Lightscline

Data Reduction AI

Teaching machines to predict using 10% important data

Problem

Sensor data

73+ Trillion GB by 2025

Infrastructure cost

$100M+ annual cloud costs

Human capital cost

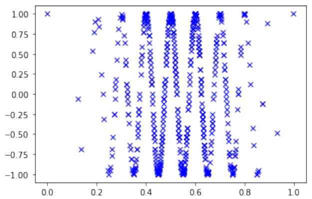
1 hour of data = **40+ hours** of analysis

Lightscline
Data Reduction AI

Solution

Lightscline's AI learns to predict using 10% data reducing

90% of AI infra & human time & costs



=

>10x faster and energy-efficient end-to-end predictions as we only analyze 10% of raw data

10x more gasoline from same amount of crude

Product

Getting started

25k lines of patented code + 4 years R&D into 4 lines of code

```
from lightscline.lightscline import LightsclineEdge
## Load data into Lightscline
ls = Lightscline(data=data,fs = SAMPLING_FREQUENCY)
## Reduce the amount of data by 70% of the original
ls.reduce_and_preprocess_data(per_reduction=70)
## Train the model
ls.train_model(verbose=True,n_iters = 1000)
## checking the results
ls.test_model()
```

- Setup within 10 mins
- On-prem / cloud hosting
- No data sharing required

Conventional vs. Lightscline AI workflow

Conventional techniques

>40 hours / dataset

Raw data files

- body_acc_x_train.txt
- body_acc_y_train.txt
- body_acc_z_train.txt
- body_gyro_x_train.txt
- body_gyro_y_train.txt

Manual feature extraction

```
555 angle(tBodyAccMean,gravity)
556 angle(tBodyAccJerkMean),gravityMean)
557 angle(tBodyGyroMean,gravityMean)
558 angle(tBodyGyroJerkMean,gravityMean)
559 angle(X,gravityMean)
560 angle(Y,gravityMean)
561 angle(Z,gravityMean)
```

Model development

- SVM
- KNN
- ANN..

Metrics

97% accuracy

Lightscline AI

~1 hour

10x less compute & infra costs

Raw data files

- body_acc_x_train.txt
- body_acc_y_train.txt
- body_acc_z_train.txt
- body_gyro_x_train.txt
- body_gyro_y_train.txt

Lightscline workflow

```
from lightscline.lightscline import LightsclineEdge
## Load data into Lightscline
ls = Lightscline(data=data,fs = SAMPLING_FREQUENCY)
## Reduce the amount of data by 70% of the original
ls.reduce_and_preprocess_data(per_reduction=70)
## Train the model
ls.train_model(verbose=True,n_iters = 1000)
## checking the results
ls.test_model()
```

Metrics

97% accuracy

>10x model size reduction

>10x faster

By predicting using just 10% important data, our lightweight models reduce your AI infra and people time and costs by 90%.

Customer case-studies

Asset Health Management

Human Activity Recognition

Enterprise energy monitoring

Asset Health Management

BOSCH

We tried the free version on other datasets and accuracies are mentioned in the below table:

S No	Dataset	Sampling Frequency	Accuracy
1.	Bearing	97656 Hz	86%
2.	Wind turbine planetary gearbox	40000 Hz	100%
3.	Unbalanced Impeller Centrifugal Blower	8000	95.2%
4.	Loose and Dent Impeller Centrifugal Blower	8000	84.6%

With just 10% data

Lightscline
Data Reduction AI

Asset Health Management

BOSCH

From
SX/BSV-TH1

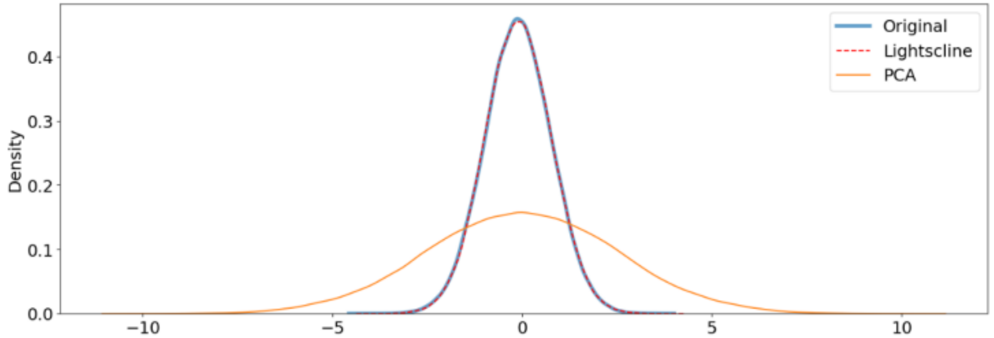
Our Reference
[redacted]

