Grenze International Journal of Engineering and Technology, June Issue

Transformer-Integrated YOLO–SSD for Real-Time Traffic Monitoring and Pedestrian Detection

Caroline Ruth C[1] and Denisha M[2]
[1-2]Division of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu
Email: carolineruth@karunya.edu.in, denisha@karunya.edu

***Abstract*— Robust real-time perception systems are critical to the future of autonomous vehicles and ensuring safety and compliance with regulations. Pedestrian detection and traffic sign recognition rank among the most important perception tasks that must operate with a high degree of accuracy in dynamic traffic systems. Challenges hindering robust recognition include diverse lighting conditions and backgrounds, occluded objects, and small-sized traffic signs. The proposed model introduces an integrated real-time object detection framework that incorporates Multiple-Object Object Detection model YOLO v11, SSD, and Transformer Based to perform simultaneous pedestrian and traffic sign recognitions. The proposed framework addresses both live-sourced video streams and recorded static images, including optimized preprocessing of image data, parallel extraction of features from images, confidence-based feature fusion algorithms, and optimized post-processing of feature populations. Evaluation of this proposal against various benchmark datasets and custom-created road-scene datasets showed that this hybrid framework achieved an overall mean average precision (MAP) of 89.6 percent compared to YOLOv11 82.4 %, SSD 85.7 % and the standalone transformer-based object detector 86.9%. While providing comparable performance across the various benchmarks, the hybrid framework maintains a real-time inference capability of 29 frames per second (FPS). The proposed hybrid detection system performed at the same time as the models and consistently outperformed the individual model detections while still providing real-time results.**

***Keywords*— Autonomous Driving, Pedestrian Detection, Traffic Sign Recognition, YOLO, SSD, Transformer-Based Object Detection.**

I. INTRODUCTION

AI has become a key technology used in autonomous driving and smart transport. There must be a large amount of real-time data collected about the driving scene to allow for the safety and efficiency of driving. The detection of pedestrians and traffic signs are critical perception tasks that help reduce accidents and to enable the driving operation to obey the traffic laws. The early computer vision techniques used HOG (Histogram of Oriented Gradients) and support vector machines for pedestrian detection; however, they did not work well with dynamic traffic scenes with complex. Deep learning and CNN (convolutional neural network) have provided significant improvements to the performance of detection systems. However, the real-world challenges of detection systems are due to differing lighting conditions, weather conditions, environmental conditions (i.e. occlusions), and having small or far away objects present at times make obtaining both the best accuracy for detection and the

Grenze ID: 01.GIJET.12.2.613_26
© Grenze Scientific Society, 2026

fastest time to return results a challenge and separate areas of research for many researchers working on autonomous driving systems today.

Researchers have recently looked into using state-of-the-art object detection methodologies in order to help meet these challenges. Redmon et al. developed a real-time object recognition methodology called YOLO, which has quick processing times but does not work well at detecting small items. [2]. Liu created a modified version of YOLO called the Single Shot Detector (SSD), which has shown some improvements in its ability to detect items at multiple scales but does not perform well in a cluttered setting [3]. Following this work, feature pyramid networks and multi-scale fusion were added to enhance the performance of detecting small items; however, these systems have resulted in higher computational loads [4]. In addition, more recently developed Transformer-based vision models have shown notable global context modeling capabilities but are still not able to function in real-time at this point due to their high computational costs [5]. Despite these advances, most current methodologies still independently address pedestrian and traffic sign detection. This indicates that there is a need for a unified and efficient detection methodology that provides and balances speed, accuracy, and scalability while detecting both pedestrians and traffic signs.

Recently, researchers have been exploring hybrid and multi-model detection strategies to alleviate the shortcomings of traditional methods by harnessing the combined strengths of different architectures. The integration of fast, single-stage detectors with multi-scale feature extractors and context reasoning modules has demonstrated the potential to enhance detection robustness in real-world driving environments. However, currently available hybrid frameworks are either limited to detecting only one class of object or impose a significant computational burden which inhibits their use in real-time applications. A major obstacle to developing a comprehensive unified pipeline capable of detecting both pedestrians and traffic signs remains unresolved in autonomous drive perception systems.

