
Machine Learning-based Anomaly Detection with Magnetic Data

Peetak P. Mitra*

Computational Science and Engineering Program,
Total EP Research & Technology USA.,
Houston, TX 77002

Denis Akhiyarov

Computational Science and Engineering Program,
Total EP Research & Technology USA.,
Houston, TX 77002
denis.akhiyarov@total.com

Mauricio Araya-Polo

Computational Science and Engineering Program,
Total EP Research & Technology USA.,
Houston, TX 77002

Daniel Byrd

TOTAL EP Research & Technology USA.,
Houston, TX 77002

Abstract

Pipeline integrity is an important area of concern for the oil and gas, refining, chemical, hydrogen, carbon sequestration, and electric-power industries, due to the safety risks associated with pipeline failures. Regular monitoring, inspection, and maintenance of these facilities is therefore required for safe operation. Large standoff magnetometry (LSM) is a non-intrusive, passive magnetometer-based measurement technology that has shown promise in detecting defects (anomalies) in regions of elevated mechanical stresses. However, analyzing the noisy multi-sensor LSM data to clearly identify regions of anomalies is a significant challenge. This is mainly due to the high frequency of the data collection, mis-alignment between consecutive inspections and sensors, as well as the number of sensor measurements recorded. In this paper we present LSM defect identification approach based on machine learning (ML). We show that this ML approach is able to successfully detect anomalous readings using a series of methods with increasing model complexity and capacity. The methods start from unsupervised learning with "point" methods and eventually increase complexity to supervised learning with sequence methods and multi-output predictions. We observe data leakage issues for some methods with randomized train/test splitting and resolve them by specific non-randomized splitting of training and validation data. We also achieve a 200x acceleration of support-vector classifier (SVC) method by porting computations from CPU to GPU leveraging the cuML RAPIDS AI library. For sequence methods, we develop a customized Convolutional Neural Network (CNN) architecture based on 1D convo-

*Work conducted during internship at TOTAL EP Research & Technology USA. He is affiliated with the Department of Mechanical and Industrial Engineering, University of Massachusetts.

lutional filters to identify and characterize multiple properties of these defects. In the end, we report scalability of the best-performing methods and compare them, for viability in field trials.

1 Introduction

Pipeline integrity is required for safe operation and it is ensured by regular monitoring, inspection, and maintenance of pipeline facilities. The undetected defects can cause significant damage to the day-to-day operations and the environment. Traditionally, intrusive inspection devices are used to detect the defects in the pipe material. These devices measure magnetic or ultrasonic signal in the surface and can disrupt or block the production. Hence, an non-intrusive external inspection is preferred to internal inspection.

Due to the induced mechanical or residual stresses, e.g. around defects, the magnetic behavior of the pipeline material changes locally due to the Villari magneto-restriction effect [1], [2]. Non-intrusive Large Stand-Off Magnetometry (LSM) technology has emerged as a promising tool in detecting small changes in the magnetic readings around defects associated with the regions of elevated stresses [1], [2], [3]. In addition to localizing the effects spatially, LSM’s stress quantification has shown to provide insights into the properties of the defect in some cases [3]. A simplified schematic of the LSM is shown in Figure 1 with a multi-axial, multi-sensor alignment. The process of LSM involves the use of an array of sensors installed on an instrument to perform periodic observations of the asset’s changes in magnetic signature [4]. Empirical correlation functions are used to differentiate the defect signals from normal regions. However, these functions have a number of limitations, such as requirement for multiple scans, reduced sensitivity to smaller defects, low localization precision, and ignoring some defect properties.

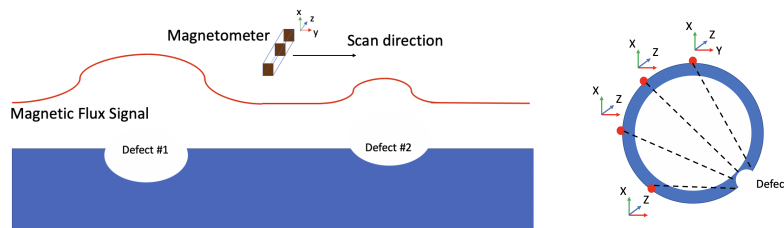


Figure 1: A schematic for the Large StandOff Magnetometry shows the essential components for this technology. The magnetometer instrument typically has a multi-sensor configuration [as shown in the inset figure], and gathers data in all three spatial dimension. When the magnetometer passes over a damaged/defect region, it records a different magnetic flux compared to the other sections of the pipeline. Due to interferences, the signals might be noisy and therefore to robustly identify anomalies becomes a challenge.

