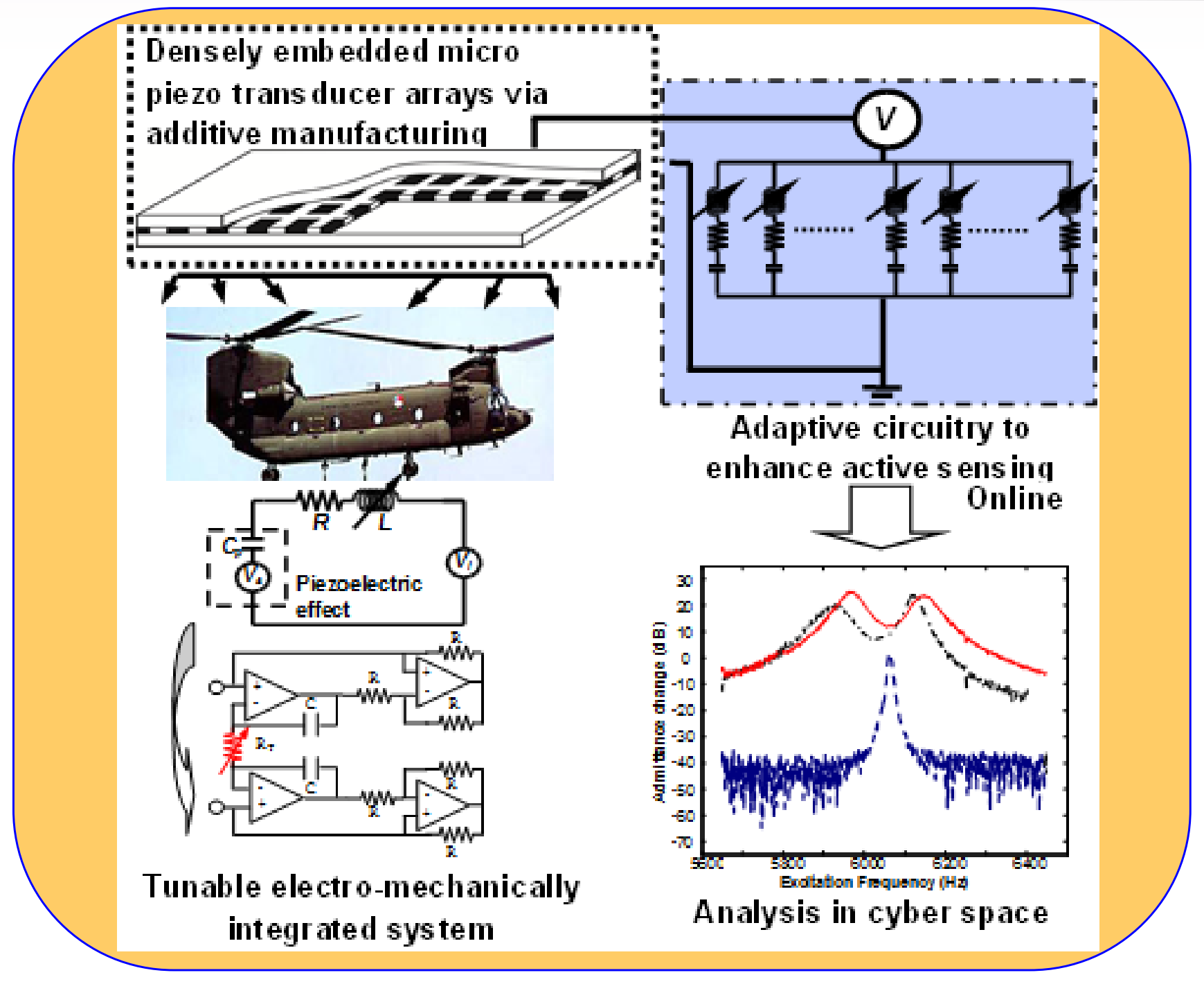
CPS/Synergy/Collab: Cybernizing Mechanical Structures through Integrated Sensor-Structure Fabrication – The Data Science Aspects

Faculty members: Yu Ding*, Jiong Tang **, and Chuck Zhang ***
Graduate research assistants: Ahmed Aziz Ezzat*; Imtiaz Ahmed*
*** Texas A&M University; ** University of Connecticut; *** Georgia Institute of Technology**

NSF Grant No.: 1545038(TAMU)/1544707(Uconn)/1544595 (GA Tech). Project Manager: Dr. Bruce Kramer, Program: CPS-CMMI