Overall, there is a pressing need for an integrated detection framework that can efficiently accommodate multiple input sources while sustaining high levels of accuracy and delivering real-time performance. In this paper we describe a unified object detection framework that incorporates YOLO, SSD, and Transformer-based feature enhancement methodologies to fulfill these criteria and facilitate reliable and scalable perception for autonomous vehicles and intelligent transportation systems.

II. Literature Review

The development of autonomous driving technology has driven advancements in detection methods. There are two major surveys on existing deep learning-based detection methods in the intelligent transportation systems arena, one by Chen [1] and the other by Zou [2]. While both cover similar ground, they both have persistent problems with detecting small objects, time- limited constraints on detecting objects (real-time), and degradation of performance under environmental conditions. The combination of these problems is why there remains a call for additional investigation into enhanced detection systems, including both hybrid systems and systems that are optimized for enhanced object detection.Most studies relating to the methodology behind the detection of objects have focused on single-model detection frameworks.Redmon. [3] explain how they developed the "you only look once" (YOLO) framework and that YOLO reformulated detection problems into single regression problems, yielding two benefits: an ability to turn detections around quickly and there were also several small object detections; however, those typically yielded poor performance (less than satisfactory).

Liu [4] further addressed the issue of improved accuracy for small objects by creating a single-shot frame (SSD) method that uses multi-scale feature maps; however, when it comes to detection within clutters of traffic scenes, the SSD still struggles. Lin. [5] introduced feature pyramid networks (FPN) to improve the multi-scale representation of their method; while yielding better performance within the class of small-sized detections than their original methods, they also yielded a greater computational overhead. Bochkovskiy. [6] explained how they improved YOLO using both architectural improvements and enhancements of the YOLO training algorithm to improve performance than their version but still maintained a dependency on the hardware being deployed for the real-time detection of most objects. Lastly, Carion [7] described how they used transformer models to create a unique object detection framework called DETR.

A number of research studies have been published that focus on combining accuracy and speed using hybrid methods. Zhang

[8] developed a combined pedestrian detector that used CNN-based detectors plus attention mechanisms to enhance their detection capabilities in highly complex environments, while Wang [9] created a multi-task learning framework for detecting both traffic signs and pedestrians simultaneously; this resulted in better operational efficiency but did have limited scalability. Kim [10] used temporal features from video streams to

provide improved detection consistency, although their overall latency increased when operating with higher resolution inputs. Additionally, other studies have examined ways to optimize detection performance; more specifically, He [11] found that by utilizing residual learning with object detection, they could improve the depth of the feature representations significantly. Ge [12] focused on lightweight detection models designed to run on edge devices and emphasized the trade-offs between computational efficiency and accuracy, while Liu. [13] introduced adaptive anchor mechanisms specifically for small object detection within traffic environments.

In relation to context and reasoning used to identify occluded traffic signs, Zhu [14] discussed the importance of this context in regard to identifying occluding factors out of context when attempting to identify the occluded traffic sign. The use of Vision Transformers (a new computer vision paradigm) has also gained traction as a possible technology to improve completed image understanding through using CNN + Transformer hybrid models as demonstrated by Dosovitskiy [15] and as a result of these advances has led to improvements within the original state-of-the-art CNN networks. Moreover, Sun [16] have recently proposed new methods of efficiently augmenting data within a given Traffic dataset(s), while Huang [17] have investigated and reported on the robustness of object detection systems by being exposed to adversely affected weather conditions. Tang [18] have examined multi-modal sensor fusion solutions for enhancing pedestrian safety, and Li [19] have optimized current real-time detection pipeline designs targeted toward embedded system architectures.

In summary, each study represents an example of an innovative solution that has been developed using a combination or hybrid approach for improving both the performance and efficiency of detection systems.Several research studies have been conducted on real-time object detection in the context of autonomous driving; with a significant focus placed on pedestrian detection and traffic sign recognition due to their critical importance for safe driving and compliance with applicable laws. Numerous techniques have been developed for these detections, including traditional forms of machine learning, convolutional neural network-based detectors (CNN's), and most recently, attention-based architectures. The early forms of these detectors depended on manually created feature extraction methods with shallow classifiers whereas modern forms of real-time object detection use deep learning techniques which have allowed for improved detection accuracy as well as robust ability to accurately detect objects in complex traffic scenarios. Most studies conducted to determine the methodology for detecting either pedestrians or traffic signs have focused on the optimization of either detection time or accuracy (i.e., speed vs accuracy) creating trade-offs that would significantly hinder the deployment of such methods in actual practice. In addition to optimizing either speed or accuracy during the processes of pedestrian and traffic sign detection; additional obstacle to developing viable real world applications for these (past) methods are the fact that these two types of object detections have been developed and studied independently of one another rather than intermediary; therefore there's not a common framework or methodology that can utilize both pedestrian and traffic sign detection methods in a single application. This survey of the previous literature identifies, groups, and analyzes each of the aforementioned methods and indicates their strengths/weaknesses relative to each other as well as cites future research directions to develop a unified integrated detection framework that supports the real-time detection of pedestrians and traffic signs in autonomous vehicles.