Tel
[redacted]

[redacted]

Version	V1.0
Topic	**Lightscline**
Description	Exploration of Lightscline Free Version

Data reduction technique was a good approach it eliminated redundant information and focused solely on relevant data. The technique was compared with Principal Component Analysis (PCA) and the distributions were found to be different. Lightscline's approach to data reduction kept the distribution of the data intact. Plots below shows the comparison with PCA.

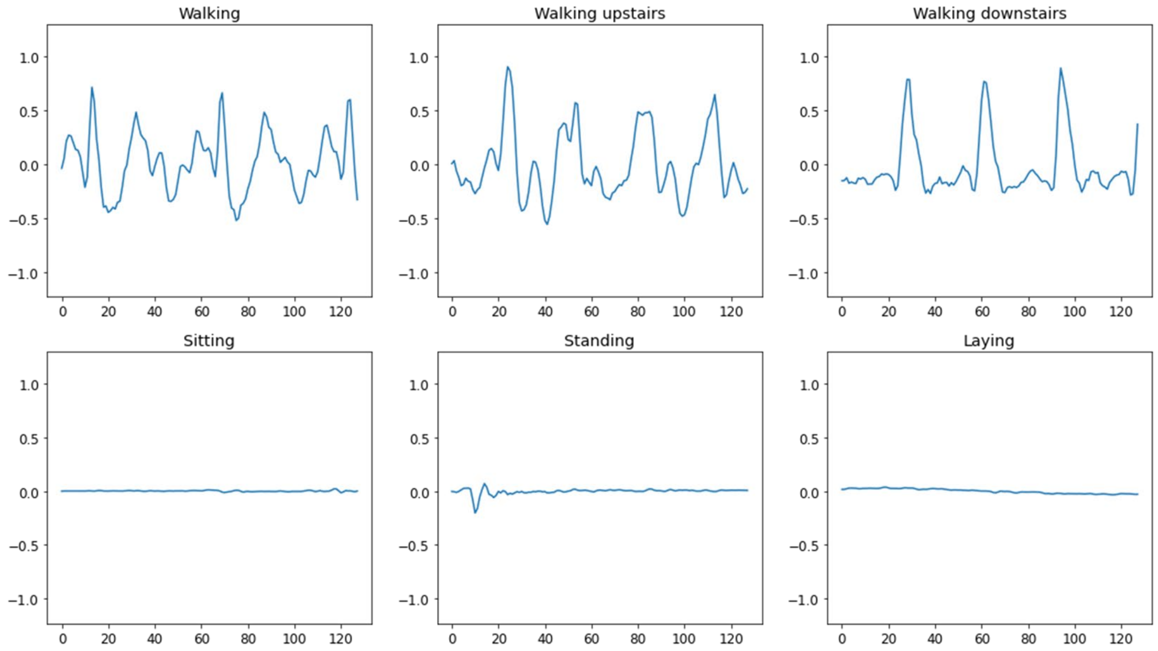


Lightscline
Data Reduction AI

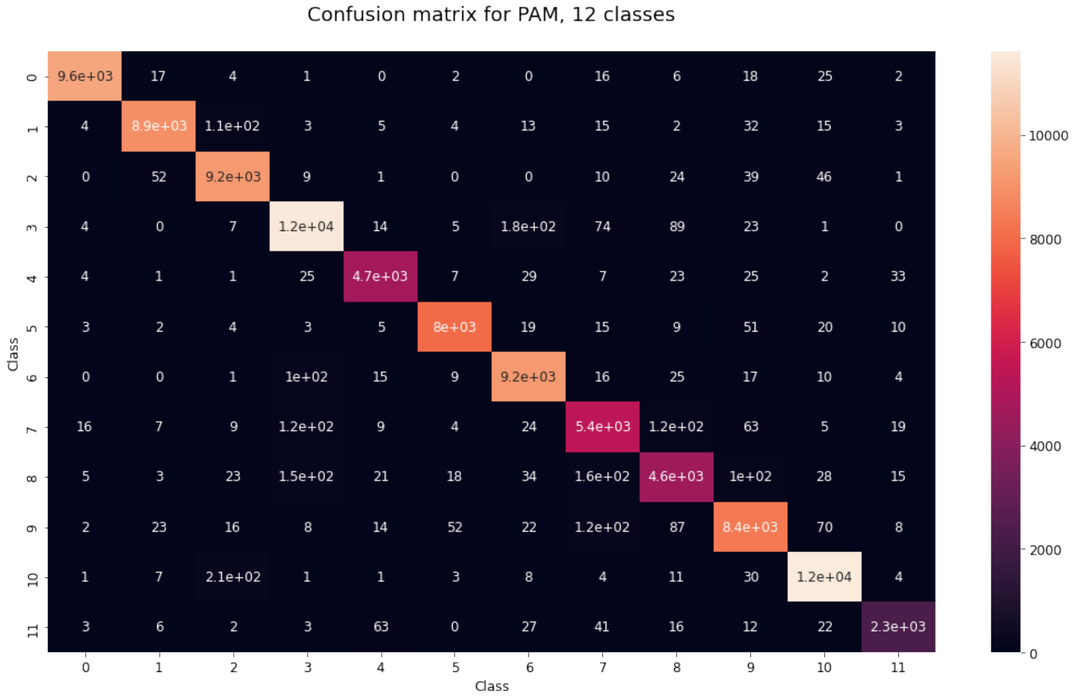
Human Activity Recognition

Model size reduction by ~9x and significant speed-ups

93% accuracy on UCI HAPT dataset (6 classes)