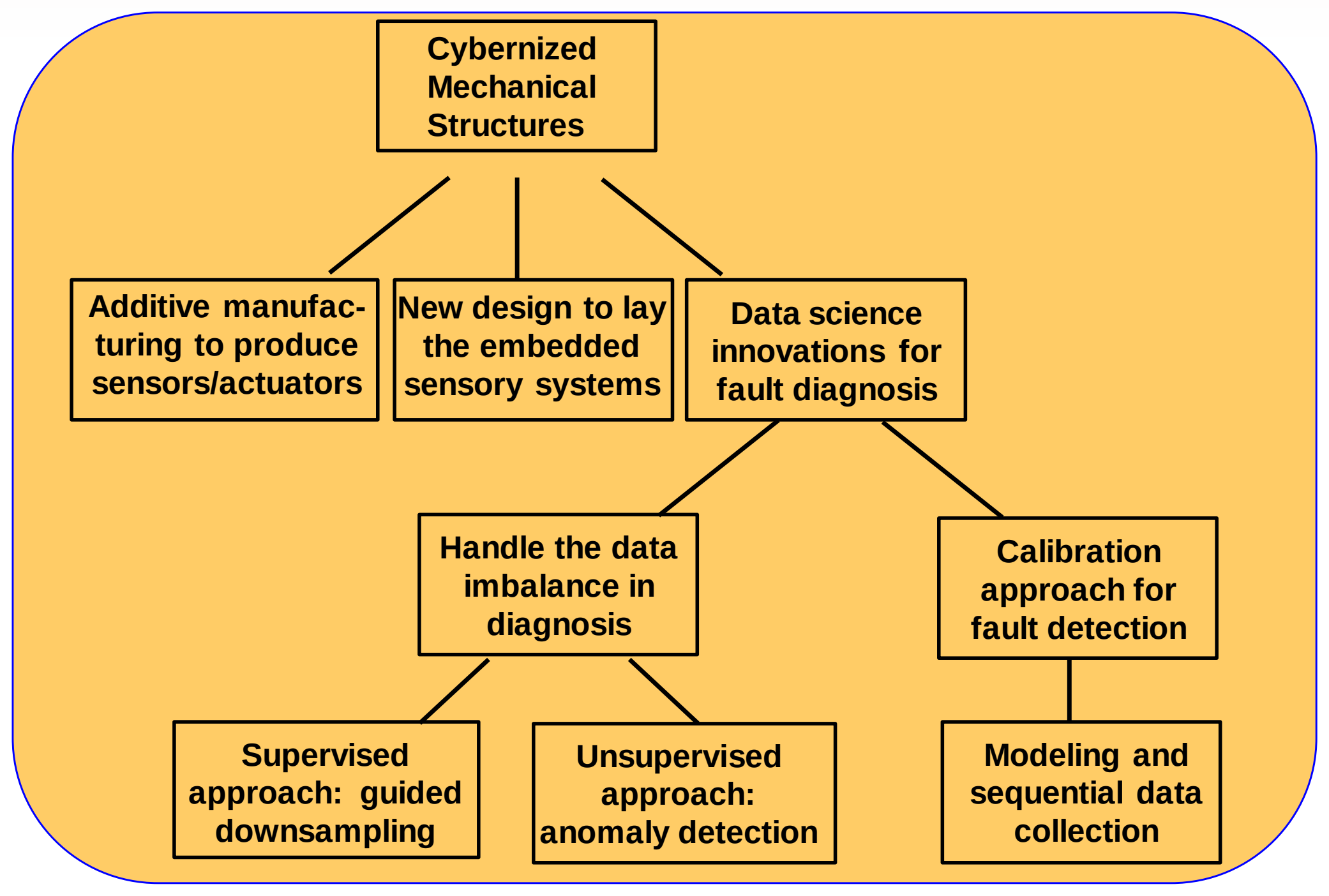
Motivation & Objectives

- Address the challenge of performing timely and accurate identification of structural faults to increase reliability and durability.
- Proposes to create a new framework of cybernizing structural system for self-diagnosis by taking advantage of additive manufacturing technology.
- Envisions that a civil structure system inserted with densely distributed active sensing elements analogous to a nerve system in a biological entity, enabling autonomous operation.
- The proposed system is cyber-physical in nature and needs data science innovations to convert raw sensor data to decision-facilitating information.