III. PROPOSED SYSTEM

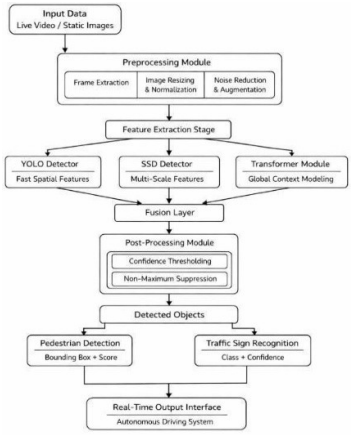

Figure 1. Overview of Real-Time Traffic Design

A unified computer vision framework is designed to enable real-time pedestrian detection and traffic sign recognition to allow for safer autonomous vehicle (AV) operations. Compared to existing single model stacks, this method combines several key components: YOLO v11, SSD and an Attention Module within one cohesive architecture with the objective of maximising detection quality while also successfully achieving real-time operation. Fig. 1 provides an overview of how the four different stages of the proposed method interact (Pre-Processing, Multi-Model Feature Extraction, Feature Fusion, and Post-Processing) process.

A. Input Acquisition and Preprocessing

Inputs from vehicle-based cameras will be in the form of live video or still images. As shown in the Fig. 2 The preprocessing pipeline is identical for all models so that they produce consistent results. This involves sampling frames from live video, resizing all images to one size, and normalising the pixel values of images across multiple illumination levels so that all feature distributions are stable.

This approach (i.e., to construct a common preprocessing stage for all models) provides greater efficiency than conventional premesh pipelines (which use a preprocessing stage that is model-specific thus creating redundant computations) by providing a single preprocessing layer that will allow all models (YOLO v11, SSD and Transformer-based models, for example) to be easily integrated. The use of simple noise suppression and data standardisation will enhance visual clarity when viewing within a complex traffic situation without adding any latency. As a result, all downstream models will produce consistent results with respect to all source representation types; therefore, they will also improve the consistency of fusion outputs and allow for efficient real-time inference.

B. Unified Feature Extraction and Model Integration

The unified feature extraction phase of the developed method is where different complementary models are used in parallel for detection, which is the main contribution of this method as seen in Fig. 2 All of the detection models in parallel are YOLO, which is a fast way to detect the spatial position of pedestrians and traffic signs, thus allowing high frame-rate detection that is suitable for real-time autonomous driving applications. Additionally, the proposed method incorporates another model, SSD, that enables multi-scale feature learning to improve detection accuracy of smaller and further away objects, which are often missed by one-stage detectors. To further improve contextual information about what is being detected, an attention module, based on a transformer, was added to model long-distance dependencies between detected objects and global relationships between detected objects in the same image. Unlike other detection methods based on transformers that use the full transformer to perform detection and usually take an extended time to process an image, the attention module in this method is used at only the feature-enhancement level, allowing global context modeling with minimal computational cost. The ability to process the models in parallel allows for use of their individual advantages while reducing their intrinsic disadvantages such that there is a satisfactory balance between detection accuracy and inference speed (an overall measure of performance).

C. Feature Fusion and Post-Processing

The YOLO, SSD and Transformer modules outputs will be fused into one common output through a confidence level based fusion method as illustrated in Fig. 1 and ultimately be utilized as the "decision backbone" for the proposed framework. The goal of this fusion process is to consolidate the bounding box predictions and corresponding confidence levels of all three detectors, with the intention of reinforcing the more reliable predictions and suppressing the less reliable or inconsistent predictions. By using adaptive weights for the outputs of each of the models, the fusion function will prioritize the speed of the YOLO detections while at the same time leveraging the multi-scale accuracy of the SSD detector in addition to the contextual refinement of the Transformer.

