Subject Code	Subject Name	Credits
26CS609	EDGE INTELLIGENCE	4

Course Objectives:

1. To design and deploy Al models on resource-constrained edge devices.
2. To explore optimization techniques for efficient edge Al systems.
3. To analyse real-world applications of edge intelligence.
4. To study emerging trends like federated learning and TinyML.

Learning Objectives:

After completion of the course the students will be able to:

1. Analyse edge computing architectures and paradigms.
2. Design ML models for resource-constrained edge devices.
3. Apply optimization techniques for efficient edge deployment.
4. Evaluate security, privacy, and distributed learning methods.
5. Develop real-time edge Al applications.

Unit I (12 Hrs.)

Foundations of edge intelligence: Cloud, Fog and Edge computing evolution - Edge Al: Concept and Definition -Edge-cloud trade-offs (latency, bandwidth, privacy) - Edge architecture - Resource constraints (compute, memory, energy) - Real-time edge applications

Unit II (12 Hrs.)

Edge-oriented machine learning: ML pipeline for edge systems - Lightweight models (MobileNet, SqueezeNet) -TinyML and embedded inference - On-device vs cloud training - Multi-objective optimization - Frameworks (TensorFlow Lite, ONNX Runtime)

Unit III (12 Hrs.)

Optimization and deployment: Quantization and pruning techniques - Knowledge distillation methods - Edge-specific Neural Architecture Search - Graph optimization and operator fusion - Edge deployment pipeline - Runtime inference optimization strategies

UNIT IV (12 Hrs.)

Distributed edge intelligence & security: Federated and split learning - Communication-efficient training - Differential privacy and encryption - Adversarial and poisoning attacks - Blockchain-based trust management - Fault tolerance and energy-aware Al

UNIT V (12 Hrs.)

Edge ai systems design: Edge Al system design - Task partitioning and scheduling - Latency and energy modeling - Resource management and scaling - Performance benchmarking - Real-world deployment cases

Reference Books:

1. Pete Warden & Daniel Situnayake, TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers, 2019.
2. Qiang Yang et al., Federated Learning: Collaborative Machine Learning with Privacy, 2019.
3. Yu Shi et al., Edge Al: On-Demand Accelerated Deep Learning Inference, 2020.
4. Rajkumar Buyya et al., Cloud, Edge, and Fog Computing: Technologies and Applications, 2021.
5. Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, 2016.