Densely embedded micro piezo transducer arrays via additive manufacturing
V
Adaptive circuitry to enhance active sensing
Online
R
L
C_p
V_s
Piezoelectric effect
V_i
R
C
R_T
Admittance change (dB)
Excitation Frequency (Hz)
Tunable electro-mechanically integrated system
Analysis in cyber space

Approaches

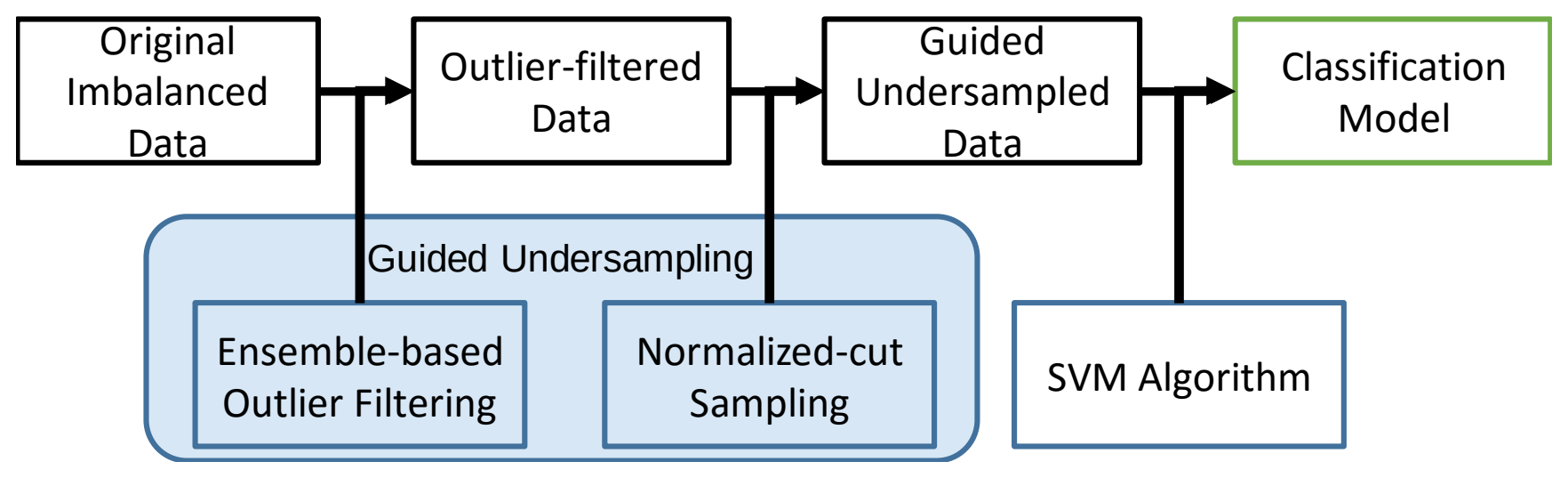
- Synthesis of new sensing modality, a duel-field electro-mechanical tailoring with tunable, integrated actuator/sensor units.
- Design of new fabrication by directly inserting sensing nerves inside of structure.
- Formulation of new data analytics – intelligent and robust inference to identify faults and to guide sensor tuning.
- On data science aspects, two parallel efforts are undertaken: an unsupervised learning approach and a calibration-based fault identification and detection, and the associated sensor tuning.


Cybernized Mechanical Structures
Additive manufac-turing to produce sensors/actuators
New design to lay the embedded sensory systems
Data science innovations for fault diagnosis
Handle the data imbalance in diagnosis
Calibration approach for fault detection
Supervised approach: guided downsampling
Unsupervised approach: anomaly detection
Modeling and sequential data collection