96.7% accuracy on PAMAP2 dataset (12/18 classes)



Lightscline
Data Reduction AI

Image cc: https://www.nasa.gov/gallery/hrp-images/

Enterprise energy monitoring

Helps data scientists select 100 useful windows from 10MM+ collected over 3 years

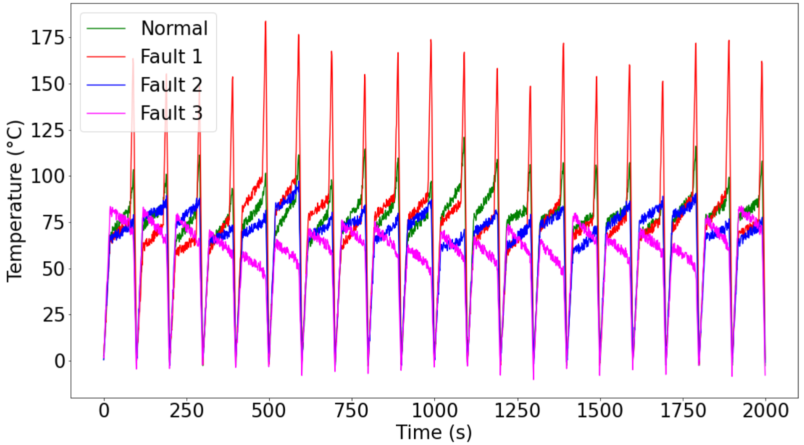
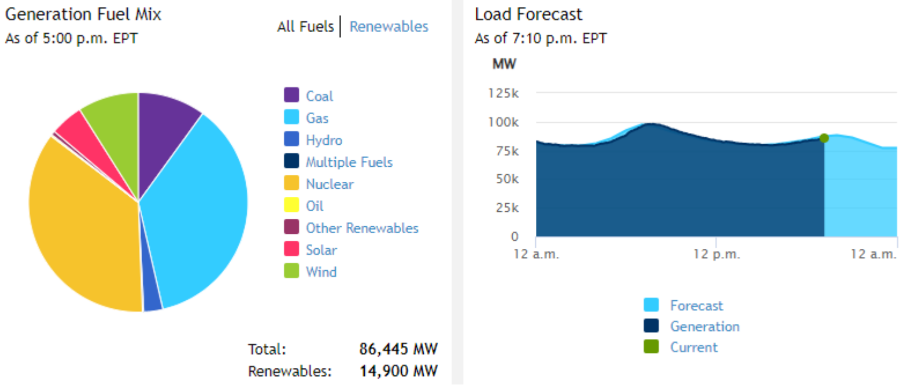


Figure out the reason for 100 MWh excess energy consumption



Analyze 100 useful windows which were previously hidden in an S3 bucket

% data used	Accuracy
8	99.9
10	99.9
15	100

*enables applications not possible today

Lightscline
Data Reduction AI

Lightscline
Data Reduction AI

By predicting using just 10% important data, our lightweight models reduce your AI infra and people time and costs by 90%.

