# A BRIEF REVIEW OF HYPERPARAMETER OPTIMIZATION METHODS FOR MACHINE LEARNING

Richard Wen
rwen@ryerson.ca
Department of Civil Engineering

November 29, 2017

Presentation for Research Methods in Computer Science Course Instructed by Dr. Cherie Ding.

Ryerson University

# OUTLINE

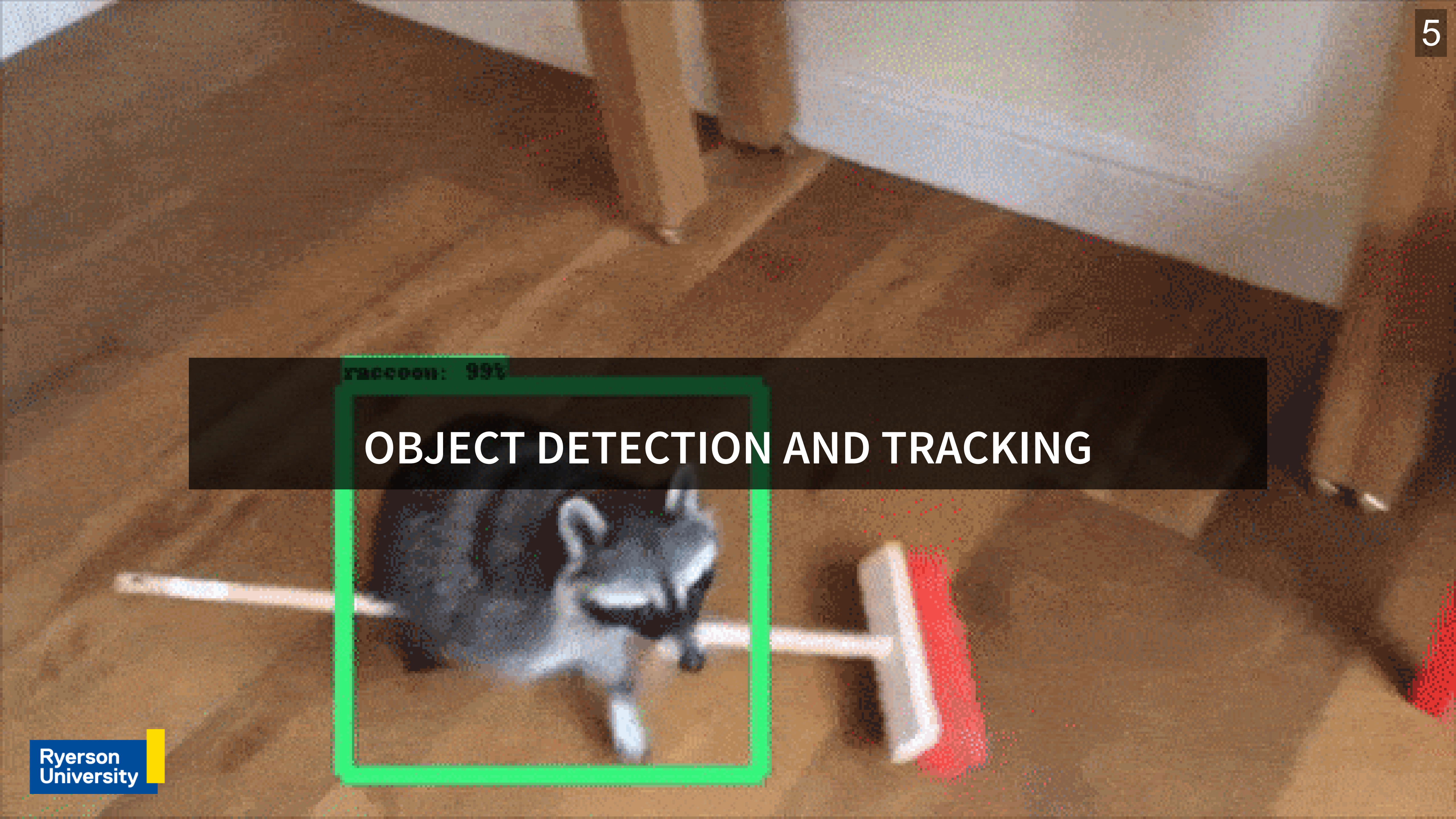
1. Introduction
2. Methods
3. Results
4. Discussion
5. Conclusion
6. References

