Surveillance Advances
Progrès dans le domaine de la surveillance

Establishing a Surveillance, Early Warning, and Forecasting System for Avian Influenza Outbreaks in Canada

Mise en place d'un système de surveillance, de détection rapide et de prévision des épidémies d'influenza aviaire au Canada

May 27 2025

12:00 – 1:00pm CT / 1:00 – 2:00pm ET

Speaker

Dr. Zahra Movahedi Nia Research Associate, PhD in Computer Engineering, York University

National Collaborating Centre for Infectious Diseases
Centre de collaboration nationale des maladies infectieuses

Public Health Agency of Canada
Agence de la santé publique du Canada

Land Acknowledgment:

Today's moderator is located in the City of Hamilton, Ontario.

Situated upon the traditional territories of the Erie, Neutral, Huron-Wendat, Haudenosaunee and Mississauga Peoples, this land is covered by the Dish With One Spoon Wampum Belt Covenant, an agreement between the Haudenosaunee and Anishinaabek to share and care for the resources around the Great Lakes.

Today, Hamilton is home to many Indigenous Peoples from across the Turtle Island. I honour the Indigenous Peoples who have lived on and cared for these lands for generations. I am grateful for the opportunity to share and call this place home.

Housekeeping

- Seminar recording and presentation slides will be available shortly after the seminar at the NCCID website: https://nccid.ca/
- If you have technical problems with Zoom, please email us at nccid@umanitoba.ca
- The chat box for participants has been disabled for this session. We will use the chat box to share additional information.
- Please use the Q&A tab to submit your questions for our speakers. You can "like" other people's questions to push them up in priority

Image by macrovector on Freepik

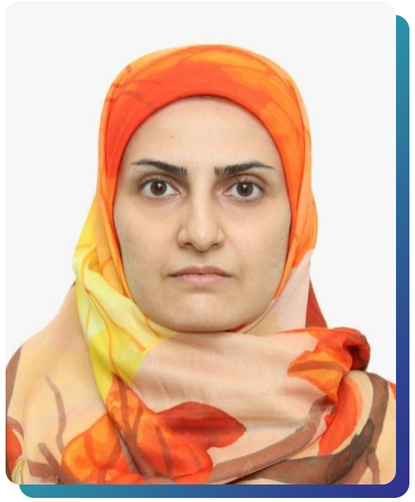
Accreditation

Surveillance Advances is a self-approved group learning activity (Section 1) as defined by the Maintenance of Certification Program of the **Royal College of Physicians and Surgeons of Canada**.

The seminar series is also approved by the Council of Professional Experience for professional development hours for members of the **Canadian Institute of Public Health Inspectors**.

If you would like a letter of participation, please complete the survey which will be shared after the seminar.

Today’s speaker



Dr. Zahra Movahedi Nia

Research Associate,
PhD in Computer Engineering,
York University

Establishing a Surveillance, Early Warning, and Forecasting System for Avian Influenza Outbreaks in Canada

Zahra Movahedi Nia
Research associate, AI4PEP

AIMLab | UNIVERSITY OF TORONTO

GLOBAL SOUTH
AI4PEP NETWORK
SUD GLOBAL
RÉSEAU AI4PEP

AFRICA - CANADA
ARTIFICIAL INTELLIGENCE AND DATA INNOVATION CONSORTIUM

UNIVERSITY OF TORONTO
Dalla Lana School of Public Health

Prof. Jude Kong

IDRC · CRDI

International Development Research Centre
Centre de recherches pour le développement international

NSERC
CRSNG

Canada

Sida

SWEDISH INTERNATIONAL
DEVELOPMENT COOPERATION AGENCY

Foreign, Commonwealth
& Development Office

New Frontiers in Research Fund
Fonds **Nouvelles frontières** en recherche

Dr. Doris Leung
Canadian Animal Health Surveillance System (CAHSS)

Dr. Murray Gillies
Canadian Animal Health Surveillance System (CAHSS)

Dr. Emma Gardner
Canadian Animal Health Surveillance System (CAHSS)

Prof. Nicola L. Bragazzi
Department of Foods and Drugs, University of Parma

Dr. Itlala Gizo
World Organization of Animal Health (WOAH)

Prof. Nelson Lee
Dalla Lana School of Public Health, University of Toronto

Dr. Emmanuel Musa
Faculty of Public Health
University of Bath

Oscar Pang
University of Toronto

Xuerong (Snow) Zhou
University of Toronto

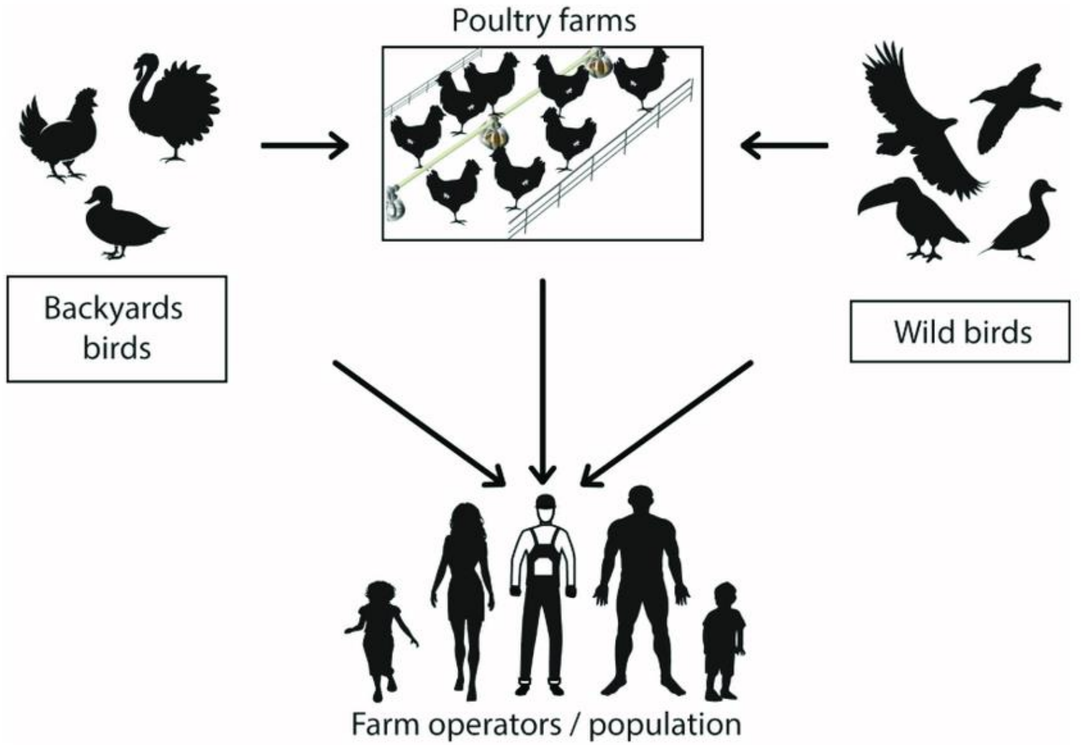
Shixun Huang
University of Toronto

Conflicts of interest

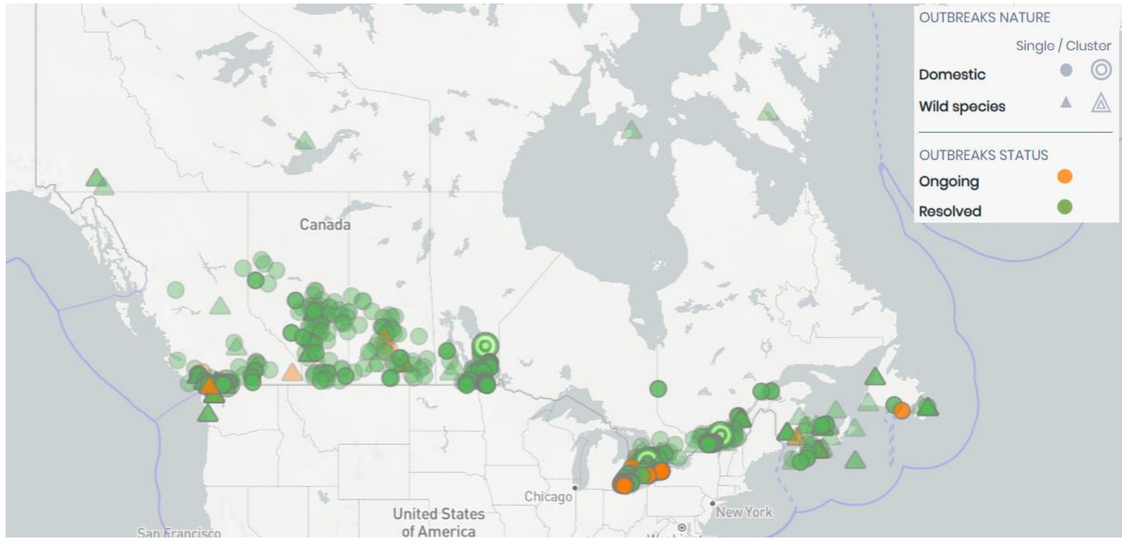
We claim no conflict of interest

Introduction

- Avian Influenza is classified into 2 categories based o its pathogenicity:
  - Low Pathogenic Avian Influenza (LPAI)
  - Highly Pathogenic Avian Influenza (HPAI)
- Highly Pathogenic Avian influenza (HPAI) is primarily found in birds and waterfowls.
- It can also be transmitted to mammals including humans.
- It has a fatality rate of about 52%.
- Symptoms include cough, sore throat, fever, and diarrhoea

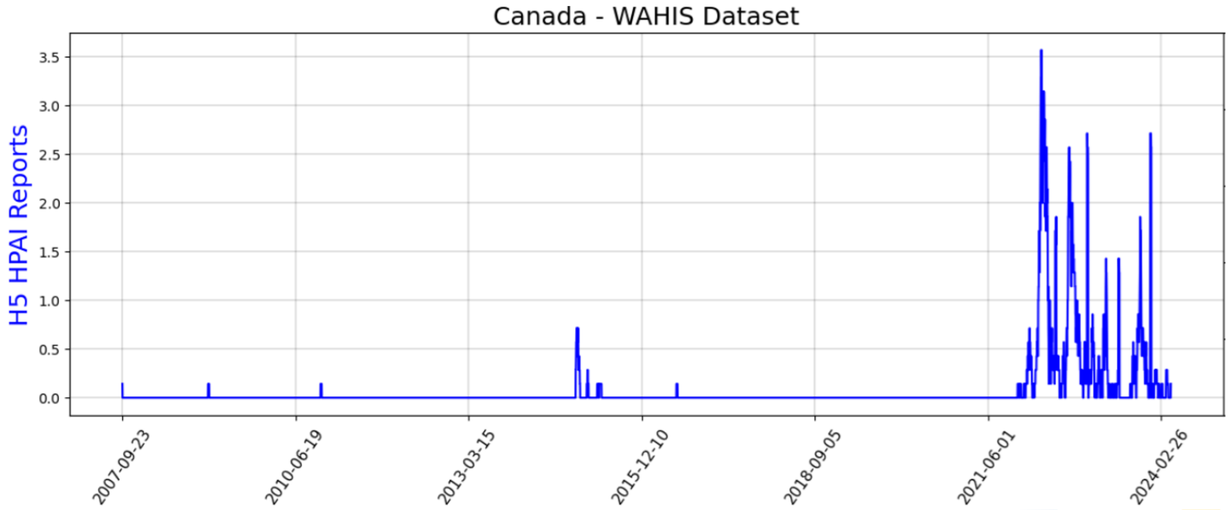


Introduction



- Since November 2021, Canada and the USA have experienced an unprecedented consecutive series of HPAI outbreaks.

