

# Artificial Intelligence (AI) in Epidemiology

## Epidemiology Seminar Series

### VTPMD 7080

**Semester:** Fall/2022

**Class location:** Zoom

**Class meeting time:** Mondays 4-5pm EST

**Grading:** S/U

**Credit Hours:** 1 Hour

### Instructor: Dr. Renata Ivanek-Miojevic

E-mail: [ri25@cornell.edu](mailto:ri25@cornell.edu)

Office: S1-026 Schurman Hall

Office phone: 607-253-4383; Cell: 607-216-2716

Office hours: After class or by appointment

### Course Rationale and Learning Outcomes:

The purpose of this course is to provide a platform for discussion of the health-related topics, with focus on the population-level inference, epidemiologic methods and principles, and the strength of evidence. This semester the discussion topic is *AI in Epidemiology*. Using machines, AI simulates human intelligence in systems programmed to think like humans and mimic their actions or support decision making by humans. AI technologies, such as, but not limited to, those based on neural networks, vision, robotics, expert systems, machine learning, natural language processing, or fuzzy logic, can enhance disease prevention and control at an individual as well as population level.

In this journal club-based course we will discuss the state of AI in epidemiology and examine existing research. At the core of the discussions will be the epidemiologic principles for study design and conduct, and the strength of evidence.

**Format, Procedures, Requirements and Grading:** Attendance at class meetings is expected. The grade for this course will be S/U and will be based on attendance, participation in discussions and moderating some class meetings over the semester. Moderating a class meeting means that you will prepare 5-10 questions for articles assigned for the meeting, send questions to the rest of the seminar participants a week before the meeting and use the questions to facilitate discussion and learning at the meeting. While knowing answers to all questions posed by the moderator is commendable, it is not necessary. To pass the course, the student is expected to attend at least 10 of 13 scheduled class meetings.\*

\*A make-up oral final exam will be available to students only if their absences were University Excused. These students will be encouraged to prepare a two-page long essay summarizing the course reading material as a preparation for the oral exam.

**Academic Integrity:** Each student in this course is expected to abide by the Cornell University Code of Academic Integrity.

**Accommodations for students with disabilities:** In compliance with the Cornell University policy and equal access laws, I am available to discuss appropriate academic accommodations that may be required for student with disabilities. Requests for academic accommodations are to be made during the first three weeks of the semester, except for unusual circumstances, so arrangements can be made. Students are encouraged to register with Student Disability Services to verify their eligibility for appropriate accommodations.

**Inclusivity Statement:** This course is committed to providing an atmosphere for learning that respects diversity.

## Epidemiology seminar schedule and readings

### Contents

Aug 29, 2022 .....	2
Sep 12, 2022.....	3
Sep 19, 2022.....	4
Sep 26, 2022.....	5
Oct 3, 2022 .....	6
Oct 17, 2022 .....	7
Oct 24, 2022 .....	8
Oct 31, 2022 .....	9
Nov 7, 2022 .....	10
Nov 14, 2022 .....	11
Nov 21, 2022 .....	12
Nov 28, 2022 .....	13
Dec 5, 2022 .....	14
Additional Reading.....	15

Aug 29, 2022

**Ethics guidelines for trustworthy AI** <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

**Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment**  
<https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

**A definition of Artificial Intelligence: main capabilities and scientific disciplines**  
<https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

Discussion questions:

1. What are the key characteristics of “artificial intelligence”?
2. What is “intelligence” (not just in the context of AI)?
3. What is “rationality”?
4. How does an AI system achieve “rationality”?
5. What is “bounded rationality”?
6. What is a “learning rational system”?
7. What do the techniques for “reasoning and decision-making” do?
8. What are the techniques for “learning” within AI systems?
9. What are the differences between “narrow” and “general” AI?
10. How can data introduce bias into an AI system? What are the data privacy issues?
11. What is a “black-box AI” and what is an “explainable” AI?
12. What is a “goal directed AI”?
13. What did you find the most interesting in the assessment list for “Trustworthy AI”?

Sep 12, 2022

Appleby, R. B., & Basran, P. S. (2022). **Artificial intelligence in veterinary medicine**, *Journal of the American Veterinary Medical Association*, 260(8), 819-824.

<https://doi.org/10.2460/javma.22.03.0093>

Discussion questions:

1. Article summary and aim
2. How has AI benefited the medical practice?
3. Is anyone familiar with the One Health approach? Can you please explain what it means
4. How does this approach tie into vet medicine and AI?
5. Do you think AI is necessary for practice? In what way? – You can refer to the article
6. What are some challenges of integrating AI in vet medicine? Mentioned in the article or outside the article
7. How can they be overcome? Or What are some possible solutions or test trials?

Basran, P. S., & Appleby, R. B. (2022). **The unmet potential of artificial intelligence in Veterinary Medicine**. *Journal of the American Veterinary Medical Association*.

<https://doi.org/10.2460/ajvr.22.03.0038>

Discussion questions:

1. What are your thoughts on Transfer learning? Has this been done in any other field? What potential might this have?
2. What application of AI in vet medicine were you most interested in and why? ( Some examples listed were companion animal care, and disease abnormalities)
3. Both articles emphasized the importance of good data. In current medicine (not just vet medicine) do you think data is stored and processed effectively? What can be improved?
4. What was the biggest takeaway from either of the articles?

Key Terms / Concepts to look out for when reading

- The three types of AI – strong, artificial narrow, and artificial superficial
- Broad categories that encompass AI
- Machine learning
- Machine learning models – supervised, unsupervised, semi supervised
- Artificial neural networks (ANN)
- Radionics
- Natural learning processing (NLP)
- Cost function
- Transfer learning

Sep 19, 2022

## **How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic**

<https://www.mdpi.com/1660-4601/17/9/3176/htm>

Discussion questions:

1. What are the three classical Vs of Big Data?
2. The article actually mentions another “V” of Big Data. What is it?
3. The reading provides some examples of types of Big Data. What are they and what distinguishes them from one another?
4. What is an example of a non-classical dataset and its use case?
5. The article provides several examples of short-term applications of AI and Big Data in COVID-19. Please pick one or two of your favorites and explain why.
6. The article provides several examples of medium-term applications of AI and Big Data in COVID-19. Please pick one or two of your favorites and explain why.
7. **For Discussion** - Accurate and effective use of AI and Big Data requires the collection of a large amount of data on as many different people as possible. However, this collection of data inevitably requires the individual to be connected to the “technology grid” in some way or another. This means that disadvantaged populations who lack reliable, personal access to technology like the internet and smart devices may be left even further behind as we move towards a future that is driven by AI and Big Data-based research and solutions.
  1. How might we be able to collect data on the populations that are not “plugged in”?
  2. Even if we are unable to collect data on this population, what are some ways they can still benefit from an AI or Big Data driven future?

## **Artificial Intelligence for infectious disease Big Data Analytics**

<https://www.sciencedirect.com/science/article/abs/pii/S2468045118301445>

Discussion questions:

1. What is “dirty data” and why is it important to the discussion of AI and Big Data?
2. What makes the application of conventional data and statistical methods to the medical and public health fields particularly difficult?
3. Which kinds of experts will need to collaborate to ensure that AI and Big Data can be applied to public health problems? What would be their roles?
4. What are some limitations of utilizing AI specifically in the context of epidemiology?

Sep 26, 2022

### **Artificial Intelligence and One Health: Knowledge Bases for Causal Modeling**

<https://link.springer.com/article/10.1007/s41745-020-00192-3#Sec2>

Discussion questions:

1. Short summary of the article and main takeaways/thoughts.
2. Can someone remind us what One Health is quickly?
3. How can AI play a role in tracing KFD?
4. Can someone tell us what non-monotonic logic models are?
5. What is causal modeling? Give examples.
6. This article was written in 2020, does anyone know of any new AI technology that has been created since then that would assist One Health scientist?

### **High-performance medicine: the convergence of human and artificial intelligence**

<https://www.nature.com/articles/s41591-018-0300-7>

Discussion questions:

1. Main thoughts and conclusions from the article.
1. What is ghost cytometry?
2. What was your biggest take away from this paper?
3. How does this paper help us understand how to use AI in a One Health perspective?

Oct 3, 2022

**“Vaccines Meet Big Data: State-of-the-Art and Future Prospects. From the Classical 3Is (“Isolate–Inactivate–Inject”) Vaccinology 1.0 to Vaccinology 3.0, Vaccinomics, and Beyond: A Historical Overview”**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5845111/>

Discussion questions:

1. Can someone remind us what big data is (think: 3 V's)?
2. What stood out to you from this article? Any connections to previous articles we've discussed?
3. This article talks about the urgency of “new vaccines” in light of globalization. Can someone talk about how human migration and globalization can lead to the emergence or reemergence of disease?
4. What is the difference between vaccinology 1.0 and 2.0?
5. What are some examples of “second generation” vaccines?
6. What is vaccinology 3.0?
7. What is vaccinomics?
8. Thoughts on the use of experimental immunology or “reverse vaccinology” to develop new vaccines? Can you think of any potential limitations of this technology?

**“Using big data to understand the online ecology of COVID-19 vaccination hesitancy”**

<https://www.nature.com/articles/s41599-022-01185-6>

Discussion questions:

1. What is vaccine hesitancy?
2. What are the 3 proposed categories of factors that affect individuals' refusal to get vaccinated?
3. How does social media facilitate vaccine discourse? Did anything stand out to you regarding the current discourse surrounding vaccination uptake on social media?
4. Can someone provide an overview of the Health Belief Model (HBM)?
5. How can the HBM provide insight into vaccine uptake?
6. Taking into consideration HBM constructs and factors noted in the article, how can public health and other health professionals use Big Data to address vaccine hesitancy and vaccine uptake?
7. Do/should social media platforms have a responsibility to censor health misinformation? Could this impact Big Data analysis and public health applications?

Oct 17, 2022

Broadbent, A., & Grote, T. (2022). **Can robots do epidemiology? Machine Learning, causal inference, and predicting the outcomes of public health interventions.** *Philosophy & Technology*, 35(1). <https://doi.org/10.1007/s13347-022-00509-3>  
<https://link.springer.com/article/10.1007/s13347-022-00509-3>

Discussion questions:

1. What role does causation play in epidemiology?
2. How is causation not considered by machine learning?
3. What data sources can be used to conduct epidemiology with machine learning? What types of conclusions can be drawn?
4. What is the difference between machine learning for computational uses and more-than-computational uses?
5. What is the relationship between prediction and causal inference?
6. What were the three “weak reasons” given to reject machine learning in epidemiology? What do these mean?
7. To what extent do you agree or disagree with the case for causal constraint when using machine learning in epidemiology?
8. What four reasons were given for causal constraint when using machine learning in epidemiology? Why do they fail?
9. What arguments can you make for and against the use of machine learning in epidemiology? What do you think the constraints of machine learning in epidemiology are?

Wiemken, T. L., & Kelley, R. R. (2020). **Machine learning in Epidemiology and Health Outcomes Research.** *Annual Review of Public Health*, 41(1), 21–36.  
<https://doi.org/10.1146/annurev-publhealth-040119-094437>  
[https://www.annualreviews.org/doi/10.1146/annurev-publhealth-040119-094437#\\_i23](https://www.annualreviews.org/doi/10.1146/annurev-publhealth-040119-094437#_i23)

Discussion questions:

1. What is epidemiology?
2. What is the curse of dimensionality?
3. What are the differences between supervised and unsupervised models?
4. What elements or steps are included in the development of a machine learning model?
5. What is pretraining or hyperparameter optimization?
6. For one of the issues described, please explain the conclusion.
7. What is one of the issues with interventional machine learning?
8. What are your thoughts on open-source and user-friendly software for machine learning models?