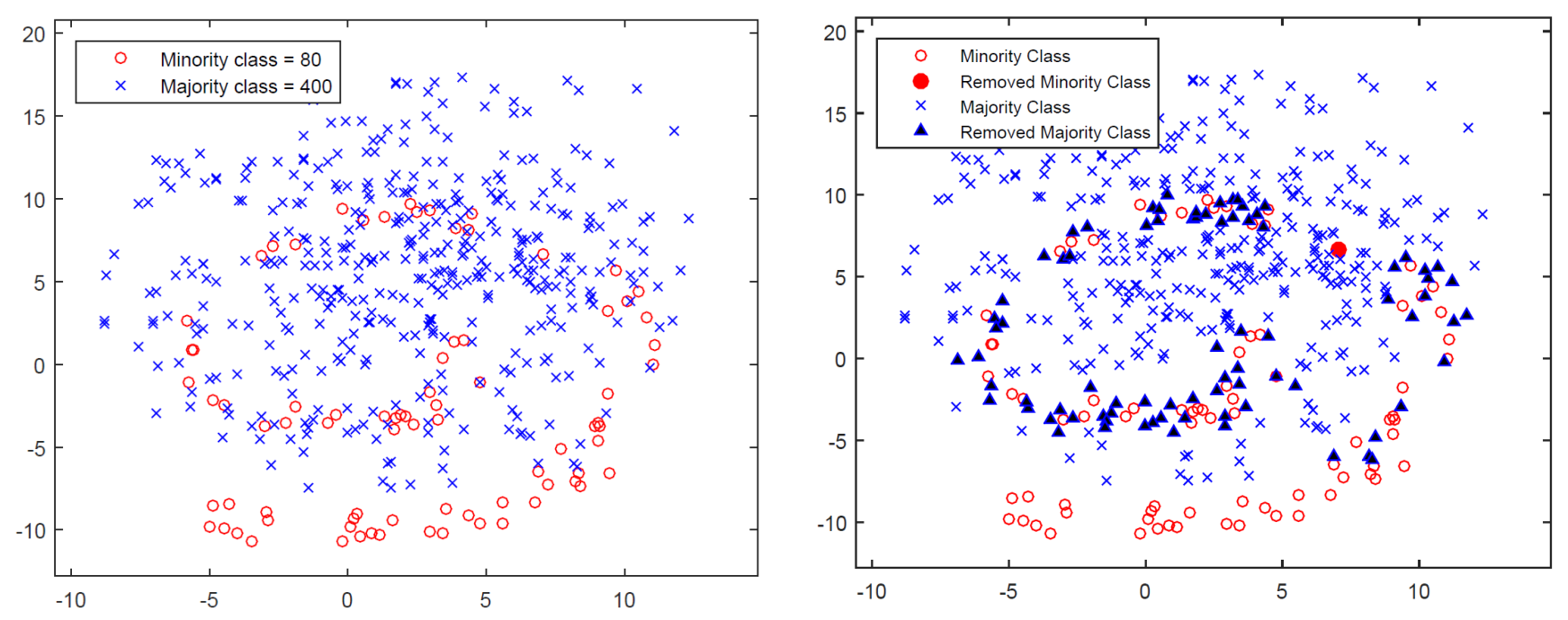
Research Efforts and Results

T1: Two-class classification method for handling data imbalance in diagnosis

- **Problem Statement**
  - When it comes to fault detection, the data amount of faulty data versus normal condition data is always imbalanced. The majority of he training data is normal, while a small amount of the training data is faulty. Supervised learning-based fault detection has an overwhelming under-detection problem, due to the data imbalance.
- **State of the Art**
  - There are three main schools of thought.
  - Cost-sensitive approach is to assign different costs to miss detections versus false positives.
  - Resampling approach is to up-sample the minority data or down-sample the majority data, to balance the data ratio.
  - Synthetic sampling is to artificially create additional Our minority data points to boost their presence.
- **Our Approach**
  - Our team develops a guided sampling approach, falling under the resampling school in general.