# INTRODUCTION

# MACHINE LEARNING

- Learn from data
- Accuracies > 99% possible
- Real-world applications

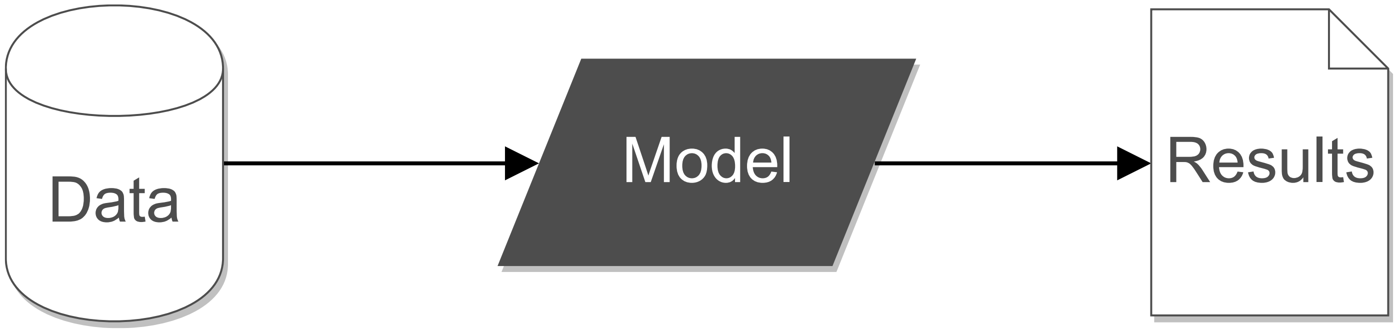
# OBJECT DETECTION AND TRACKING



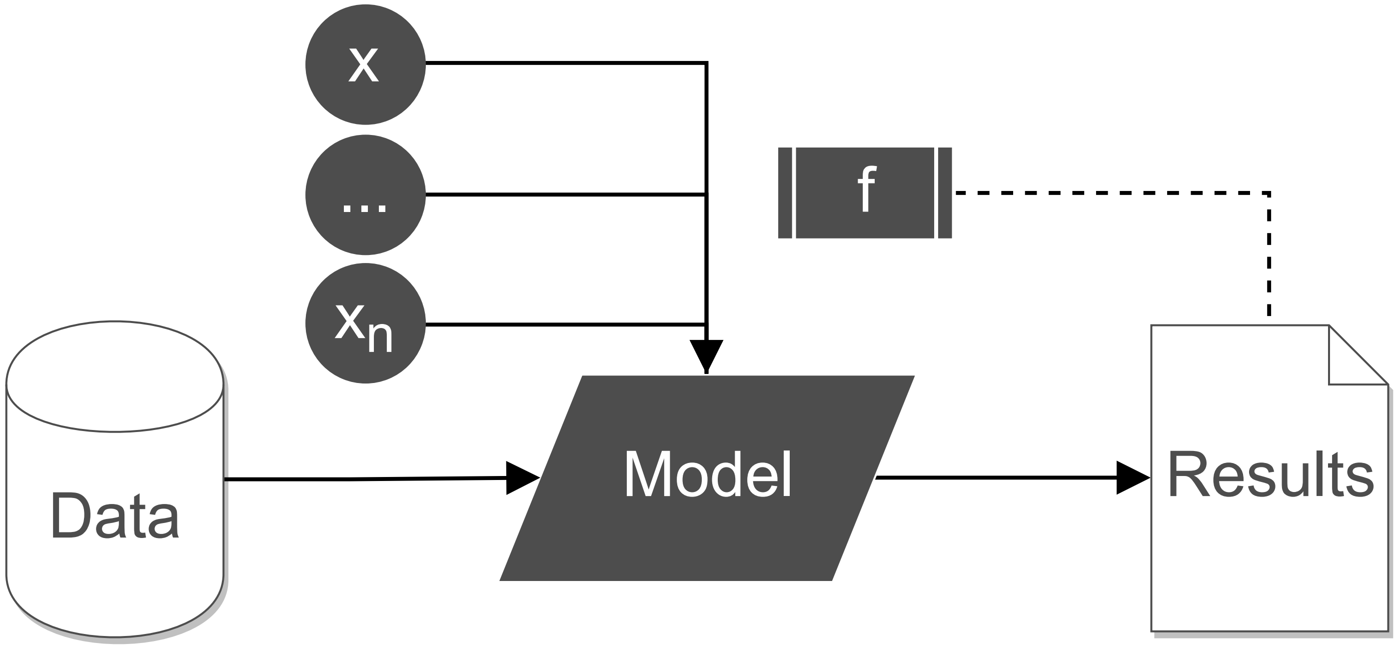
# MEDICAL IMAGE DIAGNOSES


Hurricane Harvey Flooding
Simonton, Texas
November 20, 2016 (Left)
August 30, 2017 (Right)
DISASTER MAPPING
DigitalGlobe
Ryerson University