Oct 24, 2022

**Artificial intelligence in healthcare: An essential guide for health leaders:**

<https://pubmed.ncbi.nlm.nih.gov/31550922/>

Discussion questions:

1. What are some challenges that health organizations need to overcome to bring AI to success?
2. What is machine learning?
3. Explain one type of machine learning algorithm and give an example of how this algorithm is being used in healthcare (hint: there are 4 types of algorithms mentioned)
4. Describe a situation where machine learning is less applicable or not applicable in healthcare.
5. What is natural language processing? Give an example of how natural language processing can be used in healthcare.
6. What is artificial intelligence voice technology? What are artificial intelligence assistants? Give an example of how these technologies are being used in healthcare
7. Do you use artificial intelligence assistants in your daily lives? Have they been helpful to take care of your health?
8. Which machine learning highlight stood out the most to you and why?
9. What do you think about the partnerships between governments and AI industry in healthcare? Any benefits or harm?

**Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians:**

<https://pubmed.ncbi.nlm.nih.gov/32558657/>

Discussion questions:

1. AI research in health care poses unique challenges compared to other technical domains. Which are those challenges?
2. The article mentions AI in health care has two potential advantages to human performance, which are those advantages? Do you agree with these advantages?
3. The article mentions some limitations of AI that might restrict its application in life-critical areas such as healthcare, which are those limitations? Do you agree with these limitations?
4. What is reliability and why is this concerning when using AI in healthcare?
5. Which are the factors that should be incorporated into the development of AI to achieve an optimal level of trust? Describe one of these factors. (Hint: the article mentions 3 factors)
6. Why are clinicians concerned about medical responsibility while using AI?
7. How is trust in AI different for patients compared to clinicians?

Oct 31, 2022

**The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database**

<https://www.nature.com/articles/s41746-020-00324-0>

Discussion questions:

1. Briefly, what are the promises and obstacles for AI/ML-based medical technology?
2. What are the three levels of clearance and how are they defined?
3. What is a “locked” algorithm and an adaptive algorithm and how they should be regulated differently?
4. What are the main areas (e.g., cardiology) of FDA-approved medical devices? Can you name one application from each area?
5. What do authors think FDA should improve in the regulation over AI/ML-based medical devices?
6. What other challenges in addition to regulations do you think are there in the context of commercialization of AI/ML-based medical devices?

**Machine-learned epidemiology: real-time detection of foodborne illness at scale**

<https://www.nature.com/articles/s41746-018-0045-1>

Discussion questions:

1. What is syndromic surveillance and how did it improve over years?
2. What is the current approach of determining an outbreak and what are the limitations?
3. What methods did authors use to preserve privacy of users?
4. What methods did authors use to improve accuracy of detecting symptoms related to food poisoning?
5. How did authors compare the efficacy of the application with usual inspection protocols?
6. What are the advantages of FINDER as concluded by authors?
7. What are the limitations of FINDER as concluded by authors?
8. What will be the challenges if this application is to be deployed in more cities?
9. Do you think large tech companies such as Google and Twitter are obligated to provide data that can help government determine the foodborne outbreaks?

Nov 7, 2022

### **A call for biological data mining approaches in epidemiology**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4700596/>

Discussion questions:

1. How does the use of -omic technologies change and/or benefit the field of epidemiology?
2. What is the difference between using regression models and Complex system/ABM in studying risk factors?
3. What are the limitations of using ABM?
4. How can the incorporation of biological data mining approaches address some of the limitations in Epidemiology?

### **A data mining approach to investigate food groups related to incidence of bladder cancer in the Bladder cancer Epidemiology and Nutritional Determinants International Study**

<https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/data-mining-approach-to-investigate-food-groups-related-to-incidence-of-bladder-cancer-in-the-bladder-cancer-epidemiology-and-nutritional-determinants-international-study/1CFF189D6C50D799B17F9706F60F003B>

Discussion questions:

1. What are some limitations of using conventional methods such as factor and cluster analyses?
2. What is data mining?
3. What do previous studies show when comparing the use of data mining over statistical methods in nutritional epidemiology?
4. How was a decision tree used in elucidating relationships between risk factors (diet, age, sex, smoking status) and bladder cancer?
5. In this study, do you think data mining accounts for any confounding factors that may exist?
6. What proof does the authors use to demonstrate that data mining is effective in studying diets in relation to cancer risk factors?
7. What other risk factors could you include in the study? Or what changes would you make to the study?

Nov 14, 2022

### **Prospects and Pitfalls of Machine Learning in Nutritional Epidemiology**

<https://www.mdpi.com/2072-6643/14/9/1705/htm>

Discussion questions:

1. What is nutritional epidemiology and how are the findings from nutritional epidemiology used?
2. What are the three main research areas in nutritional epidemiology?
3. What are the challenges in the process of causal inference? Briefly describe each challenge.
4. What is ANN? DNN? How are they different? (Uses, examples etc.)
5. How is ML used to collect data for nutritional epidemiology? Which example stood out to you the most? How does the collection of data impact the data quality?
6. What is dimensionality in ML and how is ML used to reduce dimensionality? (Hint: the authors mentioned two techniques)
7. Why is it important to control confounding and multicollinearity in nutritional epidemiologic studies? How is ML used to deal with confounding?
8. What are the practical recommendations that the authors provided? Briefly explain each recommendation.
9. After reading this paper, what are your thoughts on the use of ML in nutritional epidemiology? What benefits/limitations stood out to you the most? Do you think the use of ML is critical in this field? Do you have any personal experience in nutritional epidemiology research which you could have/hoped to apply ML?

