

# LHS 712 – Natural Language Processing for Health

## SYLLABUS

**Class #:** 22703

**Instructor:** V. G. Vinod Vydiswaran (vgvinodv@umich.edu)

**Meeting schedule:** Mondays, 13:00 – 15:50. Room: 6000 Taubman Health Sciences Library ([Map](#), [Location & Hours](#)), Zoom option available, if needed: <https://umich.zoom.us/j/93662640640> Password: lhs712wn24

**Dates:** Jan 22<sup>nd</sup> to April 29<sup>th</sup>, 2024

**Office Hour:** After class; Mondays, 16:00 - 16:50. HYBRID. On Zoom (same link as class) AND in Room 6000 Taubman Health Sciences Library

**GSI:** Dalton Simancek ([daltonsi@umich.edu](mailto:daltonsi@umich.edu))

**GSI Office Hours on Zoom :** Thursday, 14:00 - 14:50

**Exam Date:** April 29<sup>th</sup>, 2024 (in-person, will be used for project presentations)

Students in this course will learn advanced techniques to parse and collate information from text-rich health documents such as electronic health records, clinical notes, and peer-reviewed medical literature. In this elective course, students will be able to delve deeper into challenges in recognizing medical entities in text documents, extracting clinical information, addressing ambiguity and polysemy, and analyze approaches to efficiently and effectively query and retrieve relevant patient data. Students will develop tools and techniques to analyze new genres of health information using current, state-of-the-art approaches, and build tools to help in these tasks. Students will also participate in a semester-long project on addressing specific natural language processing challenges in real-life health data sets.

### A. LEARNING OBJECTIVES

By completion of the course, students will be able to:

1. Describe the role of medical natural language processing in improving health and healthcare.
2. Identify the major natural language processing challenges in health data.
3. Develop skills to process and extract information from health-related free-text data.
4. Apply the state-of-the-art machine-learning techniques to extract information from medical text data.

5. Analyze and critique natural language processing tools currently available for medical text processing.
6. Explore the recent trends and open directions in the field of medical natural language processing.

## **B. COURSE FORMAT AND GRADING**

This course will be taught using multiple methods, including discussions, mini-lectures, group work, case study analyses, videos, web-based technologies, and hands-on programming. The instructor and the GSI will give tutorials and lead discussions to introduce the basic principles and tasks of natural language processing on health data.

In Winter 2024, the course is offered in two modes. We will primarily meet in-person on Mondays through the term, starting January 22nd. The in-person sessions will include instructor-led lectures, student-driven discussion, small group interactions, and hands-on activities. In addition, as an accommodation for those unable to attend in-person (due to illness, quarantine, isolation, potential inclement weather, or other reasons), all sessions will be streamed live on Zoom. Further, a recording of the in-person sessions will be made by end of the day for students to review material covered in the lectures. Office hours with the instructor and the graduate student instructor will be remote on Zoom as well. Most content will be available online via Canvas and will include videos and hands-on activities that the students will complete each week. The class will also meet synchronously once a week on Mondays. The synchronous sessions will focus on student-to-student interactions, small group and whole class discussions, and tool demonstration.

### **B.1. Grading**

Grading will be based on:

- Reading and critiquing peer-reviewed papers (15%)
- Quizzes and in-class activity (10%)
- “Mid-term” examination (25%)
- Assignments based on NLP tools and techniques (25%)
- Semester-long course project (25%)

The individual components are explained in more detail below. Overall course grades are posted within 72 hours of when the final course assignment component (typically the final course project report) is due. The course grade will be a letter grade (A–E) following the weighting described above, and follow a grading scale such that median of the class is an A- and close to the A- / B+ threshold. A student who earns a final weighted score of less than 50% will receive a fail grade (E).

## **B.2. Reading and critiquing peer-reviewed papers (15%):**

Each week, starting week 2, students are expected to write a 200-word summary based on the reading assignment for the previous week. The reading assignments will cover significant papers on the topics being discussed in class. The short summaries are expected to discuss the key concepts described in the paper, rather than merely stating what the paper is about. A summary of key contributions, potential limitations, suggested improvements, and ideas for future follow-up work based on the paper are encouraged.

Over the course of the semester, students are expected to write reviews for about fifteen papers across ten weekly reading summary assignments. All summaries will be read by the instructor and/or GSI, and will be graded as Satisfactory (1) or Unsatisfactory (0). Each satisfactory review will count equally towards the overall grade.

## **B.3. "Mid-term" exam (25%):**

There will be one take-home examination and will be administered around Week 9 of the semester.