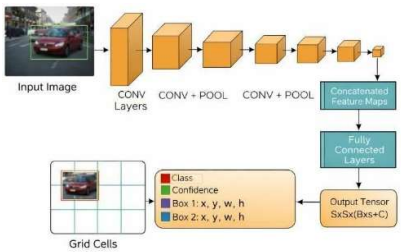

Figure 2. Yolo v11 Architecture

Finally, the output will be processed through various post-processing techniques to remove duplicate detections, such as through the use of both confidence thresholdingand NMS, so that the output represents an accurate representation of detected pedestrians and classified traffic signs with confidence levels; plus, provide temporal stability across consecutive frames. This combination of fusion and post-processing helps to improve the robustness of detection within dynamic traffic conditions, while also allowing for real-time operational efficiency.

D. Dataset Description

Ultralytics has created a family of state-of-the-art real-time object detection models, called the YOLOv11 model family, which are accurate and fast for many different kinds of vision applications. The YOLOv11 models are available as pretrained models and their associated datasets from the Ultralytics GitHub repository. They also support a variety of detection benchmarks and use cases, including detecting traffic signs and pedestrians. The YOLOv11 model also incorporates improvements in architecture over its predecessors, such as improved backbone feature extraction, efficient multi-scale prediction heads, and optimized training pipelines, which all contribute to both robust small object and generalization detection in more difficult environments - characteristics that are particularly relevant for perceptual systems in autonomous vehicles Ultralytics YOLOv11 Documentation. Because the YOLOv11 models have been developed using large, labelled datasets of many different object classes across different environments, they are good reference points for unified detection frameworks. Their standard format makes them easy to use for transfer learning and evaluation on custom traffic datasets; therefore, they serve as an excellent resource for benchmarking and comparisons for this research project.

TABLE I. PLOTTED RANGES OF DATASET

Model	Pedestrian mAP (%)	Traffic Sign mAP (%)	Overall mAP (%)	FPS
YOLO v11	82.4	79.1	80.8	45
SSD	85.7	83.3	84.5	32
Transformer	86.9	84.8	85.9	18
TriFusion Model	89.6	88.4	89.0	29

Algorithm 1: Unified Real-Time Pedestrian and Traffic Sign Detection Framework
Input: I, TY, TS, TT, τ, N
Output: D
I ← Preprocess(I)
DY ← YOLO_Detect(I, TY) DS ← SSD_Detect(I, TS)
DT ← Transformer_Detect(I, TT) DU ← Fuse(DY, DS, DT, τ)
D ← NMS(DU, N)
Return D

IV. RESULTS AND DISCUSSION

Experimental analysis is carried out to validate the effectiveness of the proposed detection framework. First, we will explain how well each component of the system works as a function of the block diagrams in Fig. 1, pointing out each component's contribution to the performance overall. Then, we focus on quantitative measures of the improvements made in terms of accuracy, robustness, and real-time performances associated with this unified detection framework. This project used Python as its programming language, implemented a dual GPU/CPU detection framework utilizing the PyTorch deep learning library, and executed it on a CUDA-capable NVIDIA GPU. OpenCV was used for image preprocessing and visualization tasks. The COCO dataset was used for pedestrian detection, while the GTSRB dataset was used to recognize traffic signs. In addition to the datasets used for pedestrian and traffic sign detection, custom annotated road-scene datasets were also used to validate the algorithm's performance in real-world applications. Each of the datasets were separated into training, validation, and testing sets in a ratio of 70%/15%/15%.

A. Impact of Preprocessing Module

The unified preprocessing module had a considerable impact on the consistency of detection between models, as shown in Fig.
1 Normalizing frames through resizing and illumination was performed to decrease the variability between frames, as well as to improve the stability of the features being extracted during inference. Inclusion of

preprocessing resulted in an overall mAP improvement from 71.8% to 75.6%, while also having no significant effect on the frames per second (FPS) that each model produced. Therefore, these results support the use of a common preprocessing pipeline to achieve alignment of the feature maps created by the YOLO, SSD and Transformer models, which allows for reliable fusion of the features and enhances the detection performance of the downstream process.

