Course Syllabus

CS581A/481A: Multi-Modal Machine Learning in Biomedicine

This syllabus is designed to equip students with practical and theoretical knowledge of multi-modal machine learning methods, empowering them to use these techniques for cutting-edge biomedical applications such as biomarker discovery and clinical decision support.

Course Overview:

This course provides an in-depth understanding of multi-modal machine learning techniques applied to biomedicine. The focus will be on integrating and analyzing multiple types of biomedical data, including multi-omics (genomics, transcriptomics, proteomics, metabolomics) and medical images (e.g., radiology, pathology). Students will learn to apply machine learning algorithms to uncover biomarkers and improve clinical decision-making and prediction models.

Class Time: T/Th: 1:15 to 2:40 PM EST, Spring 2025

Location: SL 306

Lecturer: SUNY Empire Innovation Professor Nancy Guo

Guest Lecturers: Weiying Dai, PhD (wdai@binghamton.edu)

Roger Chammas, MD, PhD (rchammas@usp.br)

Matthew Quimby, Tech Transfer (mquimby1@binghamton.edu)

Office Hours: Mondays 3 to 4 pm in EB-Q7 or by email (nguo1@binghamton.edu)

TA: Xin Zhou (xzhou11@binghamton.edu)

Prerequisites:

- Math 227 (Cal II) or Instructor Approval
- Programming: programming assignments can be implemented in WEKA or any popular programming languages, such as C, C++, Java, Python, R, or Matlab. No programming-specific issues will be covered in this course.

Course Format:

- **Lectures** (3 hours per week)
- **Hands-on Labs** (2 hours per week through TA office hours)
- **Homework Assignments** (individual)
- **Final Group Projects** (with presentations)

Credit Hours Statement:

This course is a 3-credit graduate / 4-credit undergraduate course. Students are expected to do at least 9.5 hours of course related work each week that we meet. This includes scheduled class lecture/discussion meeting times as well as time spent completing assigned readings, preparing written assignments, and other course-related tasks.

Course Outcomes / Student Learning Objectives (SLOs):

Upon successful completion of this course, students will:

1. Acquire a solid understanding of the fundamental concepts, algorithms, and techniques in machine learning and data mining (SLO1).
2. Recognize the distinctions between different machine learning and data mining methods, and understand which techniques are best suited for specific types of problems (SLO2).
3. Gain proficiency in utilizing machine learning and data mining tools (SLO3).
4. Be capable of applying machine learning and data mining methodologies to solve real-world challenges, focusing on biomedicine (SLO4).
5. Develop effective communication skills to present machine learning and data mining processes and results clearly, both in writing and orally (SLO5).

Week 1 (1/21, 1/23): Introduction to Multi-Modal Machine Learning in Biomedicine

- **Topics:**
    - Overview of multi-modal machine learning
    - Applications in biomedicine: biomarker discovery, clinical prediction, and personalized medicine
    - Key biomedical data types: multi-omics, imaging, clinical records
    - Challenges in multi-modal data integration
- **Reading:**
    - Introduction to biomedical data integration techniques (e.g., multi-omics, image-based biomarkers)

Week 2 (1/28, 1/30): Basics of Multi-Omics Data

- **Topics:**
    - Types of omics data: genomics, transcriptomics, proteomics, metabolomics
    - Data formats and preprocessing (e.g., normalization, missing data handling)

    - Techniques for dimensionality reduction (e.g., PCA, t-SNE, UMAP)
- **Lab Demo:**
    - Preprocessing and exploratory analysis of genomic and transcriptomic data
- **Reading:**
    - Key papers on multi-omics data preprocessing and integration

---

Week 3 (2/4, 2/6): Machine Learning in Multi-omics Data Analysis

- **Topics:**
    - Supervised learning techniques (e.g., decision tree, Naïve Bayes, SVM, random forests)
    - Entropy, Information Gain, and Gain Ratio
    - Evaluation metrics for classification (e.g., precision, recall, F1-score, confusion matrix)
- **Lab Demo:**
    - Build a classifier for omics data to predict disease states
- **Reading:**
    - Review on ML models used in genomic data analysis

Week 4 (2/11, 2/13): Feature Selection in Biomarker Discovery

- **Topics:**
    - Feature selection vs. feature extraction
    - Feature selection algorithms (Relief, consistency-based subset selection)
    - Application to biomarker discovery
- **Reading:**
    - Case study papers on machine learning in personalized medicine and clinical prediction

---

Week 5 (2/18, 2/20): Unsupervised learning in molecular subtype discovery

- **Topics:**
    - Unsupervised clustering (centroid-based, density-based, hierarchical clustering)

    - Applications in molecular subtyping
- **Reading:**
    - Key papers in breast cancer molecular subtyping and validation

---

Week 6 (2/25, 2/27): Rule-based unsupervised learning

- **Topics:**
    - Association rule discovery (Apriori)
- **Lab Demo:**
    - Feature selection, supervised and unsupervised classification

---

Week 7 (3/4, 3/6): Image Analysis in Biomedicine (Dr. Weiying Dai)

- **Topics:**
    - Introduction to medical imaging modalities (e.g., CT, MRI, PET, histopathology)
    - Image preprocessing and augmentation techniques
    - Deep learning models for image analysis (e.g., CNNs)
- **Lab Demo:**
    - Image preprocessing and classification using CNNs on medical image data
- **Reading:**
    - Overview of convolutional neural networks in medical image analysis

---

Week 8 (3/18, 3/20): Biomarker Discovery and Evaluation of Models in Biomedicine

- **Topics:**
    - Identifying biomarkers from multi-omics data using hybrid models
    - Model validation strategies: cross-validation, external validation, prospective clinical trials
- **Lab Demo:**
    - Classification, cross-validation, and external validation
- **Reading:**
    - Guidelines on validating machine learning models for clinical use

