राष्ट्रीय प्रौद्योगिकी संस्थान पटना / NATIONAL INSTITUTE OF TECHNOLOGY PATNA
(शिक्षा मंत्रालय, भारत सरकार के अधीन एक राष्ट्रीय महत्व का संस्थान / An Institute of National Importance under Ministry of Education, Gov. of India)
संगणक विज्ञान एवं अभियांत्रिकी विभाग / DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
अशोक राजपथ, पटना - ८००००५, बिहार / Ashok Rajpath, Patna- 800005, Bihar
Tel. No. – 0612-2372715, 2370419 (Ext-200) email- cseoffice@nitp.ac.in

*CSXX02*76: Edge AI and Federated Learning

L-T-P-Cr: 3-0-0-3

Prerequisites:

- Basic knowledge of Machine Learning and Deep Learning.
- Proficiency in Python programming.
- Understanding of basic computer networks and operating systems.
- Familiarity with linear algebra, probability, and statistics.

Course Objectives:

CO1	Understand the fundamental concepts of Edge Computing, Artificial Intelligence, and Distributed Systems.
CO2	Analyze the challenges and opportunities of deploying AI models on resource-constrained edge devices.
CO3	Apply various optimization techniques for efficient Edge AI model deployment.
CO4	Grasp the principles and motivations behind Federated Learning for privacy-preserving and collaborative AI.
CO5	Implement basic Federated Learning algorithms and understand their architectural implications.
CO6	Evaluate the privacy, security, and ethical considerations in Edge AI and Federated Learning systems.
CO7	Explore real-world applications and future trends in this rapidly evolving field.

Course Outcomes (COs) contribution to the Programme Outcomes(POs)

Strength of Correlation: **3**: High (Strong contribution), **2**: Medium (Moderate contribution) **1**: Low (Slight/Indirect contribution), **Blank** : No correlation

CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**
CO1	3	3										
CO2	3	3										
CO3	3	3										
CO4	3	3		2		2						
CO5		3		3		3						
CO6	3	2		3								

Course Units:

Unit 1: Introduction to AI, Edge Computing, and Distributed Systems (8 hours)

- **1.1 Review of Artificial Intelligence and Machine Learning:**
    - Brief overview of AI, ML, and Deep Learning paradigms.
    - Key concepts: supervised, unsupervised, reinforcement learning.
    - Introduction to neural networks and their architectures (CNNs, RNNs, Transformers - high-level).
- **1.2 Introduction to Edge Computing:**
    - Definition and characteristics of Edge Computing.
    - Comparison with Cloud Computing and Fog Computing.
    - Advantages of Edge Computing: low latency, bandwidth efficiency, privacy, reliability.
    - Challenges: resource constraints, security, management.
- **1.3 Distributed Systems Fundamentals:**
    - Concepts of distributed data processing and parallel computing.
    - Client-server architectures, peer-to-peer networks.
    - Introduction to distributed consensus and fault tolerance (briefly).

Unit 2: Fundamentals of Edge AI (10 hours)

- **2.1 AI Model Optimization for Edge Devices:**
    - **Model Quantization:** Fixed-point arithmetic, 8-bit quantization, post-training quantization, quantization-aware training.
    - **Model Pruning:** Weight pruning, neuron pruning, filter pruning.
    - **Knowledge Distillation:** Transferring knowledge from a large teacher model to a small student model.
    - **Neural Architecture Search (NAS) for Edge:** Overview of NAS techniques for finding efficient architectures.
- **2.2 Hardware for Edge AI:**
    - Overview of System-on-Chips (SoCs), Microcontrollers (MCUs).
    - Specialized AI Accelerators: Neural Processing Units (NPUs), GPUs for edge, FPGAs, ASICs.
    - Power consumption and thermal management considerations.
- **2.3 Edge AI Frameworks and Tools:**
    - **TensorFlow Lite:** Model conversion, interpreter, delegate concepts.
    - **PyTorch Mobile:** Deployment for mobile and edge devices.
    - **OpenVINO (Intel):** Optimizing and deploying models on Intel hardware.
    - Other relevant frameworks (e.g., ONNX Runtime, TVM).
- **2.4 Deployment Strategies for Edge AI:**
    - Model deployment pipelines.
    - Over-the-air (OTA) updates for edge models.
    - Monitoring and managing AI models at the edge.

Unit 3: Introduction to Federated Learning (10 hours)

- **3.1 Motivation for Federated Learning:**
    - Data privacy concerns (GDPR, HIPAA implications).
    - Bandwidth limitations and communication costs.
    - Collaborative AI without direct data sharing.
    - Data silos and regulatory compliance.
- **3.2 Core Concepts of Federated Learning:**
    - Global model vs. Local models.
    - Central server (aggregator) and participating clients.
    - The FL training loop: local training, model upload, global aggregation, model download.
- **3.3 Federated Averaging (FedAvg) Algorithm:**
    - Detailed explanation of the FedAvg algorithm steps.
    - Convergence properties and challenges.
- **3.4 Types of Federated Learning:**
    - **Horizontal Federated Learning:** Data sharing in feature space (e.g., mobile phone users).
    - **Vertical Federated Learning:** Data sharing in sample space (e.g., different organizations with common users).
    - **Federated Transfer Learning:** Leveraging pre-trained models in FL settings.

Unit 4: Advanced Topics in Federated Learning (9 hours)

- **4.1 Privacy-Preserving Techniques in FL:**
    - **Differential Privacy (DP):** Adding noise for privacy guarantees.
    - **Secure Multi-Party Computation (SMC):** Cryptographic techniques for secure aggregation.
    - **Homomorphic Encryption (HE):** Performing computations on encrypted data.
- **4.2 Communication Efficiency in FL:**
    - Model compression techniques (quantization, sparsification) for communication.
    - Client selection strategies.
    - Asynchronous FL.
- **4.3 Handling Heterogeneity in FL:**
    - **Statistical Heterogeneity:** Non-IID data distribution across clients.
    - **System Heterogeneity:** Varying computational and network capabilities of clients.
    - Personalization in FL.
- **4.4 Robustness and Fairness in FL:**
    - Addressing adversarial attacks (e.g., model poisoning, backdoor attacks).

- Byzantine robustness.
- Fairness considerations in FL: ensuring equitable performance across client groups.

Unit 5: Applications and Case Studies (8 hours)

- **5.1 Edge AI Applications:**
    - **IoT and Smart Homes:** Anomaly detection, predictive maintenance.
    - **Autonomous Vehicles:** Real-time perception, decision making.
    - **Smart Cities:** Traffic management, surveillance.
    - **Healthcare:** Wearable devices, remote patient monitoring.
    - Industrial automation.
- **5.2 Federated Learning Applications:**
    - **Mobile Keyboards:** Next-word prediction.
    - **Healthcare:** Collaborative disease diagnosis, drug discovery.
    - **Finance:** Fraud detection, credit scoring.
    - **Smart Retail:** Demand forecasting.
- **5.3 Case Studies and Practical Implementations:**
    - Discussion of prominent real-world deployments and research projects.
    - Introduction to open-source FL platforms (e.g., Flower, PySyft - high-level overview).

Unit 6: Ethical Considerations and Future Trends (5 hours)

- **6.1 Privacy, Security, and Bias:**
    - Deep dive into privacy leakage risks in FL.
    - Security vulnerabilities and attack vectors.
    - Addressing algorithmic bias in distributed AI systems.
- **6.2 Regulatory Aspects:**
    - Overview of data protection regulations relevant to Edge AI and FL (e.g., GDPR, India's Digital Personal Data Protection Act).
- **6.3 Research Challenges and Future Directions:**
    - Scalability, trust, and interpretability in FL.
    - Integration of Edge AI and FL with other emerging technologies (e.g., Blockchain).
    - Open problems and research opportunities.

4. Textbooks and References:

1. *Edge AI: Convergence of Edge Computing and Artificial Intelligence* by Arpan Pal, et al. (Wiley)
2. *Federated Learning: Privacy and Incentive* by Li, Qiang, et al. (Springer)

3. *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (MIT Press) - for foundational ML/DL.