- On November 9, 2024 a human case of H5 avian influenza was confirmed in British Columbia.



Source: World Organization for Animal Health (WOAH) and World Animal Health Information System (WAHIS)

Introduction

- There is a need for strong surveillance methods:
  - HPAI is transmissible to mammals including humans
  - Although very rare, but it is also transmissible between humans.
  - Therefore, there is a risk of a pandemic.
  - HPAI has the potential to cause billions of dollars loss for poultry farmers, producers, and public sectors.

- Moreover, Several factors in current outbreaks urges for stronger surveillance methods:
  - Unprecedented and frequent outbreaks
  - Cross-species transmission
  - Increased involvement of waterfowls and mammals
  - Wider geographic range of infection
  - Market impacts, especially in chicken and egg prices

Introduction

- Syndromic surveillance helps us with preparedness, rapid response, and recovery.

- Avian influenza surveillance methods include:
  - Collecting samples from sick or dead birds.
  - Collecting samples from domestic and para-domestic environments e.g. soil, water, air, cages, feeding area.
  - Clinical data.

- Traditional surveillance methods are expensive and time consuming.
- It may take several weeks for the result to be prepared and reported.

Introduction

- Web-based data and data collected from RESTful APIs are beneficial to conventional health related data:
  - It is easier and faster to access them.
  - They are more cost-efficient.
  - They could be accessed automatically.
  - They provide data days to weeks earlier.

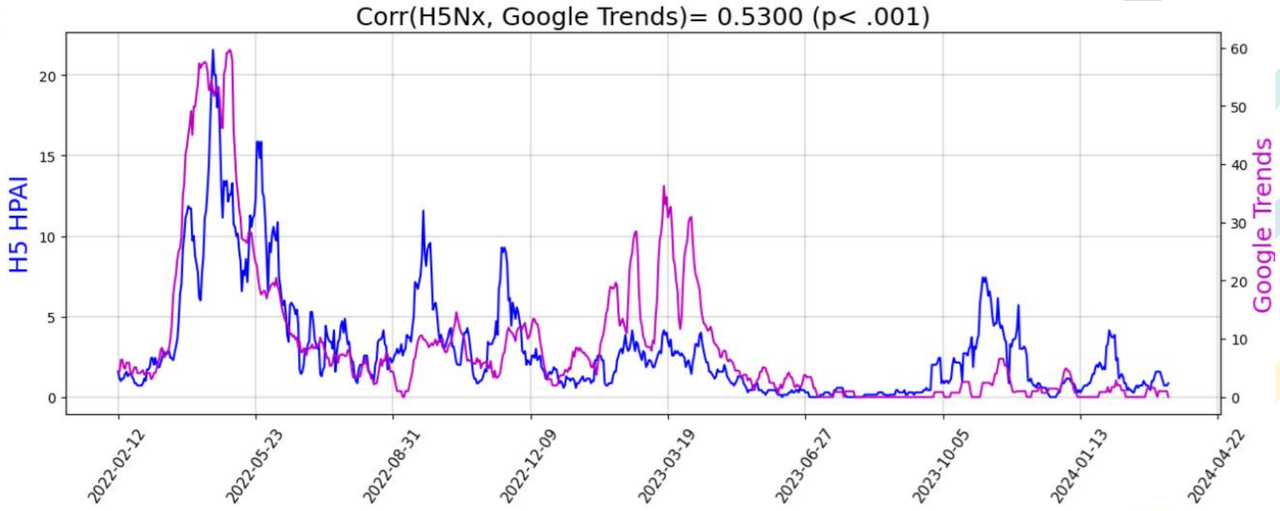
- Such unconventional data include:
  - Number of searches in search engines such as Google, (Google Trends).
  - Posts on social media platforms such as Reddit and Facebook.
  - Number of news articles released.

Introduction

- Early Warning System (EWS) of Infectious diseases:
  - Planning, designing, and implementing methods for prediction or early detection of emerging or re-emerging outbreaks

- Predictive Modeling:
  - Processing and analyzing previous events to forecast future.

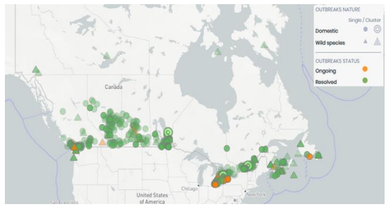
- Time series prediction is a predictive modeling technique in which incidences are viewed as time series.
  - Future could be forecasted using previous patterns.
  - Future could be forecasted using related indicators.



Learning Objectives

- Our contribution is three-fold:
  - Propose additional data sources for surveillance of HPAI.
  - Build an early warning system to forecast outbreaks on both country- and regional-level.
  - Study the importance of different data sources.

- Our research is beneficial to:
  - Preparing for, controlling and preventing HPAI outbreaks
  - Minimizing economic implications

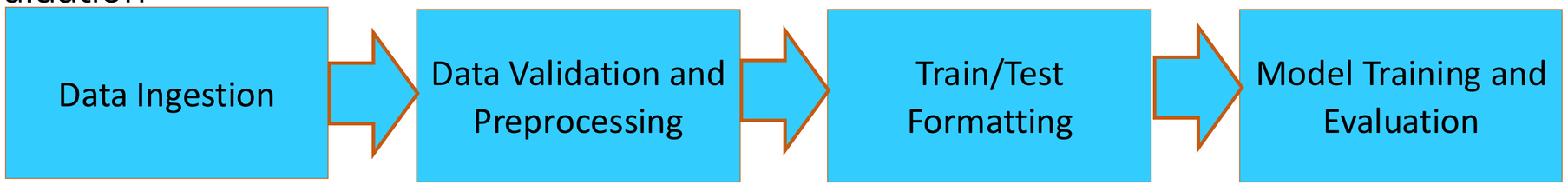


Inform Policy

Image by vectorjuice on Freepik

Method: The Machine Learning Pipeline

- The machine learning pipeline includes 4 stages:
  - Data ingestion
  - Data validation and preprocessing
  - Train/Test Formatting
  - Model Training and Evaluation



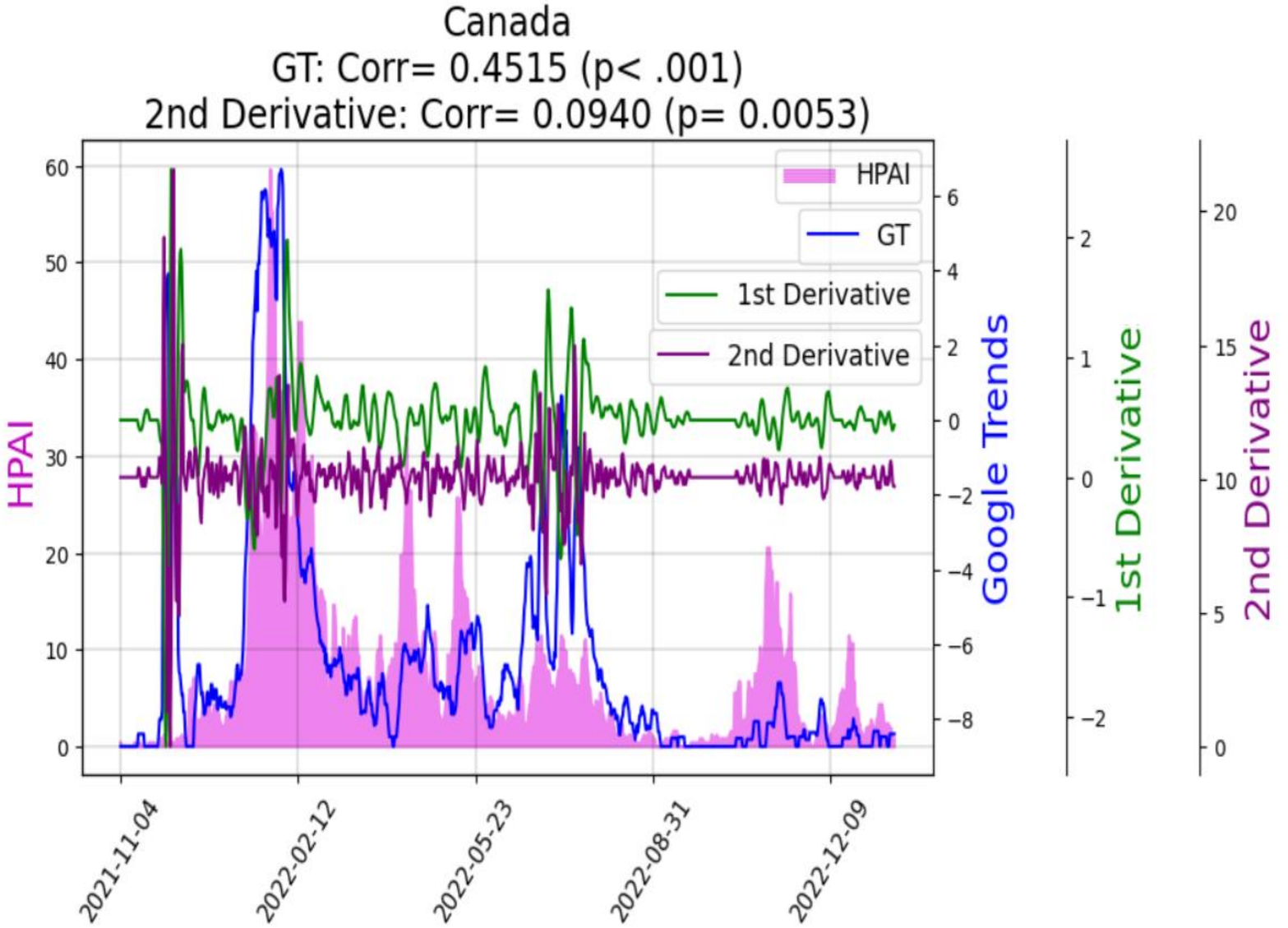
- The Data Ingestion stage combines 7 data sources:
  - Historical data
  - Google Trends
  - GDELT
  - Google News
  - Reddit
  - Facebook
  - Weather and air Quality
- All datasets are collected through RESTful APIs or webpage scraping, without human intervention, making the pipeline fully automatic.

Methods: Dataset Ingestion

- Historical data: number of HPAI cases in wild birds is gathered and provided by Canadian Food Inspection Agency (CFIA). (https://www.arcgis.com/apps/dashboards/89c779e98cdf492c899df23e1c38fdbc)
- The dataset is on daily basis
- The dataset provides number of outbreaks
- The dataset provides collection date.