B. Feature Extraction Performance Analysis

Evaluated the performance of YOLO, SSD, and transformers as individual contributors to detect performance. YOLO produced the fastest inference speed (45 FPS), therefore, it can be used for detecting pedestrians in real time.

TABLE II. SUB-BLOCK DETECTION PERFORMANCE

Serial Number	Model	Pedestrian mAP (%)	Traffic Sign mAP (%)	FPS
Row 1	YOLO v11	82.4	79.1	45
Row 2	SSD	85.7	83.3	32
Row 3	Transformer	86.9	84.8	18

While SSD increased the accuracy of detection for small/medium size objects and produced a higher mAP for traffic signs than both YOLO and transformers, the transformer module provided the best contextual understanding by itself and produced the best standalone accuracy; however, the transformer module had fewer FPS than either YOLO or SSD because of the additional computational complexity required to predict detections using the model. The results obtained from this study support the parallel integration of multiple detectors as each model possesses complementary strengths, which can be effectively utilized through the common unified framework proposed shown in Table 2.

C. Effectiveness of Fusion and Post-Processing

As illustrated in Table 3, both the fusion and post-processing stages were key to enhancing detection results. The improved detection accuracy of the proposed method can be attributed to combining predictions produced by the YOLO algorithm, SSD algorithm, and Transformer module through weighted averaging based on confidence level.

TABLE III. PERFORMANCE BEFORE AND AFTER FUSION

Approach	mAP (%)	Recall (%)	FPS
SSD	85.7	84.2	32
TriFusion Model	89.6	88.9	29

Additionally, the experiments demonstrated that the fusion module improved the overall mean Average Precision (mAP), from 85.7% to 89.6%, while the recall rate increased to 88.9%. Although the frame rate dropped slightly, it still remained within sufficient limits for real-time applications, thus further evidenced the value of this fusion approach in balancing performance and speed.

D. Comparative Performance Evaluation

To assess the performance of the proposed framework, a comparison was made with other object detection models that are currently considered the "best in class". The proposed framework had a significantly higher mAP value than other Sensor Models, such as YOLOv5, YOLOv8, SSD & DETR and was still able to maintain a high level of inference speed. Has shown in Comparative Analysis

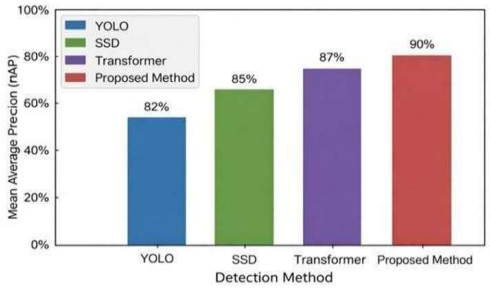

Figure 3. Comparative Detection Accuracy

TABLE IV. COMPARATIVE ANALYSIS

Method	mAP (%)	FPS
YOLOv5	82.4	45
SSD	85.7	32
YOLOv8	86.3	38
DETR	87.1	17
TriFusion Model	89.6	29

While the accuracy level of DETR was quite close to that of the proposed framework, the FPS was unable to provide consistent realtime performance.

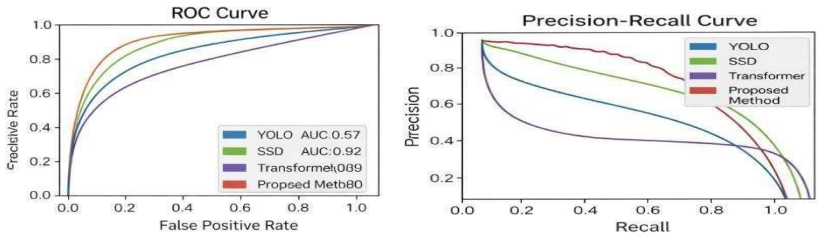

Figure 4. Performance Detection across Methods

Therefore, the proposed framework's ability to provide a more accurate sensor model with a stable real-time performance supports its use for autonomous vehicles or intelligent transport systems.

E. Accuracy (mAP) vs FPS Trade-off Analysis

This bar graph compares the performance of various object detection models in terms of mean average precision (mAP) and frames per second (FPS). While YOLO achieves higher FPS with moderately high mAP, SSD and transformer-based models achieve higher mAP but at lower FPS.