The use of Machine Learning (ML) for anomaly detection has shown great promise (e.g. [5, 6, 7]), mainly with regards to sequential data and time-series applications (e.g. [8, 9]). Some of these methods are purely data-driven techniques identifying the occasional outliers, and others are hybrid methods incorporating feature engineering as well as prior information into the system that aim to robustly characterize the anomalous behavior [8]. Scalable ML frameworks for anomaly detection ([10, 11]) have grown in importance in the recent years, for instance the Python library *pyod* [12]. ML-based structure health monitoring for asset integrity is growing in usage and has been successfully applied to civil and aerospace engineering applications [13, 14].

1.1 Our Contribution

In this work, the effectiveness of classical machine learning methods from scikit-learn [15] and *pyod* [12] Python-libraries, and customized Convolutional Neural Network (CNN) architectures from Keras [16] library for identification and characterization of anomalous magnetic field reading on the LSM experimental dataset are explored. The challenges in using multi-modal, multi-sensor information are discussed, including sequence alignment and remedies suggested. Multi-task classification is successfully achieved using a customized CNN architecture with a Conv1D filter. Scalability of these methods from lab experiments to field trials are discussed and slow performing methods are accelerated using algorithmic advancements.

2 Dataset

The datasets were obtained from controlled experiments conducted by a research facility. These experiments were conducted using multiple instruments and multiple sensor inputs. These instruments collected data in three axial directions, X, Y and Z for each sensor thereby providing multiple channels for dataset collection. The pipes were scanned with 1 mm resolution. Defects and welds were manufactured to mimic realistic field scenarios. These datasets were obtained at various circumferential positions to the defect. We present the details about two representative defects that we studied and are summarized below, in Table 1

Table 1: Defect characteristics for limited field trial run

| Defect number | Location | Volume* | Depth* | Width* |
|---------------|----------|---------|--------|--------|
| D1 | 2 ft | 0.2 | 0.77 | 0.45 |
| D2 | 76 ft | 0.6 | 0.62 | 1 |

* These quantities have been normalized

Due to the nature of data collection using spatially distant sensors, alignment of the data channels was required. We used the Dynamic-Time-Warping (DTW) algorithm to re-align the datasets such to reduce alignment error to discover anomalies based on the non-aligned consecutive sensor readings with shifts in position. Typically a DTW algorithm has a quadratic space and time complexity [17, 18] as it measures the similarity between the two sequences in the matrix form and calculates an optimal match between them. As one scales the warping with more data relevant to field applicability of this method, the cost to compute scales non-linearly and becomes computationally intractable. Therefore, we use an approximation of the DTW algorithm, known as FastDTW [19] which has a linear space and time complexity. FastDTW makes use of a multilevel approach that recursively projects a solution from a coarser resolution and refines the projected solution, thereby providing the necessary speed-up [19]. More details are in the Appendix section. Once the different data channels are aligned we apply ML techniques for anomaly detection.

3 Machine Learning Results

Anomaly detection for sequential datasets using Machine Learning (ML) has shown to perform well for a variety of applications. In this section, we start by applying off-the-shelf packages such as pyod and scikit-learn to our datasets, pre-processed with the FastDTW alignment method. We first review the "point-based" unsupervised methods, following which challenges with data leakage for supervised learning methods are discussed and solutions provided. Finally, the customized CNN architecture is introduced for multi-task classification.

3.1 Unsupervised - Point based Learning

For unsupervised point-based methods we use the pyod v0.8 [12] package. Within pyod, we apply several anomaly detection models. Models are first trained on an original pipe scan and returns a baseline score. These trained networks are then tested on the actual LSM data, thereby determining their performance. Out of methods we tried, only k-nearest neighbor (K-NN) detector works reasonable well. Results are shown in Figure 2, and it shows regions of higher outlier scores in defect vicinity, marked by the vertical dotted lines.