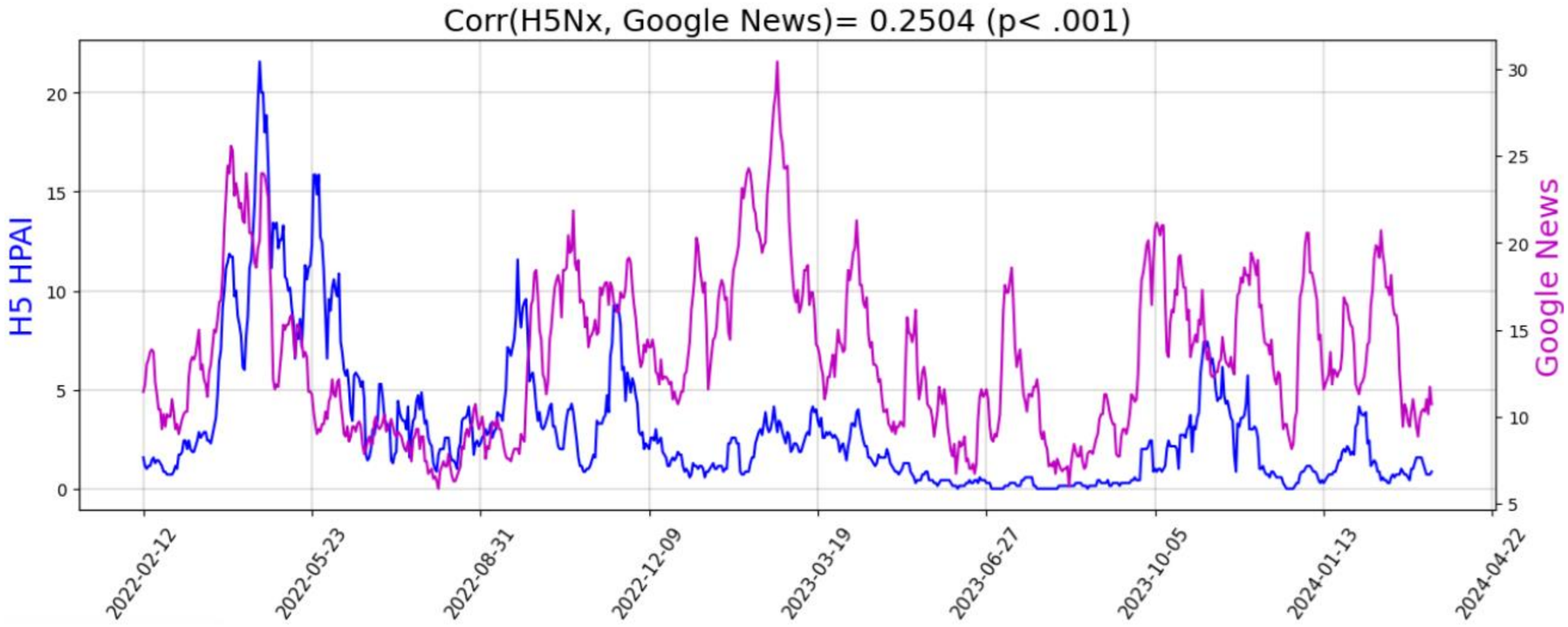
Methods: Dataset Ingestion

- Google Trends: number of times a particular term, or topic has been searched in Google.
- First derivative of Google Trends signifies if the Google trends is increasing or decreasing.
- Second derivative of Google Trends shows how fast Google trends is increasing or decreasing.



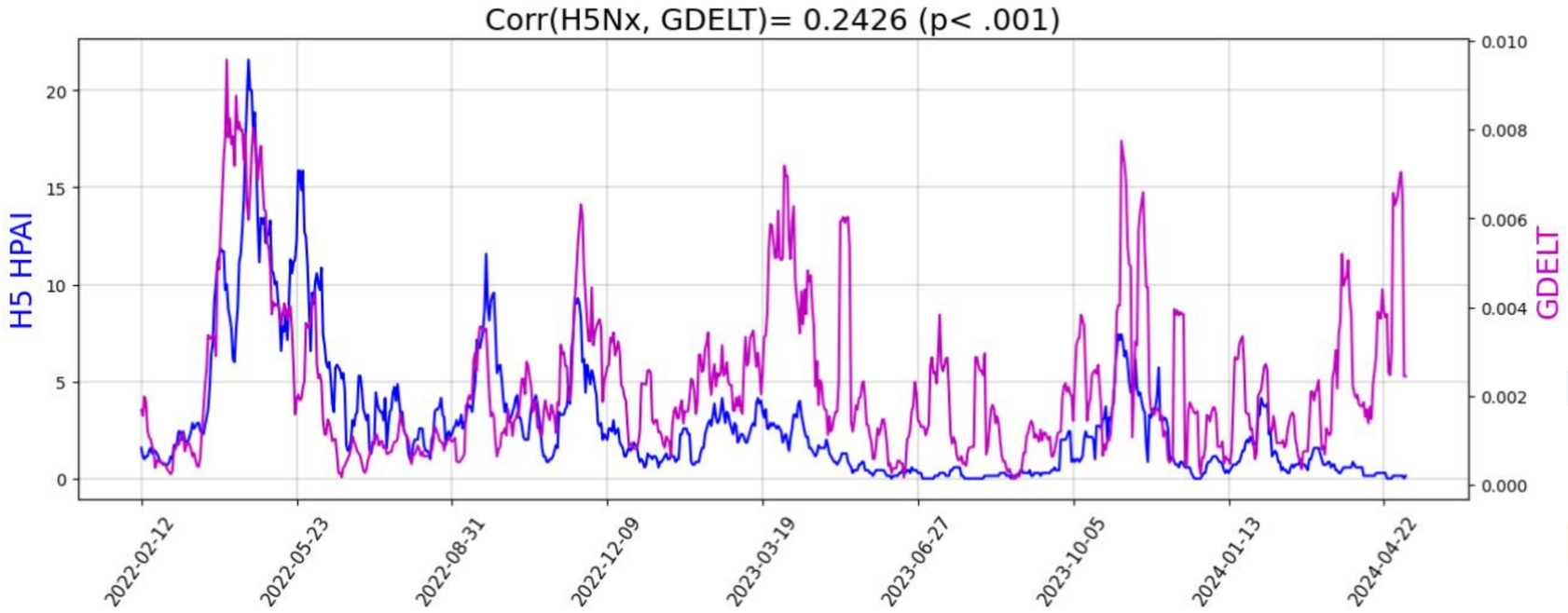
Methods: Dataset Ingestion

- Google News API returns the number of news articles released on a certain topic in a particular country.



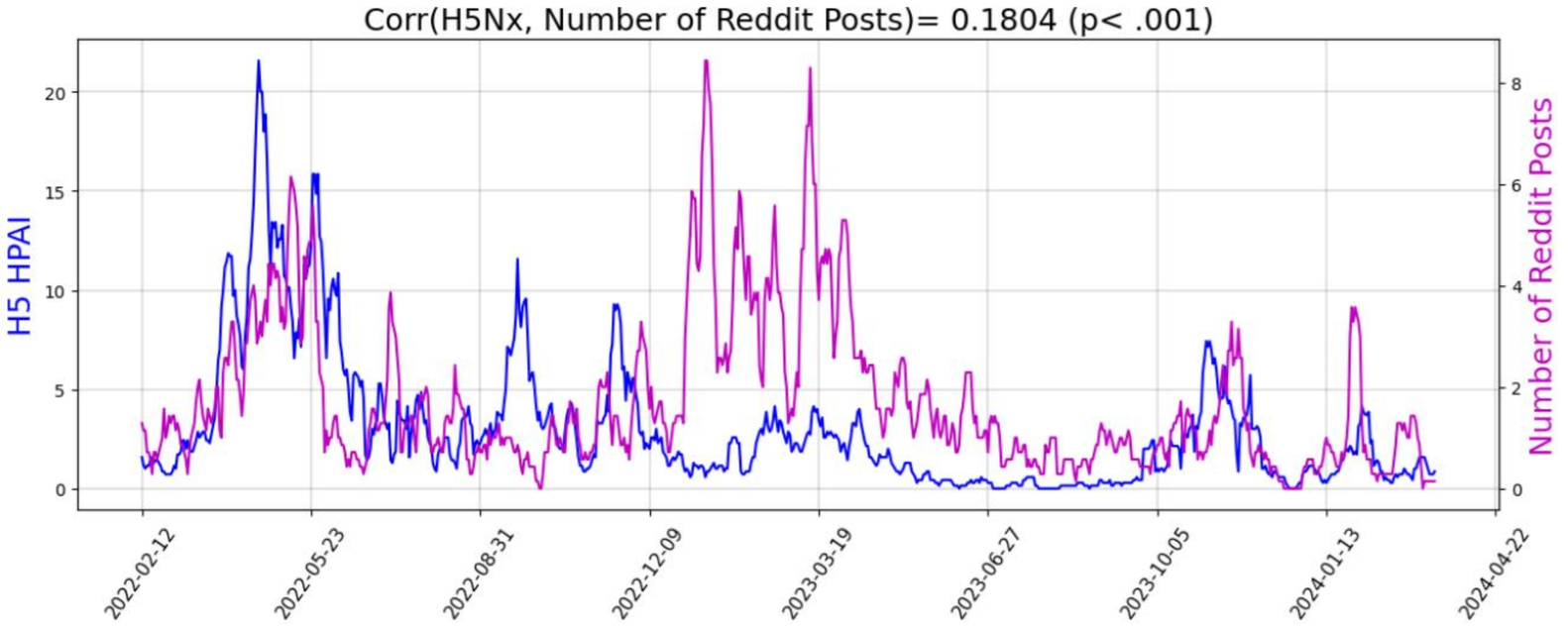
Methods: Dataset Ingestion

- GDELT provides the number of online news articles released about a certain topic in a particular country over time.
- Broadcasts, print, and web news
- Online news media
- Television and radio
- Academic and NGO reports