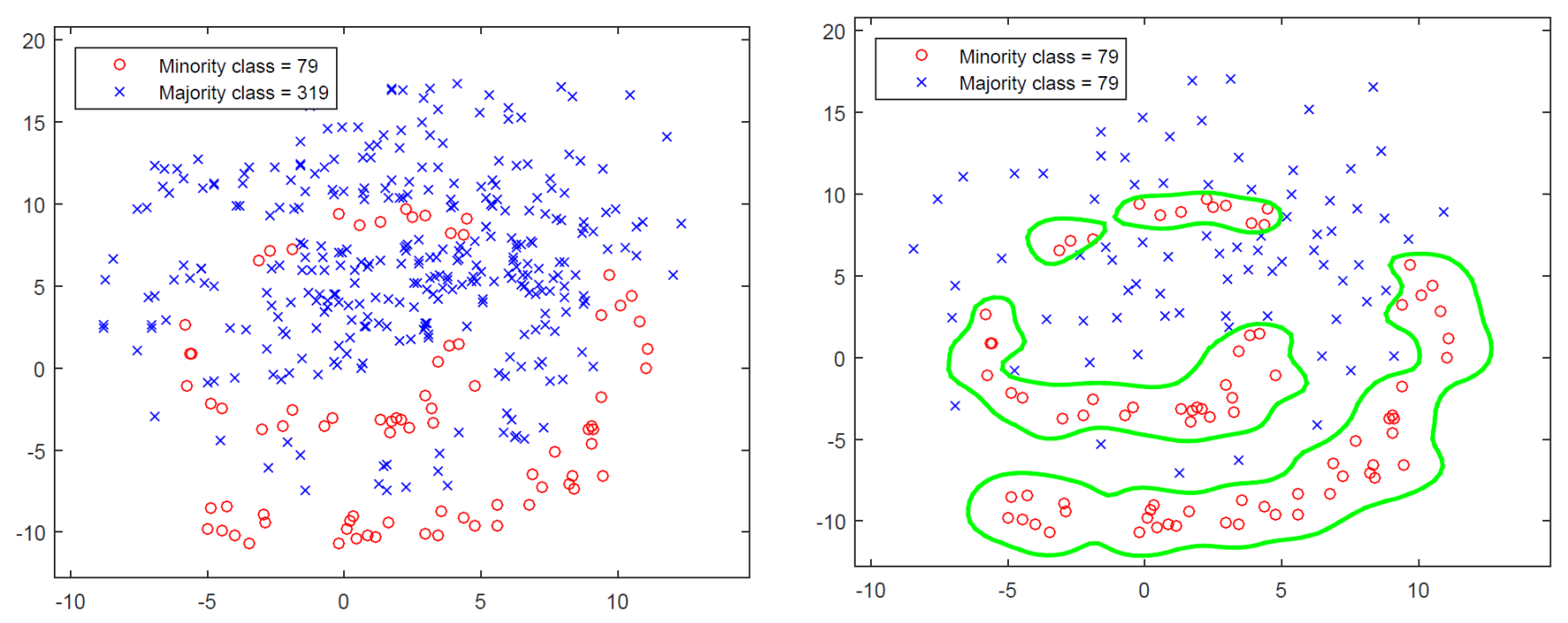
  - Our guided sampling approach involves identification and removal of outliers in the minority data points.



(a) Spiral-shaped imbalanced data

(b) Outlier instances identified by our ensemble-based outlier filtering

  - Our guided sampling approach uses a graph cut approach to down sample the majority data points with certain guarantee of uniformity in coverage after down-sampling.