The unsupervised learning point-based methods were computationally cheap to investigate, therefore we used these methods to not only determine their applicability to our data, but also to better understand how we may leverage the multi-modal aspect of data recording (namely feature engineering). Data configuration includes, combining or splitting multiple parameters such as pipe sections, sensor channels, sensor axes, and sensor pipe circumferential positions. We obtained the outlier score plots, shown in Figure 2, by individually studying each sensor input. In order to determine the sensitivity of the sensor inputs, we performed multiple experiments combining adding all data channels together, studying multiple angles together, transforming the data using normalization and spectral techniques. We observe that while the k-NN model is very robust in identifying anomalies with a higher outlier scores, studying multiple angles together has shown to outperform other feature engineering options, clearly distinguishing regions of anomaly with low false positives as seen in Figure 2. This can be attributed to a two-fold reason. 1. The orientation of the LSM instrument when positioned at different angles with respect to the defect, and 2. The aggregation of the data from all these channels is able to robustly identify anomalous magnetic field reading compared to a channel studied in isolation.

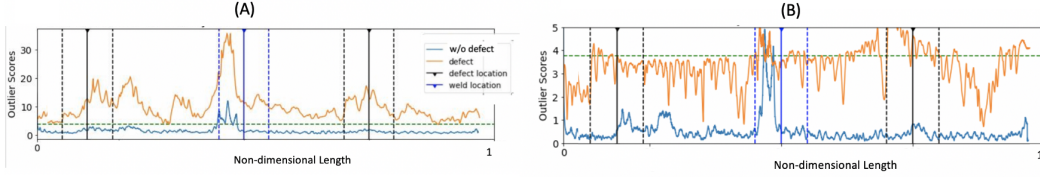


Figure 2: We show the outlier scores after using the k-NN point based method, trained on the initial scan. The regions of defect are marked by vertical black solid lines, with the corresponding dashed lines indicating the sensitivity of the instrument. The vertical blue lines indicate the regions of weld, with the same estimate for the region of magnetic field distortion. Panel (A) shows results when multiple angles are used together, and shows clear regions of higher outlier scores (denoting defects), with limited false positives. Panel (B) shows the outlier scores for individual sensors and is comparatively much more noisy and although it shows regional spikes in the defect region, it is not clearly distinguishable compared to Panel (A)

Table 2: The supervised classification methods with their Accuracy and MCC scores

| Method | Acc. Score - Train | Acc. Score - Test | MCC - Train | MCC - Test |
|----------------|--------------------|-------------------|-------------|------------|
| SVC | 0.78 | 0.87 | 0.56 | 0.75 |
| Decision Tree | 0.98 | 0.84 | 0.97 | 0.70 |
| MLP-Classifier | 0.80 | 0.83 | 0.68 | 0.73 |

Therefore, in the following sections we present our results based on aggregating multiple angles together as a single input to the ML model.

3.1.1 Supervised - Point based Learning

The dataset contained a very limited number of defects, which is a challenge for supervised learning, thereby creating a major class imbalance. The labelling was achieved by creating a binary mask in the regions with defects. The defect location and nearby region of distortion of magnetic field is found about 3 feet from the actual location in either direction based on the instrument sensitivity. For the spatial location based masking, we observed issues related to data leakage [refer to Figure 8 in the Appendix section], which has the potential to cause unintentional and non-intuitive data leakage during the randomized test train split. In order to avoid this, we masked an entire defect region as our test dataset, and the rest was used for training. As a result, our training/test set includes 35000/1000 samples with the non-defect label, and 10000/1000 samples with the defect label. The class imbalance was addressed by adding class weights, in proportion to the number of samples in each class, to our learning models.

To quantify the model performance, we report the confusion matrix (in the Appendix section, Figure 7) accuracy scores and the Matthews Correlation Coefficient (MCC). The MCC is a robust correlation coefficient for binary classification problems with imbalanced datasets [20]. A score of +1 indicates perfect correlation, -1 indicates total disagreement, and 0 indicates no better than random predictions [20]. Mathematically, MCC is defined as

$$|MCC| = \frac{TP.TN - FP.FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (1)$$

where, TP , is true positive, TN is true negative, FP is false positive, FN is false negative.

We used the scikit-learn package [15] to build a pipeline with various ML methods including Support Vector Classification (SVC), Logistic Regression, Decision Tree, k-NN, MLP-Classifier, Gaussian Process Regression, and we report the best performing models in Table 2. Among all models SVC, Decision Tree and MLP Classifier are the three best performance methods in terms high accuracy on test and train datasets with low false positives.