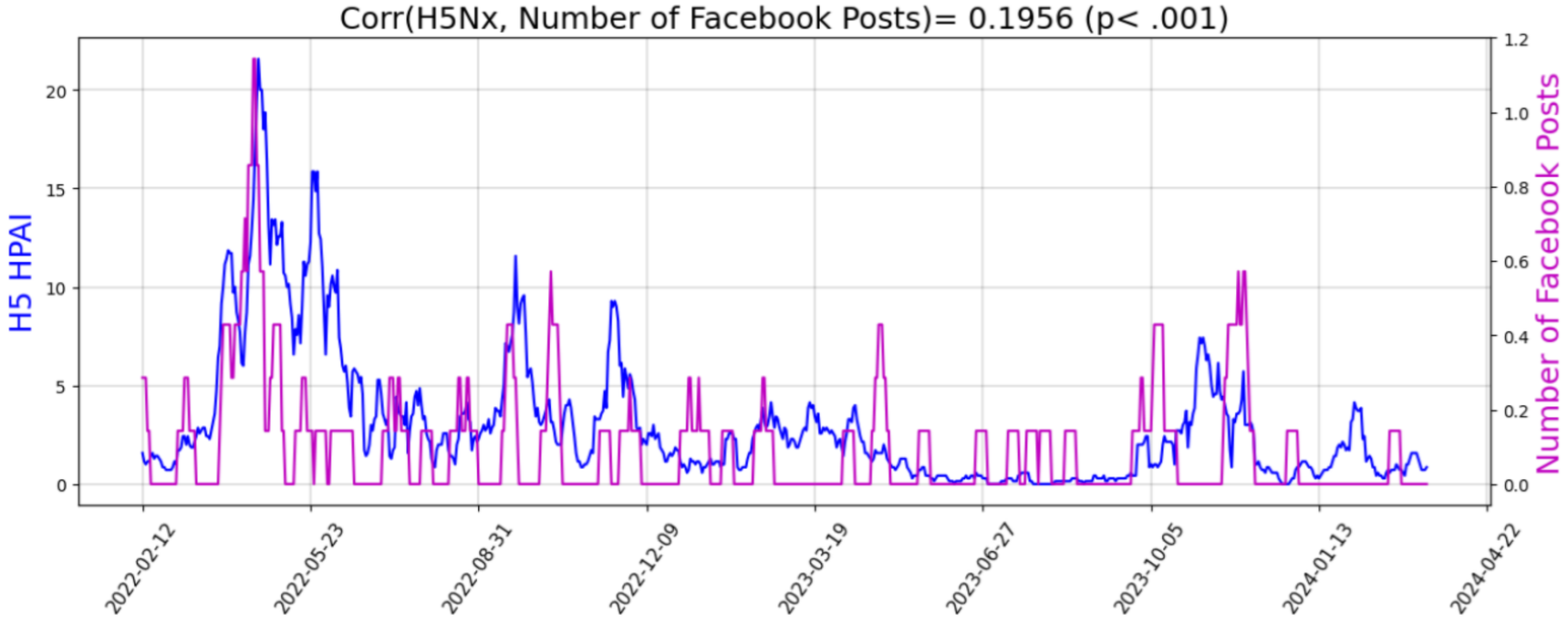
Methods: Dataset Ingestion

- Number of Reddit posts have a significant correlation with the number of cases.
  - Keywords include: bird, avian, poultry, farm, waterfowl, h5, hpai, aviaire, oiseau, sauvagine, sauvagine
  - Subreddits include: canada, canadanews, ontarion, toronto, ottawa, quebec, montreal, alberta, calgary, edmonton, britishcolumbia, vancouver, saskatchewan, regina, manitoba, winnipeg, nunavut, novascotia, halifax, newbrunswick, newfoundland_labrador, princeedwardisland, charlottetown, yukon, northwestterritories, capebreton,



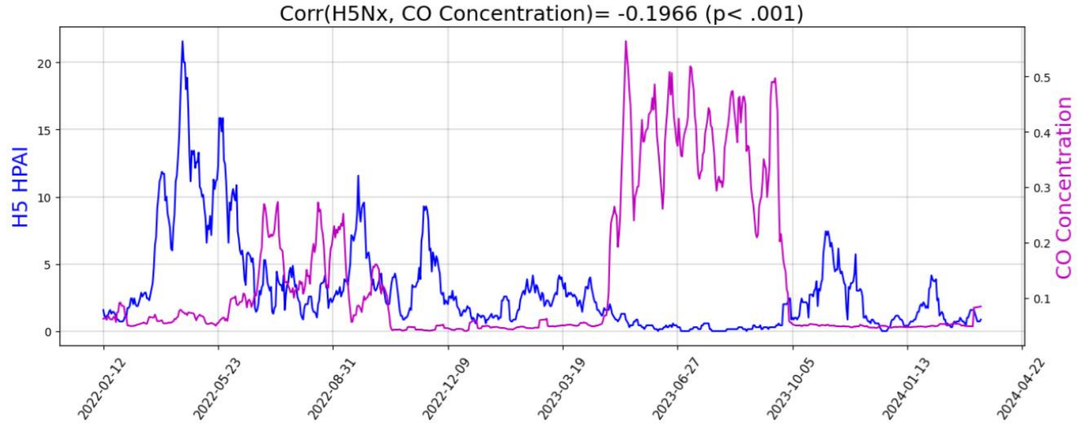
Methods: Dataset Ingestion

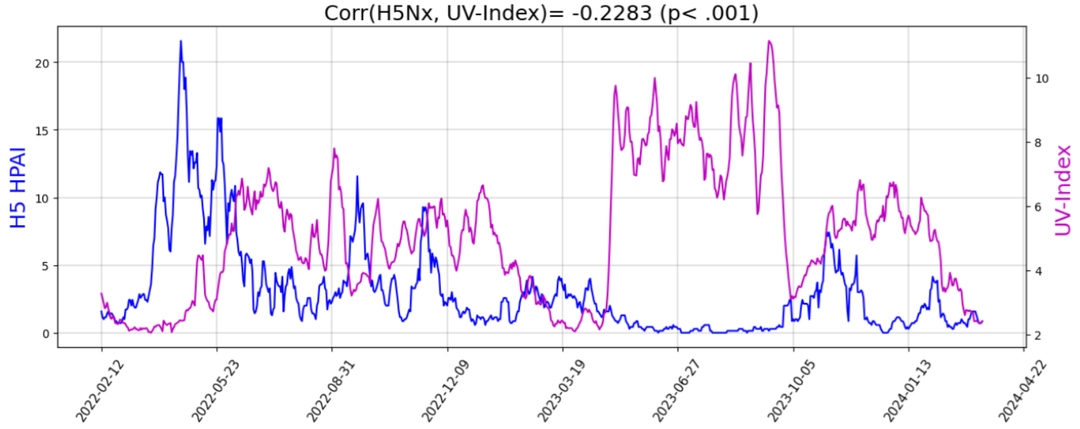
- Number Facebook posts were gathered from Facebook pages of poultry and bird farms in different provinces of Canada.
- Keywords include: bird, avian, poultry, farm, waterfowl, h5, hpai, aviaire, oiseau, sauvagine, sauvagine



Methods: Dataset Ingestion

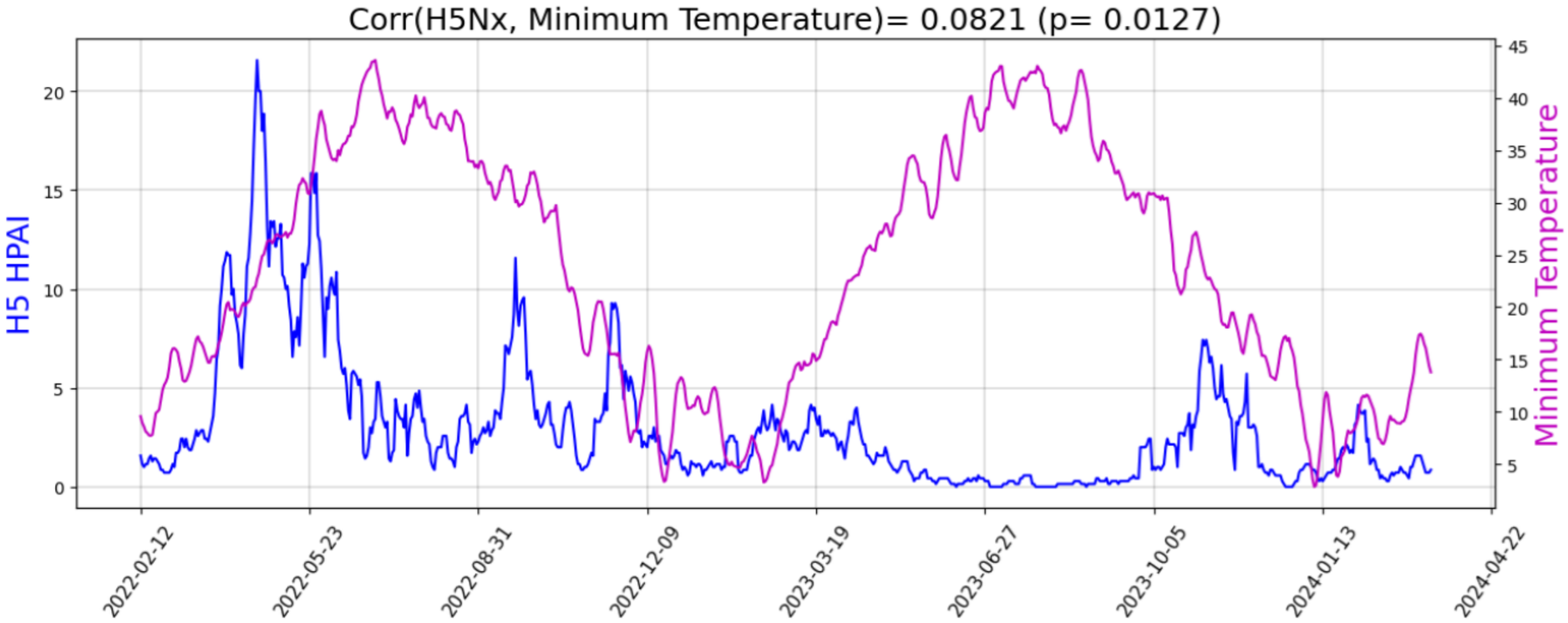
- Air quality data was gathered Sentinel-5P mission, Google Earth Engine.
- Two parameters had the best correlation: CO and UV-index.
- Negative correlation indicates that migratory birds appear when air pollution is lower.





Methods: Dataset Ingestion

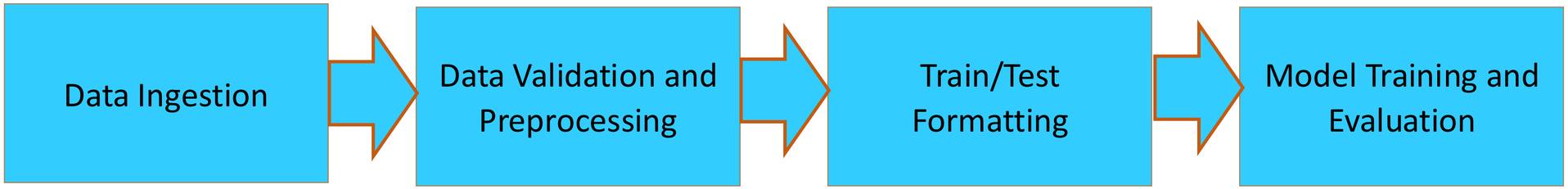
- Temperature is a good indicator for avian influenza surveillance.
- Positive correlation indicates that migratory birds appear when correlation is higher