# MACHINE LEARNING PROCESS



# MACHINE LEARNING MODEL




Ryerson University

# HYPERPARAMETERS

- Adj*u*stable val*u*es
- Affect learning performance
- Before learning process

# HYPERPARAMETER OPTIMIZATION

- Adjustable values $x \ldots x_n$
- Affect performance measure $f$
- Find optimal $x \ldots x_n$

$$\underset{x \ldots x_n}{\operatorname{argmax}} f(x \ldots x_n)$$

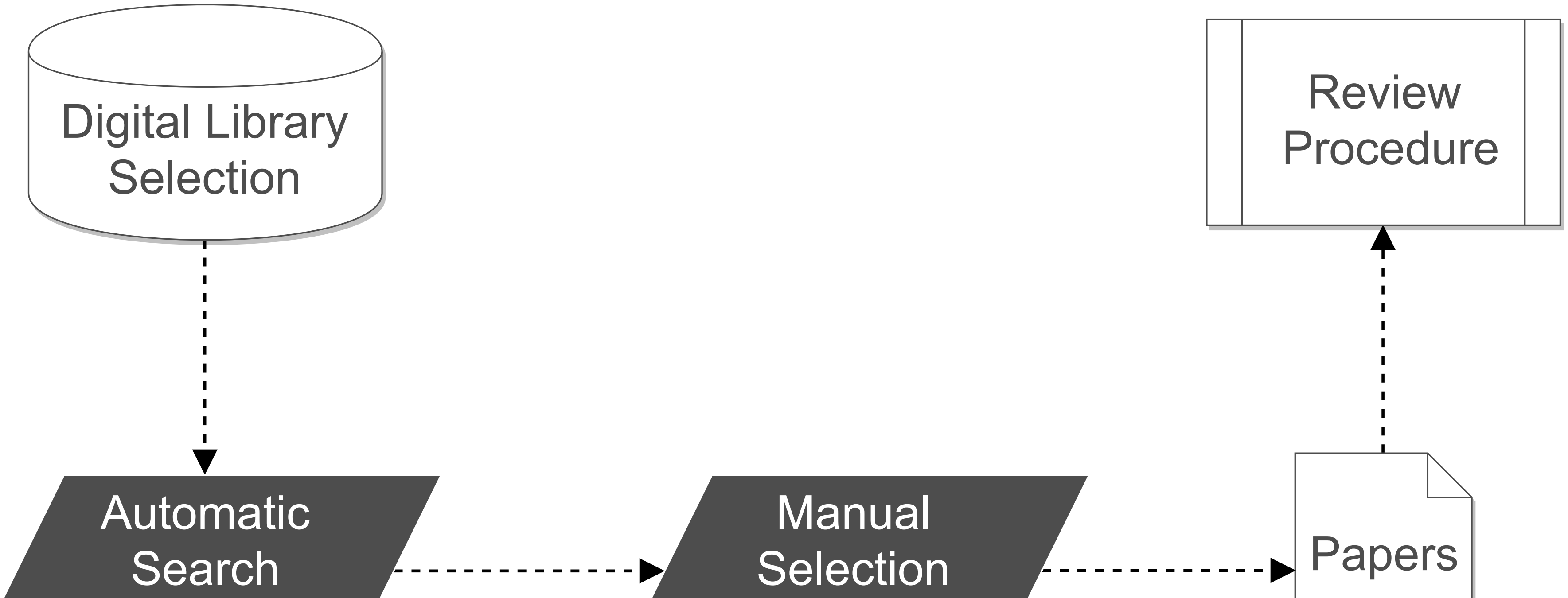
## OBJECTIVES

For hyperparameter optimization:

1. `Summarize` recent methods
2. `Discuss` limitations
3. `Propose` improvements and *future* directions

