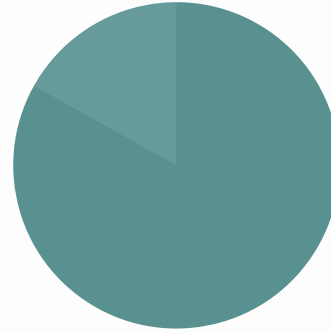


# Machine Learning- Assisted Raman Spectroscopy

Presented by  
Maryam Ghaffari

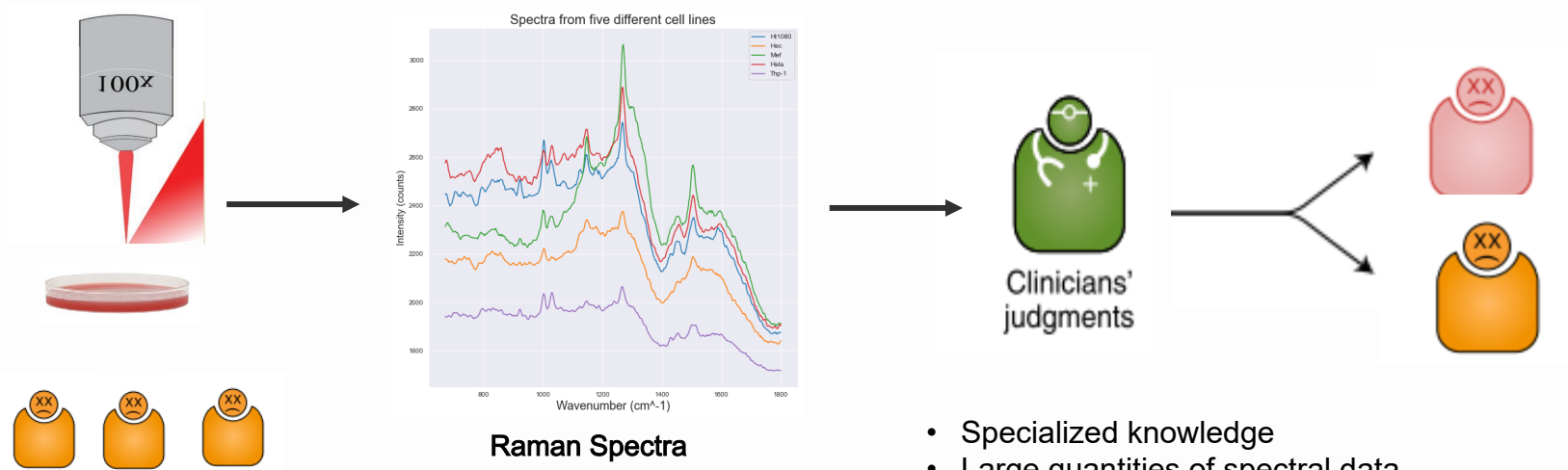




# Outline

- Problem statement
- Mission
- Methodology
  - Obtain Data
  - Scrub Data
  - Explore Data
  - Model Data
  - Interpret Model
- Conclusion
- Recommendation & future works

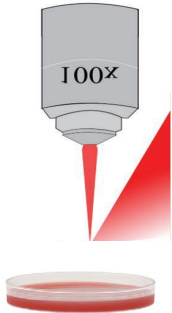
# Problem Statement



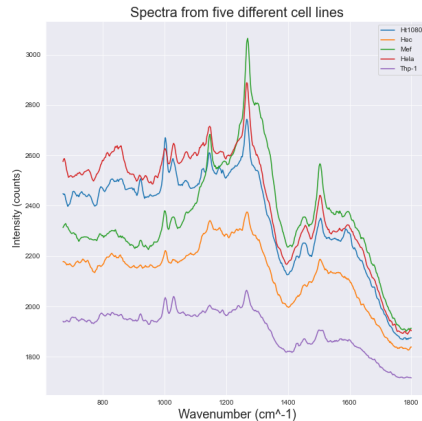
- Specialized knowledge
- Large quantities of spectral data
- Time consuming
- Weak Raman scatterers

# Mission

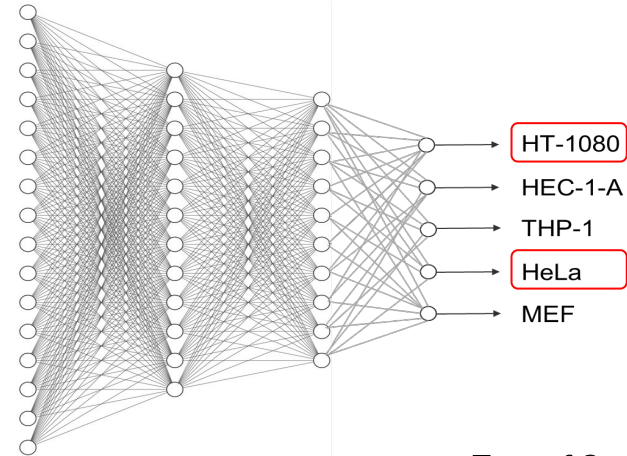
- Make Raman spectroscopy more accessible to **non-expert** users & facilitate **faster** and more **precise** data analysis.
- **Automated spectra analysis** methods based on machine learning