Methods: Data Validation and Preprocessing

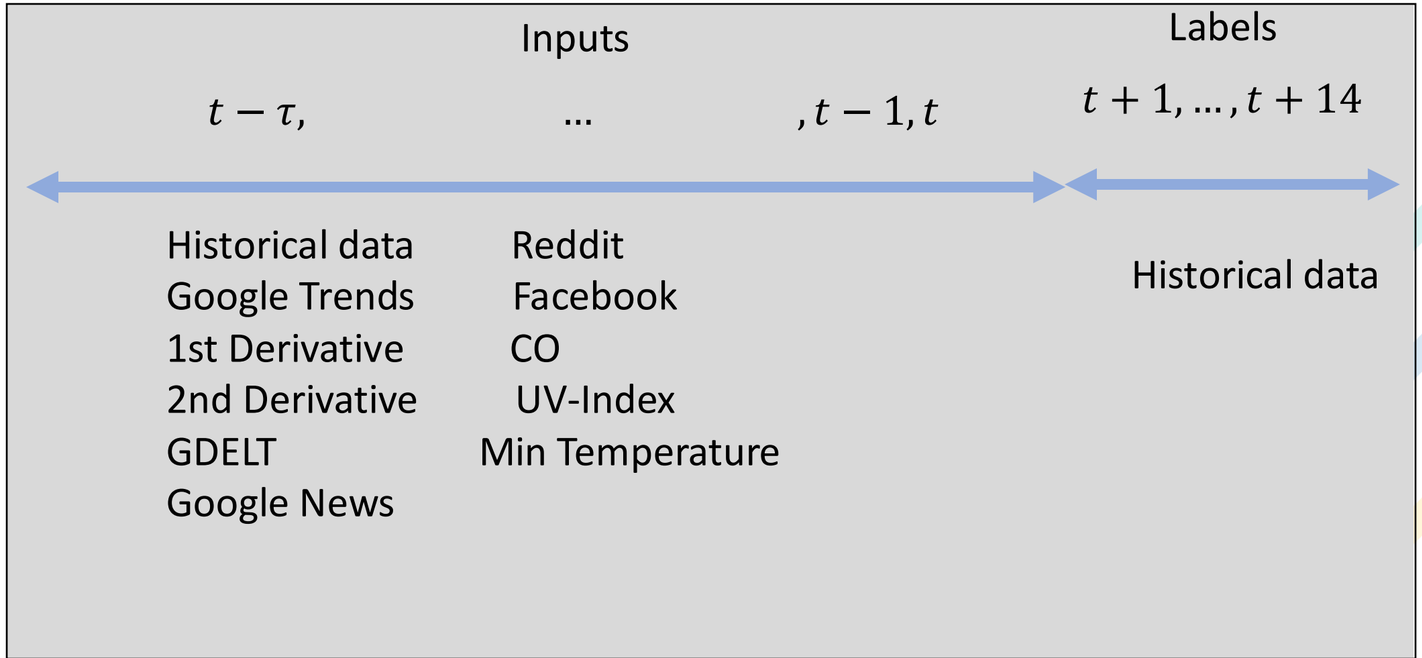
- The datasets are organized to be used for building 6 different machine learning models:
  - Gated Recurrent Unit (GRU)
  - Long-Short Term Memory (LSTM)
  - CNN-GRU
  - CNN-LSTM
  - Random Forest (RF)
  - Support Vector Machine (SVM)
  - Naïve Bayes (NB)

- All the datasets were treated as time series from November 4th 2021 to March 31st 2024.
- Missing values were filled with the average of preceding and succeeding values.
- Scaling was performed on the datasets using the min-max scaler method.
  - Random Forest does not require data scaling



Methods: Train/Test Formatting

- Observations need to be extracted from the datasets for Multivariate prediction.
- In each observation
  - A sequence of τ previous values of all the datasets is used for the input
  - The next 14 values of the historical data is used for the labels
- Principal Component Analysis (PCA) was used to unify the different variables for
  - Random forest
  - SVM
- Training set: 2/3 of the observations
- Testing set: remaining 1/3 of observations



Methods: Model Training and Evaluation

- Various models were employed to forecast the avian influenza outbreaks: GRU, LSTM, CNN-GRU, CNN-LSTM, RF, and SVM
- Forecasting was applied up to 14 days in advance.
  - For RF and SVM, 14 different models need to be trained for 14 days-ahead prediction → a different model for each day in advance
  - For deep-learning methods, only one model needs to be trained → by placing a linear NN layer on top with 14 outputs for projection
- Hyperparameters such as sequence size τ, dropout, learning rate, decay, and batch size were optimized using the WandB package.
- Two metrics were used to evaluate the performance:
  - Root Mean Square Error (RMSE) $$RMSE = \sqrt{\frac{1}{n}\sum_i (a_i - p_i)^2}$$
  - Correlation Coefficients $$Corr = \frac{n\sum_i a_i p_i - \sum_i a_i \sum_i p_i}{\sqrt{n\sum_i a_i^2 - \left(\sum_i a_i\right)^2}\sqrt{n\sum_i p_i^2 - \left(\sum_i p_i\right)^2}}$$
- Jackknife measure was used to estimate the importance of different data sources.

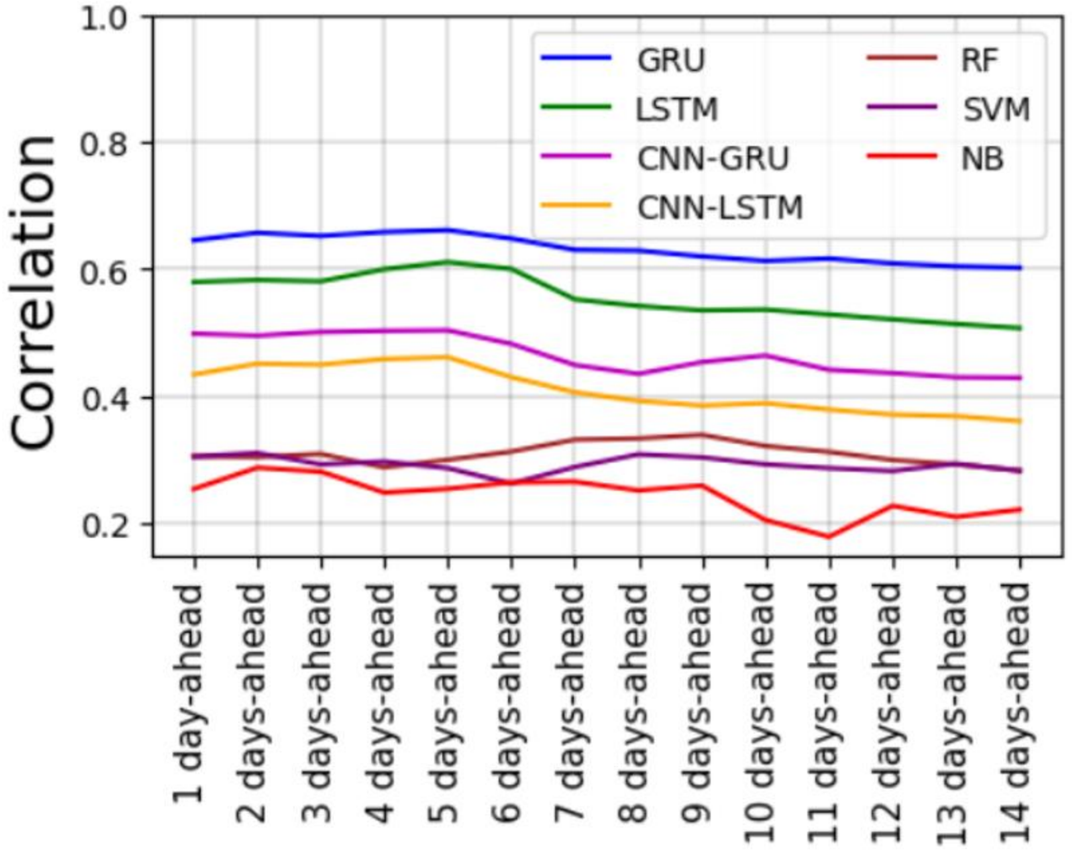
Results: Country-Level

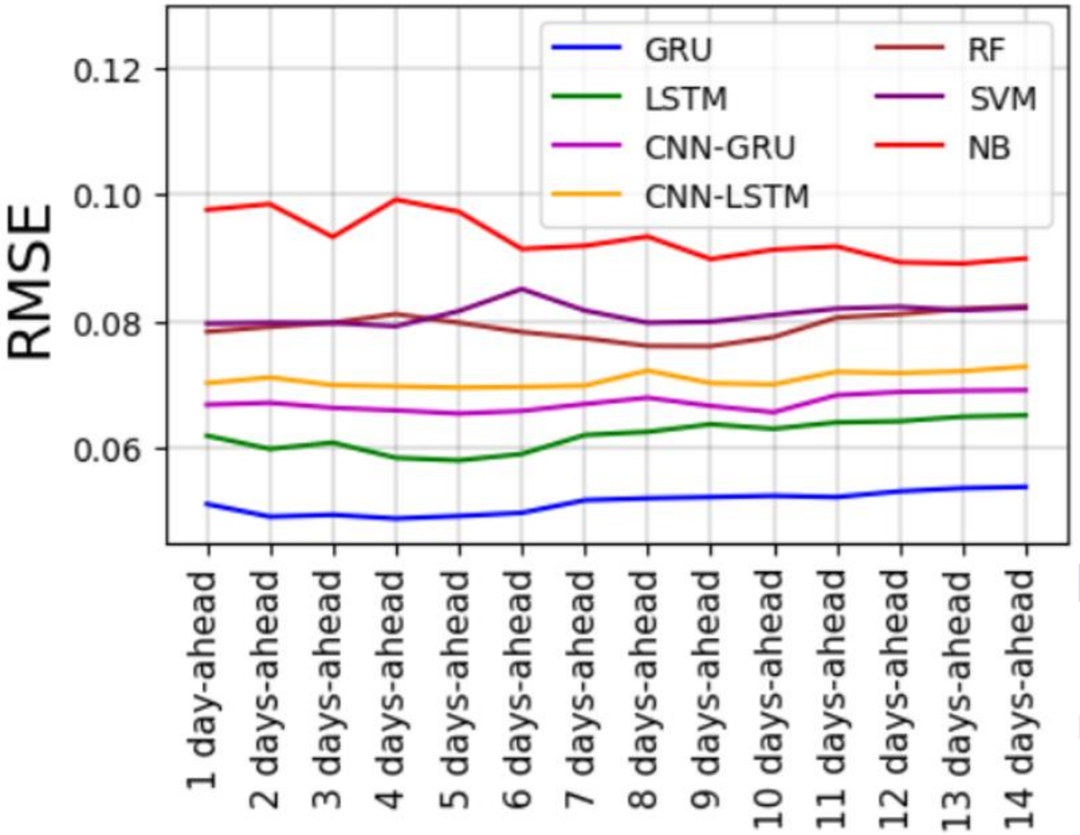
- GRU provided the best accuracy among all the different methods that was used.