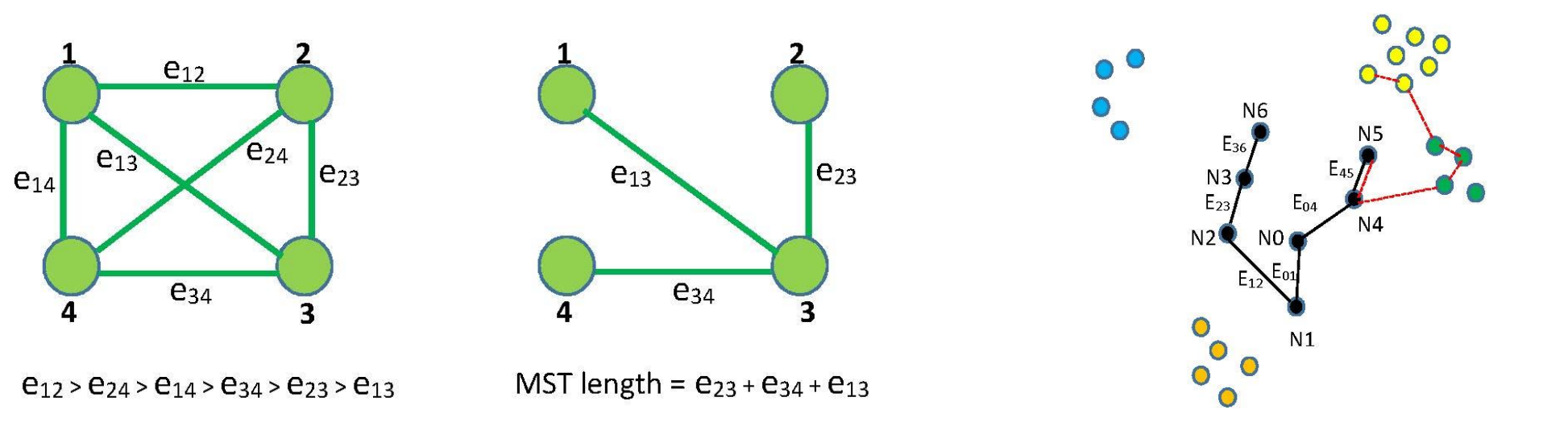


(c) Outlier-filtered data

(d) Normalized-cut under-sampling applied to the majority instances

T2: Unsupervised learning for outlier and anomaly detection

- **Problem Statement**
  - Faults could have happened without being known. Unsupervised learning is to flag the faulty data points in the existing dataset or signal the emergence of an anomaly as it is happening.
- **State of the Art**
  - One fundamental issue is to prove the distinctness of anomalous observations relative to the normal observations. The most commonly used dissimilarity metric is still the Euclidean distance and some of its statistical variants such as the Mahalanobis distance.
  - Euclidean distance and its variants lose effectiveness in high dimensional spaces, promoting the use of angle-based metric or the subspace methods.
- **Our Approach**
  - Our team proposes to use a minimum spanning tree (MST) to provide an approximation of geodesic distance in a high dimensional space and then use it as the (dis)similarity metric.
  - We model the data observations as a network of nodes where edges represent the Euclidean distance from one another. An anomalous node would be the one which is less connected to its neighboring nodes.
  - A MST is a measure that can capture the relative connectedness among nodes and approximate the geodesic dissimilarities among observations.



Minimum spanning tree is a subset of the edges in that connects all the vertices together, without any cycles and with the minimum possible total edge weight.

The total edge weight of the MST for node $N0$ is $W_{N0} = E_{01} + E_{12} + E_{23} + E_{04} + E_{45} + E_{36}$.

  - The numerical analysis in comparing our proposed method with 12 popular anomaly detection methods on 20 benchmark datasets demonstrates the superiority of the MST-based approach and supports the claimed merit.

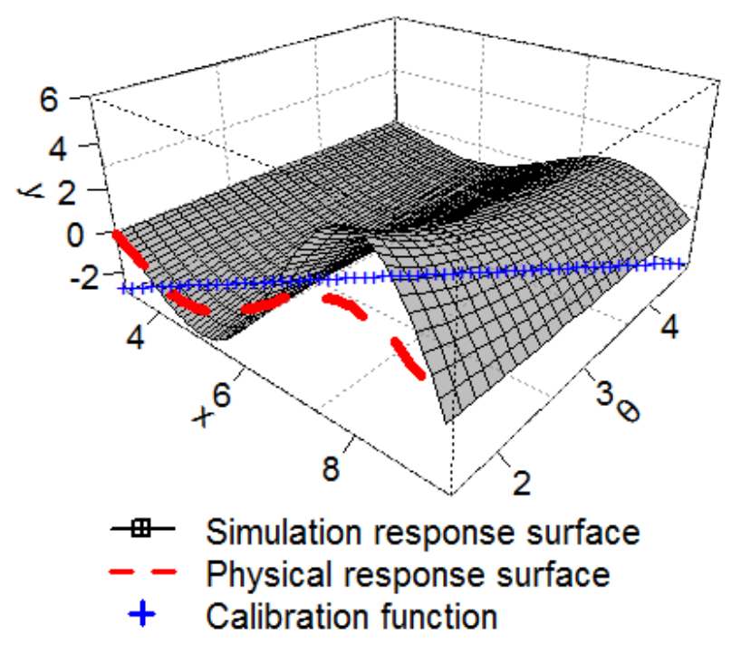
Performance \ Outlier Detection Methods	MST	COF	LDF	KNN	ODIN	LOF	KNNW	Simplified LOF	LoOP	INFLO	LDOF	Fast ABOD	KDEOS
Better(no. of datasets giving uniquely best result)	5	0	3	1	1	0	1	0	0	0	0	0	0
Equal(no. of datasets equal to the best result)	9	6	6	2	2	3	2	2	2	2	0	2	1
Close(no. of datasets, within 20% of the best result)	4	9	6	5	4	10	5	4	5	5	5	3	0
Worse(no. of datasets, not within 20% of the best result)	2	5	5	12	13	7	12	14	13	13	15	15	19