Product

- 25k lines of patented code + 4 years R&D
- No data sharing required
- On-prem/cloud hosting

```
from lightscline.lightscline import LightsclineEdge
## Load data into Lightscline
ls = Lightscline(data=data,fs = SAMPLING_FREQUENCY)
## Reduce the amount of data by 70% of the original
ls.reduce_and_preprocess_data(per_reduction=70)
## Train the model
ls.train_model(verbose=True,n_iters = 1000)
## checking the results
ls.test_model()
```

4 lines of code setup under 10 minutes

Use-cases

- 95%+ accuracies
- Results in minutes
- Ready to deploy
- 10x AI cost reduction
- 10x productivity gains
- 10x model size reduction
- Explainability
- SWaP-C ready

Asset Health Management

BOSCH

We tried the free version on other datasets and accuracies are mentioned in the below table:

S No	Dataset	Sampling Frequency	Accuracy
1.	Bearing	97656 Hz	86%
2.	Wind turbine planetary gearbox	40000 Hz	100%
3.	Unbalanced Impeller Centrifugal Blower	8000	95.2%
4.	Loose and Dent Impeller Centrifugal Blower	8000	84.6%

Human Activity Recognition

- 96.7% accuracy on PAMAP2 dataset (12/18 classes),
- 93% accuracy on UCI HAPT dataset (6 classes), using only 10% raw data
- Model size reduction by ~9x and significant speed-ups
- No need of manual feature extraction