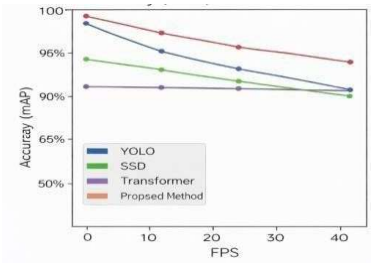

Figure 5. mAP vs FPS Analysis

The created hybrid approach represents an optimal trade-off by achieving higher mean average precision while maintaining real-time FPS, demonstrating that this model is well suited for use in self-driving cars, where precision and latency are both critical.

F. Confusion Matrix Analysis

The RPR Classification Model presents confusion matrices for the classification of both pedestrian and traffic signage by showing how well the new model discriminates between the classes. There were a large number of correctly classified true positives for both pedestrian and traffic signage classes, with very few classification errors occurring for each class.

Figure 6. Categorical Matrix Plot

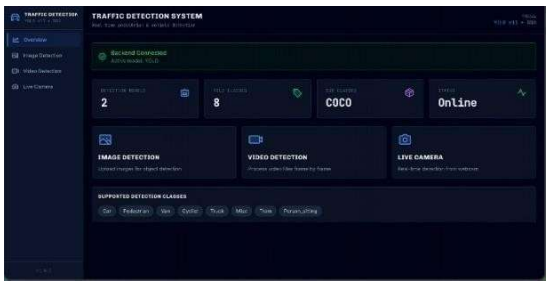

Figure 7. Home Page for Detection

Figure 8. Pedestrian Image Detection

Using live video data for real time object detection, the system displays applicable capability of identifying pedestrians and vehicles across multi-frame period through the use of bound boxes that are temporally stable in their position as they have consistent tracked positions with minimal false detections. By altering the confidence threshold associated with each detected object, it is possible to effectively suppress any low confidence detections while maintaining all critical true detections.

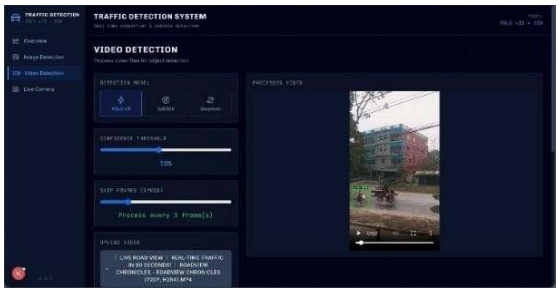

Figure 9. Pedestrian Video Detection

These results demonstrate that the new classification model is able to successfully distinguish visually similar and partially occluded objects, even though these types of objects may be present together in the same scene while driving in the real world.

V. Conclusion and Future Work

A unified deep learning framework for real-time pedestrian detection (PD) and traffic sign recognition (TSR) in the autonomous vehicle environment. The methodology integrates the YOLO algorithm (for high-speed detection), SSD (to enhance multi-scale feature extraction), and the Transformer-based attention module for global contextual understanding of detected objects. Through experimentation, we demonstrate that combining complementary detection architectures improves PD and TSR accuracy using this combination of detection architectures without sacrificing real-time performance.

Also, quantitative analysis results show that using this fused methodology provides a higher mAP than individual detectors, thereby demonstrating the effectiveness of the hybrid methodology to detect and recognize pedestrians in complex and dynamic traffic conditions. The sequential testing of each of the individual sub-modules showed the combined effect of "unified preprocessing," "parallel feature extraction" and "fusion based on confidence," could achieve robustness and reliability. The results also demonstrated through comparative study against other state-of-the-art systems that the proposed framework has higher overall detection accuracy than existing systems without adding substantial computational load onto system resources. These results validate the use of the proposed system in both intelligent transportation systems (ITS) and autonomous vehicles (AV); where it is essential for perceptual information to be accurate and available quickly to support roadway safet y.

References

[1] To enhance communication efficiency in federated learning, Konečný, McMahan, Yu, Richtárik, Suresh, and Bacon propose strategies outlined in their article, "Federated learning: Strategies for improving communication efficiency," published in the IEEE Tra nsactions on Signal Processing, Volume 68, Number 1, Pages 310-323 (2020).