T3: Calibration approach: modeling and sequential data collection

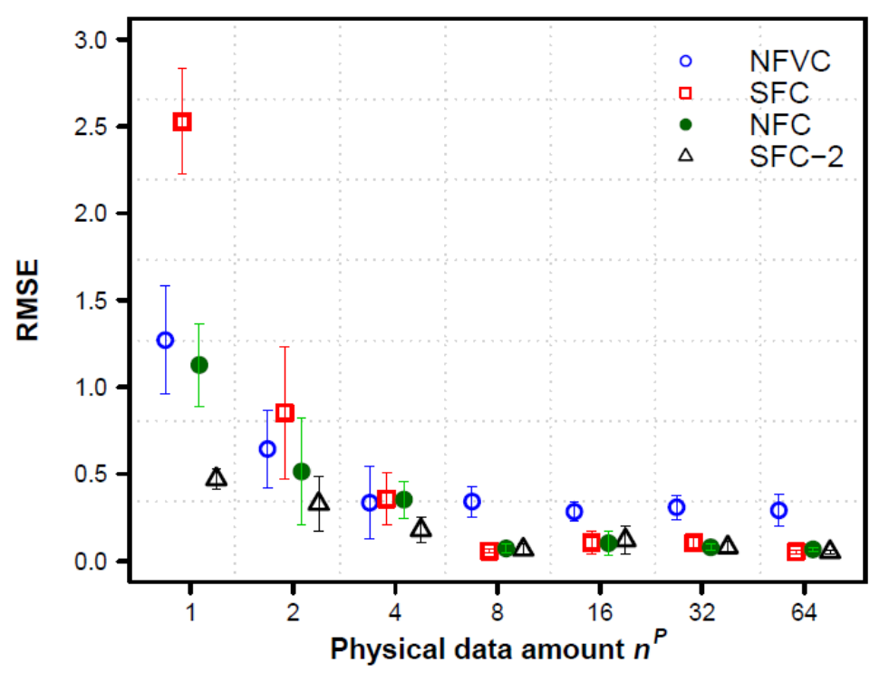
- **Problem Statement**
  - Piezo-electric impedance/admittance data, whether sensor- or model-based, can inform us about structural faults. One promising research direction is data fusion from different sources using a calibration approach.
- **State of the Art**
  - Calibration of computer models entails the estimation of a set of calibration parameters that best align a model's prediction and the underlying physical system. Widely used methods assume a parametric form for the calibration parameters and then minimize a distance metric between the model outputs and the observational data.
  - Moreover, calibration is known to be sensitive to both the quality and quantity of the integrated data streams. Sequential experimental design for a better calibration performance is an emerging research topic.
- **Our Approach**
  - We propose a general, nonparametric estimation of the calibration parameters using a Reproduced Kernel Hilbert Space approach:

$$\hat{\theta} = \underset{\theta_j \in \Theta}{\operatorname{argmin}} \frac{1}{n} \left\{ y_i^p - y_i\big(x_i, \theta(x_i)\big) \right\}^2 + \lambda \sum_{j=1}^{q} \left|\left|\theta\right|\right|^2$$

  - Through extensive numerical analysis, we conclude that sequential design methods applied separately to the integrated data streams could be more beneficial to the learning framework than other more sophisticated techniques.



Simulated dataset to test the calibration approach.