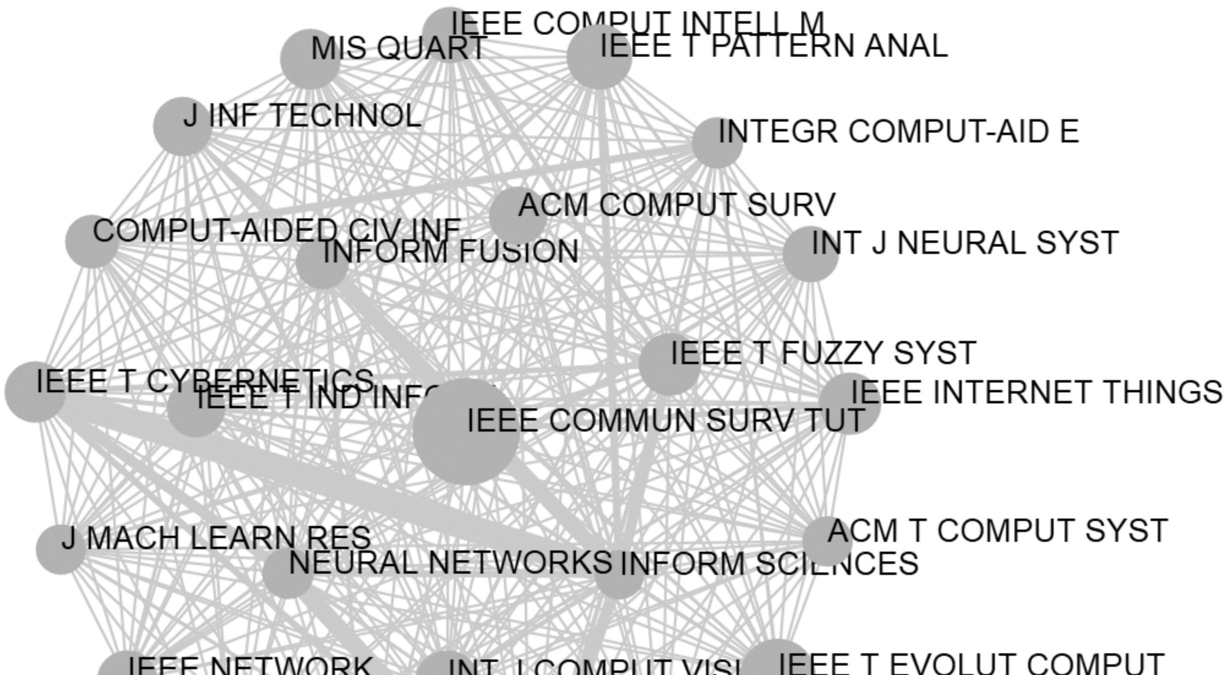
# METHODS

# METHODS PROCESS




Ryerson University

# DIGITAL LIBRARY SELECTION




Ryerson University

# AUTOMATIC SEARCH QUERIES

| Query | Value |
|---|---|
| `Publication` | **IEEE** or **ACM** |
| `Year` | 2014 to October 5, 2017 |
| `Title contains` | hyperparameter, optimization |

