Proposed title of the project

Mirna Elizondo

Your supervisor’s name

Programme

Computer Science

Texas State University

Date

Table of Contents

[Introduction 3](#_Toc7773421)

[Background and context 3](#_Toc7773422)

[Problem statement 3](#_Toc7773423)

[Research questions 3](#_Toc7773424)

[Relevance and importance of the research 3](#_Toc7773425)

[Literature review 3](#_Toc7773426)

[Key concepts, theories and studies 3](#_Toc7773427)

[Key debates and controversies 3](#_Toc7773428)

[Gaps in existing knowledge 3](#_Toc7773429)

[Research design and methods 3](#_Toc7773430)

[Aims and objectives 3](#_Toc7773431)

[Methods and sources 3](#_Toc7773432)

[Practicalities and potential obstacles 3](#_Toc7773433)

[Implications and contributions to knowledge 3](#_Toc7773434)

[Practical implications 3](#_Toc7773435)

[Theoretical implications 3](#_Toc7773436)

[Reference list 4](#_Toc7773437)

[Research schedule 5](#_Toc7773438)

[Research phase 5](#_Toc7773439)

[Objectives 5](#_Toc7773440)

[Deadline 5](#_Toc7773441)

# **Introduction**

Titles:

* RADx Long COVID Prediction Challenge: Evaluating methods for predicting the risk of developing PASC/Long COVID in patients with SARS-CoV-2 infection
* Evaluating methods for predicting the risk of developing PASC/Long COVID in patients with SARS-CoV-2 infection using NCAT’s National COVID Cohort Collaborative (N3C) Data Enclave.
* Evaluating methods for predicting the risk of developing PASC/Long COVID in patients with SARS-CoV-2 infection using a patient’s historical clinical data using NCAT’s National COVID Cohort Collaborative (N3C) Data Enclave.

The introduction should include the following elements:

Contextualize

* What is already known about the problem?
* Is the problem limited to a certain time period or geographical area?
* How has the problem been defined and debated in the scholarly literature?

Why is matters?

* How will resolving the problem advance understanding of the topic?
* What benefits will it have for future research?
* Does the problem have direct or indirect consequences for society?

## **Background and Context**

Lead the reader into the topic and scope of your research.

## **Problem Statement**

Describe the theoretical or practical [research problem](https://www.scribbr.com/dissertation-writing-roadmap/problem-statement/) that you want to address. What is already known about the problem? What is missing from current knowledge?

For patients diagnosed with the SARS-CoV-2 infection it has caused a significantly impact on their health and the American health care system. Some individuals have a risk of developing long-term effects from their infections, known as PASC/Long COVID. These conditions can last weeks or months. Current researchers are working to understand more about who is at risk for experiencing Long COVID conditions; although they are still learning to what extent certain groups are at risk and how different groups can experience it. By identifying the risk factors and the most frequent symptoms and diagnoses experience by patients with Long-COVID researchers can collaborate with professional medical groups to develop and offer guidelines for healthcare providers, patients, and the general public.

## **Research Questions**

State the specific [question(s)](https://www.scribbr.com/dissertation-writing-roadmap/research-questions/) that you aim to answer.

In this project, we plan to develop ML pipeline to provide insight into patient-level data from large amount of data, to better understand the effects of SARS-CoV-2 on patients, and to identify prognostic factors leading to PASC/Long COVID.

## **Relevance and Importance of the Research**

Make it clear what new insights you will contribute, who they are relevant to, and why the research is worth doing.

# **Literature review**

The [literature review](https://www.scribbr.com/dissertation/literature-review/) summarizes, compares and critiques the most relevant scholarly sources on the topic. There are many different ways to structure a literature review, but it should explore:

## **Key Concepts, Theories and Studies**

Compare, contrast, and establish the theories and concepts that will be most important for your project.

## **Key Debates and Controversies**

Identify points of conflict and situate your own position.

## **Gaps in Existing Knowledge**

Show what is missing and how your project will fit in.

# **Research design and methods**

Here you should explain your approach to the research and describe exactly what steps you will take to answer your questions.

## **Research design**

Explain how you will design the research. [Qualitative or quantitative](https://www.scribbr.com/methodology/qualitative-quantitative-research/)?

* Quantitative Metrics:
  + Build models on a training dataset established by challenge organizers.
  + Models will be tested on a prospectively collected holdout test dataset to establish model performance
  + Calibration metrics (mean error, brier score, etc.)
  + Area Under the Precision Recall Curve
  + Area Under the Receiver Operator Curve
    - To evaluate model generalizability, where models will be evaluated for temporal, cross-site, and demographic generalizability.
  + Precision and Recall
* Qualitative Metrics:
  + Utility
    - Feature interpretability and relevance
      * Do the factors driving the individual predictions make sense?
    - Timeliness of predictions
      * Make predictions before diagnosis
    - Context utility – can a model make accurate predictions
    - likelihood of implementation
      * likelihood of hospital could implement the model
  + Reproducibility
    - Technical reproducibility
    - Prediction reproducibility
    - Documentation reproducibility
    - Method clarity

Original data collection or [primary/secondary sources](https://www.scribbr.com/citing-sources/primary-and-secondary-sources/)? Descriptive, correlational or experimental?

## **Methods and Sources**

Describe the tools, procedures, participants and sources of the research.

When, where and how will you collect, select and analyze data?

* Models will be developed withing the N3C enclave. For each of the patients in the relevant challenge dataset, models will output scores between 0 and 1, with one score per person in the validation data.

## **Practical Considerations**

Address any potential obstacles, limitations and ethical or practical issues. How will you plan for and deal with problems?

* Using the N3C Enclave system

# **Implications and contributions to knowledge**

Finish the proposal by emphasizing why your proposed project is important and what it will contribute to practice or theory.

## **Practical Implications**

Will your findings help improve a process, inform policy, or make a case for concrete change?

## **Theoretical Implications**

Will your work help strengthen a theory or model, challenge current assumptions, or create a basis for further research?

# **References**

AuthorLastName, FirstInitial., & Author LastName, FirstInitial. (Year). Title of article. Title of Journal, Volume(Issue), Page Number(s). <https://doi.org/number>

<https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html>

<https://www.hhs.gov/civil-rights/for-providers/civil-rights-covid19/guidance-long-covid-disability/index.html#footnote10_0ac8mdc>

# **Research schedule**

|  |  |  |
| --- | --- | --- |
| **Research phase** | **Objectives** | **Deadline** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |