1. *Course number and name*

**ENGR 492: Hardware for Machine Learning**

1. *Credits and contact hours*

3 credit hours

1. *Instructor’s or course coordinator’s name*

Instructor: Hao Jiang

Course coordinator: Hao Jiang, Professor

1. *Text book, title, author, and year*

Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, Joel S. Emer, *Efficient Processing of Deep Neural Networks*, Morgan & Claypool Publishers, 2020. (ISBN: 978-1681738314)

1. *Specific course information*
2. *brief description of the content of the course (catalog description)*

Deep neural networks and their basic numerical process; neural networks processing hardware’s performance metrices and bottlenecks, popular architecture design, micro-architecture approaches, and technology choices.

1. *prerequisites or co-requisites*

Grades of C- or better ENGR 213, ENGR 353 and ENGR 356

1. *indicate whether a required, elective, or selected elective course in the program*

Elective for Electrical Engineering and Computer Engineering

1. *Specific goals for the course*
2. *Specific outcomes of instruction, ex. The student will be able to explain the significance of current research about a particular topic.*
* To investigate a variety of machine learning algorithms including k-nearest neighbor, state vector machine, multi-layer perceptron, convolution neural networks, object detections, recurrent neural networks, and transformers.
	+ To analyze the performance of a variety of machine learning accelerator architectures.
	+ To study a variety of popular machine learning accelerator hardware platforms including FPGA, GPU, ASIC and in-memory computation.
1. *Explicitly indicate which of the student outcomes listed in Criterion 3 or any other outcomes are addressed by the course.*

Course addresses ABET Student Outcome(s): 1, 2, 6, 7

1. *Brief list of topics to be covered*
* Python basics.
* Machine learning algorithms: k-nearest neighbor, state vector machine, multi-layer perceptron, convolution neural networks, object detections, recurrent neural networks, and transformers.
* Numerical process of inference, back propagation and gradient descent.
* Machine learning hardware performance metrics.
* Survey of machine learning hardware accelerator’s architecture and technology platforms.
* Micro-architecture design and performance analysis.