# MANUAL SELECTION CRITERIA

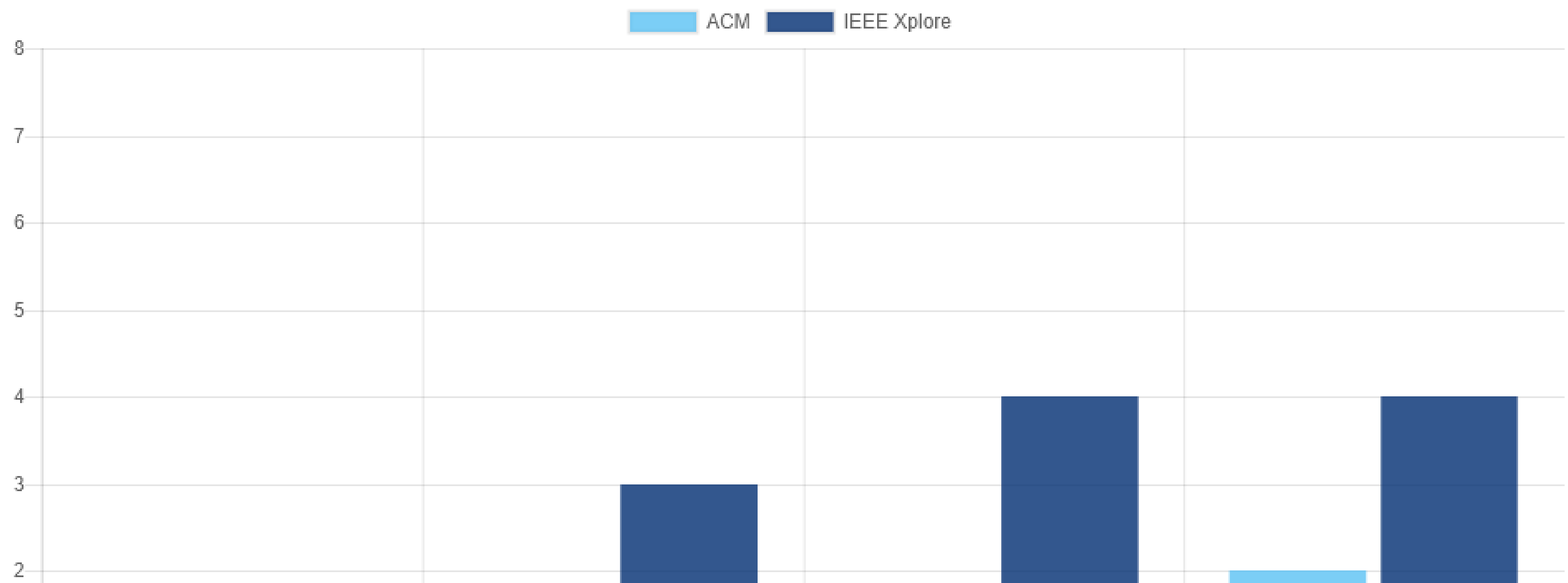
| Criteria | Description |
| --- | --- |
| `Detailed` | specific methods and results, >= 8 pages |
| `Relevant` | mention hyperparameter optimization |
| `Practical` | experiments, benchmarks, applications |

# REVIEW PROCEDURE

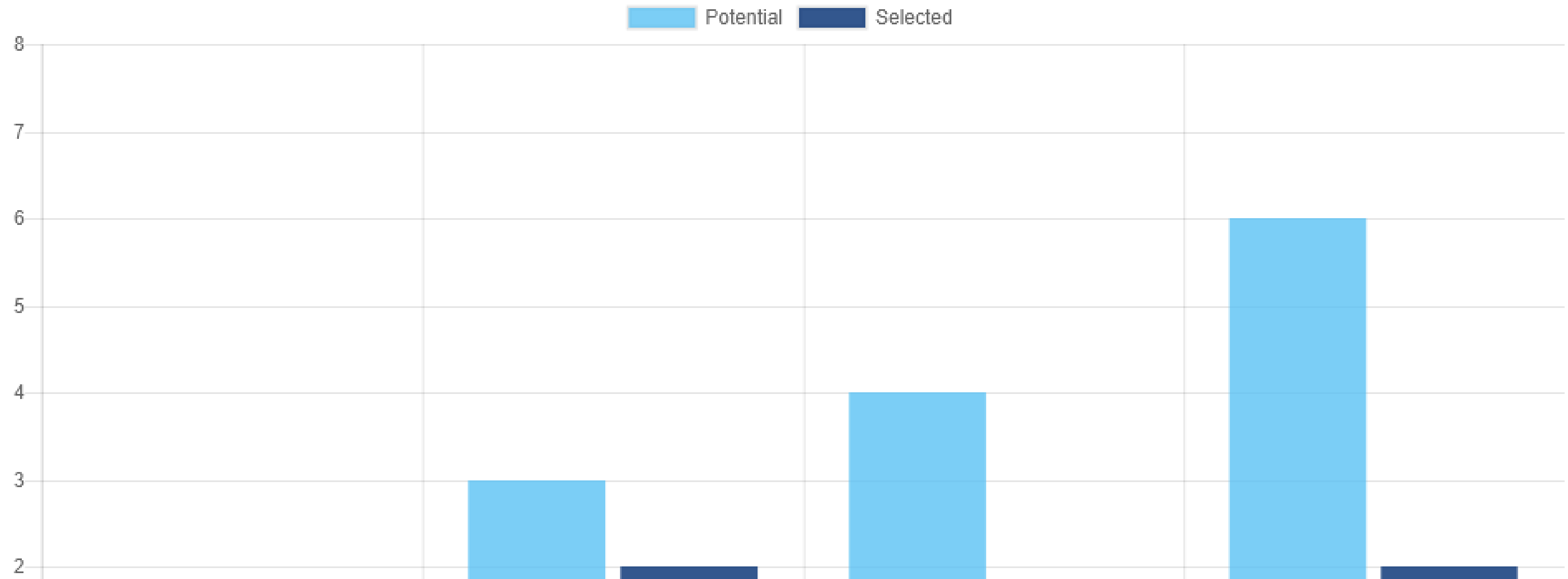
1. `Identify` methods
2. `Summarize` methods
3. `Summarize` experiments and results
4. `Discuss` limitations, improvements, directions