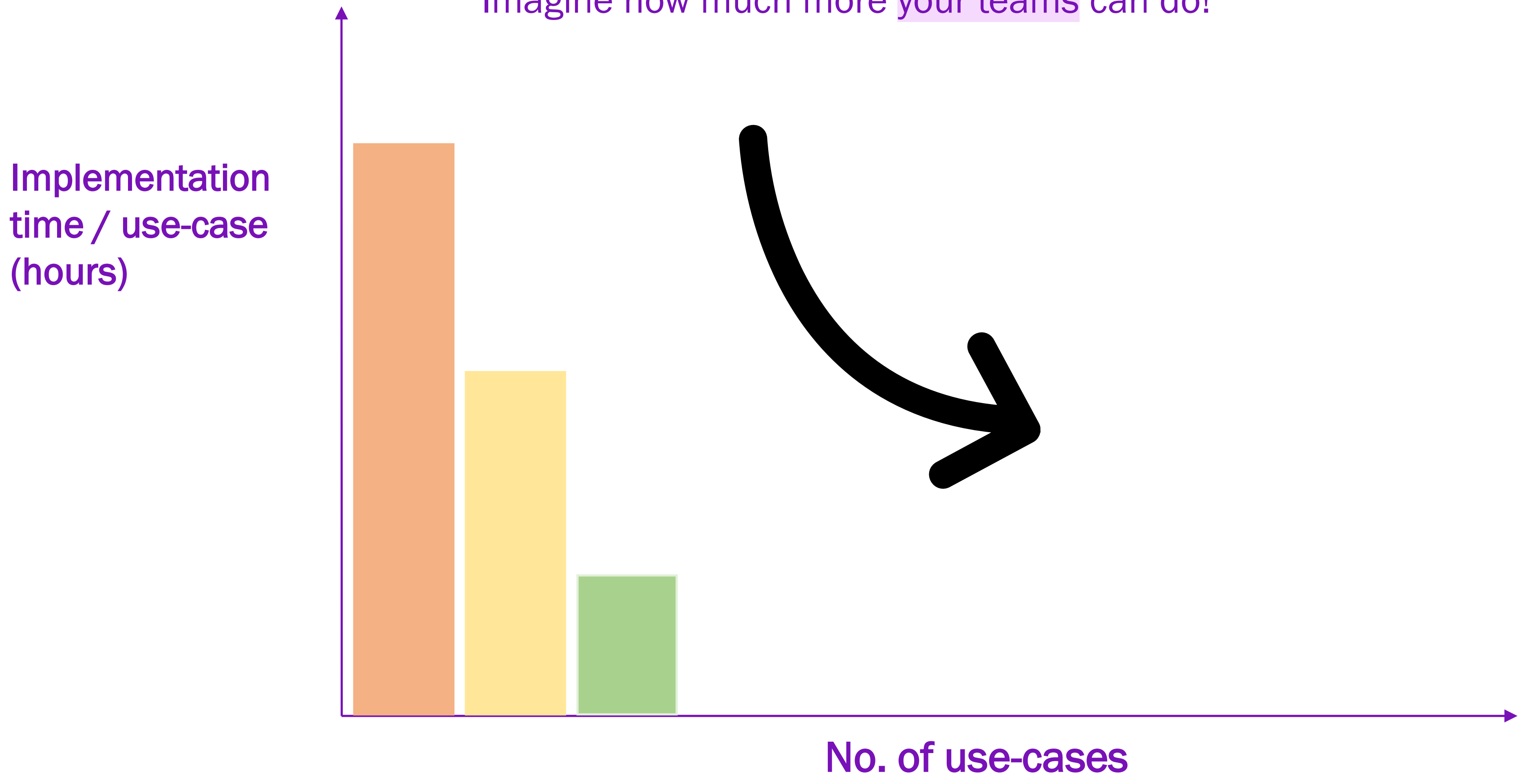
Enterprise energy monitoring

- 95%+ accuracies on energy data anomaly detection
- Automatically select the 10% important windows
- Automatic labelling on millions of windows

Lightscline
Data Reduction AI

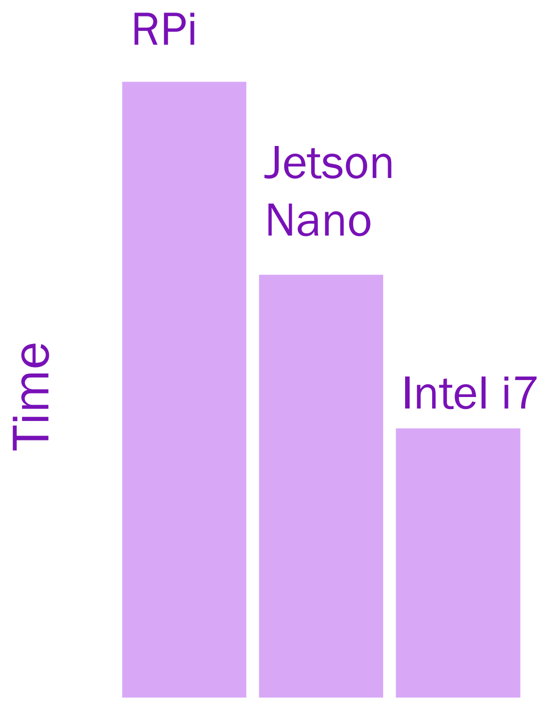
You can 10x your productivity by quickly scaling to different use-cases

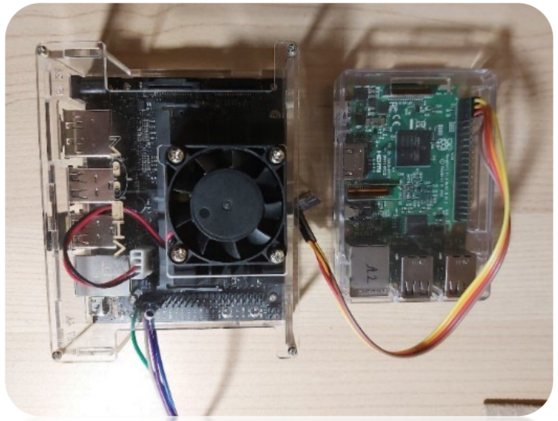
Imagine how much more your teams can do!