---

Week 9 (3/25, 3/27): Clinical Decision Support Using Machine Learning

- **Topics:**
    - Time-series longitudinal model evaluation (Cox model, Kaplan-Meier analysis)
    - Clinical decision support systems using machine learning
    - Regulatory aspects and ethical considerations in clinical ML applications
- **Reading:**
    - Papers on machine learning in clinical decision-making and biomarker discovery

---

Week 10 (4/1, 4/3): Advanced Machine Learning Techniques for Multi-Modal Data

- **Topics:**
    - Graph-based learning in bioinformatics
    - Time-series and dose-response multi-omics data analysis
    - Generative AI and large language models for biomedical applications
- **Reading:**
    - Recent trends in graph-based learning and generative AI for biomedical applications

---

Week 11 (4/8, 4/10): Multi-Modal Data Integration, Emerging Trend, and Future Direction

- **Topics:**
    - Challenges of multi-modal data fusion
    - Data fusion strategies: early, late, and hybrid integration
    - Methods for integrating multi-modal data for clinical predictions
    - Emerging techniques: federated learning and causal inference in biomedicine
- **Reading:**
    - Papers on integrating multi-omics and imaging data in biomedicine

---

Week 12 (4/15, 4/17): Research design of causal inference in biomedical validations

- **Topics**
    - Research designs and experiments for validation of causal inference in biomedicine (Dr. Roger Chammas)

    - Innovation and Translation of Research Results (Matthew Quimby)
- **Reading:**
    - Review translational research on validation of bioinformatics results

---

Weeks 13 and 14 (4/22, 4/24, 4/29, 5/1): Final Project Presentations

- **Topics:**
    - Student presentations on final projects: combining multi-modal data for clinical prediction
- **Assessment:**
    - Presentation and report submission
- **Evaluation:**
    - Peer feedback and instructor evaluation

*** The schedule may be subject to changes.**

Recommended Texts and Resources:

- Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann) 4th Edition. by Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal
    - https://ml.cms.waikato.ac.nz/weka/Witten_et_al_2016_appendix.pdf
- Introduction to Data Mining (Second Edition), Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Addison-Wesley, April 2018.
- Introduction to Machine Learning, Ethaem Alpaydin, The MIT Press, 2004.
- Machine Learning, Tom Mitchell, McGraw Hill 1997.
- Online journals and papers will be provided weekly to accompany each lecture topic.

Lecture Notes and Supplemental Materials:
Lecture notes, assignments, and additional supplemental materials, will be posted on Brightspace (generally before class). Lecture notes do not capture everything covered in class. Some significant parts of a class, including all the details in terms of explanations, analyses, and in-class exercises, may not be included in the lecture notes.

Grading (evaluations against SLOs):

Quizzes (30%)	4 quizzes on key concepts
Homework Assignments (30%)	5 assignments on exercises
Two Final Projects (30%)	○ A group presentation (3 people per group) of a literature review of current topics of multi-modal ML ○ A group project (3 people per group) focusing on a real-world biomedical application with presentation and project report
Participation and Attendance (10%)	

1. **Quizzes:** There will be 4 quizzes on key concepts covered in the lectures. The quizzes will be given during the class time. There may not be any prior notice of the quizzes.
2. **Homework Assignments:** There will be 5 assignments in the form of question answering/programming on key concepts and algorithms. They are developed to help you understand the lecture materials and prepare for 4 quizzes. All solutions of assignments must be typed (no handwritten solutions accepted) and submitted electrically.
3. **Group Presentation:** Each student will be required to form a presentation group (≤ 3 students) to give one presentation on a selected topic (a list of topics provided by the instructor). The presenters will also be responsible for conducting group discussions and answering questions.
4. **Group Project:** Each student will be required to form a project group (3-4 students) to conduct a term project on a topic proposed by the group and approved by the instructor. In the project, the group will apply machine learning algorithms and techniques learned in this class to a real-world problem. Each project group is required to submit and present in class a project progress report (in presentation slides format) during the middle of the semester and submit a final project report (in technical report format) at the end of the semester.
5. **Class Participation:** Class participation is required for this course. Active class participation is a very important part of the learning process in this course. You will be evaluated based on both the frequency and the quality of your interactions during the class. Quality interactions possess one or more of the following properties: 1) ask a good question; 2) offer a different and unique, but relevant, perspective; 3) contribute to moving the discussion and analysis forward; and 4) build on other comments.
   While your class participation grade is subjective, it will not be random or arbitrary. You will be given an excellent or very good class participation grade as long as you frequently come to class and actively contribute to the class discussion during the lectures and student presentations. Clearly, more frequent quality interactions are better than less frequent quality interactions. Frequent absence (no shows, late arrivals, or early departures) will negatively affect your participation grade.

Academic Honesty

Discussion of general concepts and questions concerning the homework assignments among students is encouraged. However, each of you is expected to work on the homework solutions on your own. Sharing of any part of solutions is prohibited. If you are unclear about the policy, please consult with the instructor before you act. Binghamton University provides explicit guidelines in the Student Academic Honesty Code (see the University Bulletin - Academic Policies and Procedures for All Students).

Late Policy
Each assignment is due at 11:59pm on the due date. Any assignment received within the next 24 hours will be penalized by 10% of the full credit; any assignment received within the time between 24 hours and 48 hours pass the deadline is penalized by 20% of the full credit; No assignment will be accepted after 48 hours pass the deadline. Rare exceptions of this policy may be made at the discretion of the instructor under demonstrably circumstances.