# RESULTS

# POTENTIAL PAPERS



# SELECTED PAPERS



## HYPERPARAMETER OPTIMIZATION METHODS

1. `Simple`: **Exhaustive** search
2. `Advanced`: **Model** or **procedural** based search

# SIMPLE METHODS

| Method | Description |
|---|---|
| `Manual Search` | Trial and error |
| `Grid Search` | Predefined range of values |
| `Random Search` | Randomized range of values |

Ref: [6]

# MANUAL SEARCH

| Hyperparameter | Values |
|---|---|
| $x_1$ | 5 |
| $x_2$ | 10 |
| $x_3$ | 15 |

# GRID SEARCH

| Hyperparameter | Values |
|---|---|
| $x_1$ | 5, 10, 15 |
| $x_2$ | 3, 6, 9 |
| $x_3$ | 2, 4, 6 |

# RANDOM SEARCH

| Hyperparameter | Values |
|---|---|
| $x_1$ | 3, 21, 22, 7 |
| $x_2$ | 5, 10, 7 |
| $x_3$ | 3, 23 |

# ADVANCED METHODS

| Method | Description |
| --- | --- |
| `Assumption` | Expert assumptions for particular cases |
| `Evolutionary` | Procedurally generated hyperparameters |
| `Sequential Model` | Sequentially guided hyperparameters |

# ASSUMPTION BASED OPTIMIZATION

- **S**pecific algorithms, data, and cases
- **S**elect $x \ldots x_n$ based on ass*u*mptions
- e.g. **D**istrib*u*tion ass*u*mptions

Ref: [1]

Ryerson University

# EVOLUTIONARY BASED OPTIMIZATION

- Mimic biological evol*u*tion
- **E**volve and nat*u*rally select $x \dots x_n$
- e.g. **G**enetic algorithms, particle swarm

Ref: [7]

# SEQUENTIAL MODEL BASED OPTIMIZATION (SMBO)

- Model performance function $f$ using $x \ldots x_n$
- Predict next best set of hyperparameters
- e.g. Bayesian optimization, random forest

Ref: [8-10]