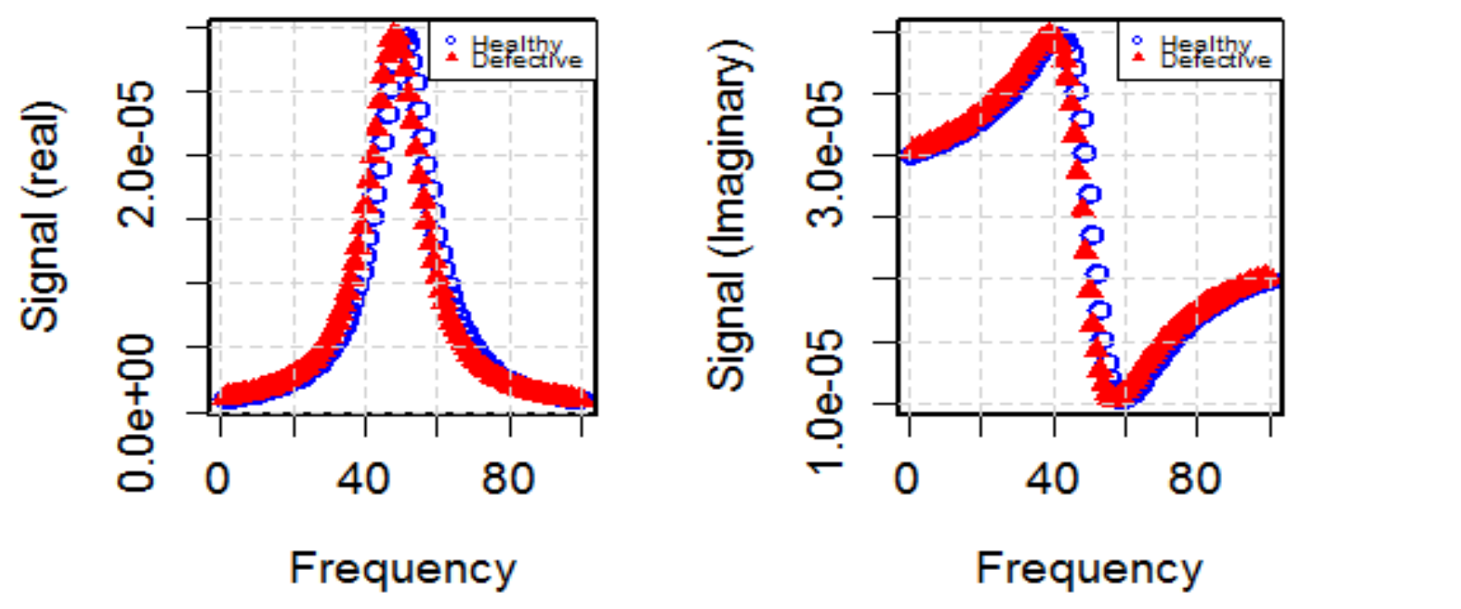


Predictive performance after using our calibration formulation. Design sequentially generated using our proposed approach.

  - Numerical analysis on a simulated set of data shows that the final predictive performance of a calibrated computer model using our calibration formulation and for which the design is sequentially generated using our proposed approach, outperforms existing methods in the literature.

T4: Calibration formulation for detection structural faults

- **Problem Statement**
  - Structural fault detection entails the identification of the location and severity of faults in a structure. Model-based methods such as Finite Element Analysis (FEA) can be used in conjunction with online measurements to detect changes in local material properties. Nevertheless, the large number of FEA segments that are susceptible to fault occurrence, compared to the number of available sensor observations, makes the problem severely under-determined.
- **State of the Art**
  - Focus has been recently directed towards data-driven methods to overcome the problem and estimate optimal fault parameters.
  - Given the large number of combinations of possible fault locations and severity levels, the parameter space could be massive and a large number of computationally intensive FEA runs deems traditional approaches for estimation computationally impractical.
- **Our Approach**
  - We propose a combination of a pre-screening method and a calibration approach to accurately estimate the fault parameters.
  - First, observational admittance/impedance data are collected via a sequential experimental design strategy.



Healthy (blue) vs. Defective (red) Admittance/Impedance observational data

  - Then, we obtain preliminary estimates of "most likely" locations and severity levels via linear approximation of the relationship between the observational data and the so-called sensitivity matrix.

$$\boldsymbol{\Delta Y = D_k s_k \rightarrow \widehat{D}_k = av\left(\frac{\Delta Y(\omega_j)}{s_{jk}}\right) \rightarrow SI_k = \mathbf{arcsin}\left(\frac{s_k^T \Delta Y}{|s_k|.|\Delta Y|} - 1\right)}$$

  - FEA runs are then performed in the vicinity of the selected candidates. The number of runs to be performed can be tuned according to the computational capability. If partial information is to be gathered, sequential design strategies could be used to solicit the most informative subset of data and then, a surrogate model interpolator is fit to the FEA data.
  - Both the observational and the model-based data are then fed into our proposed calibration framework, as described in T3 (where $\boldsymbol{\theta} \coloneqq \boldsymbol{D}$), to estimate the exact location and severity level of the fault.

Benefits

- Potentially lead to paradigm-shifting progress in structural health monitoring and self-diagnosis.
- Data science advancement to enable and facilitate accurate and robust decision-making.
- Contribute to student multidisciplinary training.

Ongoing and future work

- Non-negative matrix factorization approach, together with the minimal spanning tree, to boost the detection capability in unsupervised anomaly detection.
- Tailor the calibration formulation for structural health monitoring.
- Integrate the data science effort with the additive manufacturing effort and the sensor layout effort.

Publications related to this project

- A. A. Ezzat, A. Pourhabib, and Y. Ding, "Sequential design for functional calibration of computer models," *Technometrics*, accepted, September 2017.
- Sung, K.-S., E. Moreno-Centeno, Y. Ding, "Guided undersampling using ensemble filtering and normalized cuts for imbalanced data Classification," *Journal of Machine Learning Research*, in revision, 2017.
- A. Pourhabib, R. Tuo, S. He, Y. Ding, and J.Z. Huang, "Local calibration of computer experiments," *Journal of the American Statistical Association*, revised and re-submitted, August 2017.
- I. Ahmed, A. Dagnino and Y. Ding, "Unsupervised anomaly detection based on minimum spanning tree approximated distance measures and its application to hydropower turbines," *IEEE Transactions on Automation Science and Engineering*, submitted, Sept 2017.

For more information, contact Dr. Yu Ding (yuding@tamu.edu, 979-458-2343) or Dr. Jiong Tang (jiong.tang@uconn.edu, 860-486-5911)