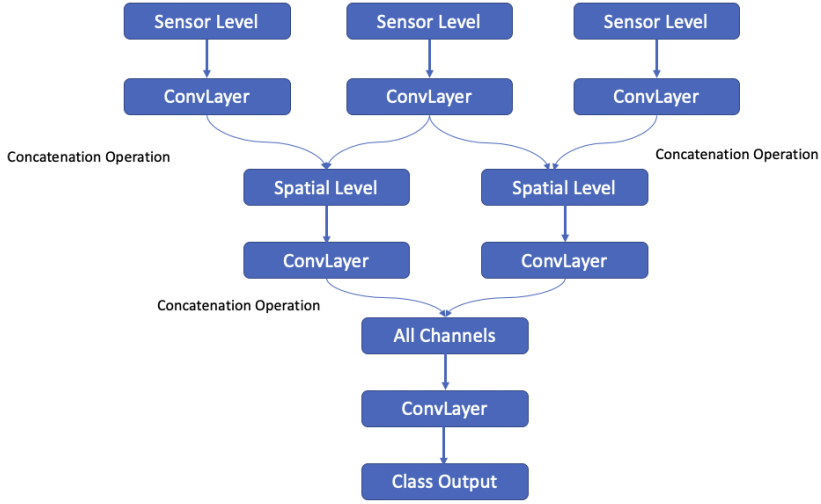


Figure 3: A schematic of the customized 1D-Convolutional Neural Network architecture, showing concatenation points at spatial dimension level. Each concatenation operation corresponds to one dimension of the multi-dimensional data

3.2 Supervised - Sequenced based Learning

The "point-based" methods both in an unsupervised and supervised learning formulation have shown promising results in identifying local defects or the defect labels successfully. However, these models are not able to characterize the defect properties such as predicting the volume or the depth of the anomaly among other features. We identify Convolutional Neural Network (CNN) as a good candidate for sequence based learning and also for application to multi-task classification. CNN based architectures have previously been successfully used to do such multi-task classification [21, 22]. We propose using sequence based CNN, to detect the spatial variation of the magnetic signal along the defect. CNN-based networks can extract spatial features ([23, 24]) from the magnetic signal. Upon building this sequence based learning model successfully, the model is extended to have multi-outputs for characterization of the defect region [21].

The challenge in using CNN for this binary classification task, is that the method is restricted in using 1D Convolutional filters (Conv1D) since there is not long enough sequence except along the profile dimension of the data. Other dimensions are limited due to acquisition constraints due to certain number of consecutive measurements, which prevents applying the convolution operator along them. The dataset has the dimension, $M \times N \times P$, where M is the number of points in the profile (based on the sensor resolution) along which we apply Conv1D filter, N , is the number of channels which we concatenate in the network in the first level, and P is the total number of angles we concatenate in the network in the second level. In order to use the current dataset as a sequence, we break it into a sub-sequence of 100 profile points which is long enough to capture a defect region. The schematic of the CNN based architecture is shown in Figure 3 where we use concatenation operations at the spatial levels to first study the effects of the multi-channel input separately, and post-concatenation operation study them together. Each Conv Layer is composed of Conv1D-filter -> Batch Normalization Layer -> Dropout [p varied between 0.2 to 0.4] layers. The total number of trainable parameters were 1,249,091. We used the Nesterov-Adam optimizer [25], with an initial learning rate of 0.001 with a decay in a plateau learning region. Binary crossentropy was used as a loss function. One of the defect region as were hold out as a test case, and the rest of the dataset is used for training the model. Figure 4 shows the probability scores from the loss layer overlaid on the actual score (mask), along the profile. It can be observed that the Conv1D network is able to accurately represent all defects successfully.