Blood sample



Raman Spectra



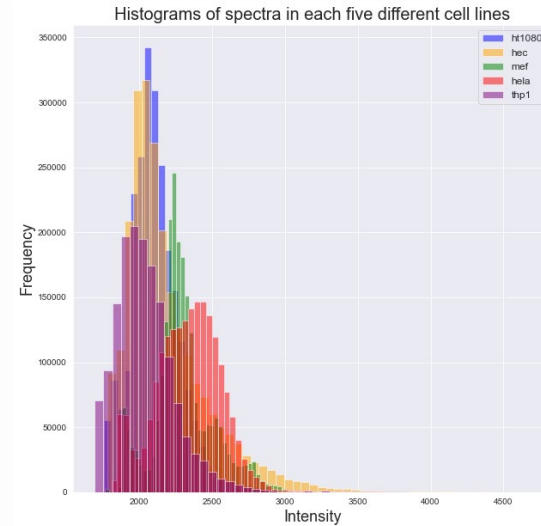
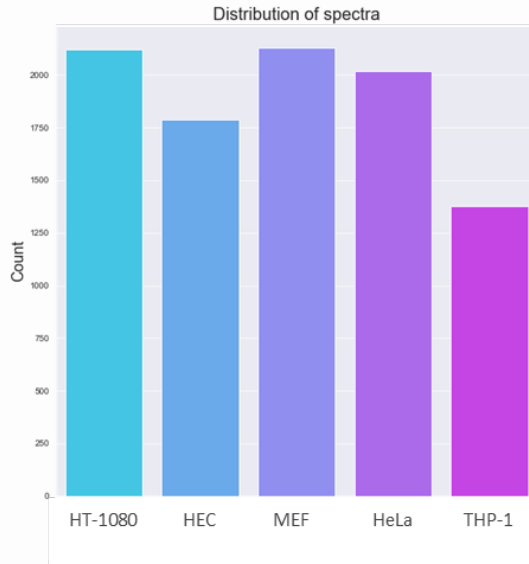
ML

Type of Cancers



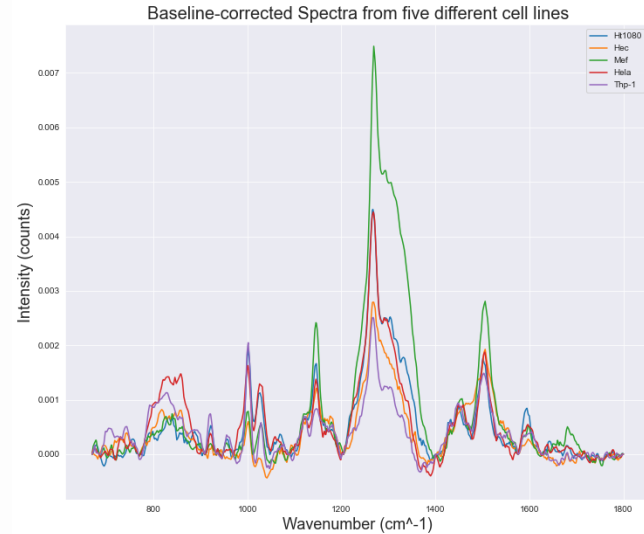
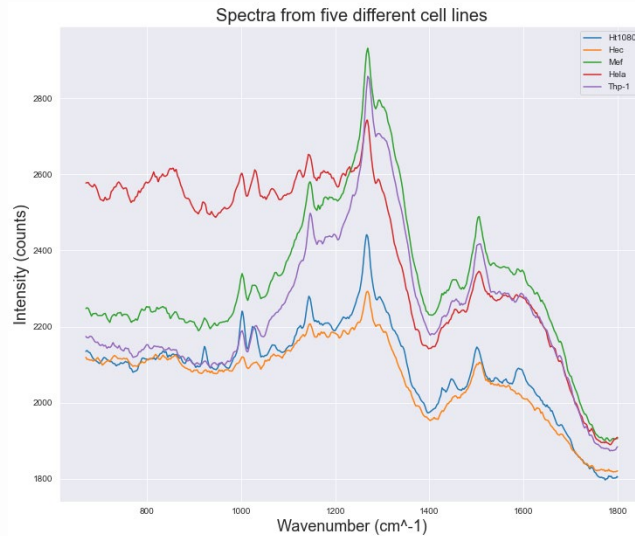
# Methodology: Obtain & Scrub data

- **Raman spectra** of biological samples from five types of cell lines were stored in a **CSV file** (rows=9467, columns=1132).