	5 days-ahead prediction		14 days-ahead prediction	
	Correlation	RMSE	Correlation	RMSE
GRU	**0.661 (p<.001)**	**0.0508**	**0.602 (p<.001)**	**0.0538**
LSTM	0.611 (p<.001)	0.058	0.508 (p<.001)	0.0651
CNN-GRU	0.504 (p<.001)	0.0654	0.429 (p<.001)	0.0691
CNN-LSTM	0.462 (p<.001)	0.0695	0.361 (p<.001)	0.0728
RF	0.3 (p<.001)	0.0797	0.283 (p<.001)	0.0823
SVM	0.287 (p<.001)	0.0816	0.282 (p<.001)	0.0826
NB	0.255 (p<.001)	0.0973	0.222 (p<.001)	0.0899

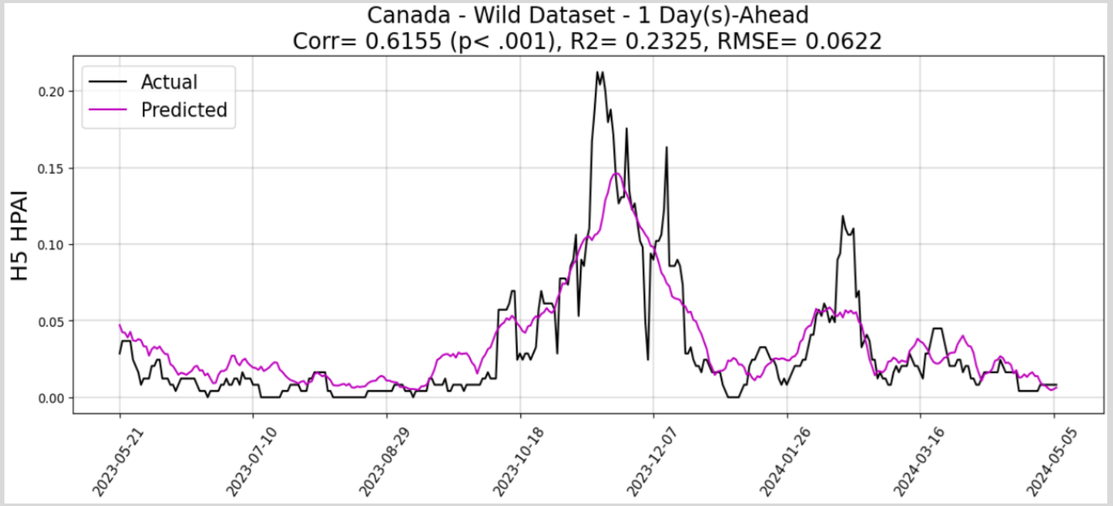
Results: Country-Level

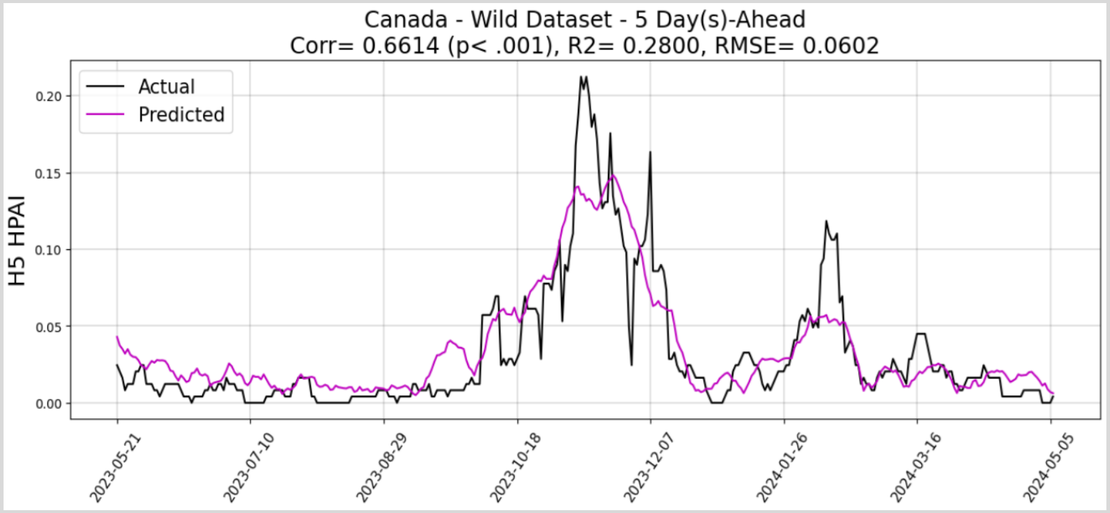
- GRU provided the best accuracy among all the different methods that was used.

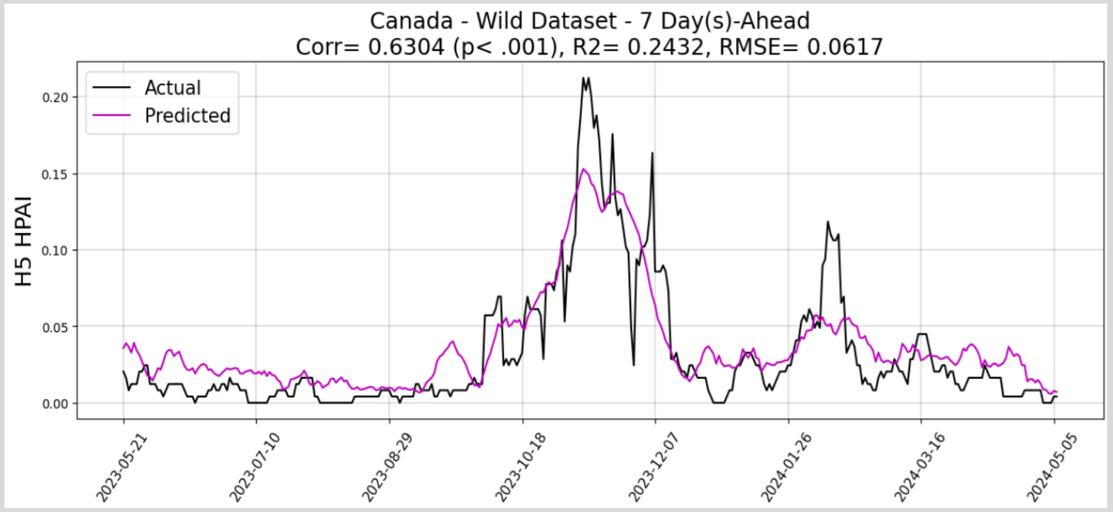


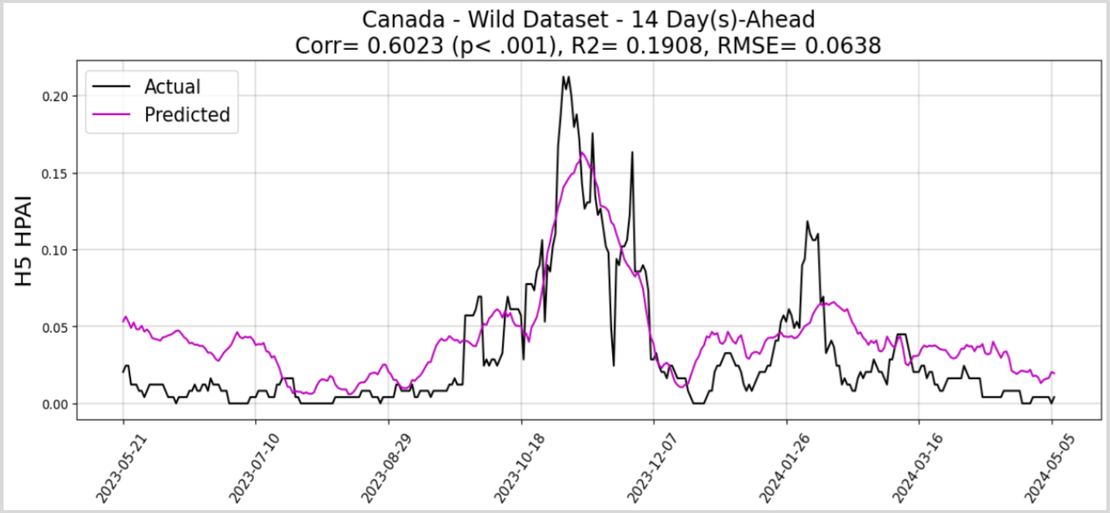


Results: Country-Level



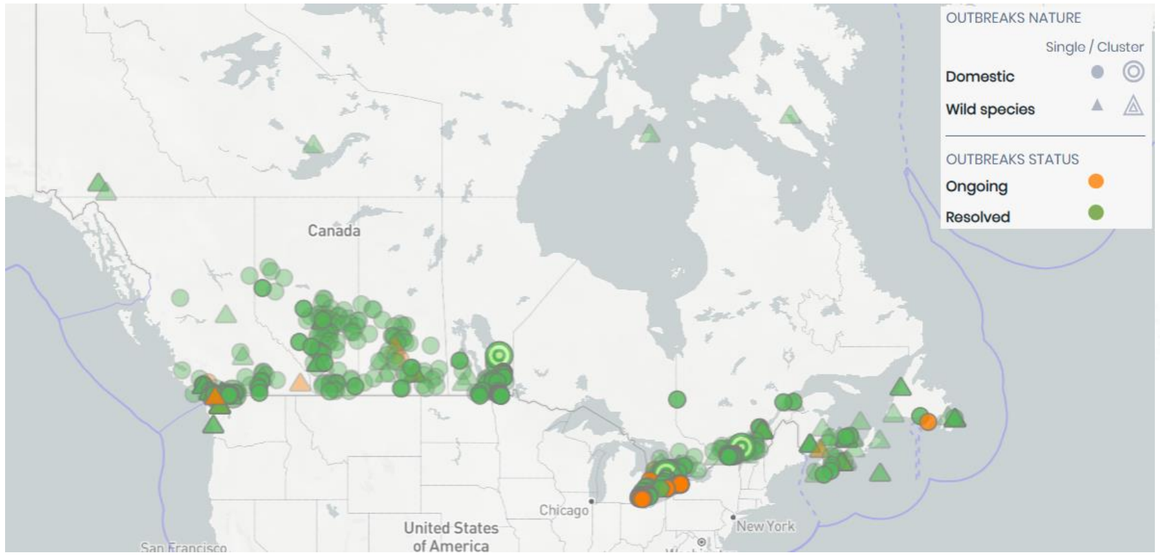






Regional-Level

- Territorial Region was removed due to lack of data and few cases of avian influenza.