3.2.1 1D-Convolutional Neural Network for Multi-Output Predictions

To characterize the defect properties such as the volume and depth [see Table 1], previous described CNN architecture is extended to include multiple outputs, for depth and volume prediction as well as

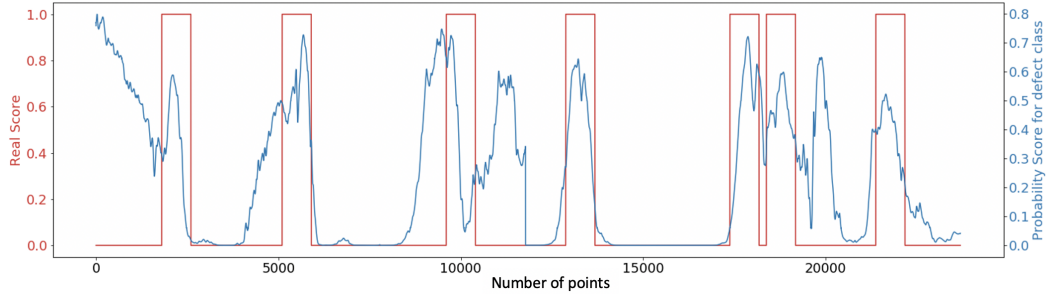


Figure 4: The plot shows the probability score from the cross entropy loss plotted as a function of the number of points and overlaid with the real score, shows spikes in probability at the defect locations for all of the defects, including the hold-out test dataset. The few false positives can be addressed by training on more data.

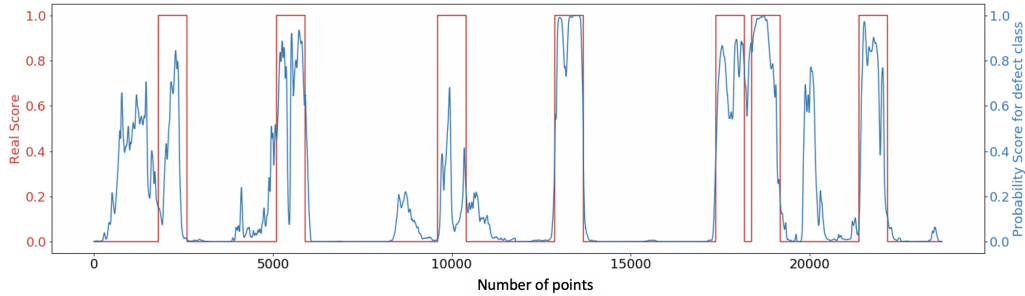


Figure 5: Similar to Figure 4, we show the probability scores from the final layer of the CNN architecture. We see the network is able to classify the defect regions, marked in red, successfully.

the mask (binary classification). The prediction accuracy for the defect characteristics with reasonable accuracy close to 90% for the depth and volume, as well as the mask [refer to Figure 5]. These results can be used to build a ML model that maximizes the information from datasets like in this study with limited labels for defect by using other defect properties to improve predictions (similar outcome to [21]).

4 Scalability

A limited Proof-of-Concept (PoC) in using Machine Learning for LSM data anomaly detection using multi-modal, lab experimental dataset were developed. However, the biggest adoption for industrial adoption of this ML approach would be the issue of scalability to large datasets coming from the field. For example, the lab experiments have about 10000 profile points, while 1-100 million profile points are captured in field acquisitions. Therefore it is important to address the time complexity performance of all these methods. A comparison of the different successful methods described in the previous section (and is shown in Table 4 in the Appendix section) identifies the slow performing methods, mainly SVC.

4.1 Algorithmic enhancement of slower methods

SVC is a robust method especially for binary classification tasks [15], however the current design of the algorithm in scikit-learn library is optimized for CPU without parallelization. Therefore as can be seen from Table 3 SVC performs poorly when facing large datasets. RAPIDS AI cuML library ([26]) is an open-source software that enables efficient use of GPUs via CUDA and numba [27] for end-to-end machine learning pipelines. For large field datasets (bigger than 100000 data points), a 200x speedup over the scikit-learn implementation of SVC is achieved as seen in Table 3.

Table 3: Observed speedup of SVC using RAPIDS-AI library. All times in seconds

| Data points | scikit-learn | RAPIDS-AI | Speedup |
|--------------------|---------------------|------------------|----------------|
| 10000 | 8.76 | 2.90 | 3 |
| 100000 | 751 | 3.75 | 200 |
| 1000000 | 18274 | 98 | 186 |