## **B.4. Assignments (25%):**

There will be 3-4 assignments during the semester based on specific health text processing tasks. The tasks will be closely related to the course material, with real-world data and gold-standard judgments provided. This may include an in-situ data mining challenge using online competition services such as Kaggle-in-Class (<http://inclass.kaggle.com>). Students can submit and resubmit their results to the competition site and get instant feedback (evaluation metrics) from the system. The tasks will likely be around information extraction, classification or "named entity" identification of various health-related concepts, such as disease characteristics, severity, or social media mentions.

## **B.5. Quizzes and In-Class activity (10%):**

All students are encouraged to actively participate in class discussions. To enable critical review of the material covered in class, there will be periodic short quizzes conducted online. Students are also encouraged to share relevant resources and tips either in class or on Canvas discussion forum, to demonstrate active engagement with the course content and interaction with peers.

## **B.6. Course Project (25%):**

A course project is required. Individual projects are preferred. Small group projects are acceptable upon justification. The grading of group members will be adjusted according to their contribution to the project.

The course project will take the format of either a software system that applies existing data mining techniques to a specific type of data, or a research experiment documented in the form of a research paper.

Examples of course projects include:

1. A de-identification tool for health records using conditional random fields
2. Retrieving information about relevant clinical trials for a given case
3. Comparing authorship networks and communities in different clinical specialties
4. Identifying high-quality consumer-centric resources

The grading for the course project will be split as follows:

1. **Proposal (15%):** A two-page proposal, describing the project topic, objectives, expected deliverable (software package, demo, and/or a technical report), and a list of team members and their expected contribution to the project.  
*Tentative deadline:* Around Week 5 (late February)
2. **Progress report (10%):** A one-page summary of the progress, any hurdles towards timely completion of the stated objectives. If there are any significant changes to the submitted proposal, the students should describe them in detail in the progress report. Consider this as a checkpoint towards achieving the stated goals of the project. There are no penalties for changes to the proposal document, rather it may be more prudent to recalibrate or clarify the expected outcomes during this stage.  
*Tentative deadline:* Around Week 9 (late March)
3. **Project Presentation (25%):** Students will give a short presentation to showcase their project in class. The focus of this presentation is to demonstrate and describe what was done, report interesting observations, present key conclusions, and discuss potential limitations of the study. Students working in teams may choose to present as a group or elect one of the team members to present on their behalf. Students will not be penalized for choosing not to present individually, as long as the project itself is showcased.  
*Tentative date:* Exam Week (Monday, April 29th, 2024)
4. **Final project deliverable and report (50%):** Students are expected to submit their project deliverable, along with a brief report. The report should include the key observations and conclusions based on the project and suggest potential follow-up studies. Teams working on the project together must also describe individual contributions of the team members.  
*Tentative deadline:* Wednesday of Exam week (Wednesday, May 1st, 2024)

## C. WEEKLY SCHEDULE

Please check [this page](#) periodically for a more detailed, accurate, and up-to-date schedule and reading list. The following list of topics and schedule is tentative and will likely change based on class dynamic.

## **Unit 1: Introduction (Typically, Week 1)**

- What is Natural Language Processing for Health Data?
- Tasks and Challenges of Big Data in Health

## **Unit 2: NLP Essentials: Basic Natural Language Processing (Typically, Week 2)**

- Dealing with words, sentences
  - tokenization, normalization
  - ngrams
  - word sense disambiguation
  - sentence boundary detection
  - Challenges due to acronyms (polysemy, synonymy)
- Processing sentences and corpora
  - part of speech tagging
  - parsing, syntax
  - semantic role labeling
  - pronoun, co-reference resolution
  - statistical NLP
- Tools: NLTK, CoreNLP (Stanford NLP)

## **Unit 3: Text classification (Typically, Weeks 4-6)**

- Classification paradigms
- Evaluation metrics
- Data sources:
  - i2b2 and MIMIC
  - SMM4H
  - Other research datasets
- Text classification methods / algorithms
  - Support Vector Machines
  - Naïve Bayes
  - Decision trees
- Input representation: features
- Term weighting (tf-idf)
- Tools: Weka, Scikit-learn

## **Unit 4: Information Extraction (Typically, Weeks 3, 7-9)**

- Named entity recognition
- Protected health information
- De-identification
- Negation detection and hedging

- Regular expressions
- Hidden Markov Models
- Conditional Random Fields
- Deep learning and Neural network models
  - Word Embedding and other input representation models
  - Convolutional Neural Networks (CNN)
  - Long Short-term Memory (LSTM) models
  - Transformer models
- Role of annotations
- Tools: NegEx, MIST, MedEx, MedXN, MedLEE, CLAMP, PyTorch