### **Role of artificial intelligence-internet of things (AI-IoT) based emerging technologies in the public health response to infectious disease in Bangladesh**

<https://www.sciencedirect.com/science/article/pii/S2405673122000307>

Discussion questions:

1. What is IoT? What role can IoT and AI technologies play in endemic countries?
2. How were IoT and AI used for the COVID-19 and Aedes mosquito outbreak? Provide examples.
3. What are your thoughts on the digital healthcare systems (telehealth, mobile health etc.)? What about smart disease surveillance system?
4. Do you think the applications of IoT and AI mentioned in this paper are realistic (especially in a setting like Bangladesh)? Any concerns?
5. What are your thoughts on data security/privacy? To what extent do you think it should be protected or do you think it should not be a concern if it is used for improving public health?

Nov 21, 2022

**Identifying and modeling the distribution of cryptic reservoirs of Ebola virus using artificial intelligence**

Kollars et al. 2018

<https://www.ajol.info/index.php/ajcem/article/view/172666>

Discussion questions:

1. What are the main goals of the study?
2. What are the possible pathways that EBOV uses to jump between species?
3. What is BioTEMS, how does it work, and what has it previously been used for?
4. What are the major findings and benefits of this study?
5. What are the limitations of this method and study?

**Artificial intelligence and avian influenza: Using machine learning to enhance active surveillance for avian influenza viruses**

Walsh et al 2019

[https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.13318?casa\\_token=NMF9Y8IdjnYAAAAA:m1DFumFQbxYrra5oM9CwPzdHHawmAD9h8POYbiraTBb-ao6gfaHX88CWKwx7dvC\\_VtrKhyifF8oV3A](https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.13318?casa_token=NMF9Y8IdjnYAAAAA:m1DFumFQbxYrra5oM9CwPzdHHawmAD9h8POYbiraTBb-ao6gfaHX88CWKwx7dvC_VtrKhyifF8oV3A)

Discussion questions:

1. What are AIVs and why are they important?
2. What makes AIVs so difficult to monitor?
3. How was machine learning used in this study? What was the overarching goal, and what method did they use? What information did they use to train the algorithm?
4. What variables were the most important predictors in isolating AIV from samples? What patterns emerged across these predictors?
5. How are these findings useful? What are some of the limitations of how they applied this method?

Nov 28, 2022

### **Artificial Intelligence in Dentistry: Past, Present, and Future**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9418762/pdf/cureus-0014-00000027405.pdf>

Discussion questions:

1. What are the key elements of an AI system?
2. Explain the mechanism of working of an AI model.
3. What are the components of the AI hierarchy system and how does it relate to dentistry?
4. What are the applications of AI in endodontics?
5. What are the other applications of AI in dentistry and how is it used?
6. What are some of the fascinating uses of AI in dentistry?
7. What is bioprinting?
8. Can AI completely replace dentists/clinicians? Why or why not?

### **The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6748906/pdf/jgc-16-08-585.pdf>

Discussion questions:

1. What are the 3 classical genres of Machine Learning?
2. What are the AI applications in the medical field?
3. What are the 4 major applications of AI in CVD?
4. What are some examples of applications of AI in CVD?
5. What are the current shortcomings of AI in medicine?
6. What is a traditional and a novel medical model?
7. What is a clinical warning system?

Dec 5, 2022

### **AI Surveillance during Pandemics: Ethical Implementation Imperatives**

<https://onlinelibrary.wiley.com/doi/10.1002/hast.1125>

Discussion questions:

1. What are some examples of how AI has been used in countries around the world to combat the COVID-19 pandemic?
2. What are some examples of statutes or regulations that address privacy concerns to protect health data for individuals, and what parts of the world are covered by these statutes or regulations?
3. What are some examples of data generated by COVID-19 AI surveillance programs that are not covered under the U.S. Health Insurance Portability act of 1996 (HIPAA)?
4. What was the major outcome of the *Jew Ho v. Williamson* federal court case?
5. The authors state that to “avoid discrimination in the use of AI, stakeholders must be transparent about which communities and individuals are being monitored.” In your opinion, do you think that using AI surveillance to selectively monitor specific communities can lead to discrimination/stigmatization of those communities? Why or why not?
6. Can you think of any applications/websites that cannot be used or installed without allowing that application/website to be granted access to your data? In such a case would you say the act of granting access to your user data is voluntary or compulsory? Explain your choice.

### ***Salmonella enterica* serovar Typhimurium isolates from wild birds in the United States represent distinct lineages defined by bird type**

<https://journals.asm.org/doi/full/10.1128/aem.01979-21>

Discussion questions:

1. Why is it important to study *Salmonella enterica*, a pathogen usually transmitted to humans through eating contaminated food, in non-food sources like wild birds?
2. What are some examples of possible interspecies transmission of *S. enterica* between wild birds and humans?
3. How did the authors obtain the 131 *S. Typhimurium* isolates from wild birds that they subjected to whole genome sequencing in this study? Was there bias in this selection process?
4. What is a “table of confusion”, and how was it used to assess the ability of the machine learning Random Forest Classifier to correctly identify the source of *S. Typhimurium* isolates?
5. How did the addition of the 131 *S. Typhimurium* isolates from wild birds into the Random Forest Classifier change the accuracy of predicting the source of *S. Typhimurium* isolates?
6. How can enhanced prediction of the source of a human pathogen (e.g., *Salmonella*) through machine learning techniques ultimately aid in improving public health outcomes?