5 Conclusions and Future Work

Large Stand-Off Magnetometry (LSM) is a promising approach that records magnetic fluxes for non-destructive and non-intrusive inspection of pipelines. We developed a proof-of-concept with limited lab experimental data, that leverages state-of-the-art machine learning algorithms for the purpose of identifying anomalous magnetic flux readings associated with the regions of defect. These datasets are multi-modal, multi-sensor sequential data that need several pre-processing steps before applying the ML methods. An approximate multi-sequence alignment technique based on Dynamic Time Warping (DTW) scalable to real-world applications was implemented. A sequence of models with increasing levels of complexity to refine on defect detection capabilities were studied. "Point-based" methods in an unsupervised and supervised learning setting performed well with high accuracy on test/train data and low false positives. A customized Convolutional Neural Networks (CNN) with 1D convolutional filters was built that reflects the dimensionality of magnetic measurements. In addition to identifying anomalies, defect properties were characterized by using a multi-output CNN network with reasonable success. The scalability of these models from research lab experiments to field data showed some slow performing methods, and a 200x speedup of RBF-kernel SVC was achieved using RAPIDS AI cuML library on 32GB V100 GPUs. One way to improve the model performance would be to augment the datasets using the principles of magneto-restriction.

6 Acknowledgements

We thank the HPC team at Total EP Research and Technology USA for providing support and Deepwater RD team for providing steering and domain expertise.

References

- [1] EA Fnaiech, M Munschy, S Corbineau, M Marzin, P Rohart, and S Takillah. Large stand-off magnetometry (lsm) for buried pipeline inspection.
- [2] C Onuoha, E Pozniak, S McDonnell, V Shankar, T Lich, M Jawed, et al. Use of large standoff magnetometry in pipeline integrity investigations. In *CORROSION 2020*. NACE International, 2020.
- [3] Tianzong Xu. Understanding quantitative performance of large standoff magnetometry in detecting live gas pipeline anomalies with stress estimation. In *International Pipeline Conference*, volume 51869, page V001T03A020. American Society of Mechanical Engineers, 2018.
- [4] Shamus McDonnell, Chukwuma Onuoha, Eric Pozniak, and Vignesh Shankar. Identifying stress concentrations on buried steel pipelines using large standoff magnetometry technology. In *International Pipeline Conference*, volume 51869, page V001T03A003. American Society of Mechanical Engineers, 2018.
- [5] Terran Lane and Carla E Brodley. An application of machine learning to anomaly detection. In *Proceedings of the 20th National Information Systems Security Conference*, volume 377, pages 366–380. Baltimore, USA, 1997.
- [6] Taeshik Shon and Jongsub Moon. A hybrid machine learning approach to network anomaly detection. *Information Sciences*, 177(18):3799–3821, 2007.
- [7] Taeshik Shon, Yongdae Kim, Cheolwon Lee, and Jongsub Moon. A machine learning framework for network anomaly detection using svm and ga. In *Proceedings from the sixth annual IEEE SMC information assurance workshop*, pages 176–183. IEEE, 2005.
- [8] Mohiuddin Ahmed, Abdun Naser Mahmood, and Md Rafiqul Islam. A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems*, 55:278–288, 2016.
- [9] Leman Akoglu and Christos Faloutsos. Anomaly, event, and fraud detection in large network datasets. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 773–774, 2013.
- [10] Nikolay Laptev, Saeed Amizadeh, and Ian Flint. Generic and scalable framework for automated time-series anomaly detection. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1939–1947, 2015.
- [11] Karima Echihabi, Kostas Zoumpatianos, and Themis Palpanas. Scalable machine learning on high-dimensional vectors: From data series to deep network embeddings. In *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*, pages 1–6, 2020.
- [12] Yue Zhao, Zain Nasrullah, and Zheng Li. Pyod: A python toolbox for scalable outlier detection. *arXiv preprint arXiv:1901.01588*, 2019.
- [13] Keith Worden and Graeme Manson. The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):515–537, 2007.
- [14] Charles R Farrar and Keith Worden. *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, 2012.
- [15] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [16] Antonio Gulli and Sujit Pal. *Deep learning with Keras*. Packt Publishing Ltd, 2017.
- [17] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