# Methodology: Scrub Data

- Preprocessing : Baseline Correction & Normalization





# Methodology: ML Models



$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$



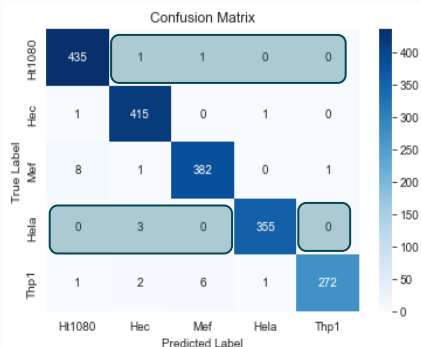
Recall => 0.9

Elapsed Time =< 60 Seconds

ML Models	Accuracy	Precision	Recall	F1 Score	Elapsed time (Seconds)	Misclassifications
Logistic Regression	0.985	0.99	0.98	0.99	3.54	27
Decision Tree	0.961	0.96	0.96	0.96	17.72	73
Decision Tree (200 top features)	0.919	0.92	0.92	0.92	4.65	152
Convolutional Neural Network (CNN)-without preprocessing	0.231	0.05	0.20	0.08	216.31	1449
Convolutional Neural Network (CNN)-with preprocessing	0.985	0.99	0.98	0.99	271.61	27
Fully connected neural network (FCNN)-without preprocessing	0.231	0.05	0.20	0.08	57.46	1450
Fully connected neural network (FCNN)-with preprocessing	0.988	0.99	0.99	0.99	57.04	22
Residual Network (ResNet)-with preprocessing	0.149	0.03	0.20	0.05	754.91	1604
Residual Network (ResNet)- with preprocessing regularization	0.149	0.03	0.20	0.05	754.91	1604

# Methodology: Model Interpretation

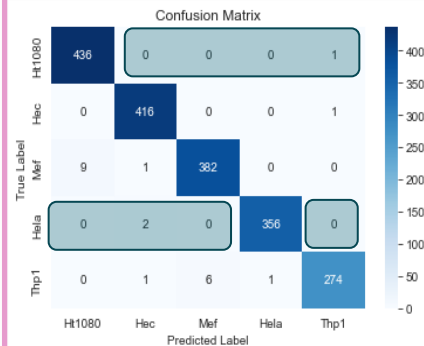
Logistic Regression



False Negative

	precision	recall	f1-score
Ht1080	0.98	1.00	0.99
Hec	0.98	1.00	0.99
Mef	0.98	0.97	0.98
Hela	0.99	0.99	0.99
Thp1	1.00	0.96	0.98

FCNN

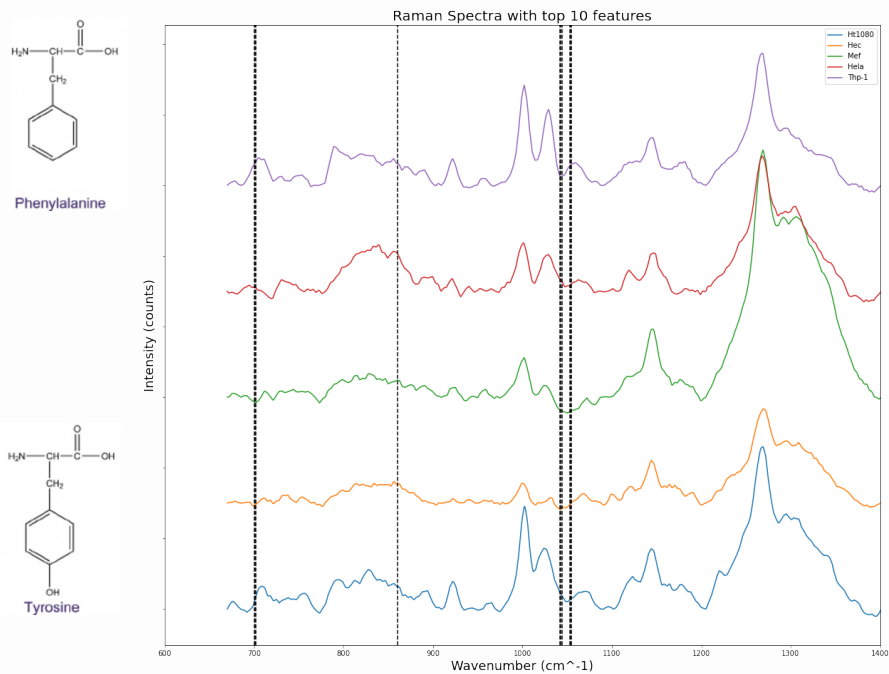


False Negative

	precision	recall	f1-score
Ht1080	0.97	1.00	0.99
Hec	0.99	0.99	0.99
Mef	0.98	0.97	0.98
Hela	0.99	0.99	0.99
Thp1	1.00	0.97	0.99



# Feature Importance





# Recommendation & Future works

- Deploy the model
- Increase dataset size
- Fine-tune the models
- Develop user-friendly interface
- Develop ensemble models
- Investigate feature extraction methods
- Further validation



# Conclusion

## FCNN Model

- ✓ Recall  $\Rightarrow$  0.9
- ✓ Low false-negative in high-risk cell types
- ✓ Elapsed Time  $\leq$  60 Seconds

# Thank You!

Email: [mar.ghaffari@gmail.com](mailto:mar.ghaffari@gmail.com)

GitHub: <https://github.com/MarGhaf/Machine>

Assisted - Raman - Spectroscopy

- Learning -