## Additional Reading

Regulation of predictive analytics in medicine: Algorithms must meet regulatory standards of clinical benefit	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6557272/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6557272/</a>
Artificial Intelligence in Public Health and Epidemiology	<a href="http://dx.doi.org/10.1055/s-0038-1667082">http://dx.doi.org/10.1055/s-0038-1667082</a>
COVID-19 pandemic and artificial intelligence: challenges of ethical bias and trustworthy reliable reproducibility?	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8495685/pdf/bmjhci-2021-100438.pdf">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8495685/pdf/bmjhci-2021-100438.pdf</a>
The epidemic volatility index, a novel early warning tool for identifying new waves in an epidemic	<a href="https://www.nature.com/articles/s41598-021-02622-3">https://www.nature.com/articles/s41598-021-02622-3</a>
Application of Neural Network Model in an Epidemiological Study	<a href="https://www.sciencepublishinggroup.com/journal/paperinfo.aspx?journalid=148&amp;doi=10.11648/j.ajam.20150304.16">https://www.sciencepublishinggroup.com/journal/paperinfo.aspx?journalid=148&amp;doi=10.11648/j.ajam.20150304.16</a>
Artificial Intelligence and Big Data in Public Health	<a href="https://www.mdpi.com/1660-4601/15/12/2796/htm">https://www.mdpi.com/1660-4601/15/12/2796/htm</a>
Strengthening public health surveillance through blockchain technology	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779606/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779606/</a>



# Epidemiology seminar schedule and readings

## Contents

Aug 29, 2022 .....	2
Sep 12, 2022.....	3
Sep 19, 2022.....	4
Sep 26, 2022.....	5
Oct 3, 2022 .....	6
Oct 17, 2022 .....	7
Oct 24, 2022 .....	8
Oct 31, 2022 .....	9
Nov 7, 2022 .....	10
Nov 14, 2022 .....	11
Nov 21, 2022 .....	12
Nov 28, 2022 .....	13
Dec 5, 2022 .....	14
Additional Reading.....	15

Aug 29, 2022

**Ethics guidelines for trustworthy AI** <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

**Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment**  
<https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment>

**A definition of Artificial Intelligence: main capabilities and scientific disciplines**  
<https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

Discussion questions:

1. What are the key characteristics of “artificial intelligence”?
2. What is “intelligence” (not just in the context of AI)?
3. What is “rationality”?
4. How does an AI system achieve “rationality”?
5. What is “bounded rationality”?
6. What is a “learning rational system”?
7. What do the techniques for “reasoning and decision-making” do?
8. What are the techniques for “learning” within AI systems?
9. What are the differences between “narrow” and “general” AI?
10. How can data introduce bias into an AI system? What are the data privacy issues?
11. What is a “black-box AI” and what is an “explainable” AI?
12. What is a “goal directed AI”?
13. What did you find the most interesting in the assessment list for “Trustworthy AI”?

Sep 12, 2022

Appleby, R. B., & Basran, P. S. (2022). **Artificial intelligence in veterinary medicine**, *Journal of the American Veterinary Medical Association*, 260(8), 819-824.

<https://doi.org/10.2460/javma.22.03.0093>

Discussion questions:

1. Article summary and aim
2. How has AI benefited the medical practice?
3. Is anyone familiar with the One Health approach? Can you please explain what it means
4. How does this approach tie into vet medicine and AI?
5. Do you think AI is necessary for practice? In what way? – You can refer to the article
6. What are some challenges of integrating AI in vet medicine? Mentioned in the article or outside the article
7. How can they be overcome? Or What are some possible solutions or test trials?

Basran, P. S., & Appleby, R. B. (2022). **The unmet potential of artificial intelligence in Veterinary Medicine**. *Journal of the American Veterinary Medical Association*.

<https://doi.org/10.2460/ajvr.22.03.0038>

Discussion questions:

1. What are your thoughts on Transfer learning? Has this been done in any other field? What potential might this have?
2. What application of AI in vet medicine were you most interested in and why? ( Some examples listed were companion animal care, and disease abnormalities)
3. Both articles emphasized the importance of good data. In current medicine (not just vet medicine) do you think data is stored and processed effectively? What can be improved?
4. What was the biggest takeaway from either of the articles?

Key Terms / Concepts to look out for when reading

- The three types of AI – strong, artificial narrow, and artificial superficial
- Broad categories that encompass AI
- Machine learning
- Machine learning models – supervised, unsupervised, semi supervised
- Artificial neural networks (ANN)
- Radionics
- Natural learning processing (NLP)
- Cost function
- Transfer learning

Sep 19, 2022

## **How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic**

<https://www.mdpi.com/1660-4601/17/9/3176/htm>

Discussion questions:

1. What are the three classical Vs of Big Data?
2. The article actually mentions another “V” of Big Data. What is it?
3. The reading provides some examples of types of Big Data. What are they and what distinguishes them from one another?
4. What is an example of a non-classical dataset and its use case?
5. The article provides several examples of short-term applications of AI and Big Data in COVID-19. Please pick one or two of your favorites and explain why.
6. The article provides several examples of medium-term applications of AI and Big Data in COVID-19. Please pick one or two of your favorites and explain why.
7. **For Discussion** - Accurate and effective use of AI and Big Data requires the collection of a large amount of data on as many different people as possible. However, this collection of data inevitably requires the individual to be connected to the “technology grid” in some way or another. This means that disadvantaged populations who lack reliable, personal access to technology like the internet and smart devices may be left even further behind as we move towards a future that is driven by AI and Big Data-based research and solutions.
  1. How might we be able to collect data on the populations that are not “plugged in”?
  2. Even if we are unable to collect data on this population, what are some ways they can still benefit from an AI or Big Data driven future?

## **Artificial Intelligence for infectious disease Big Data Analytics**

<https://www.sciencedirect.com/science/article/abs/pii/S2468045118301445>

Discussion questions:

1. What is “dirty data” and why is it important to the discussion of AI and Big Data?
2. What makes the application of conventional data and statistical methods to the medical and public health fields particularly difficult?
3. Which kinds of experts will need to collaborate to ensure that AI and Big Data can be applied to public health problems? What would be their roles?
4. What are some limitations of utilizing AI specifically in the context of epidemiology?