### **Unit 5: Information Retrieval (Typically, Week 12)**

- Vector space models
- Probabilistic models
- Index construction
- Query construction
- Ranking retrieved results
- Evaluation and Shared Tasks
- Tools: EMERSE, TREC Clinical Decision Support Task

### **Unit 6: Unified Medical Language Systems (UMLS) & Medical Ontologies (Typically, Week 10)**

- Metathesaurus
- Semantic Relationships
- Lexical Tools and SPECIALIST Lexicon
- Medical Subject Heading (MeSH)
- RxNORM
- ICD codes
- LOINC
- SNOMED CT
- Tools: MetaMap, QuickUMLS, cTAKES

### **Unit 7: Advanced methods and Applications (Typically, Week 13)**

- Question Answering
  - question classification
  - answer extraction and ranking
- Summarization
- Social media mining
- Community health informatics
- Sentiment analysis
- Tools: Watson Health

## Unit 8: Ethical considerations in Big Data for Health (Typically, Week 11)

- The power of Big Data
- “With great power comes great responsibility!”

### D. READINGS

The readings of this course will be selected from the recent literature in major journals and conference proceedings in the field of medical informatics. They include, but are not limited to, the Journal of American Medical Informatics Association (JAMIA), the Journal of Biomedical Informatics (JBI), the Journal of Medical Internet Research (JMIR), Bioinformatics, and conferences such as the Annual Meeting of the American Medical Informatics Association (AMIA). Some relevant papers published in the Computer Science venues that describe relevant methodologies for natural language tasks will also be selected. Such venues include the Association of Computational Linguistics (ACL), Empirical Methods of Natural Language Processing (EMNLP), and the Association for the Advancement of Artificial Intelligence (AAAI). It is also encouraged that students review and suggest relevant literature to add to the reading list.

#### D.1. Suggested Reading List

[This page](#) lists some seminal and noteworthy papers as a reading list for the course. Many of the papers from this list will be assigned as weekly reading assignments.

In addition, the following textbook that could be used for an *optional* supplemental reading.

1. Kevin B Cohen and Dina Demner-Fushman. Biomedical Natural Language Processing.  
*This book has a good introduction to various biomedical natural language processing tasks for those with a working knowledge of NLP.*

### E. POLICIES

#### E.1. Late submission policy

Students have 72 hours of buffer grace period for the entire semester. If necessary, students may use it to submit any of the assignments, homework, or the course project reports late without any effect on the overall grade. The grace period, however, cannot be used to submit the exams or quizzes late. A student may use it all on one assignment or use a bit of it for any number of assignments. Once the buffer grace period is used up, late submissions will not be graded.

#### E.2. Expectations around Attendance

This is an advanced elective course; and therefore, discussion and sharing your perspectives about the course content are key to learning for the entire class. There will be readings assigned each week that will prepare the student for discussions both during

the synchronous sessions and on Canvas. Much of research is problem solving; therefore, students will focus on difficult scenarios and questions that present challenging approaches. Some of the discussions will ask you to interpret the approaches and findings in prior work in your own work, and will address some of the most common decision-based scenarios students may face in the future as independent investigators and analysts.

All students are expected to complete all assigned activities for the week before the next session, including posting their responses to weekly prompts asynchronously. All students are expected to attend the synchronous sessions and participate in the discussions. This will be facilitated through small-group breakout sessions.

Attendance will not be taken during the synchronous sessions. However, class participation during the synchronous sessions and/or active engagement with online discussion throughout the semester constitutes 5% of the overall grade.

### ***E.2.1. Digital Etiquette***

Students should mute their phones/tablets/laptops before class sessions begin and turn off ringtones and other audible alerts to avoid distractions during the live sessions. Turning on the camera or video feed is optional but encouraged. However, it is expected that you are fully engaged and participating in the discussions during the synchronous sessions. If there is an emergency, please excuse yourself from the classroom and/or the Zoom meeting to attend to it.

### **E.3. Accessibility**

The university provides reasonable accommodations to qualified individuals with disabilities or accessibility concerns upon request. If you think you need an accommodation for a disability, please let the instructor know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way we teach may be modified to facilitate your participation and progress.

As soon as you make me aware of your needs, we can work with the Office of Services for Students with Disabilities (SSD) to help us determine appropriate accommodations. SSD (734-763-3000; <http://www.umich.edu/sswd/>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. I will treat any information that you provide in as confidential a manner as possible. For more information, see <https://ssd.umich.edu/article/americans-disabilities-act-ada>.

Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit <https://ssd.umich.edu/topic/our-services> for more information on student accommodations and SSD's [Frequently Asked Questions](#) page for COVID-19 accommodations updates.

#### **E.4. Religious Accommodations**

It is also the University's policy that every reasonable effort be made to help students avoid negative academic consequences when their religious obligations conflict with academic requirements. Students who expect to miss classes, examinations, or other assignments as a consequence of their religious observance are requested to contact the instructor by the course drop/add deadline (around third week of classes). For more information see [https://www.provost.umich.edu/calendar/religious\\_holidays20-21.html](https://www.provost.umich.edu/calendar/religious_holidays20-21.html)

#### **E.5. COVID-19 Statement**

This course is being run in-person in Winter 2024 with an option for live, remote participation. As we continue to address the lingering disruptions of COVID-19, we, as a community, are continuously evaluating the best approach to provide outstanding learning experience.

For the safety of all students, faculty, and staff on campus, it is important for everyone to comply with safety measures that have been put in place for our protection. We each have a responsibility for protecting the collective health of our community. Your participation in this course on an in-person basis is conditional upon your adherence to all safety measures mandated by the State of Michigan and the University, including maintaining physical distancing of six feet from others, properly wearing a face covering in class, and following instructions regarding cleaning your study space. Other applicable safety measures may be described in the [Wolverine Culture of Care](#) and the [University's Face Covering Policy for COVID-19](#). Your ability to participate in this course in-person, as well as your grade, may be impacted by failure to comply with campus safety measures.

Individuals seeking to request an accommodation related to the face covering requirement under the Americans with Disabilities Act should contact the [Office for Institutional Equity](#). If you are unable or unwilling to adhere to these safety measures while in a face-to-face class setting, you will be required to participate on a remote basis or to disenroll from the class. I also encourage you to review the [Statement of Student Rights and Responsibilities](#), which includes a COVID-related Statement Addendum.

##### ***E.5.1. Live, Remote Participation***

Remote option will be made available for all sessions that the class meets in-person, via Zoom. Meeting dates and times will be scheduled through [Canvas](#) and should appear on your calendar.

Please take the time to familiarize yourself with Zoom by visiting [U-M's Getting Started with Zoom](#) page. You may choose to use Zoom on your mobile device (phone or tablet).

Things to Know About Zoom:

- You must sign into my Zoom session using your uniquename and password (for med.umich.edu users: enter the Level 1 credentials).

- The Zoom sessions are recorded.
- Improper classroom behavior is not tolerated within Zoom sessions.
- You can contact [Information and Technology Services](#) if you have any technical issues accessing Zoom. You can also contact [Health Information Technology & Services](#) for help.

## **E.6. Academic Conduct**

### ***E.6.1. Collaboration***

The academic community, like all communities, functions best when its members treat one another with honesty, fairness, respect, and trust. The Department of Learning Health Sciences and the instructor strongly encourage collaboration while working on some assignments, such as homework problems and interpreting reading assignments as a general practice. Active learning is effective. Collaboration with other students in the course will be especially valuable in summarizing the reading materials and picking out the key concepts. You must, however, write your homework submission on your own, in your own words, before turning it in. If you worked with someone on the homework before writing it, you must list any and all collaborators on your written submission. Read the instructions carefully and request clarification about collaboration when in doubt. Collaboration is almost always forbidden for take-home and in class exams.

### ***E.6.2. Academic Integrity***

All written submissions must be your own, original work. Original work for narrative questions is not mere paraphrasing of someone else's completed answer: you must not share written answers with each other at all. At most, you should be working from notes you took while participating in a study session.

You may incorporate selected excerpts, statements, or phrases from publications by other authors, but they must be clearly marked as quotations and must be attributed. If you build on the ideas of prior authors, you must cite their work. You may obtain copy-editing assistance, and you may discuss your ideas with others, but all substantive writing and ideas must be your own or be explicitly attributed to another. Largely duplicate copies of the same assignment will receive an equal division of the total point score from the one piece of work.

Please refer to the Rackham's Graduate School Academic for the definition of plagiarism, cheating, and other academic misconduct; the consequences for intentional or unintentional plagiarism; and resources to help you avoid it. Suspected academic misconduct will be handled by the course instructor. <https://rackham.umich.edu/academic-policies/section8/>

The policy handbook is also available here: <http://www.rackham.umich.edu/current-students/policies/academic-policies>

## **E.7. Student Grading and Academic Dispute Resolution**

### ***E.7.1. Grading Dispute Resolution***

Regular feedback on graded assignments and timely resolution of any grading concerns is helpful for student learning and overall positive experience. Students are encouraged to discuss any assignment grade concerns with the instructor within 48 hours of the grades being posted on Canvas. Similarly, any concerns and grievances related to the overall course grades should be raised by contacting the faculty of record within 48 hours of the course grades being posted on Wolverine Access.