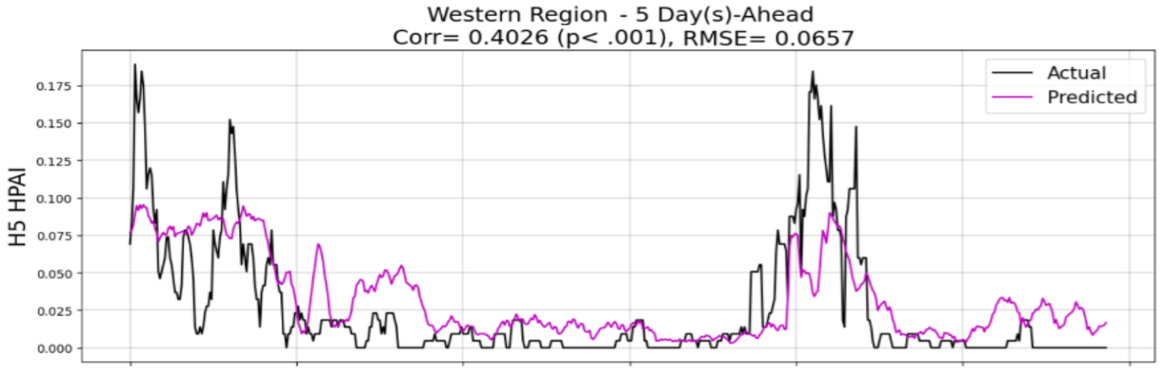


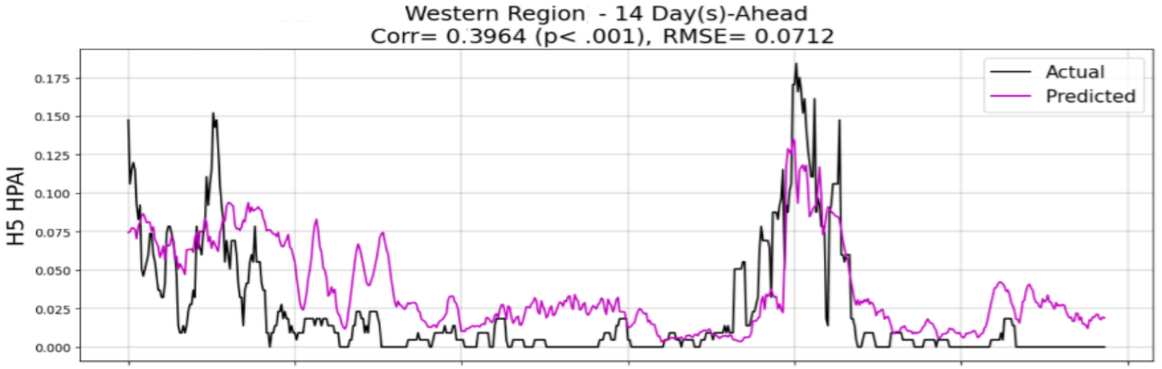
Regional-Level Dataset

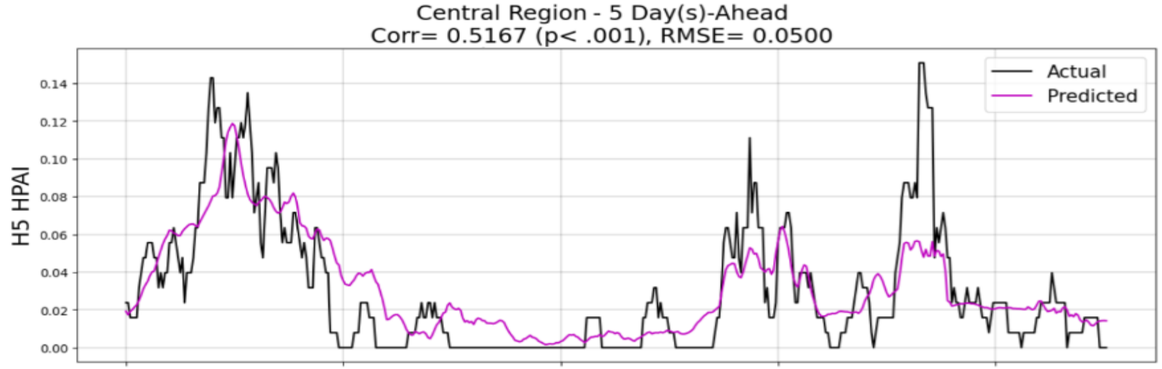
- Most of the dataset had a mild correlation with the number of avian influenza cases in different regions.

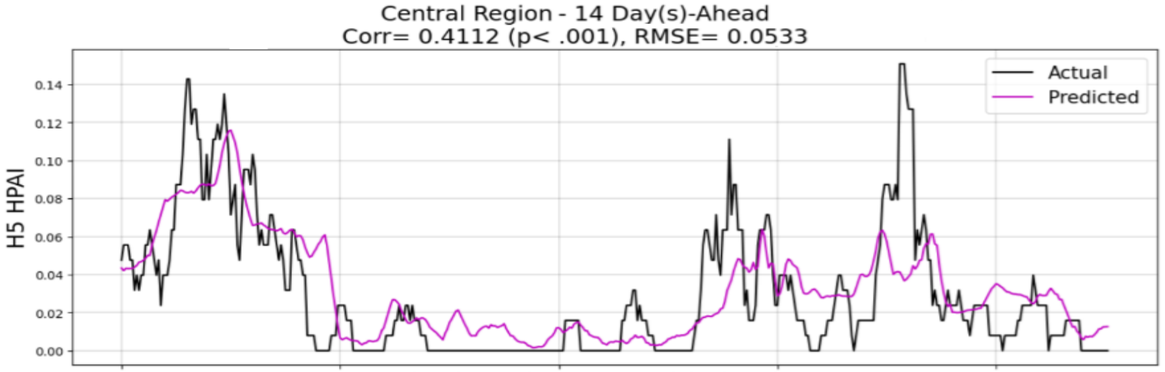
	Western Region	Central Region	Atlantic Region
Google Trends	0.2831 (p< .001)	0.1946 (p< .001)	—
1st Derivative	—	—	—
2nd Derivative	—	0.0898 (p= .0064)	—
Google News	0.1731 (p< .001)	0.1413 (p< .001)	—
Reddit	0.1741 (p< .001)	0.01682 (p< .001)	—
Facebook	0.1806 (p< .001)	0.0817 (p= .0131)	—
CO	-0.0971 (p= .0032)	-0.0988 (p= .0027)	—
UV-Index	-0.0963 (p= .0034)	-0.1283 (p< .001)	-0.0974 (p= .0031)
Min Temperature	0.1229 (p< .001)	—	—
GDELT	0.2269 (p< .001)	0.219 (p< .001)	—