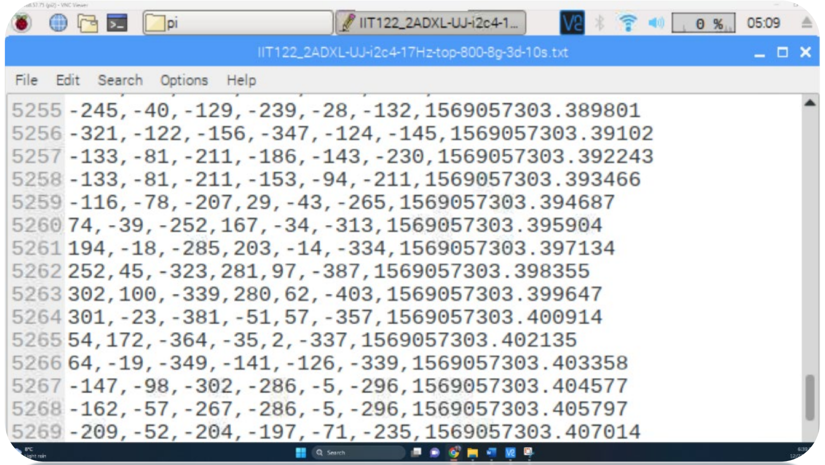


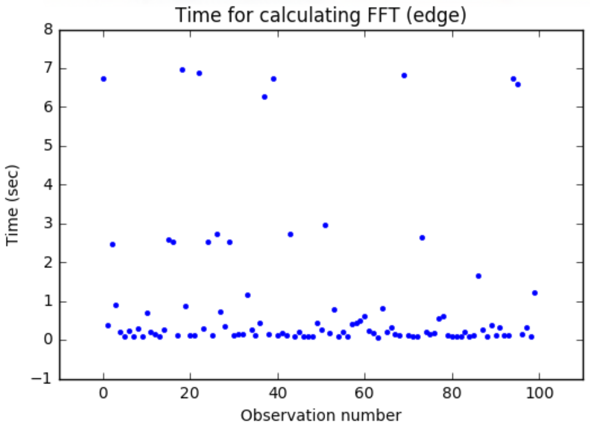
Lightscline
Data Reduction AI

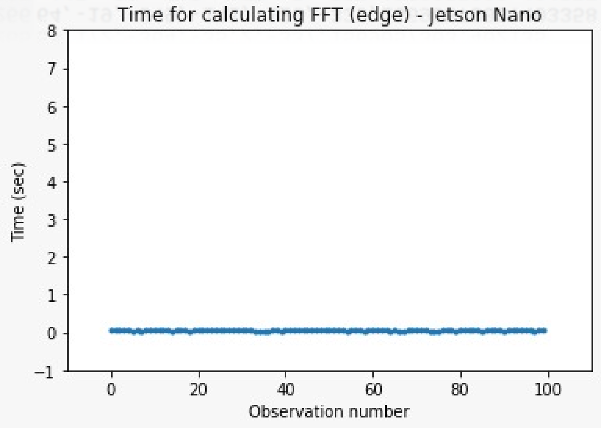
You can scale quickly across a variety of devices







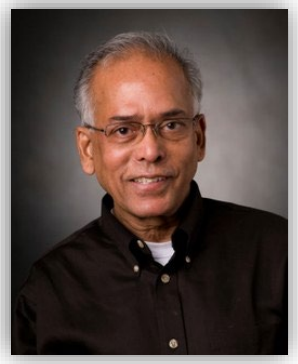




About us

Advisors

sme | Manufacturing ENGINEERING
30 UNDER 30
HONOREE



Ankur Verma

Co-Founder & CEO,

30 under 30 - Society of Mfg Engg,

PhD Candidate,

Ex-Fraunhofer, GM R&D

Ayush Goyal

Co-Founder & CTO,

MS Candidate in CS,

Ex-UHG

Prof. Soundar Kumara

Co-Founder & Chief Scientist,

Allen E. Pearce and Allen M. Pearce Professor of Industrial Engineering, Father of Smart Manufacturing

Todd Erdley

Founder at Videon,

Director & Portfolio Manager at Ben Franklin Tech Partners

Elizabeth Hay

Director, Launchbox HappyValley, Penn State

N241-025: Advanced Artificial Intelligence/Machine Learning Techniques for Automated Target Recognition (ATR) Using Small/Reduced Data Sets

DESCRIPTION:

ATR is the ability for a system or algorithm to recognize and identify targets, objects of interest or threats based on data obtained from sensors. In Navy mine countermeasure (MCM) operations, sensors collect data to identify and localize targets of interest in marine environments.