For this course, the grade dispute arbitration will happen within the Health Infrastructures and Learning Systems (HILS) program. After initial discussion with the faculty of record for the course, there is an opportunity for a second look by another HILS faculty member with appropriate expertise for resolving grade disputes. If the student and faculty are not able to come to an agreement about the grade dispute, the HILS program director or assigned designee (e.g., the Associate Program Director) will adjudicate the dispute. If the student is still dissatisfied after adjudication by the HILS program director or assigned designee, the dispute will escalate to the Conflict resolution office at the Medical School.

### ***E.7.2. Student Academic Dispute Procedures***

Rackham's Academic Dispute Resolution Policy and Procedures are available to Rackham students who have a dispute or disagreement with faculty or staff about the equity and fairness of decisions or procedures that affect their academic standing, the conduct of their research, and progress toward the degree. Such issues may arise regarding fair and equal treatment in the conduct of a class, in the pursuit of the student's research, and in the grading or evaluation of academic work and research. Other issues may concern the equity and fairness of program, department, or Rackham policies. <https://rackham.umich.edu/academic-policies/section9/>

## **E.8. Diversity, Equity, and Inclusion**

The Department of Learning Health Sciences (DLHS) is committed to developing the institutional mechanisms and norms necessary to promote the values of diversity, equity, and inclusion, both inside and outside our classrooms. To this end, DLHS upholds the expectations that all courses will: (1) be inclusive, (2) promote honest & respectful discussions, (3) follow multicultural ground rules and (4) abide by UM policies and procedures. Inclusive courses are those in which teachers and learners co-create and co-sustain environments that support and encourage all members to participate equitably. <https://rackham.umich.edu/rackham-life/diversity-equity-and-inclusion/>

## **E.9. Student mental health and wellbeing**

The University of Michigan is committed to advancing the mental health and wellbeing of its students, while acknowledging that a variety of issues, such as strained relationships, increased anxiety, alcohol/drug problems, and depression, directly impacts

students' academic performance. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at (734) 764-8312 and <https://caps.umich.edu/> during and after hours, on weekends and holidays or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at (732) 764-8320 and <https://www.uhs.umich.edu/mentalhealthsvcs>, or for alcohol or drug concerns, see [www.uhs.umich.edu/aodresources](http://www.uhs.umich.edu/aodresources). For a more comprehensive listing of other mental health resources available on and off campus, please visit <http://umich.edu/~mhealth/>.

### **E.10. Sexual Misconduct/Sexual Harassment Reporting**

Title IX prohibits sex discrimination to include sexual misconduct: harassment, domestic and dating violence, sexual assault, and stalking. If you or someone you know has been harassed or assaulted, you can receive confidential support and academic advocacy at the Sexual Assault Prevention and Awareness Center (SAPAC). SAPAC can be contacted on their 24-hour crisis line, 734-936-3333 and online at [sapac.umich.edu](http://sapac.umich.edu). Alleged violations can be reported non-confidentially to the Office for Institutional Equity (OIE) at [institutional.equity@umich.edu](mailto:institutional.equity@umich.edu). Reports to law enforcement can be made to University of Michigan Police Department at 734-763-3434.

### **E.11. Support for Food Insecurity**

Students across the country experience food insecurity at alarming rates. The Maize and Blue Cupboard at the University of Michigan provides food, kitchen and cooking supplies, personal and household items, and support services. For information about accessing their services, please visit <https://mbc.studentlife.umich.edu/>

### **E.12. Confidentiality and Mandatory Reporting**

As instructors, one of our responsibilities is to help create a safe learning environment on our campus. Instructors are required to share information regarding sexual misconduct or information about a crime that may have occurred on U-M's campus with the University. Students may speak to someone confidentially by contacting SAPAC's Crisis Line at (734) 936-3333.

### **F. Faculty of record and Instructor:**

V.G.Vinod Vydiswaran, Ph.D.

Associate Professor, Department of Learning Health Sciences, Medical School,  
University of Michigan

Associate Professor, School of Information, University of Michigan

1161F - NIB, 300 N. Ingalls Street, Ann Arbor, MI 48109

(734) 647-1207

[vgvinodv@umich.edu](mailto:vgvinodv@umich.edu) (preferred mode to reach the instructor)