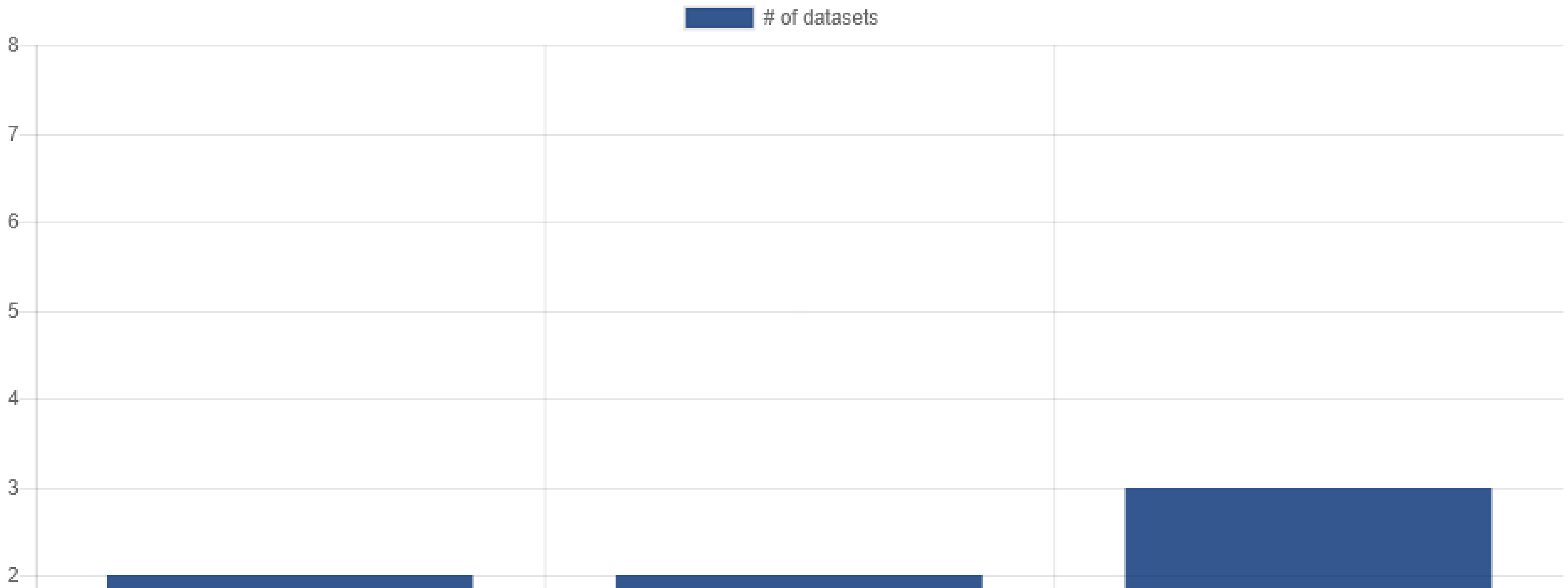
# SMBO IMPROVEMENTS

- Transfer learned hyperparameters [1]
- Discover better starting hyperparameters [2]
- Measure transfer hyperparameters influence [3]
- Consider hyperparameter inter-dependency [4]

# SMBO EXPERIMENTS

- Improved performance (e.g. ~10-20%)
- Better hyperparameters than simple methods
- Same time and iteration constraints

# SMBO DATA




Ryerson University

# DISCUSSION

# LIMITATIONS

- Man*u*al constraints selection
- Scalability
- Dataset variability

# MANUAL CONSTRAINTS

- Hyperparameters $x \ldots x_n$
- Performance measure $f$
- Set of constraints $C$
- Iteration or steps $t$

$$\underset{x \ldots x_n \in C}{\operatorname{argmax}} f(x \ldots x_n) \text{ given } t$$

Ryerson University

# IMPROVEMENTS AND FUTURE DIRECTIONS

- Automated Machine Learning
- Combining methods
- Sampling

# CONCLUSION

- Reviewed 5 papers
- **SMBO** as an effective framework
- Constraints and scalability
- *A*utomated **M**achine **L**earning

# REFERENCES

- [1] N. Schilling, M. Wistuba, L. Drumond, and L. Schmidt-Thieme, “Joint model choice and hyperparameter optimization with factorized multilayer perceptrons,” in 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), Nov 2015, pp. 72–79.
- [2] M. Wistuba, N. Schilling, and L. Schmidt-Thieme

- [3] M. Wistuba, N. Schilling, and L. Schmidt-Thieme, "Hyperparameter optimization machines," in 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Oct 2016, pp. 41–50.
- [4] J. C. Levesque, A. Durand, C. Gagne, and R. Sabourin, "Bayesian optimization for conditional hyperparameter spaces," in 2017


Ryerson University

- [5] A. Quitadamo, J. Johnson, and X. Shi, "Bayesian hyperparameter optimization for machine learning based eqtl analysis," in Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics, ser. ACM-BCB '17. New York, NY, USA: ACM, 2017, pp. 98–106. [Online]. Available:


Ryerson University

- [7] D. Whitley, S. Rana, J. Dzubera, and K. E. Mathias, "Evaluating evolutionary algorithms," Artificial intelligence, vol. 85, no. 1, pp. 245– 276, 1996.
- [8] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in Advances in neural information processing systems, 2012, pp. 2951–

- [9] F. Hutter, H. H. Hoos, and K. Leyton-Brown, “Sequential model-based optimization for general algorithm configuration.” LION, vol. 5, pp. 507– 523, 2011.
- [10] D. R. Jones, M. Schonlau, and W. J. Welch, “Efficient global optimization of expensive black-box functions.” Journal of Global optimization, vol

# THANK YOU

**R**ichard **W**en
rwen@ryerson.ca
**D**epartment of **C**ivil **E**ngineering

github.com/rrwen/slides-rmcs-litreview

Ryerson University