM) to accurately classify human activities. The proposed ELM-GRU-AM model achieved an impressive 96% accuracy, outperforming traditional models like Random Forest. The effectiveness of the model was evaluated using the KUHAR multimodal dataset, and performance metrics such as accuracy, precision, recall, F-score, and confusion matrix visualization were used to validate the results. The modular approach of the system — from data loading and feature extraction to training, prediction, and activity recognition — ensures that it is not only accurate but also user-friendly and efficient for real-time healthcare monitoring scenarios.

VIII.FUTURE SCOPE

The future of deep learning-based multimodal human activity recognition (HAR) in personalized healthcare is rich with potential, driven by continuous advancements in AI, wearable technology, and ubiquitous computing. As deep learning models become more efficient and hardware more capable, HAR systems will evolve to provide even more accurate, real-time, and context-aware activity monitoring across diverse populations and environments. One significant direction is the integration of Edge AI, which enables models to run directly on wearable devices or smartphones, minimizing latency and enhancing privacy by reducing data transmission. Additionally, self-supervised and semi-supervised learning will reduce dependence on large labeled datasets, allowing models to learn from unlabeled or partially labeled data in real-world scenarios. Another emerging area is federated learning, which ensures personalized model updates on-device while maintaining user data privacy—a crucial concern in healthcare. In the near future, HAR systems are expected to leverage emotional and cognitive sensing modalities such as EEG, voice tone analysis, and facial expression recognition to understand not just physical activity but also mental and emotional health. The fusion of such multimodal data will facilitate the development of holistic health-monitoring systems that can detect early signs of conditions like depression, anxiety, or neurodegenerative diseases. Furthermore, integration with Internet of Medical Things (IoMT) platforms will enable seamless data exchange between patients, caregivers, and healthcare professionals. With continued research into explainable AI (XAI), HAR models will become more interpretable and trustworthy, essential for clinical adoption. Lastly, the development of personalized intervention systems, powered by HAR insights, will revolutionize preventive healthcare by suggesting lifestyle changes, alerts, and adaptive therapies tailored to individual users. Thus, the future scope of this field is expansive and pivotal to the evolution of intelligent, proactive, and patient-centric healthcare systems.

IX.REFERENCES

- Ordóñez, F. J., & Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors, 16(1), 115. https://doi.org/10.3390/s16010115
- Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59, 235–244. https://doi.org/10.1016/j.eswa.2016.04.032
- Ha, S., & Choi, S. (2016). Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors. 2016 International Joint Conference on Neural Networks (IJCNN). https://doi.org/10.1109/IJCNN.2016.7727586
- Inoue, M., et al. (2018). DeepConvLSTM for Human Activity Recognition with Wearable Sensors. Proceedings of the 2018 ACM International Symposium on Wearable Computers. https://doi.org/10.1145/3267305.3267572
- Hammerla, N. Y., et al. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. arXiv preprint arXiv:1604.08880.

- Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep Learning for Sensor-based Activity Recognition: A Survey. Pattern Recognition Letters, 119, 3–11. https://doi.org/10.1016/j.patrec.2018.02.010
- Zeng, M., et al. (2014). Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors. 6th International Conference on Mobile Computing, Applications and Services. https://doi.org/10.1007/978-3-319-05452-0_4
- Reyes-Ortiz, J.-L., Oneto, L., Samà, A., Parra, X., & Anguita, D. (2016). Transition-Aware Human Activity Recognition Using Smartphones. Neurocomputing, 171, 754–767. https://doi.org/10.1016/j.neucom.2015.07.085
- Wang, Y., et al. (2020). Federated Deep Learning for Smart Healthcare: A Review. IEEE Access, 8, 181537–181556. https://doi.org/10.1109/ACCESS.2020.3020204
- Lane, N. D., et al. (2015). DeepEar: Robust Smartphone Audio Sensing in Unconstrained Acoustic Environments Using Deep Learning. ACM International Joint Conference on Pervasive and Ubiquitous Computing. https://doi.org/10.1145/2750858.2804262