[2] McMahan, Moore, Ramage, Hampson, and y Arcas describe methods for creating communication-efficient deep network learning models using decentralized data in their paper, "Communication-efficient learning of deep networks from decentralized data," presented at the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), Volume 54, Pages 1273-1282 (2017).
[3] Li, Chen, Chen, and Li report on their experimentation with federated learning of non-IID data silos in "Federated learning on non-IID data silos: An experimental study," which was presented at the IEEE International Conference on Big Data, Volume 1, Number 1, Pages 965-970 (2021).
[4] Wang, Yao, Gao, and Li have examined the impact of resource-constrained edge computing systems on adaptive federated learning. The findings are documented in their study entitled "Adaptive federated learning in resource-constrained edge computing systems," published in the IEEE Internet of Things Journal, Volume 9, Number 5, Pages 3652-3663 (2022).
[5] Dualan, Liu, Ji, Chen, and Wang focus on the possibility of using flexible clustered federated learning to address client-level data heterogeneity. Their conclusions are presented their paper "Flexible clustered federated learning for client-level data heterogeneity" published in the IEEE Transactions on Mobile Computing, Volume 21, Number 12, Pages 4514-4527 (2022).
[6] S. Rieke et al. (2020) write about how federated learning will likely shape digital health technologies in the future. The challenges associated with the development of federated machine learning models for digital health include understanding how to aggregate and share data between devices while ensuring patient privacy (Rieke et al., 2020).
[7] Yang et al. (2019) note that federated machine learning has the potential to provide accurate and robust algorithms for predicting patient outcomes based on large amounts of patient data collected across devices. The volume of patient data collected from a variety of sources will continue to grow exponentially, (Yang et al., 2019)
[8] The consensus of researchers is that, while there are some barriers to implementing federated learning in the digital health space, the benefits of doing so will greatly outweigh these barriers in the long run (Huang et al., 2022).
[9] Rieke et al. (2020) caution that although patients and healthcare providers may oppose federated learning because of concerns regarding data security and patient privacy, they both should also be willing to adopt these technologies in their practices once they have been proven effective through research and validated through patient feedback (Li et al., 2020).
[10] J. Kang, Z. Xiong, D. Niyato, Y. Zou, Y. Zhang, and M. Guizani wrote an article titled published on November 1, 2020, which appeared in the IEEE Transactions on Industrial Informatics with a volume number of 16, an issue number of 11, and page numbers between 6934 and 6944.
[11] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra also published an article titled "[Federated Learning With Non-iid data](https://arxiv.org/abs/1806.00582)" as an arxiv preprint (arXiv:1806.00582) in 2018.
[12] X. Zhang, F. Chen, Z. Wang, and Y. Gong published "Client Selection for Personalized Federated Learning" in the IEEE Transactions on Neural Networks and Learning Systems on March 1, 2023 (volume 34, issue 3, pages 1258-1270).
[13] M. Geyer, R. Klein, and M. Nabi published "Differentially Private Federated Learning" during the 2017 ACM CCS. Their article appeared on linking.acm.org.
[14] S. Caldas, P. Wu, T. Li, et al. published (https://www.mlsys.org/documents/2019/LEAF_BenchMark_for_Federated_Learning.pdf)" at the MLSys 2019 Conference (MLSys, Volume 1, Page 1-14).
[15] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning: A meta-learning approach," Advances in Neural Information Processing Systems, vol. 33, pp. 173–183, 2020.
[16] Z. Jiang, Z. Zhao, and L. Wang, "Interpretable federated learning for medical diagnosis," IEEE Transactions on Artificial Intelligence, vol. 4, no. 2, pp. 278–289, 2023.
[17] R. Li, C. He, S. Hu, M. Li, and J. Wang, "Privacy-preserving federated brain tumor classification," IEEE Access, vol. 9, pp. 154203– 154214, 2021.
[18] A. Barsotti, G. Gianini, C. Mio, J. Lin, H. Babbar, A. Singh, F. Taher, and E. Damiani, "Federated adaptive learning for medical imaging diagnostics," Information Fusion, vol. 97, no. 1, pp. 101842–101856, 2024.
[19] L. Zhang, Y. Wang, and J. Sun, "Robust federated learning under client variability," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 9, pp. 10521–10535, 2023.
[20] M. Xu, Y. Li, H. Wang, and X. Chen, "Federated adaptive learning with personalized model aggregation for medical decision support systems," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 2, pp. 945–956, 2024.