The Navy is interested in developing state-of-the art AI/ML ATR processing algorithms, or techniques to facilitate target detection and identification using smaller data sets to train the algorithms and perform ATR. The Navy's existing Minehunting systems collect data using forward-looking sonar, a pair of side-scan sonars, and a volume search sonar. Identification and localization of underwater objects is challenged by both a reliance on large, curated data from the onboard sensors that are needed to train and perform ATR and the amount of time required to conduct ATR operations. Current MCM ATR algorithms require large amounts of data (over 200 hours of acoustic video and 1,000-2,000 target images) to train the algorithms. This training data is quite costly to obtain because it must be collected in a variety of representative operational environments.

The proposed solution should demonstrate reduction in the amount of data required to train algorithms by an order-of-magnitude smaller without degradation to identification performance (Pid) and no increase in the Probability of false alarms (Pfa). If possible, the solution should incorporate advanced ML techniques such as One Shot, Multi Shot, Few Shot etc. as well as others that yield the benefits sought.

The ATR will be initially integrated into the Navy's Generalized ATR (GATR) framework to improve detection and classification performance. The capability could eventually be integrated into a towed body to support in-stride ATR.

N241-030: Acoustic Training Data Prioritization

DESCRIPTION:

Systems that detect and track submarines are migrating to AI/ML to improve the probability of detecting submarines and to limit the probability of false alerts. The current paradigm for training AI/ML is to use large sets of data. However, the cost associated with training AI/ML on large amounts of data is high and may not result in optimal training results.

There is not currently a commercial tool to assess how comprehensive a training set truly is, how much of the training data is effectively redundant, or whether some data over-represents unusual conditions. Additionally, there is not currently a tool that would enable researchers to determine a priori whether a newly collected data set would add useful diversity to the existing training data. This lack of tools for assessing training data for AI/ML algorithms results in a current state where all data is collected for training, resulting in possible excessive training costs as well as possible over-training to specific data which may not be representative of the full range of conditions in which the system will function during hostile tactical operations.

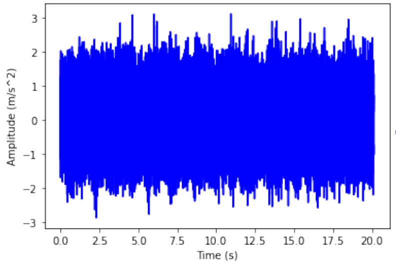
The Navy seeks a tool for analysis of acoustic data collected by undersea warfare systems to enable selection of data that is diverse, representative, and as small as practical for training of AI/ML algorithms.

Acoustic data used for detection of submarines is collected on arrays of transducers, whether towed line receive arrays such as the Multi-Function Towed Array, or hull-mounted source/receiver arrays such as the 576-element AN/SQS-53C hull-mounted sonar array. The signals from the transducers are formed into beams representing the acoustic environment as a function of bearing at any given point in time. Key characteristics of data sets will include both meta-data (e.g., season, latitude and longitude, time of day) and attributes of the data (e.g., volume reverberation levels, numbers of “clusters” associated with reflectors such as bathymetric features, marine entities, surface ships, submarines, and wakes).

The tool developed will need to demonstrate the training data prioritization technology which reduces the amount of training data used to allow the AI/ML algorithm(s) to maintain or improve performance. Performance of the system is determined by the Receiver Operating Characteristic (ROC) curve, where recorded data is run through the system to determine the number of true positives are achieved as a function of false positives.

Parallel workflow for preserving raw data

Raw data



Active real-time pipeline: 10x less training & inference costs

Lightscline workflow

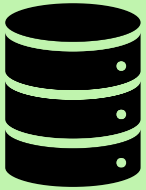
```
from lightscline.lightscline import LightsclineEdge
## Load data into Lightscline
ls = Lightscline(data=data,fs = SAMPLING_FREQUENCY)
## Reduce the amount of data by 70% of the original
ls.reduce_and_preprocess_data(per_reduction=70)
## Train the model
ls.train_model(verbose=True,n_iters = 1000)
## checking the results
ls.test_model()
```

Results

97% accuracy

Cold pipeline: store all raw data

Archival storage



Data is stored for non-real-time analysis

Lightscline
Data Reduction AI

Engagement with Lightscline

Off the shelf product:

End-to-end predictions

Customized product:

We will work with you to solve your use-case and modify the package accordingly

Customer journey

We accelerate your AI journeys by 10 years by helping data teams deploy 10x more, faster!

Pilot

(<5 licenses)

1 month

Working group

(5-20 licenses)

3 months

Enterprise contract

6 months

Lightscline
Data Reduction AI

Meeting

Customer deck

Try now