Sep 26, 2022

### **Artificial Intelligence and One Health: Knowledge Bases for Causal Modeling**

<https://link.springer.com/article/10.1007/s41745-020-00192-3#Sec2>

Discussion questions:

1. Short summary of the article and main takeaways/thoughts.
2. Can someone remind us what One Health is quickly?
3. How can AI play a role in tracing KFD?
4. Can someone tell us what non-monotonic logic models are?
5. What is causal modeling? Give examples.
6. This article was written in 2020, does anyone know of any new AI technology that has been created since then that would assist One Health scientist?

### **High-performance medicine: the convergence of human and artificial intelligence**

<https://www.nature.com/articles/s41591-018-0300-7>

Discussion questions:

1. Main thoughts and conclusions from the article.
1. What is ghost cytometry?
2. What was your biggest take away from this paper?
3. How does this paper help us understand how to use AI in a One Health perspective?

Oct 3, 2022

**“Vaccines Meet Big Data: State-of-the-Art and Future Prospects. From the Classical 3Is (“Isolate–Inactivate–Inject”) Vaccinology 1.0 to Vaccinology 3.0, Vaccinomics, and Beyond: A Historical Overview”**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5845111/>

Discussion questions:

1. Can someone remind us what big data is (think: 3 V's)?
2. What stood out to you from this article? Any connections to previous articles we've discussed?
3. This article talks about the urgency of “new vaccines” in light of globalization. Can someone talk about how human migration and globalization can lead to the emergence or reemergence of disease?
4. What is the difference between vaccinology 1.0 and 2.0?
5. What are some examples of “second generation” vaccines?
6. What is vaccinology 3.0?
7. What is vaccinomics?
8. Thoughts on the use of experimental immunology or “reverse vaccinology” to develop new vaccines? Can you think of any potential limitations of this technology?

**“Using big data to understand the online ecology of COVID-19 vaccination hesitancy”**

<https://www.nature.com/articles/s41599-022-01185-6>

Discussion questions:

1. What is vaccine hesitancy?
2. What are the 3 proposed categories of factors that affect individuals' refusal to get vaccinated?
3. How does social media facilitate vaccine discourse? Did anything stand out to you regarding the current discourse surrounding vaccination uptake on social media?
4. Can someone provide an overview of the Health Belief Model (HBM)?
5. How can the HBM provide insight into vaccine uptake?
6. Taking into consideration HBM constructs and factors noted in the article, how can public health and other health professionals use Big Data to address vaccine hesitancy and vaccine uptake?
7. Do/should social media platforms have a responsibility to censor health misinformation? Could this impact Big Data analysis and public health applications?

Oct 17, 2022

Broadbent, A., & Grote, T. (2022). **Can robots do epidemiology? Machine Learning, causal inference, and predicting the outcomes of public health interventions.** *Philosophy & Technology*, 35(1). <https://doi.org/10.1007/s13347-022-00509-3>  
<https://link.springer.com/article/10.1007/s13347-022-00509-3>

Discussion questions:

1. What role does causation play in epidemiology?
2. How is causation not considered by machine learning?
3. What data sources can be used to conduct epidemiology with machine learning? What types of conclusions can be drawn?
4. What is the difference between machine learning for computational uses and more-than-computational uses?
5. What is the relationship between prediction and causal inference?
6. What were the three “weak reasons” given to reject machine learning in epidemiology? What do these mean?
7. To what extent do you agree or disagree with the case for causal constraint when using machine learning in epidemiology?
8. What four reasons were given for causal constraint when using machine learning in epidemiology? Why do they fail?
9. What arguments can you make for and against the use of machine learning in epidemiology? What do you think the constraints of machine learning in epidemiology are?

Wiemken, T. L., & Kelley, R. R. (2020). **Machine learning in Epidemiology and Health Outcomes Research.** *Annual Review of Public Health*, 41(1), 21–36.  
<https://doi.org/10.1146/annurev-publhealth-040119-094437>  
[https://www.annualreviews.org/doi/10.1146/annurev-publhealth-040119-094437#\\_i23](https://www.annualreviews.org/doi/10.1146/annurev-publhealth-040119-094437#_i23)

Discussion questions:

1. What is epidemiology?
2. What is the curse of dimensionality?
3. What are the differences between supervised and unsupervised models?
4. What elements or steps are included in the development of a machine learning model?
5. What is pretraining or hyperparameter optimization?
6. For one of the issues described, please explain the conclusion.
7. What is one of the issues with interventional machine learning?
8. What are your thoughts on open-source and user-friendly software for machine learning models?

Oct 24, 2022

**Artificial intelligence in healthcare: An essential guide for health leaders:**

<https://pubmed.ncbi.nlm.nih.gov/31550922/>

Discussion questions:

1. What are some challenges that health organizations need to overcome to bring AI to success?
2. What is machine learning?
3. Explain one type of machine learning algorithm and give an example of how this algorithm is being used in healthcare (hint: there are 4 types of algorithms mentioned)
4. Describe a situation where machine learning is less applicable or not applicable in healthcare.
5. What is natural language processing? Give an example of how natural language processing can be used in healthcare.
6. What is artificial intelligence voice technology? What are artificial intelligence assistants? Give an example of how these technologies are being used in healthcare
7. Do you use artificial intelligence assistants in your daily lives? Have they been helpful to take care of your health?
8. Which machine learning highlight stood out the most to you and why?
9. What do you think about the partnerships between governments and AI industry in healthcare? Any benefits or harm?

**Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians:**

<https://pubmed.ncbi.nlm.nih.gov/32558657/>

Discussion questions:

1. AI research in health care poses unique challenges compared to other technical domains. Which are those challenges?
2. The article mentions AI in health care has two potential advantages to human performance, which are those advantages? Do you agree with these advantages?
3. The article mentions some limitations of AI that might restrict its application in life-critical areas such as healthcare, which are those limitations? Do you agree with these limitations?
4. What is reliability and why is this concerning when using AI in healthcare?
5. Which are the factors that should be incorporated into the development of AI to achieve an optimal level of trust? Describe one of these factors. (Hint: the article mentions 3 factors)
6. Why are clinicians concerned about medical responsibility while using AI?
7. How is trust in AI different for patients compared to clinicians?



Oct 31, 2022

**The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database**

<https://www.nature.com/articles/s41746-020-00324-0>

Discussion questions:

1. Briefly, what are the promises and obstacles for AI/ML-based medical technology?
2. What are the three levels of clearance and how are they defined?
3. What is a “locked” algorithm and an adaptive algorithm and how they should be regulated differently?
4. What are the main areas (e.g., cardiology) of FDA-approved medical devices? Can you name one application from each area?
5. What do authors think FDA should improve in the regulation over AI/ML-based medical devices?
6. What other challenges in addition to regulations do you think are there in the context of commercialization of AI/ML-based medical devices?

**Machine-learned epidemiology: real-time detection of foodborne illness at scale**

<https://www.nature.com/articles/s41746-018-0045-1>

Discussion questions:

1. What is syndromic surveillance and how did it improve over years?
2. What is the current approach of determining an outbreak and what are the limitations?
3. What methods did authors use to preserve privacy of users?
4. What methods did authors use to improve accuracy of detecting symptoms related to food poisoning?
5. How did authors compare the efficacy of the application with usual inspection protocols?
6. What are the advantages of FINDER as concluded by authors?
7. What are the limitations of FINDER as concluded by authors?
8. What will be the challenges if this application is to be deployed in more cities?
9. Do you think large tech companies such as Google and Twitter are obligated to provide data that can help government determine the foodborne outbreaks?

Nov 7, 2022

### **A call for biological data mining approaches in epidemiology**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4700596/>

Discussion questions:

1. How does the use of -omic technologies change and/or benefit the field of epidemiology?
2. What is the difference between using regression models and Complex system/ABM in studying risk factors?
3. What are the limitations of using ABM?
4. How can the incorporation of biological data mining approaches address some of the limitations in Epidemiology?

### **A data mining approach to investigate food groups related to incidence of bladder cancer in the Bladder cancer Epidemiology and Nutritional Determinants International Study**

<https://www.cambridge.org/core/journals/british-journal-of-nutrition/article/data-mining-approach-to-investigate-food-groups-related-to-incidence-of-bladder-cancer-in-the-bladder-cancer-epidemiology-and-nutritional-determinants-international-study/1CFF189D6C50D799B17F9706F60F003B>

Discussion questions:

1. What are some limitations of using conventional methods such as factor and cluster analyses?
2. What is data mining?
3. What do previous studies show when comparing the use of data mining over statistical methods in nutritional epidemiology?
4. How was a decision tree used in elucidating relationships between risk factors (diet, age, sex, smoking status) and bladder cancer?
5. In this study, do you think data mining accounts for any confounding factors that may exist?
6. What proof does the authors use to demonstrate that data mining is effective in studying diets in relation to cancer risk factors?
7. What other risk factors could you include in the study? Or what changes would you make to the study?

Nov 14, 2022

## **Prospects and Pitfalls of Machine Learning in Nutritional Epidemiology**

<https://www.mdpi.com/2072-6643/14/9/1705/htm>

Discussion questions:

1. What is nutritional epidemiology and how are the findings from nutritional epidemiology used?
2. What are the three main research areas in nutritional epidemiology?
3. What are the challenges in the process of causal inference? Briefly describe each challenge.
4. What is ANN? DNN? How are they different? (Uses, examples etc.)
5. How is ML used to collect data for nutritional epidemiology? Which example stood out to you the most? How does the collection of data impact the data quality?
6. What is dimensionality in ML and how is ML used to reduce dimensionality? (Hint: the authors mentioned two techniques)
7. Why is it important to control confounding and multicollinearity in nutritional epidemiologic studies? How is ML used to deal with confounding?
8. What are the practical recommendations that the authors provided? Briefly explain each recommendation.
9. After reading this paper, what are your thoughts on the use of ML in nutritional epidemiology? What benefits/limitations stood out to you the most? Do you think the use of ML is critical in this field? Do you have any personal experience in nutritional epidemiology research which you could have/hoped to apply ML?

## **Role of artificial intelligence-internet of things (AI-IoT) based emerging technologies in the public health response to infectious disease in Bangladesh**

<https://www.sciencedirect.com/science/article/pii/S2405673122000307>

Discussion questions:

1. What is IoT? What role can IoT and AI technologies play in endemic countries?
2. How were IoT and AI used for the COVID-19 and Aedes mosquito outbreak? Provide examples.
3. What are your thoughts on the digital healthcare systems (telehealth, mobile health etc.)? What about smart disease surveillance system?
4. Do you think the applications of IoT and AI mentioned in this paper are realistic (especially in a setting like Bangladesh)? Any concerns?
5. What are your thoughts on data security/privacy? To what extent do you think it should be protected or do you think it should not be a concern if it is used for improving public health?

Nov 21, 2022

**Identifying and modeling the distribution of cryptic reservoirs of Ebola virus using artificial intelligence**

Kollars et al. 2018

<https://www.ajol.info/index.php/ajcem/article/view/172666>

Discussion questions:

1. What are the main goals of the study?
2. What are the possible pathways that EBOV uses to jump between species?
3. What is BioTEMS, how does it work, and what has it previously been used for?
4. What are the major findings and benefits of this study?
5. What are the limitations of this method and study?