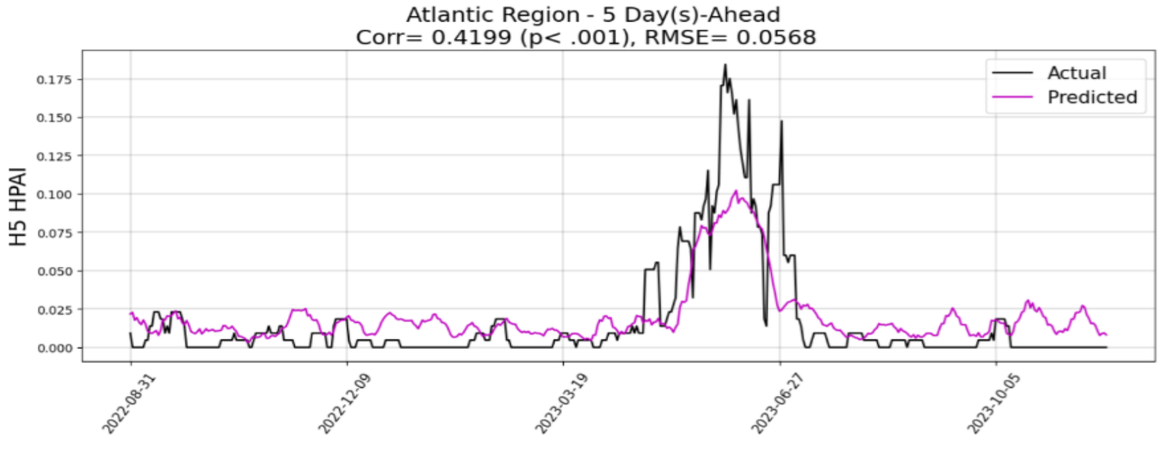
Results: Regional-Level

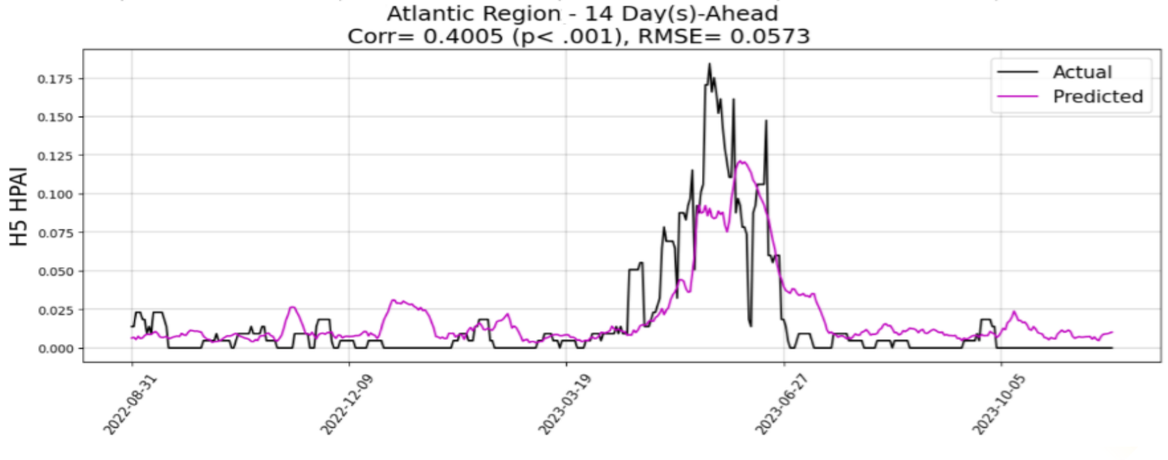












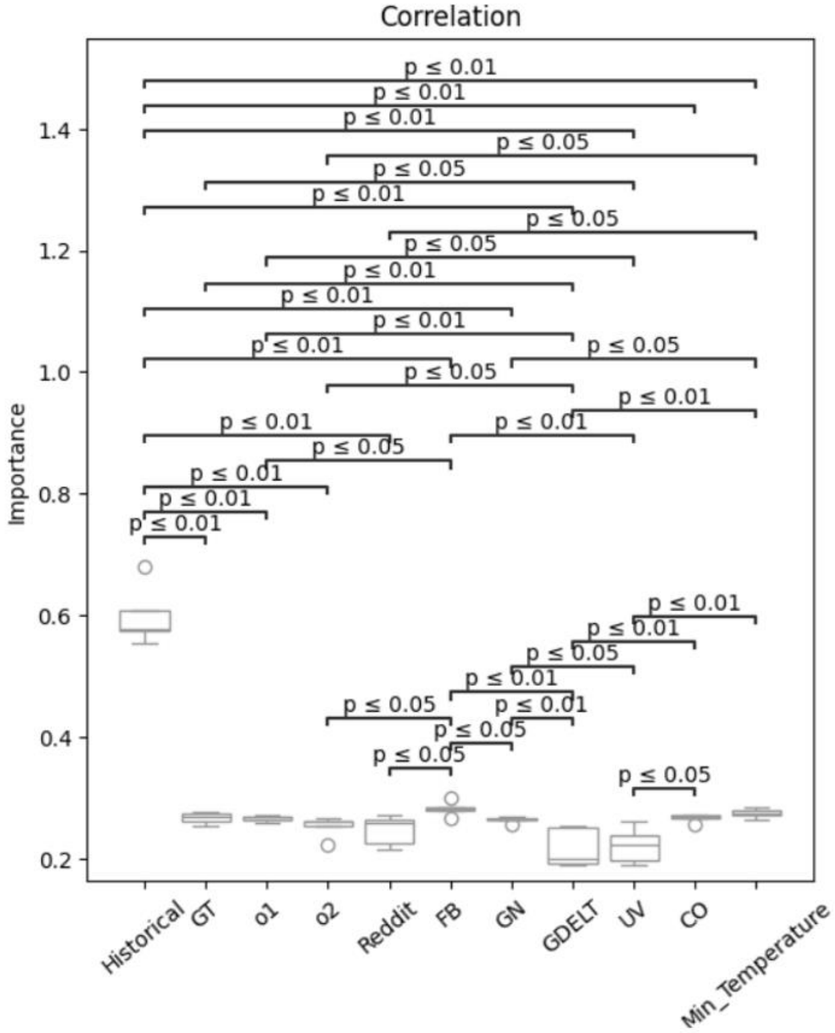
Feature Importance: Permutation Method

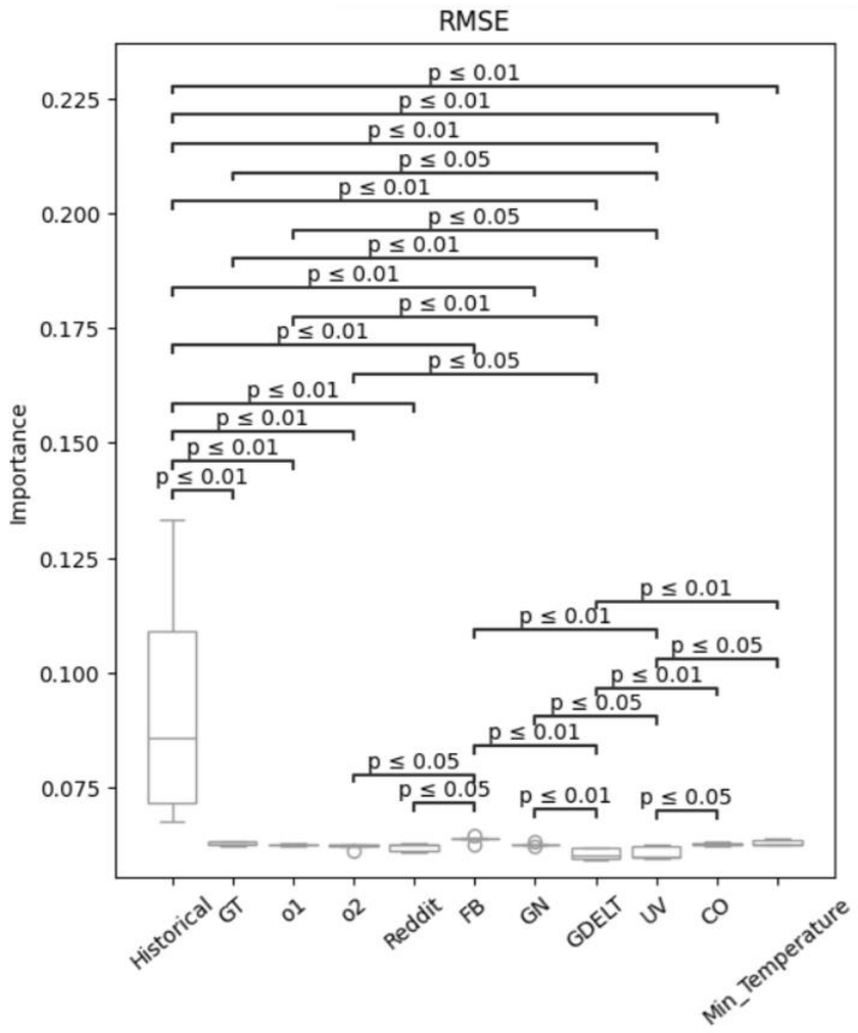
- The permutation method is used for finding feature importance.
  - In this method the training and evaluation phases are repeated several times.
  - Each time one of the features is permuted.
  - Performance degradation is computed for each evaluating metric:
    - RMSE
    - Correlation
  - The accuracy degradation indicates how important that feature is.

- Each feature was permuted 10 times.
- Performance degradations were compared using the Mann-Whitney U test.

Feature Importance: Permutation Method

- The results show that historical data has the highest importance.
- After historical data, Facebook, and Minimum Temperature also have a significantly higher importance compared to other parameters.



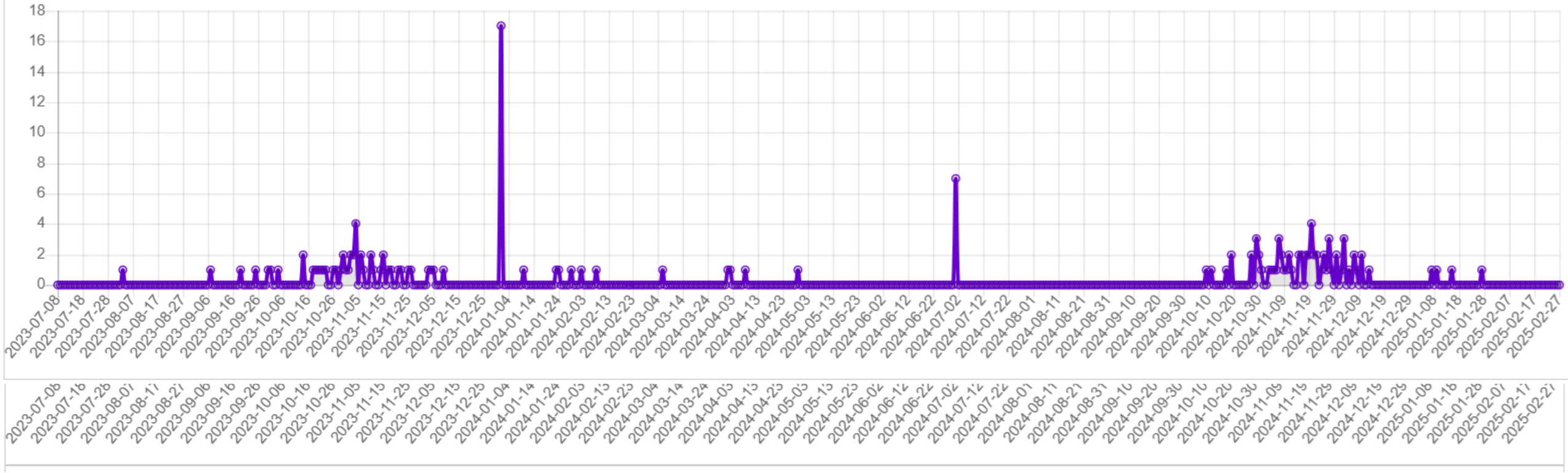


Early Warning System Framework

- An online framework was developed based on this study to illustrate avian influenza outbreaks for the next 14 days: https://aimmlab.org/early-warning-system-for-avian-influenza-outbreaks/

Various data sources could also be used for further surveillance of avian influenza outbreaks. Generally, a rise in Google trends, Google news, and the number of Reddit posts means that

Number of Cases (WAHIS Dataset)



Discussion and Conclusion

- This work introduces new sources of data for enhancing HPAI surveillance.
- A forecasting model is implemented and evaluated to assess the efficiency of the sources of data

- One big limitation of this work is not having access to enough data.
  - while our data was sourced from public platforms widely used by the general population, it may not reflect the online behavior of stakeholders most directly affected by HPAI, such as farmers and poultry industry workers.
  - Datasets on the number of avian influenza cases are provided with a great delay.
  - Wastewater data could be very helpful, yet is not available.

Discussion and Conclusion

- Another limitation is that the datasets need to be stored on fast and reliable storages and the models need to be trained and evaluated using strong systems with GPU.
- The results of the work can be consumed by resource-limited systems such as smartphones and edge-nodes.
- Future works could focus on spatial and hotspot analysis.
- In future works, social media posts could be analyzed using NLP for discovering concerns, fears, and mis/dis-information
- In future works, data from private Facebook pages and subreddits could be gathered to reflect the behaviour of farmers and poultry workers.

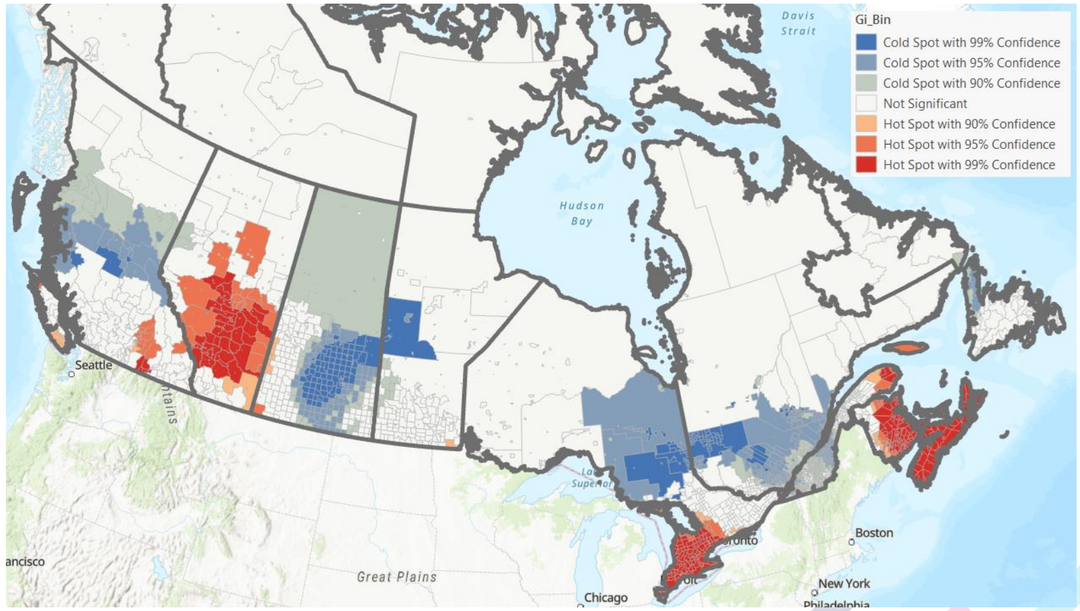
Discussion and Conclusion

- Another contribution to future work is studying the transmission of avian influenza through wild and domestic cattle and mammals, as it has been an emerging issue during the current avian influenza outbreak.
  - More data on cattle and mammal pathways is required.
  - Biological and environmental factors that increase the risk of transmission to cattle and mammals must be studied and hotspots need to be detected.
  - It is paramount to study the genomic sequences of the strains that are transmitted through mammals.
  - Measures and regulations need to be informed to evaluate, prevent or mitigate avian influenza infections in domestic cattle and mammals.

- It is essential to study how climate change is effecting spatial and temporal trends of avian influenza.
  - Climate change, urbanization, and deforestation is shifting animal, especially migratory bird, pathways.

Future Work

- Hotspot analysis:
  - Find the hot- and cold-spots of avian influenza.
  - Understand what factors play a more important role.
  - Understand the farms and poultries that are more at risk.



- Analyzing social media activity
  - Understand people’s fears and concerns using Topic Modeling and Sentiment/Emotion Analysis
  - Understand mis- and dis-information being shared on social media

Discussion Period

Any questions?

Please use the **Q&A tab** to submit your questions for our speaker. You can “**like**” other people’s questions to push them up in priority.

Closing Remarks

Thank You!

The seminar recording and presentation slides will be posted on https://nccid.ca/ in the coming weeks.

Please complete our **post-seminar evaluation survey** by scanning its QR code. Today's survey will also be distributed to you shortly after the seminar.

Join us on **June 24** (1:00-2:00pm ET) for the next seminar.

Visit https://nccid.ca/surveillance-advances-seminar-series/ for more information about the Surveillance Advances seminar series.