- [18] Donald J Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *KDD workshop*, volume 10, pages 359–370. Seattle, WA, USA., 1994.
- [19] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5):561–580, 2007.
- [20] Sabri Boughorbel, Fethi Jarray, and Mohammed El-Anbari. Optimal classifier for imbalanced data using matthews correlation coefficient metric. *PloS one*, 12(6):e0177678, 2017.
- [21] Ryan Poplin, Avinash V Varadarajan, Katy Blumer, Yun Liu, Michael V McConnell, Greg S Corrado, Lily Peng, and Dale R Webster. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, 2(3):158, 2018.
- [22] Abrar H Abdulnabi, Gang Wang, Jiwen Lu, and Kui Jia. Multi-task cnn model for attribute prediction. *IEEE Transactions on Multimedia*, 17(11):1949–1959, 2015.
- [23] Shaohui Mei, Jingyu Ji, Junhui Hou, Xu Li, and Qian Du. Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4520–4533, 2017.
- [24] Radhesyam Vaddi and Prabukumar Manoharan. Cnn based hyperspectral image classification using unsupervised band selection and structure-preserving spatial features. *Infrared Physics & Technology*, 110:103457, 2020.
- [25] Timothy Dozat. Incorporating nesterov momentum into adam. 2016.
- [26] Sebastian Raschka, Joshua Patterson, and Corey Nolet. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *arXiv preprint arXiv:2002.04803*, 2020.
- [27] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: A llvm-based python jit compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, pages 1–6, 2015.

7 Appendix

In this section, we present the supporting results referring to the text in the main document.

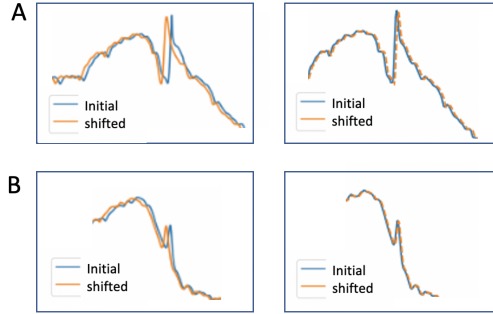


Figure 6: The magnetic field readings are obtained using an instrument with sensors and channels. The first scan refers to the pipe scan with an initial configuration, the second scan refers to the scan with shifted configuration. Figure (a) shows the differences between both scans taken from the same sensor. We clearly see some phase shift between the peaks and regions of major misalignment, have been highlighted. Figure (b) Upon applying the FastDTW algorithm, we are able to reduce the phase shift and thereby align the two scans appropriately, as a preprocessing step.

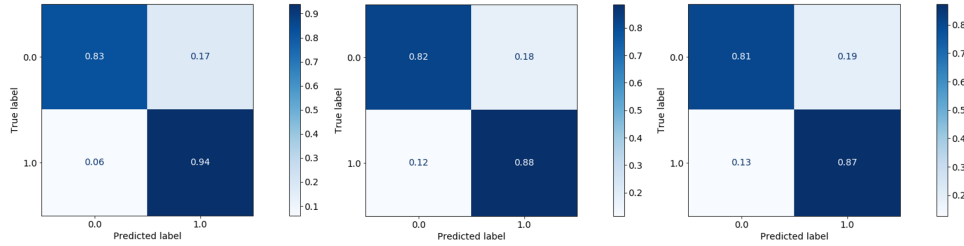


Figure 7: The confusion matrix for the best performing supervised learning point based methods are reported. From (L-to-R), Support Vector Classification, Decision Trees and MLP Classifier method results are shown. All methods perform very well on the train and test data sets with few false positives

Table 4: Time to train the algorithms to number of datapoints. All times in seconds.

| Algorithm | N | 10*N | 100*N |
|----------------|------|------|-------|
| k-NN | 1.47 | 11.2 | 140 |
| SVC[rbf] | 8.76 | 751 | 18274 |
| Decision Tree | 0.33 | 3.86 | 131 |
| MLP Classifier | 7.9 | 69 | 772 |

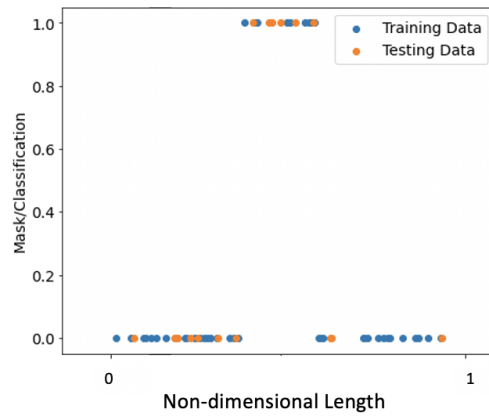


Figure 8: The issue of data leakage is seen with randomized test/train split, where there is significant overlap in the testing and training distribution, thereby creating overly optimistic models