**Artificial intelligence and avian influenza: Using machine learning to enhance active surveillance for avian influenza viruses**

Walsh et al 2019

[https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.13318?casa\\_token=NMF9Y8IdjnYAAAAA:m1DFumFQbxYrra5oM9CwPzdHHawmAD9h8POYbiraTBb-ao6gfaHX88CWKwx7dvC\\_VtrKhyifF8oV3A](https://onlinelibrary.wiley.com/doi/abs/10.1111/tbed.13318?casa_token=NMF9Y8IdjnYAAAAA:m1DFumFQbxYrra5oM9CwPzdHHawmAD9h8POYbiraTBb-ao6gfaHX88CWKwx7dvC_VtrKhyifF8oV3A)

Discussion questions:

1. What are AIVs and why are they important?
2. What makes AIVs so difficult to monitor?
3. How was machine learning used in this study? What was the overarching goal, and what method did they use? What information did they use to train the algorithm?
4. What variables were the most important predictors in isolating AIV from samples? What patterns emerged across these predictors?
5. How are these findings useful? What are some of the limitations of how they applied this method?

Nov 28, 2022

### **Artificial Intelligence in Dentistry: Past, Present, and Future**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9418762/pdf/cureus-0014-00000027405.pdf>

Discussion questions:

1. What are the key elements of an AI system?
2. Explain the mechanism of working of an AI model.
3. What are the components of the AI hierarchy system and how does it relate to dentistry?
4. What are the applications of AI in endodontics?
5. What are the other applications of AI in dentistry and how is it used?
6. What are some of the fascinating uses of AI in dentistry?
7. What is bioprinting?
8. Can AI completely replace dentists/clinicians? Why or why not?

### **The primary use of artificial intelligence in cardiovascular diseases: what kind of potential role does artificial intelligence play in future medicine?**

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6748906/pdf/jgc-16-08-585.pdf>

Discussion questions:

1. What are the 3 classical genres of Machine Learning?
2. What are the AI applications in the medical field?
3. What are the 4 major applications of AI in CVD?
4. What are some examples of applications of AI in CVD?
5. What are the current shortcomings of AI in medicine?
6. What is a traditional and a novel medical model?
7. What is a clinical warning system?

Dec 5, 2022

### **AI Surveillance during Pandemics: Ethical Implementation Imperatives**

<https://onlinelibrary.wiley.com/doi/10.1002/hast.1125>

Discussion questions:

1. What are some examples of how AI has been used in countries around the world to combat the COVID-19 pandemic?
2. What are some examples of statutes or regulations that address privacy concerns to protect health data for individuals, and what parts of the world are covered by these statutes or regulations?
3. What are some examples of data generated by COVID-19 AI surveillance programs that are not covered under the U.S. Health Insurance Portability act of 1996 (HIPAA)?
4. What was the major outcome of the *Jew Ho v. Williamson* federal court case?
5. The authors state that to “avoid discrimination in the use of AI, stakeholders must be transparent about which communities and individuals are being monitored.” In your opinion, do you think that using AI surveillance to selectively monitor specific communities can lead to discrimination/stigmatization of those communities? Why or why not?
6. Can you think of any applications/websites that cannot be used or installed without allowing that application/website to be granted access to your data? In such a case would you say the act of granting access to your user data is voluntary or compulsory? Explain your choice.

### ***Salmonella enterica* serovar Typhimurium isolates from wild birds in the United States represent distinct lineages defined by bird type**

<https://journals.asm.org/doi/full/10.1128/aem.01979-21>

Discussion questions:

1. Why is it important to study *Salmonella enterica*, a pathogen usually transmitted to humans through eating contaminated food, in non-food sources like wild birds?
2. What are some examples of possible interspecies transmission of *S. enterica* between wild birds and humans?
3. How did the authors obtain the 131 *S. Typhimurium* isolates from wild birds that they subjected to whole genome sequencing in this study? Was there bias in this selection process?
4. What is a “table of confusion”, and how was it used to assess the ability of the machine learning Random Forest Classifier to correctly identify the source of *S. Typhimurium* isolates?
5. How did the addition of the 131 *S. Typhimurium* isolates from wild birds into the Random Forest Classifier change the accuracy of predicting the source of *S. Typhimurium* isolates?
6. How can enhanced prediction of the source of a human pathogen (e.g., *Salmonella*) through machine learning techniques ultimately aid in improving public health outcomes?

## Additional Reading

Regulation of predictive analytics in medicine: Algorithms must meet regulatory standards of clinical benefit	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6557272/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6557272/</a>
Artificial Intelligence in Public Health and Epidemiology	<a href="http://dx.doi.org/10.1055/s-0038-1667082">http://dx.doi.org/10.1055/s-0038-1667082</a>
COVID-19 pandemic and artificial intelligence: challenges of ethical bias and trustworthy reliable reproducibility?	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8495685/pdf/bmjhci-2021-100438.pdf">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8495685/pdf/bmjhci-2021-100438.pdf</a>
The epidemic volatility index, a novel early warning tool for identifying new waves in an epidemic	<a href="https://www.nature.com/articles/s41598-021-02622-3">https://www.nature.com/articles/s41598-021-02622-3</a>
Application of Neural Network Model in an Epidemiological Study	<a href="https://www.sciencepublishinggroup.com/journal/paperinfo.aspx?journalid=148&amp;doi=10.11648/j.ajam.20150304.16">https://www.sciencepublishinggroup.com/journal/paperinfo.aspx?journalid=148&amp;doi=10.11648/j.ajam.20150304.16</a>
Artificial Intelligence and Big Data in Public Health	<a href="https://www.mdpi.com/1660-4601/15/12/2796/htm">https://www.mdpi.com/1660-4601/15/12/2796/htm</a>
Strengthening public health surveillance through blockchain technology	<a href="https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779606/">https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779606/</a>